THE CAPACITY OF TIME VARYING MULTIPLE USER CHANNELS IN WIRELESS COMMUNICATIONS

by

Muriel Médard

B.S., Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 1989

B.S., Mathematics
Massachusetts Institute of Technology, 1989

B.S., Humanities
Massachusetts Institute of Technology, 1991

M.S., Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 1991

Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Science in Electrical Engineering

at the
Massachusetts Institute of Technology
September 1995

Signature of Author...
Department of Electrical Engineering and Computer Science

Certified by ....
Professor Robert G. Gallager
Thesis Supervisor

Accepted by ....
Professor Frederic R. Morgenthaler
Chairman, Committee on Graduate Students

Acknowledgement.

I wish to thank my supervisor. Professor Robert Gallager, for his patience and guidance throughout this thesis and my graduate education at MIT. His devotion to his students and enthusiasm for his work was an inspiration.The members of my thesis committee, Professor Mitchell Trott and Professor Jeffrey Shapiro, were aways available and eager to share their insight. In particular, Professor Mitchell Trott spent many hours with me discussing my thesis and related topics.

My family and friends have also been very much a part of this undertaking. My husband, John, with his unfailing love and patient support, put up with years of being married to a graduate student. My mother provided me with the encouragement, help and love without which I could never have finished my doctorate. My late father always encouraged my academic undertakings and believed I could do whatever I set out to do. My daughter Vivian has not only brought great joy into my life but imaginatively modified some files in my thesis. Finally, I wish to thank all my friends at LIDS for their enthusiasm and fellowship, as well as all my other friends, in particular my dear friend Sue Patterson and all my sisters of AXΩ, without whom MIT would not have been the place where I was glad to do most of my growing up.

This research was partially supported by the DARPA MDA972-92-J-1038, by ARO-Wireless Communication - AASERT under Grant DAAH04-95-1-0265 and by NSF under Grant 9314341-NCR.

THE CAPACITY OF TIME VARYING MULTIPLE USER CHANNELS IN WIRELESS COMMUNICATIONS.

by

Muriel Médard

Submitted to the Department of Electrical Engineering and Computer Science on September 1995 in partial fulfillment of the requirements for the Degree of Doctor of Science in Electrical Engineering

ABSTRACT

We derive a model for wireless communication channels without feedback for one or several users. The model takes into account the fact that the channels vary in time. We establish the effect of time and bandwidth spreading when a user's channel is perfectly known. We derive the capacity of the system in the single and multiple user case when the channel is known at the sender and the receiver. We establish a multiple access coding theorem which uses interleaving to achieve an error upper bound which decreases exponentially in the block lengths. We study the effect upon capacity of not knowing the channel at the sender and the receiver and derive results relating the variance of the channel measurement error at the receiver to the loss in capacity with respect to the case where the channel is perfectly known. We show that, for a time invariant channel, we may achieve the same rates as in the case where the channel is perfectly known *apriori* at the receiver. For the μ^{th} order Markov channel model, we derive limit results relating the loss in capacity due to not knowing the channel to the channel characteristics. We give explicit bounds for the effect of the channel variations when we have a Gauss-Markov model for the channel taps. We show that, as the Doppler Spread goes to 0, the mutual information for the Gauss-Markov case approaches that attainable when the channel is perfectly known at the receiver. We establish some results concerning the fact that we wish to spread at least until the SNR per degree of freedom drops to a certain value. We give examples where spreading is not desirable below a certain threshold of SNR per degree of freedom. We use our results for Gauss-Markov channels to establish an upper bound to the minimum SNR per degree of freedom until which spreading is desirable.

Thesis Supervisor : Professor R. G. Gallager

Title : Professor of Electrical Engineering.

Chapter I- Introduction.

I-1- Purpose of the thesis.

The increasing applications of wireless communications have spawned much research and debate about the best manner to utilize the available spectrum, and have prompted more spectrum to be opened to commercial uses. The basic model for mobile wireless communications is that of mobile users in two-way communication with some central network. In current systems, the mobiles communicate with stationary transmitter/receiver units, called base stations, which in turn are connected to the public switched telephone network (PSTN) or some alternative carrier, such as a satellite link or the cable system. The communications link from the base station to the mobile is called the forward or downlink and the reverse link is called the uplink. The figure below gives a schematic of the network we are considering, with the downlink shown in full lines and the uplink shown in dashed lines. Note that the representation of the mobile user as a vehicle is simply for convenience, for the user could also be a pedestrian. The presence of obstacles is unavoidable, particularly in dense metropolitan areas, and may even preclude a line of sight between mobile and base station. The reflections off various obstacles leads to the reception of echoes of the original transmission. This phenomenon is known as multipath, and each reflection is called a path. These paths are typically time-varying, particularly when there is relative motion of the obstacles, the mobile, and/or the base station. For instance, in the figure below, as the mobile approaches the base station, the line of sight path becomes shorter while the path involving a reflection off the obstacle becomes longer. As we shall see later, phase changes accompany these path variations.

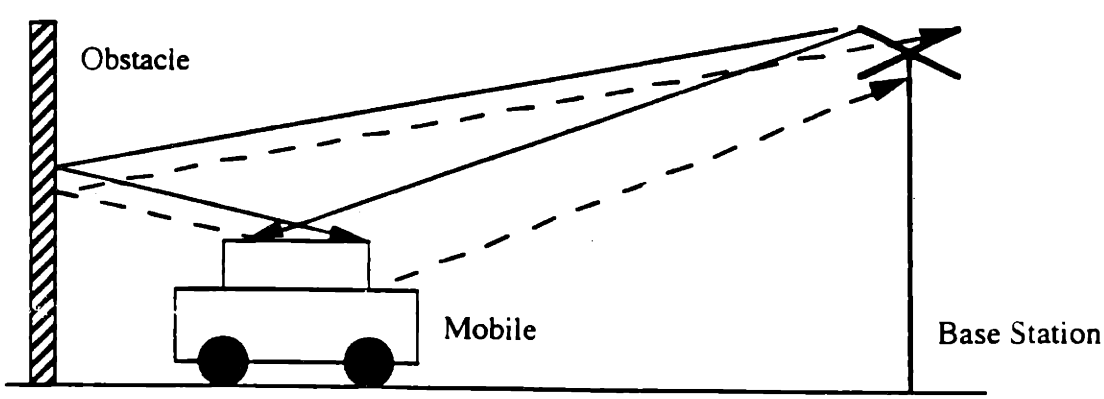


Figure I.1 : Uplink and downlink in a mobile communications environment.

The geographical area over which a particular base station communicates with mobiles is referred to as the cell corresponding to that base station, hence the name cellular telephony. Depending on the fading conditions, the cell or area of coverage of any base station may change in time and may overlap with neighboring cells. Such overlap is desirable so that calls may be transferred smoothly from one cell to another. This transfer is called hand-off. Improper hand-off may lead to a call being lost and, hence, to very poor quality of service. The cells are often represented schematically by a honeycomb pattern of contiguous hexagons as shown below in dashed lines.

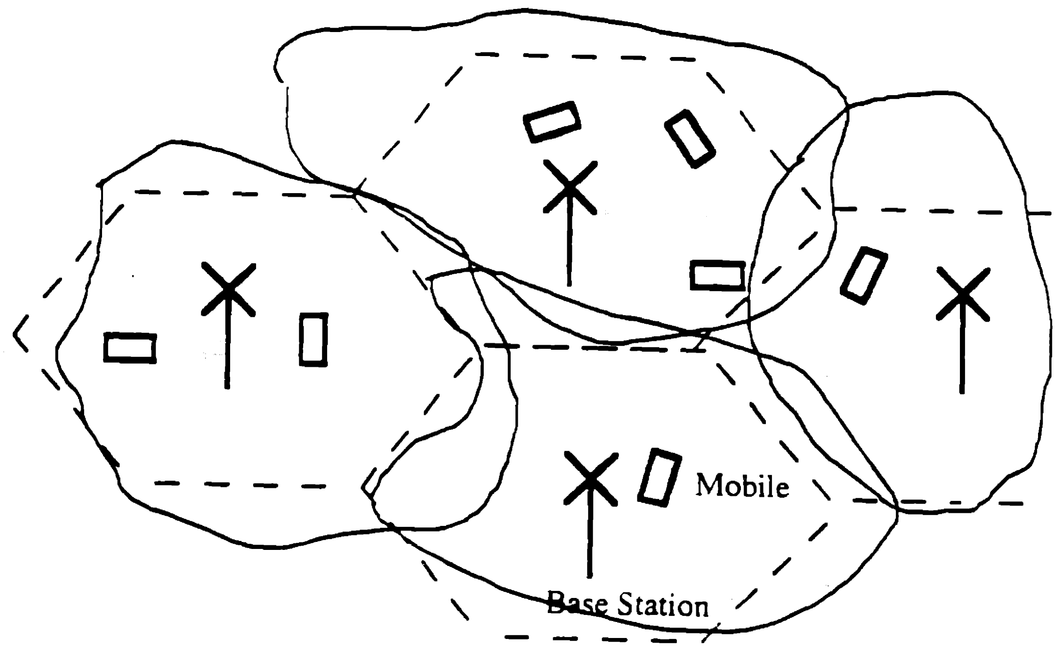


Figure I.2 : Cellular structure.

Within the general framework we have sketched, there are many proposed standards and many claims are made by these standards' respective proponents. We concentrate on one of the most important features of these standards, namely the transmission scheme used in the uplink and downlink. Although they are important, we do not address such issues as the hand-off scheme, the compression of the data or the interaction among different cellular providers. We give below a short overview of the main existing standards in commercial systems. It is difficult to evaluate these different strategies individually without some adequate model for the problem. Several simplified models exist but fall short of considering the time-varying character of the channel. Most tend to consider the channel through a statistical description only. Another extreme in the information theoretic literature is to consider only the time-spreading nature of the channel but not its time-varying nature. When the problem of multiple access is added, most models usually consider the other users as interference if they occupy the same spectrum at the same time, or simplify the problem by requiring the users to be made mutually orthogonal by some means. The issue of how much feedback is given by the base station to the mobiles renders the problem even more confusing. While everybody agrees that some form of power control, albeit rudimentary, needs to be performed, the amount of bandwidth, power and time devoted for such power control seems to be the result of *ad hoc* schemes. In the extreme case of full feedback, the amount of bandwidth devoted to the feedback would be comparable to the amount of bandwidth used to send information bits. Therefore, we see the utility of proposing a model for time-varying channels which allows us to determine the capacity of the channel under multiple-access conditions. Capacity considerations are not purely academic. Not only do they provide us with an optimum against which to gauge any system, they also yield insight as to how to achieve such optimality. While any specific transmission scheme will eventually become obsolete and we cannot predict what the ingenuity of applications engineers may create next, we may at least attempt to determine in what direction progress will occur.

In order to motivate our need for further research, this chapter gives a brief overview of the transmission schemes presently used. From the above discussion, we see that there are three areas which need to be addressed and which are a source of confusion and difficulty. First, the time-varying nature of the channel, which introduces memory and distortion, must be dealt with appropriately. Such is the topic of Chapter II, where we consider the single-user time-varying channel without feedback. Second, the issue of multiple access is overlaid onto time-varying channels in Chapter III. Finally, in Chapter IV, we present the conclusions of our research and present some directions for further work.

I-2- Current schemes.

I-2-1- Frequency Division Multiple Access.

The currently most used scheme in North America is Frequency Division Multiple Access (FDMA), which is used in the AMPS standard. Different users transmit over different channels, or portions of the spectrum, so that they transmit signals that are orthogonal to each other. In the FDMA scheme, bandwidth is divided into channels which are reutilized at certain spatial intervals such that the interference from one cell where a certain channel is used to the next is low enough to allow adequate detection ([Lee], [Lee86]). Typically the re-use interval is considered to be a cell, so that only contiguous cells may not use the same channels. In practice, the use of directional antennas may somewhat relax this constraint. Channel re-use is only allowed when there is 10 dB attenuation between separate transmissions on the same channel, regardless of whether that channel is being used in a contiguous cell. Therefore, the power control consists of keeping each user's channel at a power high enough that each user has adequate reception at the base station and low enough that each user does not interfere substantially with other users transmitting over the same channel in another location. Two cells which use the same channel simultaneously are usually called "co-channel" cells. The interference in this context is usually measured in terms of carrier to interference (C/I) ratio or of signal to noise ratio (SNR). Clearly, the interference from co-channel cells depends on the strength of the signals, the topology of the area (which affects fading) and the distance among the cells ([Lee89], [Lee]). The power assignment in this case has the advantage of being simple, since each user has a designated channel at a constant power. However, the many restrictions placed on how spectrum, time and power are assigned probably do not make for an efficient use of the available resources.

Several techniques have been proposed and are utilized for the efficient use of channels. We may roughly divide them into techniques which consider co-channel interference as a simple exclusion constraint and techniques which attempt to perform some sort of simultaneous power control and channel assignment. Let us first consider the simplified case where the co-channel interference is simply viewed as being unacceptable below a certain carrier to interference ratio and acceptable above. In this way, the problem reduces to a discrete optimization problem where the constraints are given by the unacceptable co-channel interference pairs. The problem of optimal static channel assignment can be reduced to that of graph coloring, as shown by Hale ([Hal80]), and is therefore NP-hard. Such a problem, however, does not bear much relevance to the FDMA situation, since moving mobiles make the channel assignment problem a dynamic one. Most of the techniques proposed rely on heuristics for channel assignment. Most of the techniques keep some portion of the channels statically assigned to certain cells, while the remaining channels are dynamically allocated among the cells according to some rule. These techniques seek reasonable channels whose bandwidth and power are fixed. Their performance. determined by the available rate to the users. is not measured against achievable capacity regions. The performance evaluation of these methods also depends on heuristics and simulations. Since keeping record of the dynamic channels in a large system rapidly becomes cumbersome ([Ekh86]), schemes for borrowing channels have been proposed in which channels are borrowed only from neighboring cells. Such is the principle of directed retry, proposed by Ekhlund and Karlsson ([KE89],[Ekh86]). Finally, cost-minimization models for FDMA have been proposed. Cost reduction is proposed by Anderson, who uses it to reflect the state of the cell which is most adversely affected by a proposed borrowing ([And73]) , and by Yum and Zhang ([ZY89]). The use of simulated annealing ([DKR93]) and neural networks ([Kun88]) are also examples of minimization of a cost function approach to achieve overall optimization. Dynamic channel allocation through centralized control has been studied, often through simulation ([DV94]). Dynamic channel allocation can however be subject to certain undesirable instability ([SG93]). All the above refinements require a substantial increase in hardware and complexity. Performance limits for graph-coloring approaches have been given in ([MS94]), where the problem is reduced, in the limit of an infinite number of channels, to a linear optimization problem. It is interesting to note that the problem of the capacity region in mobile communications, which is continuous in the first place, is reduced to a simpler discrete problem for channel allocation in FDMA, which is itself solved by a continuous approximation.

Methods for considering channel assignment and power control simultaneously abound. Some rely on cell customizing. One such proposed scheme considers statically assigning different frequencies within a cell according to the distance from the center of the cell and controlling the power to reduce interference among the cells ([Lee87]). A dynamic method of power control, which is used in urban systems, is subdividing cells into "micro-cells" to respond to changing traffic conditions. More general methods for joint power control and channel assignment have been studied in [Zan92] and [CNW94]. The first work considers the static assignment for given attenuation matrices within and between cells. The second work overviews several heuristic algorithms which seek to balance the carrier to interference ratio, or pursue some sort of greedy approach ([GGV93]) in terms of carrier to interference, etc... Whichever cost minimization approach we choose, solving the problem of channel assignment always reduces to some sort of dynamic programming problem which is difficult and whose relevance to our actual problem will depend strongly on our model's assumptions.

I-2-2- Time Division Multiple Access.

Division in time rather than frequency is the basis for TDMA (time division multiple access) where different users are allocated time slots which they may use for transmission. Practical systems which use TDMA usually incorporate it into other schemes, in particular FDMA. For instance, an FDMA channel might be divided among several users through TDMA. An important example is the standard that has been adopted in Europe by the GSM (Groupe Spécial Mobile), established in 1982 by the CEPT (Conference of European Postal and Telecommunications Administrations). In the U.S., Hughes Networks Systems has created enhanced TDMA (E-TDMA), which features compression of the voice data. In Japan, TDMA is used in the Japan Digital Cellular (JDI) mobile communications standard. In TDMA, users contend for time slots, releasing them when they are not using the channel. The problem of optimally assigning time slots among users for a given traffic pattern is, as in the case of FDMA, combinatorial in nature. The problem of minimizing the number of transmissions (hence the maximum aggregate delay) for a certain number of messages to be sent to a certain number of users is NP-complete. The problem reduces to a traditional job-scheduling problem, with bin-packing (which is NP-hard) as a subproblem ([GG92], [GS76]). However, as with FDMA, this job-scheduling problem is not germane to the TDMA case, because we are very unlikely to know in detail the future traffic pattern of the users. Therefore, as in the case of channel assignment in FDMA, algorithms have been proposed which rely on heuristics to achieve suboptimal assignments. We may note that packet schemes where collision is not allowed are also a form of time sharing and reduce to the same problem as the TDMA problem outlined above. TDMA requires good synchronization to keep the users orthogonal and usually some portion of the transmission is devoted to achieving synchrony.

I-2-3-Spread Spectrum.

Another approach to division of the spectrum is the overlaying of users in spread-spectrum (SS) systems. A definition, as given in [PMS82] and quoted in [LM] is: "Spread-spectrum is a means of transmission in which the signal occupies a bandwidth in excess of the minimum necessary to send the information; the band spread is accomplished by means of a code that is independent of the data, and a synchronized reception with the code at the receiver is used for despreading and subsequent data recovery". However, this definition does not capture another aspect of spread-spectrum: the average transmitted power remains the same. Otherwise, with unlimited power, even a very small bandwidth would be enough to transmit any amount of information. The bandwidth expansion factor is called the processing gain. SS systems have been in use for many years, in particular in defense and transportation applications. SS techniques have a long history in military systems because of their low detectability to eavesdroppers and their superior ability to withstand jamming if the jammers are limited in power ([PSM82], [Lee91]). Spread-spectrum commercial applications include Qualcomm's Omnitracs, which tracks and communicates with trucks. In the past few years, the interest in SS techniques for civilian purposes has been renewed ([Goo91], [PMS91]). This interest was largely sparked by the Qualcomm [Qua91] system for personal communications, which uses direct-sequence spread-spectrum. Although the IS-95 standard uses SS transmission, the applications of SS to public communications systems have been very limited.

The two most common approaches to SS are direct sequence Code Division Multiple Access (referred to in what follows as CDMA) and Frequency Hopping (FH). In the former scheme, each user is assigned a code and shares the bandwidth with all other users. The users are therefore distinguished from each other by their codes. Each original bit of a user is transmitted as a string of bits called "chips". If all users have orthogonal codes and the channel is well-behaved, they do not interfere with each other. In general, codes with low cross-correlations, such as Gold codes, are deemed sufficient to preclude most multi-user interference ([PSM82]). However, channels in practice distort the signal, for instance through multipath, so that interference may occur among users despite the use of orthogonal codes. Therefore, unless we perform some form of joint decoding of the users, each user appears as noise to all other users. The power control for existing schemes such as the Qualcomm ([Qua91]) system and the IS-95 standard consists of maintaining the power levels of all users so that each is received at the same power. Thus, the system strives to maintain the same amount of interference at the receiver for each user.

In FH, the users use small portions of bandwidth and "hop" from one portion to another, thus spreading the users over all the available bandwidth. If the hopping patterns of the different users are selected in such a way that no two users ever use a portion of bandwidth simultaneously, i.e. if users are assigned altogether disjoint hopping patterns, then this method reduces to an FDMA problem with added diversity

([VMS84]). If CDMA and FH are used in such a way that signature codes and hopping patterns are chosen so that the interference among users is low rather than precluded. the problem of power control becomes closely linked to the detection method used and the fading experienced by the users. Error probabilities for FH are usually computed according to the probability of a "hit". i.e. of several users using the same portion of the overall bandwidth simultaneously. Frequency hops may occur more slowly than the information bit rate (slow hopping) or more rapidly than the data rate (fast hopping). The overlapping may occur because of bursty traffic, such as packet traffic, which causes users with overlapping hopping patterns to interfere if they are both transmitting ([MP90]), or because of continuous traffic in which the patterns can overlap, ([HS90]). Probabilities of error involving partial hits versus full hits refine the analysis of FH systems ([Ger90]) at the expense of complexity. Similarly, the analysis of fast hopping systems is more complex than that of slow hopping systems. Such analysis has been carried out in the slow hopping case to take into account fading channels ([GP82], [GP82,1]) and unequal power levels of users ([Ger90], [VMS84]). The complexity of such analyses, which are often modeled by extensive Markov chain subsystems, clearly shows the need for theoretical capacity results which may hold independently of the detailed system and model.

It might seem that the greater complexity associated with reception for SS systems renders it unattractive. The primary advantage of spread-spectrum schemes is that they afford greater flexibility. For instance, the Qualcomm system makes use of voice activity to stop transmission when the users are silent during the normal course of spoken conversation. If we were to re-use frequency or time slots in FDMA or TDMA when users are silent. it would require extensive re-allocation of resources and users might be dropped in the middle of a conversation. Since calls are fairly frequently dropped when a mobile changes cells and must therefore change channels, one could imagine that extensive re-assignment of channels could be disastrous. However, E-TDMA dynamically re-assigns channels during some inactive talk times. In the CDMA case, the renewed voice activity of a new user will simply cause a slight increase in interference to all users. Indeed, since partial interference is allowed, we may allow a new user to degrade the overall service of all users whereas in the FDMA or TDMA case, the new user might have been blocked while other users saw the quality of their service unchanged. Fitting a new user in an FDMA system with no available channels. for instance, requires modifying the very structure of the channels. We can imagine that reducing the bandwidth of each user by equal amounts to gain a new channel for the new user would be unattractive and complicated. In practice. when the capacity of the FDMA system is insufficient for the demand. cell-splitting may occur. i.e. the cells are made smaller in order to increase the possibility to re-use frequency. Cell-splitting is at best an *ad hoc* solution with high costs associated with it. in particular because each cell requires its own hardware and users have to be re-assigned to cells. In TDMA. having to change the timing of the system to reduce the length of the time slots could lead to timing difficulties and cell-splitting is no more attractive than in the FDMA case. Therefore. SS allows a new user to be admitted without requiring other users to modify their transmission mode.

Such a feature is not only important in allowing more users to be admitted into the system but could also allow for better differentiation in service quality. Instead of having equal service for every user, price discrimination could be performed by guaranteeing to some users better carrier to interference ratios than to others. In the case of FDMA or TDMA, such guarantees would entail unwanted dynamic changes in the frequency re-use pattern or in, respectively, the frequency or the time division. In CDMA, one could simply set the carrier to interference level at which a certain category of service would operate. Such a feature would surely be attractive in an industry suffering from high churn and which already practices every other sort of price discrimination to palliate for the excessive churn. The feasibility of such a scheme clearly depends on its sensitivity to operating conditions.

Our discussion above has not considered the type of communication that is carried. All of the systems mentioned above serve primarily as voice channels. If we have voice communication, delay may be a more stringent constraint than quality of transmission, whereas the reverse might hold for file transfers. Packetized transmission may therefore be attractive for data and is the basis of the Cellular Digital Packet Data (CDPD) standard developed by IBM. It utilizes the existing cellular networks to transmit at rates of up to 19,200 bits/sec ([Sch94]). The demand for packetized data or a continuous data stream and the allocation of resources between such data transmissions are outside the scope of this thesis. Therefore, we wish to consider what rates are achievable, without regard for the burstiness or the delay constraints of different data. Once certain rates are achievable, the question of how to allocate rates to different types of traffic may be considered separately.

I-3- Thesis outline.

The preceding discussion motivates the approach proposed for this thesis. We see that there is a vast array of implementation techniques and that it is difficult to evaluate their performance. Many fundamental questions remain unanswered and these questions have a direct impact on how to design and operate systems. Given the wide range of standards, either existing or proposed, it is useful to have a framework from which general statements may be made which are applicable to any standard. Therefore, we take an information theoretic approach to obtain results regarding the optimal achievable rates, Such an approach is general and allows us to be removed from the complicated issues of implementation. Moreover, it provides insight into how system features affect performance. We may then better isolate the issues which intrinsically limit achievable rates versus the issues which arise simply because of difficulty of implementation. Eventually, we interpret our results in terms of general properties of, and guidelines for, systems. We may divide the main thrust of the thesis into the following three topics:

- What are the optimal rates that are available under the following conditions: amount of bandwidth, transmission scheme, number of users, channel response and power assignments? The considerations which affect a system for a given number of users are the channel response and the multiple access effects. Both must be included simultaneously to obtain a plausible model. Under fully known conditions, we find the capacity of the system for a single user in Chapter II and for multiple users in Chapter III.

-What effect does uncertainty in our knowledge of the system parameters have upon capacity? In any real system, knowledge of the channel can only be acquired through measurement, and can therefore never be altogether certain. Therefore, we next consider the effect of the uncertainty in the system regarding the set of parameters mentioned above. The transmission scheme, number of users, amount of bandwidth and other such parameters which are set *apriori* are known with certainty. On the other hand, the channel response and, hence, power levels may be determined only with some uncertainty. The effect of this unavoidable uncertainty upon the capabilities of the system is investigated for single user systems in Chapter II and for multiple user systems in Chapter III.

Chapter II- The single user case in a time-varying channel without feedback.

II-1- Introduction.

This chapter looks at the mutual information between input and output for a single user transmitting on a time-varying channel. This single-user case is not of direct interest for the real applications of mobile communications, but it should allow us to deal with the issues of time-varying channels without having to contend with multiple access interference. We consider the reverse link, i.e. the link from the mobile user to the base station. The reverse link involves signals that are often weaker than the ones transmitted by the base station, since the power that may be emitted by mobiles is bounded by the limitations of portable power sources and by regulations governing the power of emissions allowed in hand-held appliances. In later chapters, we shall consider the fact that the multiple access effects are more severe for the reverse link.

The main thrust of the chapter is to examine the effect upon capacity of not knowing the channel exactly. We first present the idealized model where the channel is perfectly known at the sender and the receiver. We establish the capacity of an arbitrary but known time-varying multiplicative channel with additive white Gaussian noise (AWGN), since only time-invariant capacity results were previously available. The main contributions of the chapter are for the case where the channel is not known at the sender (as would occur when there is no feedback from the receiver to the senders) but partially known at the receiver. When the channel is known at the receiver with an error of fixed variance, the loss in mutual information due to not knowing the channel at the receiver may be bounded tightly in the variance of the channel measurement error. The uncertainty in the measurement of the channel is a consequence of its time-varying nature. Indeed, we could always obtain an arbitrarily good description of a static channel by probing it initially. We show that, if the channel is time-invariant but unknown at the receiver, we may achieve the same rates as if the channel were known at the receiver *apriori*. Therefore, we consider the intrinsic channel measurement error in terms of the channel parameters. For channels that may be represented by m^{th} order Markov processes, we establish exact limits for the loss in terms of mutual information per symbol. In particular, we relate the loss in capacity to measures of channel variability for a Gauss-Markov model for the channel. We finally consider the issue of spreading, when we spread by increasing the bandwidth for transmission while retaining the same total power for the transmission. We study the effect of the channel uncertainty upon spreading and show that the channel variations may limit the extent to which spreading is beneficial in terms of capacity.

II-2- Model.

In the following section, we present our model for the channel. Mobile communications channels are subject to many changes which cannot be controlled or predicted. Obstacles appear and disappear as the mobile moves in its environment, leading to partial blocking (shadowing) or to echoes (multipath). In examining the problems of radio propagation for mobile communications, the issue of multipath is very important. Multipath accounts for most of the deep fades and may change far more rapidly than other fading phenomena. Moreover, it is altogether dependent on the surroundings of the base station and the mobile and must be considered in real time. Multipath is, therefore, relatively ornery and cannot be simply modeled by the log-normal fading which usually describes slow fading phenomena. We present a channel model based on a general multipath situation. We establish equivalent continuous and discrete-time models for this general multipath situation. In light of these results, we review some of the most common models for narrow band multipath systems and relate them to our general results, which are applicable to both narrow and wide band situations.

II-2-1- Model for the multipath channel.

We first derive a model, both in continuous time and discrete time, for a single user multipath channel and then incorporate AWGN into our model. The approach will be extended in the next chapter to several users. We denote random variables by capital letters and sample values by lower case letters, unless we explicitly define a variable otherwise.

II-2-1-a- Continuous time model.

Let us look at the channel response at baseband for a multipath channel. We assume that the signal is contained in a bandwidth W_{input} centered around a carrier frequency f_0, so that the baseband input signal is constrained to $[-W_{input}/2, W_{input}/2]$ Although our whole argument could be carried out at passband, it would complicate the discussion. The effects of the carrier frequency are taken into account when appropriate. The goal is to develop a continuous-time baseband model and a tapped-delay line model. In order to derive such models, we first consider the continuous-time response of a channel and then look at its samples. The samples are chosen so that, using the Nyquist theorem, they are sufficient to reconstitute the continuous-time response. Equivalently, the sampling rate is greater than the frequency W to which the response is bandlimited. We consider later how to choose such a sampling rate. We present in this section the model in terms of sample values, without considering (unless otherwise specified) the type of random variable of which they are samples. The model is therefore applicable to any sort of random process and we shall later examine specific examples. We first look at a system without noise, in order to deal solely with the time-varying multiplicative part of the channel. We shall later include the effect of additive noise. The continuous-time system which we consider (for the present without noise) is represented in Figure II.1. A discrete data stream from the user is passed through a modulator, whose output is a continuous signal.

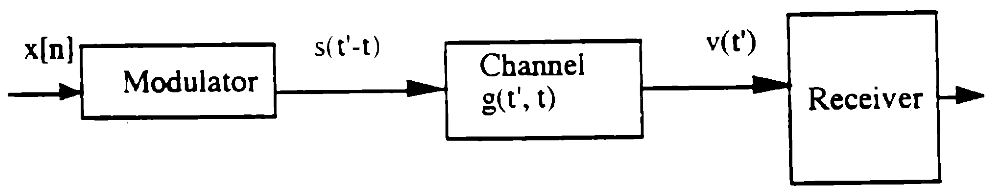


Figure II.1 : Continuous-time multipath system.

We denote the channel by a linear time-varying filter of impulse response g rather than the more familiar h to avoid confusion with the entropy function which we shall use later. The impulse response of the multipath channel seen at time t' for a transmission sent at time t'- t is taken to be:

$$g(t', t) = \sum_{\text{all paths } m} g^{m}(t', t) \qquad \text{II-2.[1]}$$

where $g^m(t, t')$ is the response of the multipath channel seen at time t' for a transmission sent at time t'-t. If we had a continuum of paths rather than several specular paths, we could easily adapt our discussion by using integration rather than summation in II-2.[1].

<u>II-2-1-b- The Doppler effect.</u>

We may note that, since g^m is time-varying, any frequency spreading may be incorporated into this complex function. Since we would in reality be transmitting at some passband, the frequency spread would be a function of the carrier frequency, denoted by f_0 in Hz, although we consider a baseband representation. Although we usually consider our model at baseband, it is more natural to discuss Doppler spreading at passband, since it is closely related to the carrier frequency. If we have reflections off objects in motion relative to the transmitter, the received signals will experience a Doppler spread in frequency ([Gal64], [Ken], [Lee]), denoted by $B_{Doppler}$. Indeed, different paths will correspond to reflections off different objects and these paths will be subject to different Doppler shifts in frequency, depending on the motion of the obstacles relative to the transmitter and the receiver. The spread among these shifts in frequency is the Doppler spread. In general, we will observe a Doppler spread in a mobile communications environment, since we have mobiles moving with respect to many fixed and moving objects, especially in an urban environment.

Let us derive the form of the Doppler shift. Let us consider a receiver and a sender moving away from it at velocity v. At time 0, the sender and the receiver are at the same position. A transmission sent at time t_S is received at time t_R. The sender transmits a sinusoid at frequency f_S. The receiver receives a sinusoid at frequency f_R. The schematics below show our model:

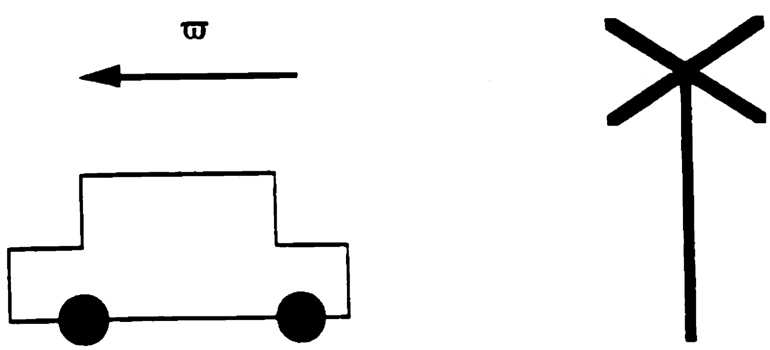


Figure II.2 : Sender and receiver in relative motion.

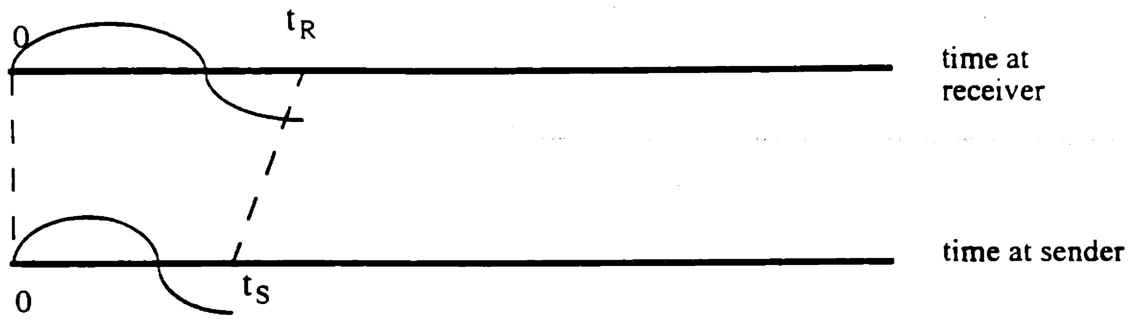


Figure II.3 : Times and transmissions at the sender and the receiver.

Suppose that the sender and the receiver are at the same location at time 0. From our model, we see that a transmission is received after having traveled a distance $t_R\varpi$. Therefore, the delay between transmission and reception is $\frac{t_R\varpi}{c}$, i.e. $t_R - t_S = \frac{t_R\varpi}{c}$. The reception time may be expressed as $t_R = \frac{t_S}{1-\frac{\varpi}{c}}$.

The number of cycles received by time t_R is the same as the number of cycles sent by time t_S and is equal to $T_S f_S$. Therefore, the frequency of the sinusoid at the receiver is given by the number of cycles of the sinusoid over the time elapsed to receive these cycles, namely $\frac{t_S f_S}{t_R} = f_S\left(1-\frac{\varpi}{c}\right)$. Therefore, the frequency shift is $f_S\frac{\varpi}{c}$.

The analysis carried out above holds for each path m corresponding to an obstacle for which the path length from the mobile to the base station is changing at speed v^m. The associated Doppler shift in Hertz is B^m given by

$$B^m = f_0\frac{\varpi^m}{c}$$

II-2.[2].

The difference between the smallest and largest Doppler shift B^m over all paths gives $B_{Doppler}$ at passband. The delay associated with path m is:

$$\tau^m(t') = \tau^m + \frac{\varpi^m}{c} t'$$
$$= \tau^m + \frac{B^m}{f_0} t'$$

II-2.[3]

where τ^m is the original time shift.

To evaluate the extent of the Doppler effect in mobile communications, we may look at its magnitude in a simple application. If we have a vehicle traveling at 55 mph, transmitting at around 800 MHz (the currently available region for wireless telephony), the shift B^m for a single path is 65.6 Hz. From our discussion of the Doppler spread, $B_{Doppler}$ is of the order of the individual B^m. $B_{Doppler}$ is therefore small in comparison with the channel spacings typically considered. For instance, in the IS-54 standard, channels are spaced 30 kHz apart. In the GSM system, they are spaced 200 kHz apart. The effect of the Doppler spread is therefore negligible when we consider its bandwidth broadening effect. $B_{Doppler}$ is of the order of the roll-off on a bandlimiting filter.

However, even if the actual spreading may be considered to be negligible, the time-varying nature of the channel may not be neglected. The importance of the Doppler spread in our application is that it indicates the speed at which the phase changes. The fact that we have frequency shifts indicates that the channel is time-varying and the magnitude of the frequency shifts gives the speed of these variations. The channel is not linear time invariant and thus simple frequency and time domain methods for general LTI systems no longer apply.

II-2-1-c- Sampling in a time-varying channel.

We have established that the channel changes in time at a rate which is bounded by a known quantity. In order to derive a discrete-time model for the system, we must determine how to sample the channel output. Since we are using a time-varying model for the channel, we need to take into account that the output bandwidth may be greater than the input bandwidth. A broadening of the bandwidth is due mainly, as we have discussed earlier, to the Doppler effect. If the channel were not time-varying, we could use a LTI model and the issue of bandwidth would be easier. For such a channel, the bandwidth of the noiseless output would be the smaller of the bandwidths of g(t) and s(t), and therefore any rate sufficient to sample s(t) would be sufficient to sample the noiseless output. The Doppler spread represents the rate at which the channel varies. As we have seen, $B_{Doppler}$ is typically small with respect to the bandwidth over which we are sending, i.e. the channel varies slowly in comparison to the time needed to send a datum. In general, we know that the bandwidth of the output of a linear time-varying filter will be bandlimited to the sum of the input bandwidth plus the filter variation bandwidth ([Kai59]). Therefore, the output may be bandlimited to $W_{input} + B_{Doppler}$ at passband. Let us derive the time-varying model in frequency to illustrate this point.

In continuous time, s(t) is the complex baseband signal transmitted by the user. This signal might typically be a data stream modulated by a modulating waveform for transmission on the channel. The signal s(t) would be modulated to some passband for transmission, although we consider a baseband model here, rather than the passband model we used for our discussion of Doppler spread. For our single user case, the response due to the m^{th} path of the channel at time t' is

$$v^{m}(t') = \int_{t=-\infty}^{\infty} s(t'-t)g^{m}(t', t)\, dt \qquad \text{II-2.[4]}$$

where

$$g^{m}(t', t) = a^{m}(t')\,\delta\left(\tau^{m}(t') - t\right) \qquad \text{II-2.[5].}$$

where $a^m(t')$ is the complex multiplicative factor associated with the m^{th} path (accounting for fading and phase shift) and, as in II-3.[3], $\tau^m(t')$ is the delay associated with the m^{th} path. The difference between the smallest and the largest τ^m is called the time spread, or T_{spread}. T_{spread} represents the interval of time over which an impulse at the sender will be received at the receiver.

Let us indicate how the Doppler effect may be included in the baseband model of II-2.[1]. The variables a^m are complex with phase changing at rate $2\pi B^m$ and with amplitude denoted by $|a^m|$. Therefore, we may write that :

$$g^m(t' - t) = a^m(t')\,\delta\left(\tau^m(t') - t\right) = \left|a^m(t')\right| \exp\left(j\left(\theta^m + 2\pi B^m t'\right)\right) \delta\left(t - \tau^m - \frac{B^m}{f_0} t'\right)$$

II-2.[6]

where j is the square root of -1 and θ^m is the phase shift at t'=0. We understand that the channel baseband output obtained in this manner would require being filtered to [-W/2, W/2].

We may note that for f_0 of the order of 800 MHz and B^m of the order of 60 Hz, the rate of change of delay given in II-2.[3] is proportional to that of the phase, but much smaller. Since $\left|a^m(t')\right|$ changes slowly with the delay, we may use the common assumption ([Pro], pg. 705) that the amplitude $\left|a^m(t')\right|$ changes much more slowly than the phase and therefore we take the amplitude to be fixed while the phase changes with rate B^m. We shall not discuss the path fading models. A discussion of such models may be found in [GH92]. Most of these models, including the widely used Hata model, involve regressions run on extensive measurements and are difficult to discuss theoretically.

Let the superscript FT denote Fourier Transform and the subscript refer to the order in which we take the transforms. The derivation of 2-dimensional Fourier transforms is well known, but we present it here for the sake of illustration. We can express the above relation by taking transforms in frequency thus :

$$V^{m^{FT}}(f') = \int_{t'=-\infty}^{+\infty} v^m(t')e^{-j\,2\pi f' t'}dt'$$

(substituting II-2.[5])

$$= \int_{t'=-\infty}^{+\infty}\int_{t=-\infty}^{+\infty} e^{-j\,2\pi f' t'} s(t'-t) g^m(t', t)\, dt\, dt'$$

(changing the variables of integration)

$$= \int_{t'=-\infty}^{+\infty} \int_{t=-\infty}^{+\infty} e^{-j\,2\pi f' t'} s(t) g^{m}(t', t'-t)\, dt\, dt'$$

(replacing s(t) by the integral of its Fourier transform)

$$= \int_{t'=-\infty}^{+\infty} \int_{t=-\infty}^{+\infty} e^{-j\,2\pi f' t'} \int_{f=-\infty}^{+\infty} S^{FT}(f) e^{j\,2\pi f t}\, df\; g^{m}(t', t'-t)\, dt\, dt'$$

(changing the order of integration)

$$= \int_{t'=-\infty}^{+\infty} \int_{f=-\infty}^{+\infty} S^{FT}(f) e^{-j\,2\pi f' t'} \int_{t=-\infty}^{+\infty} e^{j\,2\pi f t} g^{m}(t', t'-t)\, dt\, dt'\, df$$

(multiplying by exp (j2πft') and its inverse)

$$= \int_{t'=-\infty}^{+\infty} \int_{f=-\infty}^{+\infty} S^{FT}(f) e^{-j\,2\pi f' t'} e^{j\,2\pi f t'} \int_{t=-\infty}^{+\infty} e^{j\,2\pi f t} e^{-j\,2\pi f t'} g^{m}(t', t'-t)\, dt\, dt'\, df$$

(replacing the innermost integral by its Fourier transform expression)

$$= \int_{t'=-\infty}^{+\infty} \int_{f=-\infty}^{+\infty} S^{FT}(f) e^{-j\,2\pi f' t'} e^{j\,2\pi f t'} G^{m\,FT}_{2}(t', f)\; dt'\, df$$

(regrouping exponent terms)

$$= \int_{f=-\infty}^{+\infty} \int_{t'=-\infty}^{+\infty} s^{FT}(f) e^{-j2\pi(f'-f)t'} G^{m\,FT}_{2}(t', f)\ dt'\,df$$

(changing the order of integration and replacing the innermost integral by its Fourier transform expression)

$$= \int_{f=-\infty}^{+\infty} s^{FT}(f)\, G^{m\,FT}_{2,1}(f'-f, f)\ df$$

II-2.[7].

Let us now compute $G^{m\,FT}_{2,1}(f'-f, f)$ for g given by II-2.[6] where $|a^m(t')|$ is constant.

$$G^{m\,FT}_{2}(t', f) = |a^m|\, e^{j\left(\theta^m + 2\pi B^m t'\right)}\, e^{-j2\pi f\left(\tau^m + \frac{B^m}{f_0}t'\right)}$$

(regrouping exponent terms)

$$= |a^m|\, e^{j\theta^m}\, e^{-j2\pi f\tau^m}\, e^{-j2\pi t'\left(-B^m + \frac{fB^m}{f_0}\right)}$$

II-2.[8]

and hence

$$G^{m\,FT}_{2,1}(f', f) = |a^m|\, e^{j\theta^m}\, e^{-j2\pi f\tau^m}\, \delta\left(f' - B^m + \frac{fB^m}{f_0}\right)$$

II-2.[9].

If our signal is narrowband with respect to the original carrier frequency, then the term $\frac{fB^m}{f_0}$ is negligible. We see that the double Fourier transform of the response due to the m^{th} channel is an impulse centered approximately at B^m. Henceforth, we shall neglect the terms $\frac{fB^m}{f_0}$ and consider that, for each path, there is a frequency shift of B^m. Let us now look at the response to all the paths combined, i.e. at the response from the channel. We may write that

$$v(t') = \int_{t=-\infty}^{+\infty} s(t'-t)g(t', t)\, dt$$

(using II-2.[1] and II-2.[5])

$$= \sum_{\text{all paths } m} \int_{t=-\infty}^{+\infty} s(t'-t)a^m(t)\, \delta\left(\tau^m(t') - t\right) dt$$

II-2.[10].

Therefore, from the linearity of Fourier transforms,

$$G^{m\,FT}_{2,1}(f', f) = \sum_{\text{all paths } m} G^{m\,FT}_{2,1}(f', f)$$

II-2.[11].

We have already stated that $g(t', t)$ gives the effect of the channel at time t' from inputs t time units earlier. Similarly, G^{FT} gives the response at frequency f' to an input at frequency f. The effect of each path is to give at passband a response shifted by B^m in frequency, and therefore the total response of the channel is a sum of responses with frequency shifts ranging over all the possible values that B^m takes. Hence, the response of the whole channel at passband is spread in frequency over $B_{Doppler}$. Therefore, from II-2.[9], we see that the double Fourier transform of the channel response at baseband in II-2.[11] is nil unless f-f' is in the interval $[-B_{Doppler}/2, B_{Doppler}/2]$, i.e. the input at a certain frequency f has effects on the output only within $B_{Doppler}$ of that frequency. The range of f at baseband is $[-W_{input}/2, W_{input}/2]$. Therefore, from II-2.[10]-[11], we see that the output is limited at baseband to $[-B_{Doppler}/2 - W_{input}/2, B_{Doppler}/2 + W_{input}/2]$ in frequency. Therefore, it is sufficient to sample at passband at the Nyquist rate of $W = W_{input} + B_{Doppler}$. Henceforth, we shall denote this Nyquist rate by W. Such a result is compatible with what we know about sampling in a time-varying channel ([Kai59]), but our treatment above illustrates the effect of the time-variation in the particular case of a simple multipath model.

In terms of the Bello functions for the description of time-varying channels. $g(t', t)$ is the input delay-spread function ([Bel63], [GH92], pg. 128). The function $G^{FT}_{2,1}(f'-f, f)$ is the output Doppler-spread function in terms of Bello functions. We could have taken transforms in the reverse order and still have obtained the same function ([GH92], pg. 133).

The question arises as to whether we truly need to sample at this rate W when we know the channel exactly. Would it not be possible for the receiver to invert the effect of the channel and reconstruct the original information signal sent over a bandwidth of W_{input} at passband? The problem arises because of the AWGN component, which is unknown at the receiver. Communication in the presence of noise is not about signal reconstruction but about information resolution. To take an extreme example, suppose the channel simply transmits the signal unchanged over the frequency interval $[-W_{input}/2, W_{input}/2]$ at baseband and duplicates the signal with a shift in frequency of W_{input} over the band $[W_{input}/2, 3W_{input}/2]$. In that case, we could simply bandlimit the received signal to $[-W_{input}/2, W_{input}/2]$ to reconstruct the original signal. If, however, the channel adds noise over the whole spectrum, by bandlimiting to $[-W_{input}/2, W_{input}/2]$ and ignoring the received signal over $[W_{input}/2, 3W_{input}/2]$, we are in effect foregoing the benefit of having a repetition code. Therefore, because of the presence of noise, we cannot, strictly speaking, sample at passband at W_{input} even if the multiplicative part of the channel is known and invertible.

II-2-1-d- Discrete time model.

Let us now derive an expression for the sampled system. We assume that the input $s(t)$ and the response $v(t')$ at baseband are bandlimited to $[-W/2, +W/2]$, since we consider the system at baseband. The response $v(t')$ may be sampled at time intervals of $1/W$, by the Nyquist sampling theorem. We thus obtain a discrete-time sequence defined as

$$v[k] = v(k/W) \qquad \text{II-2.[12]}$$

and we may similarly define

$$s[k] = s(k/W) \qquad \text{II-2.[13].}$$

We wish to determine a discrete channel impulse response to relate v[k] to the sequence s[k]. By the Nyquist sampling theorem, we know that

$$s(t) = \sum_{n=-\infty}^{+\infty} s[n] \operatorname{sinc}\left\{\pi W\left(t - \frac{n}{W}\right)\right\}$$

II-2.[14]

where sinc(x) is sin(x)/x. Therefore, by substitution in II-2.[10], we obtain

$$v(t') = \sum_{\text{all paths } m} \int_{t=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} s[n] \operatorname{sinc}\left\{\pi W\left(t' - t - \frac{n}{W}\right)\right\} a^{m}(t')\, \delta\left(\tau^{m}(t') - t\right) dt$$

(integrating the impulse functions)

$$= \sum_{\text{all paths } m} \sum_{n=-\infty}^{+\infty} s[n] \operatorname{sinc}\left\{\pi W\left(t' - \tau^{m}(t') - \frac{n}{W}\right)\right\} a^{m}(t')$$

II-2.[15].

Hence, when we sample v(t'), we obtain

$$v[k] = \sum_{\text{all paths } m} \sum_{n=-\infty}^{+\infty} s[n] \operatorname{sinc}\left\{\pi W\left(\frac{k}{W} - \tau^{m}\left(\frac{k}{W}\right) - \frac{n}{W}\right)\right\} a^{m}\left(\frac{k}{W}\right)$$

II-2.[16].

We may therefore write

$$v[k] = \sum_{\text{all paths } m} \sum_{n=-\infty}^{\infty} s[k-n]\, g^{m}[k, n]$$

II-2.[17]

where

$$g^m[k, n] = a^m\left(\frac{k}{W}\right) \operatorname{sinc}\left\{\pi W\left(\frac{n}{W} - \tau^m\left(\frac{k}{W}\right)\right)\right\} \qquad \text{II-2.[18].}$$

Expression II-2.[17] is a discrete-time model for the channel's response, with $g^m[k, n]$ being a time-varying, discrete-time impulse response for the m^{th} path. By our initial assumption that W is large enough that v(t) may be fully reconstituted from v[k] through the Nyquist sampling theorem, II-2.[17] yields a complete characterization of the channel response to the given input. Figure II.4 shows the discrete-time system, which is equivalent to the continuous-time one represented in Figure II.1.

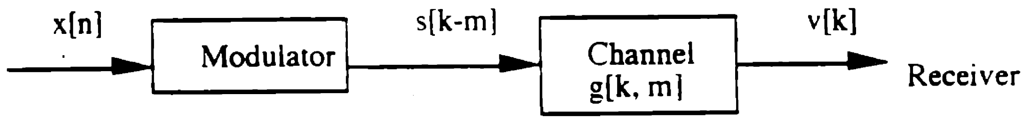


Figure II.4 : Discrete-time multipath system

For a causal channel, we may replace the $-\infty$ limit of summation in II-2.[17] by 0. Such causality agrees with our physical understanding of the channel. However, we cannot have strict causality and strict bandlimiting simultaneously. We shall address the issue of our simultaneous time and bandlimiting in section II-2-1-f.

<u>II-2-1-e- The presence of noise.</u>

Our channel model has only taken into account the linear time varying model with discrete multipath. We next take additive noise into account. Figure II.5 shows the AWGN continuous-time model. Define the baseband output, y(t'), of a multipath channel with AWGN to be the baseband output v(t') found in the section above with the addition of a term n(t'), where n(t') is the result of passing the sample value of a complex white Gaussian noise of double sided spectral density N_0 through a passband bandlimiting filter over [-W/2, W/2]. Although white Gaussian noise has a power spectrum which is non-zero over all possible frequencies, only the effect of the noise at frequencies for which v(t') is non-zero is of any interest. Indeed, noise components outside those frequencies are independent of the noise components in the bandwidth of interest. Therefore, we may bandlimit y(t') at baseband to [-W/2, W/2] without losing anything but noise which is independent of the quantities of interest. The power spectrum of N(t') is the rectangular function given by

$$S_N(f) = \begin{cases} N_0 \text{ for } |f| \le \frac{W}{2} \\ 0 \text{ otherwise} \end{cases}$$

II-2.[19]

and the autocorrelation function is

$$R_N(t) = N_0 \frac{\sin(\pi W t)}{\pi t}$$

II-2.[20].

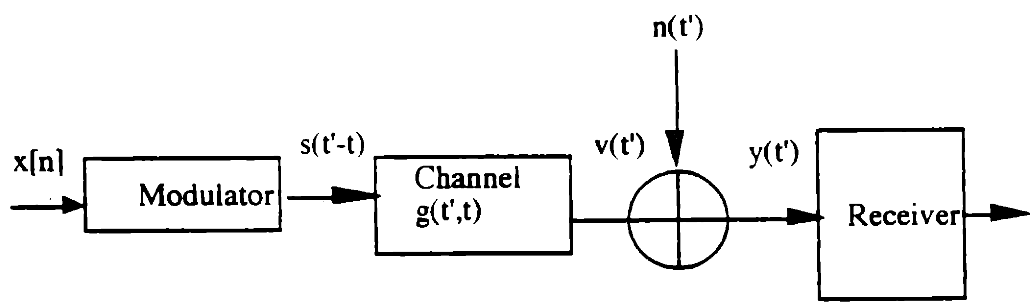


Figure II.5 : Continuous time model with AWGN.

We may sample y(t) at rate W and still recover y(t), by the Nyquist sampling theorem. For the discrete-time model, y[k] is v[k] + n[k]. The discrete quantity n[k] is defined as the sampled quantity of n(t) as in II-2.[12] and II-2.[13]. Figure II.6 shows the addition of noise with respect to Figure II.4.

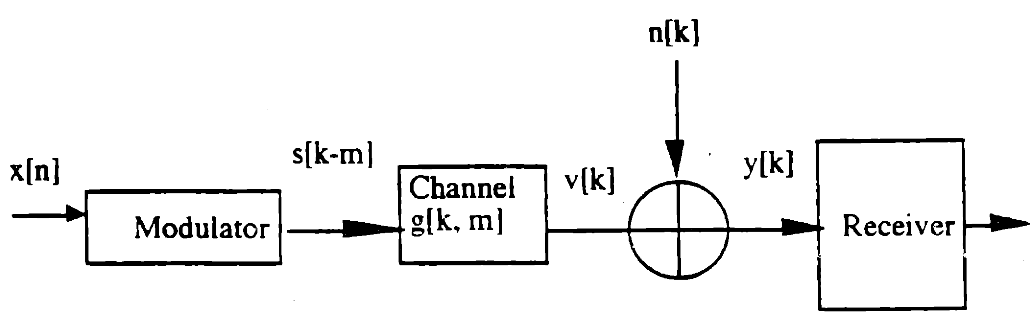


Figure II.6: Discrete time model with AWGN.

We know ([LM], pg. 46), or we may see immediately from II-2.[20], that the sampled process for the filtered complex white Gaussian process has an autocorrelation function given by

$$R_N[k] = N_0 W \delta_{0,k} \qquad \text{II-2.[21]}$$

were δ is a Kronecker delta function. Therefore, the variance of the sampled process is given by $R_N[0] = N_0W$. The real and complex component of each sample of the white noise each have variance $\sigma_N^2 = \frac{N_0 W}{2}$.

II-2-1-f- Matrix representation of the system.

In our previous discussion, the sampled version of the system requires an infinite number of samples unless we make certain assumptions, which we now discuss. Moreover, we are interested in channels that are causal, yet limited in bandwidth. We may note from expression II-2.[18] that there are infinitely many terms $g^m[k, n]$, whereas in the continuous case there is, for any given t', at most a single t which yields a non-zero value for $g^m(t', t)$. For our purposes, we wish to establish an expression for the output of the channel which is a product of finite matrices in the discrete-time case. Expression II-2.[18] shows that the terms $g^m[k, n]$ become very small with n for k fixed. Figure II.7 below shows the fact that $g^m[k, n]$ may be neglected for large n.

We may note that we have used the sinc function for our sampling reconstruction, but we could instead use a function with significantly faster roll-off in time to reconstitute the continuous time system from the sampled system. Since we must always consider that a signal starts at some time, we cannot have true bandwidth limitation. We filter through a causal filter such that a small percentage of the signal energy is outside the bandwidth of the filter. Such an assumption of approximate simultaneous band limiting and time limiting in order to obtain a matrix representation of a time-varying system is similar to that found in [Spi65]. In effect, if we choose appropriate functions for sample reconstruction of the signal, we lose an arbitrarily small percentage of the signal energy by assuming that the signal is simultaneously time and frequency limited. A more detailed discussion is found in [Spi65].

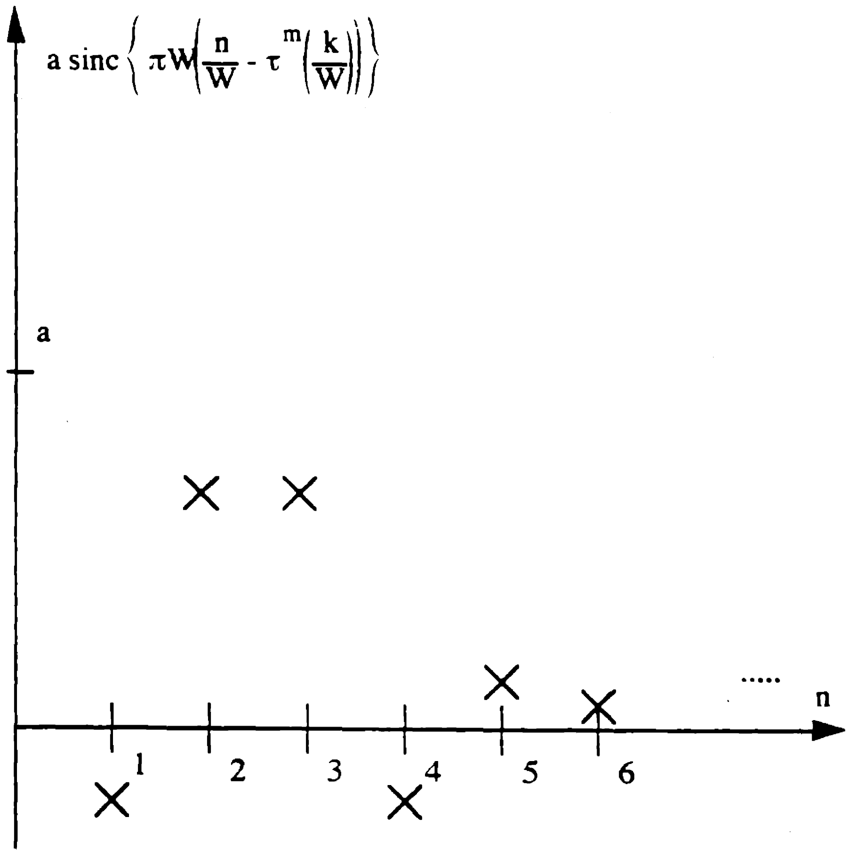


Figure II.7 : Behavior of $g^m[k, n]$ with n for the case $t^m = 5/(2W)$.

Let us consider what happens when we send a single pulse. If we arbitrarily choose the delay associated with the first path to be 0, then we see that the received signal will be vanishingly small after WT_{spread} time samples. Let us now consider that we are sending a continuous stream of data and determine which inputs affect y[k]. The input signal samples s[j] for $j < k\text{-}WT_{spread}$ have a negligible effect on y[k]. Therefore, we may choose some artificial cut-off in n such that $g^m[k,n]$ is approximated to be zero beyond that point. Therefore, we may approximate the sampled channel output as :

$$y[k] = \left[g[k,\Delta] \; \ldots \; g[k,0]\right] \begin{bmatrix} s[k-\Delta] \\ \ldots \\ s[k] \end{bmatrix} + n[k]$$

II-2.[22]

where Δ is some integer satisfying $\Delta > WT_{spread}$ and

$$g[k,i] = \sum_{\text{all paths } m} g^{m}[k,i]$$

II-2.[23].

If we wish to examine a string of outputs, say $\underline{y}_k$, where $\underline{y}_k$ is $[y[1], \ldots y[k]]^T$, we may use II-2.[19] to write the vector expression we seek. In the sequel, a subscript k after a vector ν indicates that we are considering $[\nu[1], \ldots \nu[k]]^T$, a pair of subscripts j,k where j<k indicates that we are considering $[\nu[j], \ldots \nu[k]]^T$. If we are considering a matrix, the superscript will indicate the range of the columns in the same manner. If we indicate the range of the columns but not that of the rows, it means that we are considering the whole range of rows for a square matrix. Therefore, $\underline{\nu}_k$ is a shortened notation for $\underline{\nu}^{1,1}_{1,k}$ and $\underline{\nu}^k$ is a shortened notation for $\underline{\nu}^{1,k}_{1,k}$.

We assume that s[n] for any n≤0 is zero. This implies that y[1], ..., y[k] will depend only on inputs s[1], ..., s[k]. Let us also assume, as before, that we have chosen $\Delta > WT_{spread}$ such that we take $g[k,n] = 0$ for $n > \Delta$. In that case, we may write that

$$\underline{y}_k = \underline{f}^k \underline{s}_k + \underline{n}_k$$

II-2.[24]

where $\underline{f}^k$ is the complex matrix with entries

$$\begin{cases} \underline{f}^k[j,i] = g[j,j-i] \text{ for } 0 \le j-i \le \Delta \\ 0 \text{ otherwise} \end{cases}$$

II-2.[25].

We have therefore a vector expression for the output over k unit intervals, each of length 1/W. Although the vector expression is approximate, we know that we may make this approximation arbitrarily good by the choice of Δ. We shall use this expression later when trying to compute the capacity of the system. Throughout this section, we have considered all variables as being known. We shall first relax the assumption that the input and noise are known and thus find the mutual information when the multiplicative part of the channel is known. Next, we shall relax the assumption that the channel is known in order to find mutual information results when there is uncertainty about the channel at the receiver.

II-2-2- Comparison to statistical multipath models.

In light of our model of section II-1, we may now discuss some of the existing models for multipath. As we have seen, we may describe the channel with a tapped delay line and therefore the problem of characterizing the channel reduces to the problem of determining the taps. In general, we may be able to sample at a Nyquist rate such that the effect of several paths which are close together in delay is seen only as an aggregate effect. Indeed, $W_{input} > W_{Doppler}$ because of our channel dynamics. In a narrowband model, i.e. in a case where 1/W is larger than T_{spread} or at least than the spread between a subset of paths, the effect of several paths will be combined at a single tap. If two paths are separated in time delays by more than 1/W, their primary effects appear at different taps and are therefore said to be resolved. If we are transmitting over a band which is very narrow with respect to the reciprocal of the multipath spread, then we may even describe the whole channel by a single tap which captures the effect of all the physical paths.

To illustrate the issue of resolvability, let us consider the case where we have three paths, two of which are close together in delay with respect to 1/W and a third whose delay is much greater than 1/W with respect to the first two. Figure II.8 below shows the impulse response of a channel and the received signal (at baseband) when an impulse bandlimited to W is sent. Part c of Figure II.8 shows that the effect of the two physical paths that are close together in part a may be fully described at one tap by an aggregate effect. Therefore, when we have a signal which is fairly narrowband with respect to the inverse of the delay spread, we need to consider aggregate multipath effects at the taps. When several paths affect a single tap, it is reasonable to expect that we may provide some probabilistic description of the behavior of that tap. Such is the basis for the common statistical multipath models which we discuss in this section. We first examine models which deal with looking at the distribution of a single tap and then consider the issue of observing several taps.

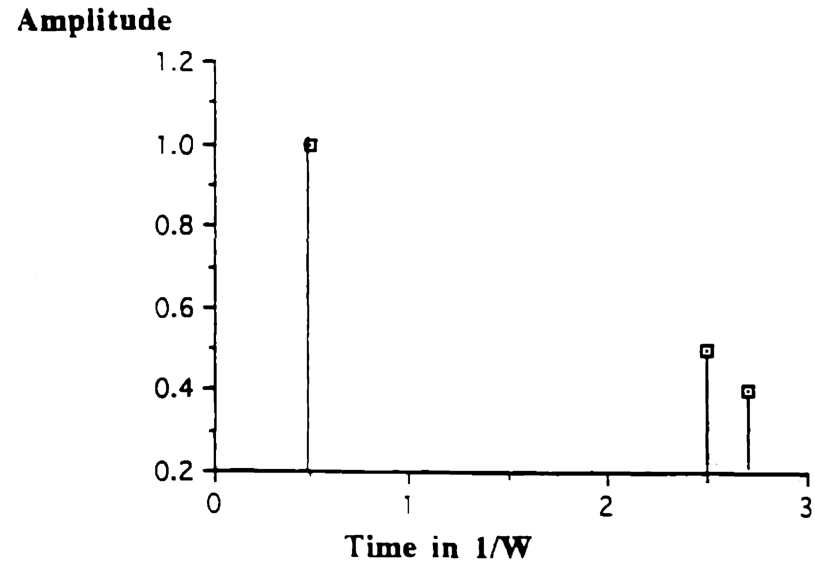


Figure II.8.a : Impulse response of the channel.

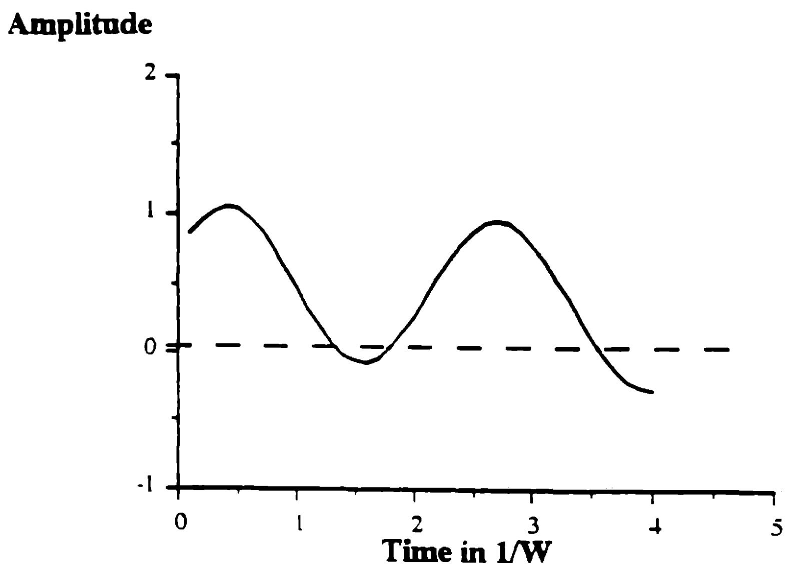


Figure II.8.b : Response of the bandlimited channel.

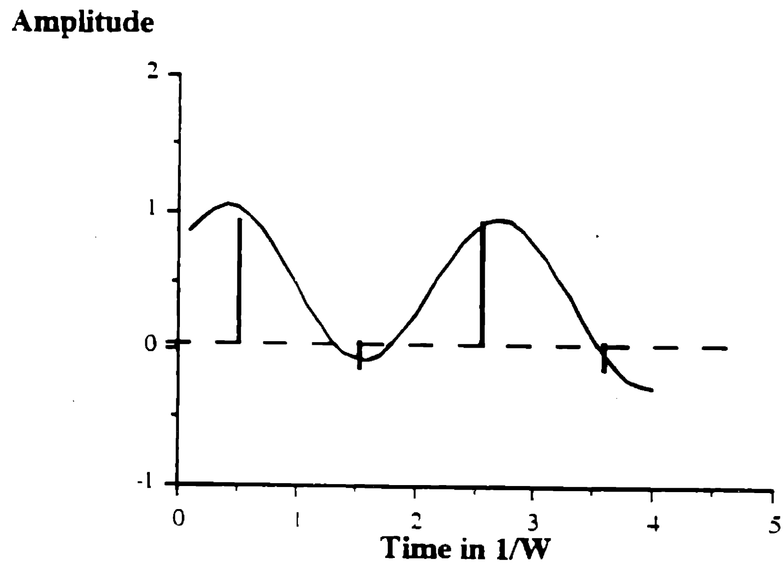


Figure II.8.c : Sampling of the bandlimited response of the channel.

II-2-2-a- Statistical multipath models for a single tap

Since each tap may be determined by the aggregate effect of several paths, it is reasonable to assume that it would not be useful to track the behavior of each path present at a tap and deduce the behavior of the tap. Therefore, statistical models for the behavior of a single tap are derived and commonly used. In light of our discussion, all the models we present below are valid for a single tap at which many paths are unresolved, i.e. $W << 1/T_{spread}$. The most common statistical models for such multipath channels are Rayleigh fading, Rician fading and Nakagami fading.

In the literature, the most common way of modeling fast fades in a mobile radio environment is to use Rayleigh fading. While we refer to the fades due to the multipath as fast fades, these fades are still slower than the rate of change of the signal. They are fast with respect to fading linked to the general terrain and atmospheric conditions, generally described by log-normal distributions. The Rayleigh fading model is not usually expressly specified to be for aggregate multipath as opposed to several distinguishable paths from the mobile transmitter to the receiver. However, in light of our previous discussion, we know that we are indeed dealing with an aggregate multipath model, i.e. a narrowband model where the paths are not resolved. The Rayleigh fading model may be attained in different ways. One way of deriving it, which presents the most physical justification, is to consider that the impulse response is the superposition of a large number of complex paths. If we have a sufficient number of paths. we may use the Central Limit Theorem (CLT). We then find that the aggregate multipath can be approximated as a random variable which has Gaussian distributed IID real and imaginary parts. The distribution of the envelope of the aggregate is then Rayleigh distributed while the phase is uniformly distributed between 0 and 2π. The Rician model is that of a single specular path, for instance a line of sight path, overlaid over a Rayleigh fading channel.

More generally, the strength of a tap for the aggregate multipath case may be modeled using Nakagami distributions ([Cha79], [Nak58]). The Nakagami distribution has often been put forward as a better approximation to the fading encountered in wireless communications than Rayleigh or Rician fading. Rayleigh is a special case of Nakagami. Therefore, given more freedom in choosing the distribution, it is not surprising that Nakagami does better. This, however, does not explain whether there is a meaningful physical link between the Nakagami distribution and the fading seen in wireless communications. Referring to the original work by Nakagami [Nak58], we see that the Nakagami distribution is derived as a distribution for aggregate multipath fading. The distribution is valid when there are many paths and when certain approximations on the coefficients of an orthonormal expansion for the amplitude of the impulse response hold. Under such conditions, it is clear why Rayleigh is a special case of Nakagami, since Rayleigh was derived using the CLT for multipath off a large number of reflectors such that the phases of the reflections off these reflectors are uniformly distributed. However, there is no physical explanation for why, if ever, the approximations taken by Nakagami hold. While the Rayleigh fading model may be derived from the CLT for a given multipath model, the Nakagami model is really a heuristic. The parameter, m, of the Nakagami distribution depends only on the first and second moments and the cross-correlation of the real and imaginary parts of the multipath expressed as a sum of amplitudes multiplying complex exponentials. The Rayleigh distribution is obtained when the parameter m is set to 1. The assertion sometimes made that m represents the number of paths forming the aggregate multipath is, therefore, not correct, although m will usually increase with the number of paths. We have more degrees of freedom in fitting the distribution to the data when we use Nakagami rather than simply Rayleigh distributions, and the extra degree of freedom is given by the parameter m. The Nakagami distribution is sometimes claimed to be useful when we are modeling a large number of paths which are not individually known. It has been reported ([Cha79]) that replacing the Rayleigh model at the receiver by a better fitting Nakagami distribution leads commonly to differences of 3 dB in the performance evaluations of systems if the error probability versus SNR is being evaluated. However, as m is increased, it would appear that the effect of changing the parameter of the Nakagami distribution decreases sharply [MG92]. The complexity of the model, however, is greatly increased, so that Nakagami distributions are rarely used.

II-2-2-b- Statistical multipath models for several taps.

None of the models considered above describe the behavior in time of the channel. The distribution of any tap may be modeled as having a Rayleigh, Rician or Nakagami distribution, but this does not indicate how this tap is correlated with the rest. If the channel varies very slowly with respect to the signaling rate, then two consecutive samples of a tap will be strongly correlated. If, on the contrary, the channel varies at a rate comparable to the signaling rate, two consecutive samples of a tap may be almost uncorrelated. In general, it is of interest to know the correlation among these samples of a tap. For instance, we probably would want to know whether the presence of a deep fade at one time sample predicts a deep fade at the next time sample. Clearly, in computing capacities for channels, such an issue is important.

A statistical model of the channel which takes into account the correlation of the channel in time and frequency is the scattering function ([Gal64], [Ken]). The scattering function s has as parameters both time, t, and frequency, f. The scattering function represents the average normalized amplitude of the component of the channel which has delay t and Doppler shift f. Such a component may be created by reflections off single large objects or off a cluster of small reflectors. Usually, in mobile communications, they will be off large objects consisting of a cluster of small reflectors, for instance a rough wall. Such a multitude of reflections might be difficult to characterize, but the overall effect can be described in part by the scattering function. The scattering function provides us with a measure of spread ([Ken]) of the channel in time and frequency which is lacking in single tap models. In particular, we may write that the average Doppler spread is given by

$$\left[\int \left(\int \sigma(t, f)\, dt\right)^2 df\right]^{-1}$$

and similarly the average time spread may be given by

$$\left[\int \left(\int \sigma(t, f)\, df\right)^2 dt\right]^{-1}.$$

The average total spread may be expressed as

$$\left[\int \int (\sigma(t, f))^2\, df\, dt\right]^{-1}.$$

Such a model gives us some information about how the channel evolves in time and frequency and does not rely on narrowband assumptions for its validity. It gives a description of the second order statistics of the channel. If we are dealing with a cloud of scatterers where we cannot hope to distinguish dominant paths, then it is as good a description as we may wish for. Indeed, the channel then has, by the CLT, a Gaussian distribution, and therefore second order statistics describe it fully. The capacity of a channel described by its scattering function is derived in [Gal], pp. 431-438, for the case of infinite bandwidth. However, second order statistics are not sufficient if we are to consider that we measure a multipath channel with good accuracy. The scatterer model is not consistent with a multipath channel where, say, dominant paths can be resolved, their relative delays are changing, they are interfering constructively and destructively, etc... These conditions violate the WSSUS assumptions of a $\sigma(t, f)$ model. For our purposes, we shall be more interested in measuring the channel than in characterizing it by a scattering function or more complex model.

II-2-3- Capacity in the case of a perfectly known channel.

II-2-3-a- Constant single path channel.

For the sake of a simple illustration, let us consider the case of a single path, where a^m is a constant a for all time and where the associated delay $\tau^m(t)$ is always 0. We know that such a flat fading model is a very poor approximation for real multipath models ([GB81]) but we wish first to establish some extremely simple framework which will be used in the time-varying channel. For this constant single path case, the channel does not change in time, and therefore the output bandwidth and the input bandwidth are the same. Let us constrain the input to power *P*, as made explicit later. This is the classic bandlimited AWGN channel. Our motivation in deriving this well-known example is to establish notation and to present a structure which will be extended in the next section for known specular paths. For complex transmission over time T using an input bandwidth W, the mutual information between input and output is given by:

$$I(\underline{Y}_k ; \underline{S}_k) = h(\underline{Y}_k) - h(\underline{N}_k)$$

II-2.[26]

where $k = \lfloor TW \rfloor$ and h denotes differential entropy, since we are dealing with pdfs. We actually have 2k degrees of freedom since we are dealing with complex random variables. In order to simplify our simplify our manipulations, we shall use the random vectors $\underline{S}'_{2k}$, $\underline{Y}'_{2k}$ and $\underline{N}'_{2k}$, whose first k components and last k componenets are the real and complex parts, of the corresponding vectors $\underline{S}_{2k}$, $\underline{Y}_{2k}$ and $\underline{N}_{2k}$.

For a given covariance matrix for the output, we know that $h(\underline{Y}'_{2k})$ is maximized when $\underline{Y}'_{2k}$ is Gaussian. Hence, we are interested in determining the covariance matrix of $\underline{Y}'_{2k}$ and then we shall attempt to construct an $\underline{S}'_{2k}$ such that $\underline{Y}'_{2k}$ is Gaussian. The bandlimited noise has zero mean at all times and is uncorrelated with the input. The output of the channel may be expressed as

$$\underline{Y}'_{2k}[i] = a\,\underline{S}'_{2k}[i] + \underline{N}'_{2k}[i]$$

II-2.[27].

The covariance matrix of $\underline{Y}'_{2k}$ has entries

$$\Lambda_{\underline{Y}'_{2k}}[i, j] = a^2 E_{S'}\left[S'[i]S'[j]\right] + E_{N'}\left[N'[i]N'[j]\right]$$

$$= a^2 E_{S'}\left[S'[i]S'[j]\right] + \frac{N_0}{2} W\,\delta[i-j]$$

II-2.[28].

It is well known that if we choose S(t) to have a Gaussian distribution, then the output of the channel will have a Gaussian distribution with covariance matrix given by II-

2.[28]. The entropy of $\underline{Y'}_{2k}$ is given by $k \ln\left(2\pi\left|\Lambda_{\underline{Y'}_{2k}}\right|\right)$ where $|\,|$ denotes absolute value of the determinant. By Hadamard's inequality, the above entropy will be maximized if we take the off-diagonal terms of the covariance matrix to be 0. We may indeed construct a Gaussian zero-mean signal S which is bandlimited to [-W/2, +W/2] and has the desired property that

$$E_S[S[i]S[j]] = E_S[S[k+i]S[k+j]] = WP\ \delta[i-j]$$

II-2.[29]

for all i and j smaller than k.

Indeed, our previous discussion concerning the noise shows that S could be constructed by taking complex white Gaussian noise with spectral density P and bandlimiting it to [-W/2, +W/2]. This construction is possible because $W_{input} = W$. The signal power per degree of freedom is a^2PW. The noise power per degree of freedom is $\frac{N_0}{2}$ Therefore

$$I(\underline{Y}_k; \underline{S}_k) = I(\underline{Y'}_{2k}; \underline{S'}_{2k}) = \lfloor TW \rfloor \ln\left(\frac{\left(a^2P + WN_0\right)}{WN_0}\right)$$

$$\approx TW \ln\left(1 + \frac{a^2P}{W N_0}\right)$$

II-2.[30]

where the approximation becomes exact as $WT \rightarrow \infty$. The expression in II-2.[30] is the well-known expression for the capacity of the bandlimited AWGN channel and is convex in W. The benefit of having extra bandwidth is therefore close to linear in W when W is small. II-2.[30] approaches a limit as W goes to infinity, and therefore the benefit of extra bandwidth is negligible for W large. Extra bandwidth cannot be detrimental, since we could always transmit over less bandwidth than the available bandwidth and thus come back to the case where we have less bandwidth.

Since we have a memoryless channel with additive Gaussian noise, we know ([Gal], pg. 337) that II-2.[30] gives the capacity for the channel.

II-2-3-b- Known specular paths.

Let us extend the simple flat fading case to the more realistic case, for terrestrial mobile communications, of several known time-varying specular paths. We assume that both the sender and the receiver know the channel. We have total bandwidth [-W/2, W/2], where W is given by the sum of the input bandwidth, W_{input}, and the Doppler spread, $B_{Doppler}$. The time we consider is the time over which we signal, T, plus Δ, which takes into account the multipath spread, T_{spread}, and the decay of the sampling function. In effect, we consider that we transmit a nil signal outside of W_{input} and that we transmit an almost nil signal before time 0 and after time T, within approximations to allow for the fact that time and bandwidth limited signals are not achievable. Let us denote by k the product of the total bandwidth W and the total time T+Δ that we consider. The output can be derived from expressions II-2.[22]:

$$\underline{Y}[i] = \sum_{n=0}^{\Delta} \sum_{\text{all paths } m} S[i-n]\, a^m\left(\frac{i}{W}\right) \operatorname{sinc}\left(\pi W\left(\frac{n}{W} - \tau^m\left(\frac{i}{W}\right)\right)\right) + N[i]$$

II-2.[31].

Our vectors are complex, since we are at baseband. The variables a^m are complex with phase changing at a rate $2\pi B^m$, where B^m is the Doppler shift for path m.

Let us again define the real vectors $\underline{S'}_{2k}$, $\underline{N'}_{2k}$ and $\underline{Y'}_{2k}$. Performing a simple maximization using the determinant of the correlation matrix of the output $\underline{Y'}_{2k}$ as in II-2-3-a is not straightforward. We therefore choose another approach. We decompose the channel into orthonormal components and apply water-filling techniques to these channel components to obtain capacity. The constraints we have are that the input is bandlimited to [$-W_{input}/2$, $W_{input}/2$] and that the input variance is limited to *P*. However, we cannot have exactly time and band limited signals. In our previous discussion, we have considered that the input was time limited. In capacity arguments, we allow T to become arbitrarily large, so that the time limitation constraint is relaxed.

For the model in II-2.[31], the output sample in time is given by the vector equation

$$[Y'[1] \ldots Y'[2k]]^T = \underline{F}^{2k}[S'[1] \ldots S'[2k]]^T + [N'[1] \ldots N'[2k]]^T$$

II-2.[32]

where $\underline{f}'^{2k}$ is almost zero for all entries except the entries $\underline{f}'[i, j]$ such that $\Delta > i\text{-}j \geq 0$. in which case

$$\underline{f}[i, j] = \begin{cases} \sum_{\text{all paths } m} \mathrm{Re}\left(a^m\left(\frac{i}{W}\right)\right) \mathrm{sinc}\left(\pi W\left(\frac{i-j}{W} - \tau^m\left(\frac{i}{W}\right)\right)\right) \\ \qquad \text{for } \Delta \geq i-j \geq 0,\ i \leq k,\ j \leq k \\ \sum_{\text{all paths } m} -\mathrm{Im}\left(a^m\left(\frac{i}{W}\right)\right) \mathrm{sinc}\left(\pi W\left(\frac{i-j}{W} - \tau^m\left(\frac{i}{W}\right)\right)\right) \\ \qquad \text{for } \Delta \geq i-(j-k) \geq 0,\ i \leq k,\ j > k \\ \sum_{\text{all paths } m} \mathrm{Im}\left(a^m\left(\frac{i}{W}\right)\right) \mathrm{sinc}\left(\pi W\left(\frac{i-j}{W} - \tau^m\left(\frac{i}{W}\right)\right)\right) \\ \qquad \text{for } \Delta \geq (i-k)-j \geq 0,\ i > k,\ j \leq k \\ \sum_{\text{all paths } m} \mathrm{Re}\left(a^m\left(\frac{i}{W}\right)\right) \mathrm{sinc}\left(\pi W\left(\frac{i-j}{W} - \tau^m\left(\frac{i}{W}\right)\right)\right) \\ \qquad \text{for } \Delta \geq (i-k)-(j-k) \geq 0,\ i > k,\ j > k \end{cases}$$

II-2.[33]

from II-2.[18], II-2.[23] and II-2.[25].

Therefore, the effect of the multipath spread is seen in the matrix $\underline{f}'^{2k}$. Complex inputs which are spaced by roughly more than the multipath spread do not affect the same outputs. This is altogether consistent with our physical model.

We have already discussed the fact that, in most situations, the effect of the Doppler spread in widening the bandwidth of the received signal with respect to the bandwidth of the signal sent is negligible. For large enough bands of spectrum, the effect of the Doppler spread is comparable to that of the roll off at the edges of the bandlimiting filter at the receiver. However, we consider here the general case where the effect of the Doppler spread in changing W_{input} to W may not be negligible. We may still describe the effect of the channel by a linear transformation on the sampled input, even if the input is sampled at W_{input} rather than at W. Let us denote by $\tilde{S}[k']$ the sample at time T/W_{input} of the continuous input S(t). We know from the Nyquist sampling theorem that we may write

$$S(t) = \sum_{n=-\infty}^{-\infty} \tilde{S}[n] \text{sinc}\left\{ \pi W_{input}\left(t - \frac{n}{W_{input}} \right) \right\}$$

II-2.[34].

Hence, by substitution of II-2.[34] into II-2.[10], the output of the channel before noise may be written as

$$V(t') = \sum_{\text{all paths } m} \int_{t=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \tilde{S}[n] \text{sinc}\left\{ \pi W_{input}\left(t' - t - \frac{n}{W_{input}} \right) \right\} a^m(t')\, \delta\left(\tau^m(t') - t\right) dt$$

(integrating the impulse functions)

$$= \sum_{\text{all paths } m} \sum_{n=-\infty}^{+\infty} \tilde{S}[n] \text{sinc}\left\{ \pi W_{input}\left(t' - \tau^m(t') - \frac{n}{W_{input}} \right) \right\} a^m(t')$$

II-2.[35].

Hence, when we sample Y(t') = V(t') + N(t'), we obtain

$$Y[i] = \sum_{\text{all paths } m} \sum_{n=-\infty}^{+\infty} \tilde{S}[n] \text{sinc}\left\{ \pi W_{input}\left(\frac{i}{W} - \tau^m\left(\frac{i}{W}\right) - \frac{n}{W_{input}} \right) \right\} a^m\left(\frac{i}{W}\right) + N[i]$$

II-2.[36].

We may therefore write

$$Y[i] = \sum_{\text{all paths } m} \sum_{n=-\infty}^{\infty} \tilde{S}[i-n]\, \tilde{g}^m[i, n] + N[i]$$

II-2.[37]

where

$$\tilde{g}^{m}[i, n] = a^{m}\left(\frac{i}{W}\right) \mathrm{sinc}\left\{\pi W_{input}\left(\frac{i}{W} - \tau^{m}\left(\frac{i}{W}\right) - \frac{i-n}{W_{input}}\right)\right\}$$

II-2.[38].

Therefore, making assumptions of approximate simultaneous time and bandwidth limiting, if we take S' to be 0 before sample 1 and we choose a $\tilde{\Delta}$ large enough, we may replace matrix expression II-2.[32] by

$$[Y[1] \ldots Y[k]]^T = \underline{\tilde{f}}^{k}_{k'} [\tilde{S}[1] \ldots \tilde{S}[k']]^T + [N[1] \ldots N[k]]^T$$

II-2.[39]

where

$$k' = \lfloor W_{input} T \rfloor \text{ and } k = \lfloor W(T + \tilde{\Delta}) \rfloor$$

II-2.[40]

and, as in II-2.[25],

$$\begin{Bmatrix} \tilde{f}[i,j] = \tilde{g}[i, i-j] \text{ for } 0 \leq i-j \leq \tilde{\Delta} \\ 0 \text{ otherwise} \end{Bmatrix}$$

II-2.[41].

The matrix $\underline{\tilde{f}}^{k}_{k'}$ gives, in terms of a linear transformation, both the effect of sampling input and output at different rates and the effect of the channel. Let us define $\underline{\tilde{S}'}_{2k'}$ from $\underline{\tilde{S}}_{k'}$ similarly to the way we defined $\underline{S'}_{2k}$ from $\underline{S}_{k}$. The expression equivalent to II-2.[39] is

$$\underline{Y'}_{2k} = \underline{\tilde{f}'}^{2k}_{2k'} \underline{\tilde{S}'}_{2k'} + \underline{N'}_{2k}$$

II-2.[42]

where

$$\tilde{f}[i,j]=\begin{cases}\sum_{\text{all paths } m} \operatorname{Re}\left(a^m\left(\frac{i}{W}\right)\right)\operatorname{sinc}\left(\pi W_{input}\left(\frac{i}{W}-\frac{j}{W_{input}}-\tau^m\left(\frac{i}{W}\right)\right)\right) & \text{for } \Delta \geq i-j \geq 0,\ i \leq k,\ j \leq k' \\ \sum_{\text{all paths } m} -\operatorname{Im}\left(a^m\left(\frac{i}{W}\right)\right)\operatorname{sinc}\left(\pi W_{input}\left(\frac{i}{W}-\frac{j}{W_{input}}-\tau^m\left(\frac{i}{W}\right)\right)\right) & \text{for } \Delta \geq i-(j-k') \geq 0,\ i \leq k,\ j > k' \\ \sum_{\text{all paths } m} \operatorname{Im}\left(a^m\left(\frac{i}{W}\right)\right)\operatorname{sinc}\left(\pi W_{input}\left(\frac{i}{W}-\frac{j}{W_{input}}-\tau^m\left(\frac{i}{W}\right)\right)\right) & \text{for } \Delta \geq (i-k)-j \geq 0,\ i > k,\ j \leq k' \\ \sum_{\text{all paths } m} \operatorname{Re}\left(a^m\left(\frac{i}{W}\right)\right)\operatorname{sinc}\left(\pi W_{input}\left(\frac{i}{W}-\frac{j}{W_{input}}-\tau^m\left(\frac{i}{W}\right)\right)\right) & \text{for } \Delta \geq (i-k)-(j-k') \geq 0,\ i > k,\ j > k' \end{cases}$$

II-2.[43].

Let us consider the 2k' x 2k' matrix $\tilde{\underline{f}}_{2k'}^{2k\,T}\tilde{\underline{f}}_{2k'}^{2k}$. Let λ_1, ..., $\lambda_{2k'}$ be the eigenvalues of $\tilde{\underline{f}}_{2k'}^{2k\,T}\tilde{\underline{f}}_{2k'}^{2k}$. These eigenvalues are real ([Str], pg. 222) and non-negative. From [Gal], Theorem 8.4.1, there exist eigenvectors $\underline{\varphi}_{i\,2k'}$ for each λ_i such that:

$$\tilde{\underline{f}}_{2k'}^{2k\,T}\tilde{\underline{f}}_{2k'}^{2k}\underline{\varphi}_{i\,2k'}=\lambda_i\,\underline{\varphi}_{i\,2k'}$$

II-2.[44]

and the vectors $\underline{\varphi}_{i\,2k'}$ are orthonormal. Moreover, there exist 2k' orthonormal vectors $\underline{\theta}_{i\,2k}$ such that:

$$\sqrt{\lambda_i}\,\underline{\theta}_{i\,2k}=\tilde{\underline{f}}_{2k'}^{2k}\underline{\varphi}_{i\,2k'}$$

II-2.[45].

Let us express the input random vector $\underline{\tilde{S}'}_{2k'}$ as a linear combination of the vectors $\underline{\varphi}_{i\,2k'}$:

$$\underline{\tilde{S}'}_{2k'} = \sum_{i=1}^{2k'} U_i\, \underline{\varphi}_{i\,2k'}$$

II-2.[46]

where the coefficients U_i are real random variables.

Let us create an orthonormal basis for vectors of dimension 2k such that the first 2k' elements of the basis are the vectors $\underline{\theta}_{i\,2k}$. We shall denote the other 2k-2k' elements as $\underline{\theta}_{i\,2k}$ also. We may then express the noise as

$$\underline{N'}_{2k} = \sum_{i=1}^{2k} \nu_i\, \underline{\theta}_{i\,2k}$$

II-2.[47]

where the random coefficients ν_i are IID complex zero-mean Gaussian random variables with mean $WN_0/2$. The noiseless output may always be expressed only in terms of the first 2k' vectors $\underline{\theta}_{i\,2k}$:

$$\underline{\tilde{\Gamma}}^{2k}_{2k'} \underline{\tilde{S}'}_{2k'} = \sum_{i=1}^{2k'} U_i \sqrt{\lambda_i}\, \underline{\theta}_{i\,2k}$$

II-2.[48].

We have therefore decomposed the channel into 2k parallel independent channels where, along the last 2k-2k' channels, nothing is transmitted and only noise is received. We may perform our maximization along the first 2k' channels subject to the constraint

$$\mathrm{tr}\left(E\left[\underline{\tilde{S}'}_{2k'} \underline{\tilde{S}'}_{2k'}{}^{T}\right]\right) \leq T\,P\,W$$

II-2.[49].

Since we have used an orthonormal basis to decompose $\underline{\tilde{S}'}_{2k'}$, II-2.[49] is equivalent to

$$\sum_{i=1}^{2k'} E\left[U_i^2\right] \leq TPW$$

II-2.[50].

We are carry out water-filling arguments similar to [CT] on these eigenvectors under our average energy constraints. To maximize capacity, we choose U_i and γ such that

$$E\left[U_i^2\right] = \left(\gamma - \frac{WN_0}{2\lambda_i}\right)^+$$

II-2.[51]

$$\sum_{i=1}^{2k'} \left(\gamma - \frac{WN_0}{2\lambda_i}\right)^+ = WTP$$

II-2.[52].

The capacity is $\frac{1}{2T}\sum_{i=1}^{2k'} \ln\left(1 + \frac{\lambda_i\left(\gamma - \frac{WN_0}{2\lambda_i}\right)^+}{\frac{WN_0}{2}}\right) = \frac{1}{2T}\sum_{i=1}^{2k'} \ln\left(1 + \frac{E\left[U_i^2\right]\lambda_i}{\frac{WN_0}{2}}\right)$ in terms of achievable rate per second.We do not need to calculate first derivative results to see that the mutual information is increasing, maybe not strictly, in W_{input}, since we could always transmit over less than the total available bandwidth. However, the incremental benefit of having more bandwidth may or may not be decreasing. Indeed, if we transmit over a portion of the channel which is severely faded because of frequency selective fading, and add an available portion of bandwidth with small fading, then the incremental benefit from that extra portion of bandwidth may be greater than the incremental benefit from a faded portion of bandwidth. In terms of our eigenvectors, adding a channel with very small associated eigenvalue may actually not even increase capacity at all. We may note that, given our construction, we need not make use of any coding theorem other than the memoryless channel coding theorem, although theorems for channels with limited memory exist. We may also note that our results do not give a compact expression for capacity when $T \rightarrow \infty$.

II-3- The effect of an unknown channel.

II-3-1- The effect upon maximum achievable mutual information of an unknown channel at the sender and an unknown or partially unknown channel at the receiver

In part II-2, we did not establish how we would determine the channel at both the receiver and the sender. In this section, we consider that the sender does not know the channel and the receiver does not know the channel, or has an imperfect estimate of the channel. We shall use the notation of II-2.[31] and II-2.[33] unless otherwise specified Such is the situation if, for instance, we are transmitting without feedback from the receiver to the sender. The sender has no way of knowing the channel. The receiver must either estimate the channel from the signal that is sent or from some separate sounding signal. If the channel is not perfectly known, it is reasonable to model it as a known part with a probabilistic additive component. Such a description of the channel is compatible with the common situation where the channel is measured and there is some zero-mean Gaussian noise of known variance in the measurement. We would also, obviously, not know *a priori* the signal s but know the distribution of the random variable S[n] of which s[n] is a sample value (we consider the discrete-time model for ease of exposition, since we have shown in II-2 that it may be freely interchanged with the continuous-time model). Similarly, we would know the distribution of $\underline{F}^{2k}$ of which $\underline{f}^{2k}$ is a sample value. The distribution of Y[k] would be given by the measurement and the measurement error distribution. Therefore, we would also know the distribution of Y[k].

In this section, we consider the effect on mutual information of the unknown channel at the receiver. The sender transmits as though the channel were an AWGN channel. For the mutual information to give some indication of the reliability of feasible communication schemes, we must establish an appropriate coding theorem. [VH94] gives a general formula for channel capacity for a single user. In appendix B, we establish a coding theorem for multiple access channels where the channel decorrelates in a manner which is defined in the appendix. The channel may have infinite memory in the outputs but we require the memory to decrease sufficiently rapidly. We also require certain ergodicity assumptions.

In the following, replacing II-2.[24] with random variables, we have

$$\underline{Y}_{2k} = \underline{F}^{2k}\underline{S}_{2k} + \underline{N}_{2k}$$ II-3.[1].

The mutual information between input and output is, by definition:

$$I(\underline{Y}_{2k}; \underline{S}_{2k}) = h(\underline{S}_{2k}) - h(\underline{S}_{2k} \mid \underline{Y}_{2k})$$ II-3.[2]

where h is the differential entropy function.

Similarly, the mutual information given perfect knowledge of the channel is

$$I\left(\underline{Y}_{2k}; \underline{S}_{2k} \mid \underline{F}^{2k}\right) = h\left(\underline{S}_{2k} \mid \underline{F}^{2k}\right) - h\left(\underline{S}_{2k} \mid \underline{F}^{2k}, \underline{Y}_{2k}\right)$$ II-3.[3].

Since $\underline{S}_{2k}$ and $\underline{F}^{2k}$ are independent.

$$h(\underline{S}_{2k}) = h\left(\underline{S}_{2k} \mid \underline{F}^{2k}\right)$$ II-3.[4].

Therefore,

$$I\left(\underline{Y}_{2k}; \underline{S}_{2k} \mid \underline{F}^{2k}\right) - I(\underline{Y}_{2k}; \underline{S}_{2k}) = h(\underline{S}_{2k} \mid \underline{Y}_{2k}) - h\left(\underline{S}_{2k} \mid \underline{F}^{2k}, \underline{Y}_{2k}\right)$$
$$= I\left(\underline{S}_{2k} ; \underline{F}^{2k} \mid \underline{Y}_{2k}\right)$$ II-3.[5]

or alternatively

$$I\left(\underline{Y}_{2k}; \underline{S}_{2k} \mid \underline{F}^{2k}\right) - I(\underline{Y}_{2k}; \underline{S}_{2k}) = I\left(\underline{F}^{2k} ; (\underline{S}_{2k}, \underline{Y}_{2k})\right) - I\left(\underline{F}^{2k}; \underline{Y}_{2k}\right)$$ II-3.[6]

$$I\left(\underline{Y}_{2k}; \underline{S}_{2k} \mid \underline{F}^{2k}\right) - I(\underline{Y}_{2k}; \underline{S}_{2k}) = h\left(\underline{F}^{2k} \mid \underline{Y}_{2k}\right) - h\left(\underline{F}^{2k} \mid \underline{S}_{2k}, \underline{Y}_{2k}\right)$$ II-3.[7].

The effect of not knowing the channel may be interpreted as the difference between the information about the channel that we obtain from both the output and the input, and the information about the channel that we obtain from the output only. Although the above expression is simple, computing it in any particular case may be difficult. However, we can make some general statements about $I\left(\underline{Y}_{2k}; \underline{S}_{2k} \mid \underline{F}^{2k}\right) - I\left(\underline{Y}_{2k}; \underline{S}_{2k}\right)$. We have the intuition that the difference in how well we can use the channel between the known and unknown channel cases depends on how well we can measure the channel. In other words, we suspect that $I\left(\underline{Y}_{2k}; \underline{S}_{2k} \mid \underline{F}^{2k}\right) - I\left(\underline{Y}_{2k}; \underline{S}_{2k}\right)$ will depend on how well the channel can be measured at the output solely from $\underline{Y}_k$. Indeed, in the limit where the channel can be measured perfectly, the difference $I\left(\underline{Y}_{2k}; \underline{S}_{2k} \mid \underline{F}^{2k}\right) - I\left(\underline{Y}_{2k}; \underline{S}_{2k}\right)$ is 0 since, from II-3.[5], $I\left(\underline{Y}_{2k}; \underline{S}_{2k} \mid \underline{F}^{2k}\right) - I\left(\underline{Y}_{2k}; \underline{S}_{2k}\right) = I\left(\underline{S}_{2k} ; \underline{F}^{2k} \mid \underline{Y}_{2k}\right)$.

Expression II-3.[5] has a further interpretation. We might in general question whether it matters that we do not know the channel, for we may perform equalization. Equalization uses the symbols already decoded to help decode the next symbols. However, equalization usually supposes that the past input can be perfectly detected from the output ([Qur85]). Such a supposition is generally thought to be reasonable, because if we did not decode the data properly anyway, we have a failure in the system. If the input is perfectly determined from the output, then the RHS of II-3.[5] is nil. Hence, if we claim that we may perform equalization perfectly, we are in essence already claiming that we do not suffer, in terms of mutual information, from not knowing the channel. Therefore, in order to perform equalization, we may be sending at a lower rate than the maximum achievable rate. Equalization assumes a slow variation of the channel so that we may measure the channel and use our measurements. In later sections, we shall see how the rate of variation of the channel relates to II-3.[6], thus giving some quantitative support to our intuitive understanding that equalization is reasonable under slowly changing channel conditions.

In the next section, we consider the case where the channel is measured at the receiver with some error of known variance. Such a model gives us bounds concerning the loss of mutual information when we have a particular error variance in the receiver. We may thus consider the fundamental detrimental effect of a measurement error without restricting ourselves to a specific measuring scheme. This allows us in later sections to consider the general loss in mutual information per symbol due to not knowing the channel at the receiver when the channel is a stochastic process with some stationarity assumptions.

II-3-2- Channel known with a small error of known variance.

II-3-2-a- Single symbol case.

Let us consider the case where the channel is known at the receiver with some mean square error. We consider for simplicity that the measurement has some known error power. We first look at the single real sample case to be able to view with ease the effect of the measurement error without having to contend with matrices. The purpose of looking at the single real sample case is to establish a framework for the more useful multiple sample case. The random variables we usually consider are complex. in keeping with our baseband model. However, in our single symbol model, we consider real random variables for ease of manipulation. When we go to the several symbol case, the case of complex random variables will be included.

We have the following model :

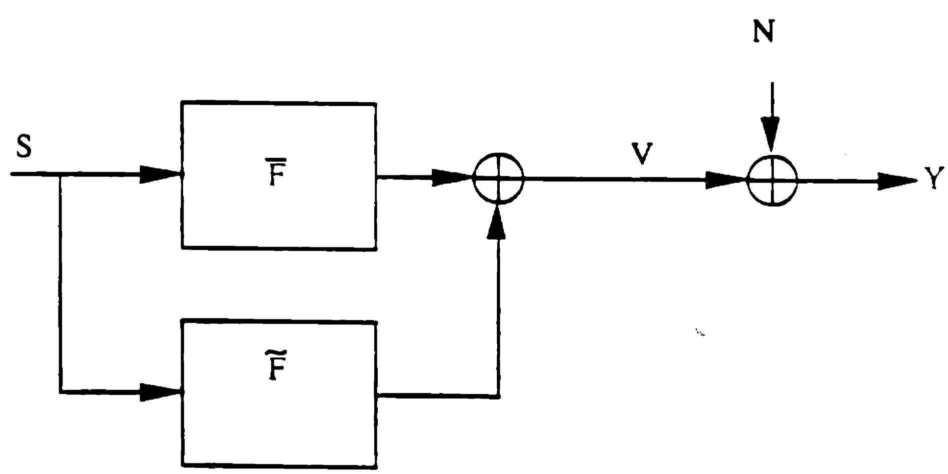


Figure II.9 : System model for channel known with error

where

Y is the output of the channel

$\overline{F}$ is a constant which is equal to the measurement of the channel

$\tilde{F}$ is a zero-mean measurement error of the channel at the receiver, with variance $\sigma_{\tilde{F}}^2$

N is AWGN with variance σ_N^2.

The input S to the channel is constrained to have mean square value at most σ_S^2. When the multiplicative noise is nil, we know that the optimal choice for the distribution of S is zero-mean Gaussian ([Gal], page 336). It is difficult in general to determine the optimal distribution of S for a general channel distribution. Our bounds are applicable for any distribution on $\tilde{F}$. We may note that we do not make any assumptions about the distribution of the measurement error. The distribution would in practice depend on the manner in which we perform the measurement and the channel distribution. Since the measurement error has zero mean, the measurement of the channel is also its mean value. We know only $\overline{F}$, i.e. the measurement, and the variance of $\tilde{F}$.

The mutual information between the output, Y, and the input, S is

$$I(Y;S|\overline{F}) = h(Y|\overline{F}) - h(Y|S, \overline{F}) \qquad \text{II-3.[8]}$$

where the entropies are differential entropies since we are dealing with continuous valued random variables. Let us examine how we may maximize II-3.[8]. In the rest of this section, we shall not explicitly write our conditioning on $\overline{F}$, since we have assumed that we have an *a priori* measurement independent of the inputs and outputs over the time during which we perform our observation.

Let us first consider whether the methods that we use to find the capacity of an additive Gaussian noise channel may be applied here. Since

$$Y = S\overline{F} + S\tilde{F} + N \qquad \text{II-3.[9]}$$

for a given value of s we may write that

$$h(Y \mid S = s) = h(s\tilde{F} + N)$$

II-3.[10]

since entropy is invariant under translation. Therefore, the second term in the RHS of II-3.[8] can be expressed as

$$h(Y|S) = \int_{-\infty}^{+\infty} p_S(s)\, h(s\tilde{F}+N)\, ds$$

II-3.[11]

The above expression cannot be calculated if we do not have a distribution for $\tilde{F}$. Even if we do know the distribution for $\tilde{F}$, whether II-3.[11] can be calculated easily depends on the form of the distribution of $\tilde{F}$.

We wish to find a lower bound to the maximum achievable mutual information I(Y;S) for the model given in Figure II.9. We may do so by using a zero-mean Gaussian input S with variance σ_S^2. In order to determine how tight this lower bound is, we also find an upper bound to the maximum achievable I(Y;S). We determine that the lower bound and the upper bound converge as $\frac{\sigma_F^2 \sigma_S^2}{\sigma_N^2} \rightarrow 0$.

<u>II-3-2-a-a- A lower bound on the maximum achievable mutual information between Y and S.</u>

To find a lower bound on the maximum achievable I(Y;S) = h(S) - h(S|Y), we may choose S to be Gaussian, even though the Gaussian distribution may not be the one that maximizes mutual information for the specific measurement noise distribution. Thus, we fix the value of h(S). We next find an upper bound on h(S|Y), which holds for all possible distributions for S. The difference between h(S) and the bound on h(S|Y) immediately yields a lower bound on I(Y;S).

By definition,

$$h(S \mid Y) = \int h(S \mid Y=y)\, p_Y(y)\, dy$$

II-3.[13].

Since adding a constant does not change differential entropy,

$$h(S \mid Y=y) = h\left(S - \alpha y \mid Y= y\right)$$

II-3.[14]

thus

$$h(S \mid Y) = h(S - aY \mid Y)$$

II-3.[15]

for any real a. Since conditioning always decreases entropy, we have that

$$h(S - aY \mid Y) \leq h(S - aY)$$

II-3.[16].

Therefore

$$h(S \mid Y) \leq h(S - aY)$$

(using the fact that the entropy of a random variable with given variance is upper bounded by the entropy of a Gaussian random variable with the same variance)

$$\leq \frac{1}{2}\ln\left(2\pi e \mathrm{Var}\left(S-\alpha Y\right)\right)$$

II-3.[17]

for any α. Therefore, II-3.[17] also holds when we minimize the RHS over a[1]. Therefore, we wish to take a so that αY is the linear minimum variance estimate of S in terms of Y ([AM], pg. 93),

$$\alpha = \frac{E[SY]}{E\left[Y^2\right]}$$

[1] This approach was suggested by Dr. S. Shamai.

$$= \frac{\bar{F}\sigma_S^2}{\bar{F}^2\sigma_S^2 + \sigma_S^2\sigma_F^2 + \sigma_N^2}$$

II-3.[18]

since S and Y are zero mean. Therefore, the variance of $S - \alpha Y$, maximized over α, is given by

$$\mathrm{Var}(S - \alpha Y) = \sigma_S^2 - \frac{\left(\bar{F}\sigma_S^2\right)^2}{\bar{F}^2\sigma_S^2 + \sigma_S^2\sigma_F^2 + \sigma_N^2}$$

$$= \frac{\sigma_S^4\sigma_F^2 + \sigma_N^2\sigma_S^2}{\bar{F}^2\sigma_S^2 + \sigma_S^2\sigma_F^2 + \sigma_N^2}$$

II-3.[19].

Therefore, we have that

$$h(S|Y) \le \frac{1}{2}\ln\left(2\pi e\frac{\sigma_S^4\sigma_F^2 + \sigma_N^2\sigma_S^2}{\bar{F}^2\sigma_S^2 + \sigma_S^2\sigma_F^2 + \sigma_N^2}\right)$$

II-3.[20].

The mutual information between S and Y may therefore be lower bounded by

$$I(S;Y) \ge \frac{1}{2}\ln\left(2\pi e\sigma_S^2\right) - \frac{1}{2}\ln\left(2\pi e\frac{\sigma_S^4\sigma_F^2 + \sigma_N^2\sigma_S^2}{\bar{F}^2\sigma_S^2 + \sigma_S^2\sigma_F^2 + \sigma_N^2}\right)$$

$$= \frac{1}{2}\ln\left(1 + \frac{\bar{F}^2\sigma_S^2}{\sigma_S^2\sigma_F^2 + \sigma_N^2}\right)$$

II-3.[21].

The above bound may be interpreted as saying that the worst effect that the measurement noise can have is to behave as AWGN. We see that the above bound is equal to the capacity of the channel that would result from sending a Gaussian signal with variance $\overline{F}^2 \sigma_S^2$ in an AWGN channel with noise variance $\sigma_S^2 \sigma_F^2 + \sigma_N^2$.

II-3-2-a-b- An upper bound on the maximum achievable mutual information between Y and S.

We obtain an upper bound to I(Y;S) by using the fact established in section II-3-i that

$$I(Y;S) \leq I(Y;S \mid F) \qquad \text{II-3.[22].}$$

For F known to be f, the maximum mutual information I(Y; S | F=f), over all possible choices for S, is given by $\frac{1}{2}\ln\left(\frac{f^2\sigma_S^2 + \sigma_N^2}{\sigma_N^2}\right)$. Hence, the RHS of II-3.[22] is equal to

$$\frac{1}{2}E_F\left[\ln\left(\frac{\overline{F}^2\sigma_S^2 + \tilde{F}^2\sigma_S^2 + \sigma_N^2}{\sigma_N^2}\right)\right].$$

Therefore, using the convexity of the ln function, we may write that

$$I(Y;S) \leq \frac{1}{2}\ln\left(1 + \frac{\overline{F}^2\sigma_S^2 + \sigma_S^2\sigma_F^2}{\sigma_N^2}\right) \qquad \text{II-3.[23].}$$

We see that the above bound is equal to the capacity of the channel that would result from sending a Gaussian signal with variance $\overline{F}^2\sigma_S^2 + \sigma_S^2\sigma_F^2$ in an AWGN channel with noise variance σ_N^2.

We may interpret the upper bound II-3.[23] as indicating that the channel has an upper bound given by the case where the effect of the measurement error is only that of extra transmission power. Intuitively, we may see the upper bound as the case where the effect of the measurement noise is altogether useful and the lower bound as the case where the measurement error is wholly detrimental. We should expect that, as $\frac{\sigma_F^2\sigma_S^2}{\sigma_N^2} \rightarrow 0$, I(Y;S) should have some limiting behavior. Indeed, from II-3.[21] and [23], we see that

$$I(Y;S) \rightarrow \frac{1}{2}\ln\left(1 + \frac{\overline{F}^2\sigma_S^2}{\sigma_N^2}\right)$$

II-3.[24]

i.e. the mutual information converges to the case where there is no measurement error. Such behavior was to be expected and the bounds II-3.[21] and [23] simply offer an indication as to how that convergence occurs.

We may point out that in the derivation of the upper bound, our analysis is valid for $\tilde{F}$ a zero-mean random variable of arbitrary finite variance. We may note that our bounds are tighter than the pessimum mutual information upper and lower bounds given by Blachman ([Bla62]) for a band-limited channel perturbed by statistically dependent interference. This is to be expected since we are not looking at a worst-case statistical dependence. We may proceed to the case of greater interest with multi-dimensional inputs and outputs.

II-3-2-b- Extension to the multiple symbol case.

The extension of the previous results to the multiple symbol case is fairly straightforward. Such an extension is useful to analyze our model as given in matrix form in section II-2. Indeed, unless we consider several time intervals, we cannot hope to capture the memory inherent in multipath systems where the inverse of the transmission bandwidth is smaller than the multipath spread. Let us assume that we know the covariance matrix of the input $\underline{S}$ and that the covariance matrix is not singular. We have, as in section II-2-3-b, that the dimension of the input vector is k' and the dimension of the output vector is k if we consider complex notation (if we are considering complex random variables expressed as real vectors. the dimensions are, respectively, 2k and 2k' in the notation of II-2-3-b).

II-3-2-b-a- A lower bound to the maximum achievable mutual information between $\underline{Y}$ and $\underline{S}$.

Let us first determine a lower bound to the achievable maximum mutual information. Assume we have a Gaussian input with mutually independent Gaussian components. Such a distribution will yield a lower bound on the maximum achievable mutual information between input and output. If $\underline{S}_{2k'}$ is 2k'-dimensional Gaussian ([CT], pg. 234), then

$$h(\underline{S}_{2k'}) = \frac{1}{2}\ln\left((2\pi e)^{2k'}\left|\Lambda_{\underline{S}_{2k'}}\right|\right)$$

II-3.[25]

where | | denotes the absolute value of the determinant of a matrix and $\Lambda_{\underline{S}_{2k'}}$ is the covariance matrix of $\underline{S}_{2k'}$. Denote the cross-correlation matrix of $\underline{S}$ and $\underline{R}$, say, by $\Lambda_{(\underline{S},\,\underline{R})}$. We take the correlation matrix of S to be given. We have that $\underline{\overline{F}}$ is a 2k x 2k' known matrix with components given by II-2.[33]. $\underline{\tilde{F}}$ is a zero mean random matrix representing the measurement noise on each component of $\underline{\overline{F}}$ which is not outside the ranges given in II-2.[33], depending on which model we choose.

To find a lower bound on the maximum achievable $I(\underline{Y};\,\underline{S}) = h(\underline{S}) - h(\underline{S}|\,\underline{Y})$, we proceed as in the single-use case with some modifications. We shall consider 2k and 2k' to be constant in the derivation of our lower bound. Therefore, we shall omit 2k and 2k' as subscripts and superscripts in order to simplify the notation. We still have that, as in II-3.[17]

$$h(\underline{S}|\,\underline{Y}) = h(\underline{S} - \underline{\alpha Y}|\,\underline{Y})$$

$$\leq h(\underline{S} - \underline{\alpha Y})$$

(using the fact that the entropy of a random variable with a given covariance matrix is upper bounded by the entropy of a Gaussian random variable with the same covariance matrix)

$$\leq \frac{1}{2} \ln\left((2\pi e)^{2k'} \left|\Lambda_{\underline{S}-\underline{\alpha}\underline{Y}}\right|\right)$$

II-3.[26]

for $\underline{\alpha}$ any real 2k' x 2k matrix. Using the same reasoning as in the one-dimensional case, we wish to find that $\underline{a}$ which gives the linear minimum variance estimator of $\underline{S}$ in terms of $\underline{Y}$. We shall derive the information form of the MVE error because it provides insight into the meaning of our bounds. Expression II-3.[18] becomes ([AM], pg. 93)

$$\underline{\alpha} = \Lambda_{(\underline{S},\, \underline{Y})} \Lambda_{\underline{Y}}^{-1}$$

(using the fact that the signal, the measurement noise and the additive Gaussian noise are all mutually independent and have zero mean)

$$= \Lambda_{(\underline{S},\, \underline{\overline{F}S})}\left(\Lambda_{\underline{\overline{F}S}} + \Lambda_{\underline{\tilde{F}S}} + \Lambda_{\underline{N}}\right)^{-1}$$

II-3.[27].

The matrix $\Lambda_{\underline{\overline{F}S}} + \Lambda_{\underline{\tilde{F}S}} + \Lambda_{\underline{N}}$ is indeed invertible, because it is positive definite. The matrix $\Lambda_{\underline{N}}$ is, from our assumptions, positive definite.

The estimation error of the input from the output is given by $\underline{E} = \underline{S} - \underline{\alpha}\underline{Y}$. Thus, in II-3.[26], we can replace $\underline{S} - \underline{\alpha}\underline{Y}$ in the RHS by $\underline{E}$. The covariance matrix of $\underline{E}$ is given by ([AM], pg. 93)

$$\Lambda_{\underline{E}} = \Lambda_{\underline{S}} - \Lambda_{(\underline{S},\, \underline{\overline{F}S})}\left(\Lambda_{\underline{\overline{F}S}} + \Lambda_{\underline{\tilde{F}S}} + \Lambda_{\underline{N}}\right)^{-1} \Lambda_{(\underline{\overline{F}S},\, \underline{S})}$$

II-3.[28].

The above matrix $\Lambda_{\underline{\underline{E}}}$ cannot be singular. If it were singular, we would have an entropy of minus infinity on the RHS of II-3.[20], and therefore we could send an infinite amount of information. Multiplying each side of II-3.[28] by $\Lambda_{\underline{\underline{E}}}^{-1}$, we obtain

$$\underline{I} = \Lambda_{\underline{\underline{E}}}^{-1}\Lambda_{\underline{S}} - \Lambda_{\underline{\underline{E}}}^{-1}\Lambda_{(\underline{S},\,\underline{\overline{F}S})}\left(\Lambda_{\underline{\overline{F}S}} + \Lambda_{\underline{\bar{\bar{F}}S}} + \Lambda_{\underline{N}}\right)^{-1}\Lambda_{(\underline{\overline{F}S},\,\underline{S})} \Leftrightarrow$$

$$\Lambda_{\underline{\underline{E}}}^{-1}\Lambda_{\underline{S}} = \underline{I} + \Lambda_{\underline{\underline{E}}}^{-1}\Lambda_{(\underline{S},\,\underline{\overline{F}S})}\left(\Lambda_{\underline{\overline{F}S}} + \Lambda_{\underline{\bar{\bar{F}}S}} + \Lambda_{\underline{N}}\right)^{-1}\Lambda_{(\underline{\overline{F}S},\,\underline{S})}$$

II-3.[29].

Let us derive $\Lambda_{\underline{\underline{E}}}^{-1}$ in the case where k=k' and $\underline{\overline{F}}$ is invertible. We shall later show that the same form extends to the case where k≠k' or $\underline{\overline{F}}$ is not invertible. Under the assumption that k=k' and $\underline{\overline{F}}$ is invertible, all the terms in the second term of the RHS of II-3.[29] have an inverse. Rewriting the second term in the RHS of II-3.[29], we have that

$$\Lambda_{\underline{\underline{E}}}^{-1}\Lambda_{(\underline{S},\,\underline{\overline{F}S})}\left(\Lambda_{\underline{\overline{F}S}} + \Lambda_{\underline{\bar{\bar{F}}S}} + \Lambda_{\underline{N}}\right)^{-1}\Lambda_{(\underline{\overline{F}S},\,\underline{S})}$$

$$= \left(\Lambda_{(\underline{\overline{F}S},\,\underline{S})}^{-1}\left(\Lambda_{\underline{\overline{F}S}} + \Lambda_{\underline{\bar{\bar{F}}S}} + \Lambda_{\underline{N}}\right)\Lambda_{(\underline{S},\,\underline{\overline{F}S})}^{-1}\Lambda_{\underline{\underline{E}}}\right)^{-1}$$

(substituting II-3.[28] for the last term)

$$= \left(\begin{array}{c}\Lambda_{(\underline{\overline{F}S},\,\underline{S})}^{-1}\left(\Lambda_{\underline{\overline{F}S}} + \Lambda_{\underline{\bar{\bar{F}}S}} + \Lambda_{\underline{N}}\right)\Lambda_{(\underline{S},\,\underline{\overline{F}S})}^{-1} \\ \left(\Lambda_{\underline{S}} - \Lambda_{(\underline{S},\,\underline{\overline{F}S})}\left(\Lambda_{\underline{\overline{F}S}} + \Lambda_{\underline{\bar{\bar{F}}S}} + \Lambda_{\underline{N}}\right)^{-1}\Lambda_{(\underline{\overline{F}S},\,\underline{S})}\right)\end{array}\right)^{-1}$$

(distributing over the sum in the last term)

$$= \left(\Lambda_{(\underline{\overline{F}S},\,\underline{S})}^{-1}\left(\Lambda_{\underline{\overline{F}S}} + \Lambda_{\underline{\bar{\bar{F}}S}} + \Lambda_{\underline{N}}\right)\Lambda_{(\underline{S},\,\underline{\overline{F}S})}^{-1}\Lambda_{\underline{S}} - \underline{I}\right)^{-1}$$

(distributing over the remaining sum)

$$
= \begin{pmatrix} \Lambda_{(\underline{\bar{F}S},\,\underline{S})}^{-1} \Lambda_{\underline{\bar{F}S}} \Lambda_{(\underline{S},\,\underline{\bar{F}S})}^{-1} \Lambda_{\underline{S}} + \Lambda_{(\underline{\bar{F}S},\,\underline{S})}^{-1} \Lambda_{\underline{\tilde{\bar{F}}S}} \Lambda_{(\underline{S},\,\underline{\bar{F}S})}^{-1} \Lambda_{\underline{S}} \\ + \Lambda_{(\underline{\bar{F}S},\,\underline{S})}^{-1} \Lambda_{\underline{N}} \Lambda_{(\underline{S},\,\underline{\bar{F}S})}^{-1} \Lambda_{\underline{S}} - \underline{I} \end{pmatrix}^{-1}
$$

II-3.[30].

Let us simplify the first term in the parentheses of the RHS of II-3.[30]:

$\Lambda_{(\underline{S},\,\underline{\bar{F}S})} = \underline{\bar{F}}\,\Lambda_{\underline{S}}$ and $\Lambda_{(\underline{S},\,\underline{\bar{F}S})} = \Lambda_{\underline{S}}\,\underline{\bar{F}}^{T} \Rightarrow$

$$
\Lambda_{(\underline{\bar{F}S},\,\underline{S})}^{-1} \Lambda_{\underline{\bar{F}S}} \Lambda_{(\underline{S},\,\underline{\bar{F}S})}^{-1} \Lambda_{\underline{S}} = \Lambda_{\underline{S}}^{-1} \underline{\bar{F}}^{-1} \underline{\bar{F}} \Lambda_{\underline{S}} \underline{\bar{F}}^{T} \left(\underline{\bar{F}}^{T}\right)^{-1} \Lambda_{\underline{S}}^{-1} \Lambda_{\underline{S}}
$$

$$
= \underline{I}
$$

II-3.[31].

The second term in the parentheses of the RHS of II-3.[30] may similarly be rewritten as:

$$
\Lambda_{(\underline{\bar{F}S},\,\underline{S})}^{-1} \Lambda_{\underline{\tilde{\bar{F}}S}} \Lambda_{(\underline{S},\,\underline{\bar{F}S})}^{-1} \Lambda_{\underline{S}} = \Lambda_{\underline{S}}^{-1} \underline{\bar{F}}^{-1} \Lambda_{\underline{\tilde{\bar{F}}S}} \left(\underline{\bar{F}}^{T}\right)^{-1} \Lambda_{\underline{S}}^{-1} \Lambda_{\underline{S}}
$$

$$
= \Lambda_{\underline{S}}^{-1} \underline{\bar{F}}^{-1} \Lambda_{\underline{\tilde{\bar{F}}S}} \left(\underline{\bar{F}}^{T}\right)^{-1}
$$

II-3.[32].

Finally, the third term in II-3.[30] may be replaced by:

$$
\Lambda_{(\underline{\bar{F}S},\,\underline{S})}^{-1} \Lambda_{\underline{N}} \Lambda_{(\underline{S},\,\underline{\bar{F}S})}^{-1} \Lambda_{\underline{S}} = \Lambda_{\underline{S}}^{-1} \underline{\bar{F}}^{-1} \Lambda_{\underline{N}} \left(\underline{\bar{F}}^{T}\right)^{-1} \Lambda_{\underline{S}}^{-1} \Lambda_{\underline{S}}
$$

$$
= \Lambda_{\underline{S}}^{-1} \underline{\bar{F}}^{-1} \Lambda_{\underline{N}} \left(\underline{\bar{F}}^{T}\right)^{-1}
$$

II-3.[33].

Substituting II.[31]-[33] in the parentheses of the RHS of II-3.[30] yields

$$\Lambda_{\underline{E}}^{-1}\Lambda_{(\underline{S},\,\underline{\bar{F}S})}\left(\Lambda_{\underline{\bar{F}S}}+\Lambda_{\underline{\tilde{F}S}}+\Lambda_{\underline{N}}\right)^{-1}\Lambda_{(\underline{\bar{F}S},\,\underline{S})}$$

$$=\left(\Lambda_{\underline{S}}^{-1}\,\underline{\bar{F}}^{-1}\Lambda_{\underline{\tilde{F}S}}\left(\underline{\bar{F}}^{T}\right)^{-1}+\Lambda_{\underline{S}}^{-1}\,\underline{\bar{F}}^{-1}\Lambda_{\underline{N}}\left(\underline{\bar{F}}^{T}\right)^{-1}\right)^{-1}$$

II-3.[34].

Substituting II-3.[34] into II-3.[29] yields

$$\Lambda_{\underline{E}}^{-1}\Lambda_{\underline{S}}=\underline{I}+\left(\Lambda_{\underline{S}}^{-1}\,\underline{\bar{F}}^{-1}\Lambda_{\underline{\tilde{F}S}}\left(\underline{\bar{F}}^{T}\right)^{-1}+\Lambda_{\underline{S}}^{-1}\,\underline{\bar{F}}^{-1}\Lambda_{\underline{N}}\left(\underline{\bar{F}}^{T}\right)^{-1}\right)^{-1}$$

(distributing)

$$=\underline{I}+\left(\Lambda_{\underline{S}}^{-1}\,\underline{\bar{F}}^{-1}\left(\Lambda_{\underline{\tilde{F}S}}+\Lambda_{\underline{N}}\right)\left(\underline{\bar{F}}^{T}\right)^{-1}\right)^{-1}$$

$$=\underline{I}+\underline{\bar{F}}^{T}\left(\Lambda_{\underline{\tilde{F}S}}+\Lambda_{\underline{N}}\right)^{-1}\underline{\bar{F}}\,\Lambda_{\underline{S}}$$

II-3.[35].

Therefore, we may write that

$$\Lambda_{\underline{E}}^{-1}=\Lambda_{\underline{S}}^{-1}+\underline{\bar{F}}^{T}\left(\Lambda_{\underline{\tilde{F}S}}+\Lambda_{\underline{N}}\right)^{-1}\underline{\bar{F}}$$

II-3.[36].

Expression II-3.[36] is for the case where k=k' and $\underline{\bar{F}}$ is invertible only. Let us now show that II-3.[36] also applies to the general case. From II-3.[28], we may write that

$$\left(\Lambda_{\underline{S}}^{-1}+\underline{\bar{F}}^{T}\left(\Lambda_{\underline{\tilde{F}S}}+\Lambda_{\underline{N}}\right)^{-1}\underline{\bar{F}}\right)\Lambda_{\underline{E}}$$

$$= \left(\Lambda_{\underline{S}}^{-1} + \bar{\underline{F}}^{T} \left(\Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right)^{-1} \bar{\underline{F}} \right)$$

$$\times \left(\Lambda_{\underline{S}} - \Lambda_{(\underline{S},\, \bar{\underline{F}}\underline{S})} \left(\Lambda_{\bar{\underline{F}}\underline{S}} + \Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right)^{-1} \Lambda_{(\bar{\underline{F}}\underline{S},\, \underline{S})} \right)$$

(distributing)

$$= I - \bar{\underline{F}}^{T} \left(\Lambda_{\bar{\underline{F}}\underline{S}} + \Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right)^{-1} \bar{\underline{F}} \Lambda_{\underline{S}} + \bar{\underline{F}}^{T} \left(\Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right)^{-1} \bar{\underline{F}} \Lambda_{\underline{S}}$$

$$- \bar{\underline{F}}^{T} \left(\Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right)^{-1} \bar{\underline{F}} \Lambda_{\underline{S}} \bar{\underline{F}}^{T} \left(\Lambda_{\bar{\underline{F}}\underline{S}} + \Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right)^{-1} \bar{\underline{F}} \Lambda_{\underline{S}}$$

$\left(\text{replacing } \bar{\underline{F}} \Lambda_{\underline{S}} \bar{\underline{F}}^{T} \text{ by } \Lambda_{\bar{\underline{F}}\underline{S}} + \Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} - \left(\Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right) \text{ in the last term} \right)$

$$= I - \bar{\underline{F}}^{T} \left(\Lambda_{\bar{\underline{F}}\underline{S}} + \Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right)^{-1} \bar{\underline{F}} \Lambda_{\underline{S}} + \bar{\underline{F}}^{T} \left(\Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right)^{-1} \bar{\underline{F}} \Lambda_{\underline{S}}$$

$$- \bar{\underline{F}}^{T} \left(\Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right)^{-1} \left(\Lambda_{\bar{\underline{F}}\underline{S}} + \Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} - \left(\Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right) \right)$$

$$\times \left(\Lambda_{\bar{\underline{F}}\underline{S}} + \Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right)^{-1} \bar{\underline{F}} \Lambda_{\underline{S}}$$

(distributing)

$$= I + \bar{\underline{F}}^{T} \left(\Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right)^{-1} \bar{\underline{F}} \Lambda_{\underline{S}} - \bar{\underline{F}}^{T} \left(\Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right)^{-1} \bar{\underline{F}} \Lambda_{\underline{S}}$$

$$= I$$

II-3.[37].

For a given cross-correlation matrix, entropy is maximized for a Gaussian distribution. Therefore, from II-3.[28]

$$h(\underline{S} \mid \underline{Y}) \leq h(\underline{E})$$

$$\leq k' \ln(2\pi e) + \frac{1}{2} \ln\left(\left| \Lambda_{\underline{E}} \right| \right)$$

(using II-3.[36])

$$\leq k' \ln(2\pi e) - \frac{1}{2}\ln\left(\left|\Lambda_{\underline{S}}^{-1} + \underline{\bar{F}}^{T}\left(\Lambda_{\underline{\tilde{F}S}} + \Lambda_{\underline{N}}\right)^{-1}\underline{\bar{F}}\right|\right)$$

II-3.[38].

The mutual information between $\underline{S}$ and $\underline{Y}$ may therefore be lower bounded by the following vector analog to II-3.[21]. We manipulate our expression so as to obtain a form which will easily be comparable to the upper bound:

$$I(\underline{S}\,;\underline{Y}) \geq \frac{1}{2}\ln\left(\left|\Lambda_{\underline{S}}\right|\left|\Lambda_{\underline{S}}^{-1} + \underline{\bar{F}}^{T}\left(\Lambda_{\underline{\tilde{F}S}} + \Lambda_{\underline{N}}\right)^{-1}\underline{\bar{F}}\right|\right)$$

(using the fact that |AB| =|A| |B|)

$$= \frac{1}{2}\ln\left(\left|\underline{I} + \underline{\bar{F}}^{T}\left(\Lambda_{\underline{\tilde{F}S}} + \Lambda_{\underline{N}}\right)^{-1}\underline{\bar{F}}\Lambda_{\underline{S}}\right|\right)$$

II-3.[39].

The interpretation of the above bound is as in the single sample case. The RHS of II-3.[39] is the mutual information between the input $\underline{S}$ to a known channel $\underline{\bar{F}}$ with independent additive Gaussian noise $\underline{\hat{N}}$ of correlation $\Lambda_{\underline{\tilde{F}S}} + \Lambda_{\underline{N}}$. Indeed, let us call $\underline{\hat{Y}}$ the output to that channel, shown in Figure II.10.

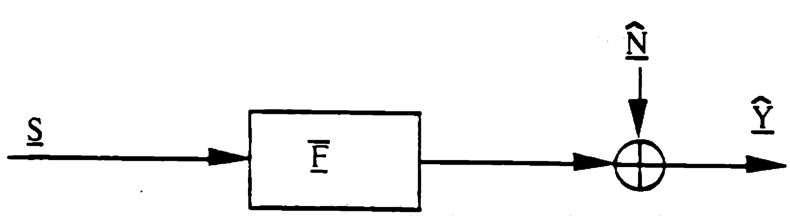


Figure II.10 : Channel with mutual information between input and output equal to the RHS of II-3.[39].

The mutual information of the above channel may be expressed as

$$I(\underline{S}\,;\underline{\hat{Y}}) = h(\underline{S}) - h(\underline{S} \mid \underline{\hat{Y}})$$

$= h(\underline{S}) - h(\hat{\underline{E}})$

II-3.[40]

where $\hat{\underline{E}}$ is the error in estimating $\underline{S}$ from $\hat{\underline{Y}}$. The correlation matrix of $\hat{\underline{E}}$ may be expressed, from standard estimation arguments, as

$$\Lambda_{\hat{\underline{E}}}^{-1} = \Lambda_{\underline{S}}^{-1} + \underline{F}^T \left(\Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}}\right)^{-1} \underline{F}$$

II-3.[41].

Combining II-3.[41] and II-3.[40] yields the RHS of II-3.[39]. Note that we may also write the mutual information between $\underline{S}$ and $\hat{\underline{Y}}$ as

$$I(\underline{S} ; \hat{\underline{Y}}) = h(\hat{\underline{Y}}) - h(\hat{\underline{Y}} \mid \underline{S})$$

$$= \frac{1}{2} \ln\left(\left| \underline{F}\Lambda_{\underline{S}}\underline{F}^T + \Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right| \right) - \frac{1}{2} \ln\left(\left| \Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}} \right| \right)$$

$$= \frac{1}{2} \ln\left(\left| \left(\Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}}\right)^{-1} \underline{F}\Lambda_{\underline{S}}\underline{F}^T + \underline{I} \right| \right)$$

II-3.[42].

Therefore, from our interpretation of II-3.[39], we may rewrite the RHS as

$$I(\underline{S} ; \underline{Y}) \geq \frac{1}{2} \ln\left(\left| \left(\Lambda_{\tilde{\underline{F}}\underline{S}} + \Lambda_{\underline{N}}\right)^{-1} \underline{F}\Lambda_{\underline{S}}\underline{F}^T + \underline{I} \right| \right)$$

II-3.[43].

II-3-2-b-b- An upper bound on the maximum achievable mutual information between $\underline{Y}$ and $\underline{S}$.

Obtaining an upper bound to $I(\underline{Y};\underline{S})$ is a straightforward extension of the single sample case. Expression II-3.[22] holds with the appropriate modifications to take into account that we are in the multi-dimensional case. The RHS of II-3.[22] is $k \ln(2\pi e) + \frac{1}{2} \ln\left(\left| \Lambda_{(\underline{F} + \tilde{\underline{F}})\underline{S}} + \Lambda_{\underline{N}} \right|\right)$ if we are in the real case. The expression analogous to II-3.[23] is

$$I(\underline{Y}\,;\underline{S}) \leq \frac{1}{2}\ln\left(\frac{\left|\Lambda_{(\underline{F}+\underline{\tilde{F}})\underline{S}} + \Lambda_{\underline{N}}\right|}{\left|\Lambda_{\underline{N}}\right|}\right)$$

(using the fact that the inverse of the determinant is the determinant of the inverse)

$$= \frac{1}{2}\ln\left(\left|\Lambda_{\underline{N}}^{-1}\Lambda_{(\underline{F}+\underline{\tilde{F}})\underline{S}} + \underline{I}\right|\right)$$

(rewriting $\Lambda_{(\underline{F}+\underline{\tilde{F}})\underline{S}}$)

$$= \frac{1}{2}\ln\left(\left|\Lambda_{\underline{N}}^{-1}\underline{F}\,\Lambda_{\underline{S}}\,\underline{F}^{T} + \Lambda_{\underline{N}}^{-1}\Lambda_{\underline{\tilde{F}}\underline{S}} + \underline{I}\right|\right)$$

II-3.[44].

The above bound is equal to the capacity of the channel that would result from sending a Gaussian signal with covariance matrix $\Lambda_{(\underline{F}+\underline{\tilde{F}})\underline{S}}$ in an AWGN channel with noise variance $\Lambda_{\underline{N}}$. Therefore, as in the single dimensional case, we may interpret the upper bound II-3.[44] as the case where the only effect of the measurement error is that of extra transmission power. In the limit as $\Lambda_{\underline{\tilde{F}}\underline{S}} \rightarrow 0$, from II-3.[43] and II-3.[44], we have

$$I(\underline{Y}\,;\underline{S}) \rightarrow \frac{1}{2}\ln\left(\left|\left(\Lambda_{\underline{N}}^{-1}\underline{F}\,\Lambda_{\underline{S}}\underline{F}^{T} + \underline{I}\right)\right|\right)$$

II-3.[45].

Note that the condition $\Lambda_{\underline{\tilde{F}}\underline{S}} \rightarrow 0$ may seem to be a fairly strong condition. However, an equivalent condition is that $\mathrm{tr}\left(\Lambda_{\underline{\tilde{F}}\underline{S}}\right) \rightarrow 0$. Since $\Lambda_{\underline{\tilde{F}}\underline{S}}$ is positive semi-definite, requiring that $\mathrm{tr}\left(\Lambda_{\underline{\tilde{F}}\underline{S}}\right) \rightarrow 0$ is equivalent to requiring each individual term in the diagonal of $\Lambda_{\underline{\tilde{F}}\underline{S}}$ to go to 0. Indeed, if the diagonal terms of $\Lambda_{\underline{\tilde{F}}\underline{S}}$ are 0 then, since $\Lambda_{\underline{\tilde{F}}\underline{S}}$ is a covariance matrix, the off-diagonal terms must go to 0 also.

In the following section, we shall use these results for channels with a measurement error to determine some general properties regarding the loss of mutual information from not knowing the channel at the receiver.

II-3-3- The relation between the rate of change of the channel and loss of mutual information from not knowing the channel at the receiver.

In our previous discussion, we have simply considered that a channel was perfectly known or known with some error without making any assumptions about how the channel changes in time. In particular, in section II-3-2, we simply assumed that we have a measurement error on the channel without concerning ourselves with the origin of the error. In this section, we make some assumptions regarding the general structure of the random process by which the channel evolves. We obtain results where the loss in mutual information due to not knowing the channel at the receiver is related to parameters of the channel. In particular, we relate the coherence time of the channel to the loss in mutual information from not knowing the channel at the receiver.

Our intuition suggests that, if we know the general structure of the channel, the speed at which the channel changes determines to some extent the error that we have in the measurement. In this section, we therefore separate the case where the channel is constant or is time-varying in some deterministic manner from the case where the channel is a random process. In the first case, we can hope to measure the channel and thus compute the capacity by considering our measured channel. We must restrict ourselves to certain types of channels, because we could construct an arbitrary time-varying deterministic channel which would thwart our efforts at measurement. When the channel is a random process, we cannot hope that repeated measurements will lead to an arbitrarily accurate estimation of the channel because, as time elapses between measurements, the channel changes stochastically. Again, an arbitrary stochastic process is unlikely to yield interesting results, and therefore we concentrate on a fairly general model, the μ^{th} order Markov channel, which includes the common Gauss-Markov model. We consider the taps of the channel to be real for ease of exposition, but all the results can be readily extended to the complex case.

II-3-3-a- Case of an unknown time-invariant channel with known distribution.

Expression II-3.[7]

$$I\left(\underline{Y}_{2k}; \underline{S}_{2k} \mid \underline{F}^{2k}\right) - I\left(\underline{Y}_{2k}; \underline{S}_{2k}\right) = h\left(\underline{F}^{2k} \mid \underline{Y}_{2k}\right) - h\left(\underline{F}^{2k} \mid \underline{S}_{2k}, \underline{Y}_{2k}\right)$$

II-3.[7]

indicates that the difference in mutual information between input and output when we know and do not know the channel depends on the difference between the accuracy of the channel measurement when the output only is considered versus the accuracy of the channel measurement when both input and output are considered. How well we can measure the channel will depend partly on how it varies. Let us consider the extreme case where the channel is unknown, with some *apriori* distribution, but each tap that describes the channel does not change with time. Intuitively, we know that if we were faced with such a channel, we would first measure it at the receiver with arbitrarily good precision by repeatedly sending a known signal and then transmit data once we had a satisfactory measure of the channel at the receiver. Such is the principle behind the actual practice of "sounding" the channel. As the number of data symbols, 2k, becomes very large, the effect of the symbols "lost" to the channel measurement will become negligible when we look at average mutual information per symbol. However, there is a residual loss per symbol, owing to the fact that the initial measurement of the channel is not perfect. We shall now quantify the notions of vanishing loss from the initial measurement and of per symbol residual loss from the measurement error. We assume that the channel may be described by a finite number Δ of contiguous taps, that the channel has an *apriori* distribution, and that its entropy is not infinitely large.

Theorem II-3.1 : Let $\underline{F}^{2k}$ be an unknown time-invariant channel with a known finite number of taps Δ, known *apriori* distribution and additive Gaussian noise. Let $\underline{S}_{2k}$ be a stationary, Gaussian distributed and power limited input to the channel with an arbitrary correlation matrix. For any positive ε, we may choose a probabilistic input $\underline{S}_{2k}'$ whose corresponding output we denote by $\underline{Y}_{2k}'$ such that for all k large enough

$$\left(\frac{I\left(\underline{Y}_{2k}; \underline{S}_{2k} \mid \underline{F}^{2k}\right) - I\left(\underline{Y}_{2k}^{'}; \underline{S}_{2k}^{'}\right)}{2k} \right) \leq \varepsilon$$

II-3.[46].

Proof:

We choose the signal $\underline{S}_{2k}'$ sent by the sender, which does not know the channel, to be such that at any time i S'[i] is either a fixed sounding pattern or S[j], where j is defined below. The sounding pattern is known at the receiver. For simplicity, the sounding pattern we choose consists of a constant tone of power ΔP for one sample followed by a nil signal for $(\Delta-1)$ time samples. Since our channel is time-invariant, $B_{Doppler}$ is nil, so we may choose W and W_{input} to be the same. Therefore the sender can send this sounding pattern. Such a pattern of constant tones satisfies the average power constraint. The sender repeats this pattern k_0 times. The transmitter thus spends $k_1 = k_0 D$ time samples transmitting a sounding signal. Figure II.11 shows the real part of the sampled received signal without noise, $\underline{v}_{k_1}$, for the first k_1 samples. It then transmits using $S'[k_1+i] = S[i]$ for $i>0$. Thus, the output Y' to the input S' is identically distributed with the output Y shifted by k_1 in time. We assume the receiver knows the sounding pattern and k_1, but does not know the input sample values after time sample k_1.

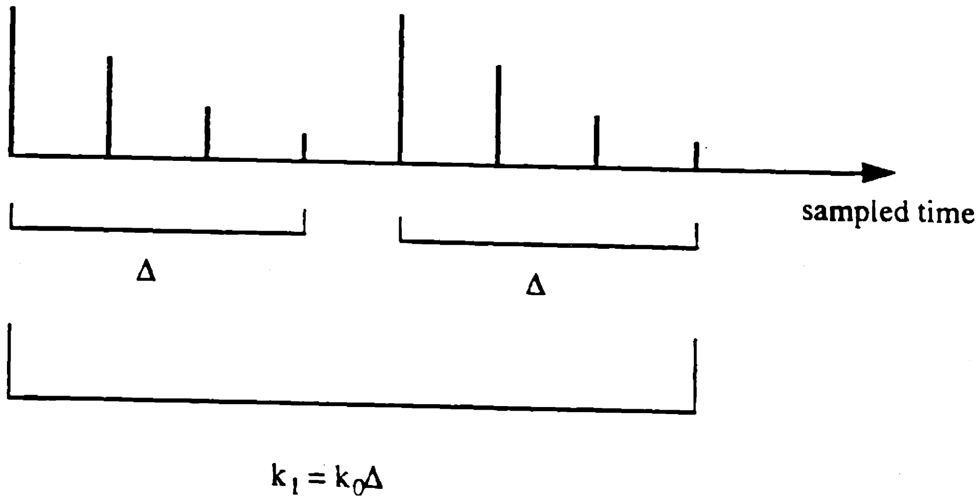


Figure II.11 : Real part of the received signal $\underline{v}_{k_1}$ (before noise) for the first k_1 samples.

Since the channel does not change, all rows of $\underline{F}^k$ are identical aside from the shift of terms. Similarly, all rows of $\underline{F}^{k,2k}$ are identical aside from the shift in terms. Let us denote by $\underline{Y}'_{2k}$ the output corresponding to the input $\underline{S}'_{2k}$. By averaging the statistically independent channel outputs $\underline{Y}'[1]$, $\underline{Y}'[\Delta+1]$, ..., $\underline{Y}'[(k_0-1)\Delta+1]$, we may obtain the real part of the first channel tap, $\underline{F}^{\Delta,\Delta}[1]$, an estimate which we denote by

$$\hat{F}^{\Delta,\Delta}[1] = \frac{\sum_{j=0}^{k_0-1} \underline{Y}'[j\Delta + 1]}{k_0\sqrt{\Delta P}} \qquad \text{II-3.[47].}$$

The variance of the estimate is the variance of the error:

$$\operatorname{Var}\left(\hat{F}^{\Delta,\Delta}[l] \mid F^{\Delta,\Delta}[l]\right) = \frac{\sum_{j=0}^{k_0-1} \operatorname{Var}\left(\underline{Y'}[l+j\Delta] \mid F^{\Delta,\Delta}[l]\right)}{k_0 \Delta P}$$

$$= \frac{\sigma_N^2}{k_0 \Delta P}$$

II-3.[48].

We may similarly obtain the error variance of our estimate for each component of each tap of the channel. Let us denote the error by the matrix $\tilde{\underline{F}}$, in keeping with the notation of section II-3-2-b. We shall prove several inequalities which, together, will yield II-3.[46].

Let us show our first inequality. Since k_1 is fixed, for all k large enough, we may write that:

$$\frac{I\left(\underline{Y}_{2k}; \underline{S}_{2k} \mid \underline{F}^{2k}\right)}{2k} - \frac{I\left(\left(\underline{S}_{k-k_1}, \underline{S}_{k,2k-k_1}\right); \left(\underline{Y}_{k-k_1}, \underline{Y}_{k,2k-k_1}\right) \mid \left(\underline{F}^{k-k_1}, \underline{F}^{k,2k-k_1}\right)\right)}{2k-2k_1} \leq \frac{\varepsilon}{3}$$

II-3.[49]

since the input distribution is stationary.

Let us now show our second inequality. For large enough k, we may write that:

$$\left(\begin{array}{l} \dfrac{I\left(\left(\underline{S'}_{k_1,k}, \underline{S'}_{k+k_1,2k}\right); \left(\underline{Y'}_{k_1,k}, \underline{Y'}_{k+k_1,2k}\right) \mid \left(\hat{\underline{F}}^{k_1,k}, \hat{\underline{F}}^{k+k_1,2k}\right)\right)}{2k-2k_1} \\ -\dfrac{I\left(\left(\underline{S'}_{k_1,k}, \underline{S'}_{k+k_1,2k}\right); \left(\underline{Y'}_{k_1,k}, \underline{Y'}_{k+k_1,2k}\right) \mid \left(\hat{\underline{F}}^{k_1,k}, \hat{\underline{F}}^{k+k_1,2k}\right)\right)}{2k} \end{array}\right) \leq \frac{\varepsilon}{3}$$

II-3.[50].

Moreover.

$$I\left(\underline{S}'_{2k}\,;\underline{Y}'_{2k}\right) = I\left(\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right);\underline{Y}'_{2k}\right)$$

(since the first k_1 inputs are known)

$$= I\left(\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right);\left(\underline{Y}'_{k_1,k},\underline{Y}'_{k+k_1,2k}\right)\mid\left(\underline{Y}'_{k_1},\underline{Y}'_{k,k+k_1}\right)\right)$$
$$+ I\left(\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right);\left(\underline{Y}'_{k_1},\underline{Y}'_{k,k+k_1}\right)\right)$$

(using the fact that the second term is zero)

$$= I\left(\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right);\left(\underline{Y}'_{k_1,k},\underline{Y}'_{k+k_1,2k}\right)\mid\left(\underline{Y}'_{k_1},\underline{Y}'_{k,k+k_1}\right)\right)$$

(since $\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right)$ and $\left(\underline{Y}'_{k_1},\underline{Y}'_{k,k+k_1}\right)$ are independent)

$$= h\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right) - h\left(\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right)\mid\underline{Y}'_{2k}\right)$$

(since $\underline{Y}'_{2k}$ fully determines $\left(\underline{\hat{F}}^{k_1,k},\underline{\hat{F}}^{k+k_1,2k}\right)$)

$$= h\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right)$$
$$- h\left(\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right)\mid\left(\left(\underline{Y}'_{k_1,k},\underline{Y}'_{k+k_1,2k}\right),\left(\underline{\hat{F}}^{k_1,k},\underline{\hat{F}}^{k+k_1,2k}\right)\right)\right)$$

(since $\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right)$ is independent of $\left(\underline{\hat{F}}^{k_1,k},\underline{\hat{F}}^{k+k_1,2k}\right)$)

$$= I\left(\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right);\left(\underline{Y}'_{k_1,k},\underline{Y}'_{k+k_1,2k}\right)\mid\left(\underline{\hat{F}}^{k_1,k},\underline{\hat{F}}^{k+k_1,2k}\right)\right)$$

II-3.[51].

Dividing the RHS and LHS of II-3.[51] by 2k, we obtain from II-3.[50] that, for large enough k:

$$\left(\frac{I\left(\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right);\left(\underline{Y}'_{k_1,k},\underline{Y}'_{k+k_1,2k}\right)\mid\left(\hat{\underline{F}}^{k_1,k},\hat{\underline{F}}^{k+k_1,2k}\right)\right)}{2k-2k_1} - \frac{I\left(\underline{S}'_{2k};\underline{Y}'_{2k}\right)}{2k}\right) \le \frac{\varepsilon}{3}$$

II-3.[52].

Expressions II-3.[49] and II-3.[52] may be interpreted as stating that, regardless of whether the channel is perfectly known or estimated through measurement, the effect of the first k_1 symbols is negligible when we average over sufficiently many symbols. Let us now show our last inequality which will relate II-3.[49] and II-3.[52] so as to yield II-3.[47]. We now must show that

$$\left(\frac{I\left(\left(\underline{S}_{k-k_1},\underline{S}_{k,2k-k_1}\right);\left(\underline{Y}_{k-k_1},\underline{Y}_{k,2k-k_1}\right)\mid\left(\underline{F}^{k-k_1},\underline{F}^{k,2k-k_1}\right)\right)}{2k-2k_1} - \frac{I\left(\left(\underline{S}'_{k_1,k},\underline{S}'_{k+k_1,2k}\right);\left(\underline{Y}'_{k_1,k},\underline{Y}'_{k+k_1,2k}\right)\mid\left(\hat{\underline{F}}^{k_1,k},\hat{\underline{F}}^{k+k_1,2k}\right)\right)}{2k-2k_1}\right) \le \frac{\varepsilon}{3}$$

II-3.[53]

for large enough k_1 and k.

Since a Gaussian distribution maximizes entropy for a given covariance matrix, we have that:

$$I\left(\left(\underline{S}_{k-k_1},\underline{S}_{k,2k-k_1}\right);\left(\underline{Y}_{k-k_1},\underline{Y}_{k,2k-k_1}\right)\mid\left(\underline{F}^{k-k_1},\underline{F}^{k,2k-k_1}\right)\right)$$

$$\le \frac{1}{2}E_{\underline{F}}\left[\ln\left(\left|\Lambda_{\underline{N}}^{-1}\underline{F}\,\Lambda_{\underline{S}}\,\underline{F}^{T}+\underline{I}\right|\right)\right].$$

$$= \frac{1}{2} E_{\hat{\underline{F}}} E_{\tilde{\underline{F}} \mid \hat{\underline{F}}} \left[\ln \left(\left| \Lambda_{\underline{N}}^{-1} (\tilde{\underline{F}} + \hat{\underline{F}}) \Lambda_{\underline{S}} (\tilde{\underline{F}} + \hat{\underline{F}})^T + \underline{I} \right| \right) \right]$$

(since $\tilde{\underline{F}}$ and $\hat{\underline{F}}$ are independent)

$$= \frac{1}{2} E_{\hat{\underline{F}}} E_{\tilde{\underline{F}}} \left[\ln \left(\left| \Lambda_{\underline{N}}^{-1} (\hat{\underline{F}} + \tilde{\underline{F}}) \Lambda_{\underline{S}} (\hat{\underline{F}} + \tilde{\underline{F}})^T + \underline{I} \right| \right) \right]$$

(from the concavity of the ln function and the fact that $\hat{\underline{F}}$ is zero mean)

$$\leq \frac{1}{2} E_{\hat{\underline{F}}} \left[\ln \left(\left| \Lambda_{\underline{N}}^{-1} \left(\hat{\underline{F}} \Lambda_{\underline{S}} \hat{\underline{F}}^T + \Lambda_{\underline{\tilde{F}S}} \right) + \underline{I} \right| \right) \right]$$

II-3.[54].

Moreover, from II-3.[43], substituting $\hat{\underline{F}}$ for $\underline{F}$, since $\hat{\underline{F}}$ is an unbiased estimate of $\underline{F}$

$$I\left(\left(\underline{S}'_{k_1,k}, \underline{S}'_{k+k_1,2k} \right); \left(\underline{Y}_{k-k_1}, \underline{Y}_{k,2k-k_1} \right) \mid \left(\hat{\underline{F}}^{k-k_1}, \hat{\underline{F}}^{k,2k-k_1} \right) \right)$$

$$\geq \frac{1}{2} E_{\hat{\underline{F}}} \left[\ln \left(\left| \left(\Lambda_{\underline{\tilde{F}S}} + \Lambda_{\underline{N}} \right)^{-1} \hat{\underline{F}} \Lambda_{\underline{S}} \hat{\underline{F}}^T + \underline{I} \right| \right) \right]$$

II-3.[55]

since S is Gaussian.

From II-3.[54] and II-3.[55], we see that, to prove II-3.[53], it is sufficient to show that

$$\frac{1}{4(k-k_1)} \ln \left(\left| \left(\Lambda_{\underline{\tilde{F}S}} + \Lambda_{\underline{N}} \right)^{-1} \underline{F} \Lambda_{\underline{S}} \underline{F}^T + \underline{I} \right| \right)$$

$$\geq \frac{1}{4(k-k_1)} \ln \left(\left| \Lambda_{\underline{N}}^{-1} \left(\hat{\underline{F}} \Lambda_{\underline{S}} \hat{\underline{F}}^T + \Lambda_{\underline{\tilde{F}S}} \right) + \underline{I} \right| \right) - \frac{\varepsilon}{3}$$

II-3.[56]

for large enough k_1 and k.

In order to prove II-3.[56], let us show that

$$\frac{1}{2}\ln\left(\left|\left(\Lambda_{\underline{\tilde{F}S}} + \Lambda_{\underline{N}}\right)^{-1}\underline{\hat{F}}\Lambda_{\underline{S}}\underline{\hat{F}}^{T} + \underline{I}\right|\right)$$

$$= \frac{1}{2}\ln\left(\left|\Lambda_{\underline{N}}^{-1}\left(\underline{\hat{F}}\,\Lambda_{\underline{S}}\,\underline{\hat{F}}^{T} + \Lambda_{\underline{\tilde{F}}\underline{S}}\right) + \underline{I}\right|\right) - \frac{1}{2}\ln\left(\left|\Lambda_{\underline{N}}^{-1}\Lambda_{\underline{\tilde{F}}\underline{S}} + \underline{I}\right|\right)$$

II-3.[57].

To prove II-3.[57], we may consider the following channel. The channel has independent Gaussian inputs $\underline{S}_1$ and $\underline{S}_2$ with covariance matrices $\underline{F}\,\Lambda_{\underline{S}}\,\underline{F}^{T}$ and $\Lambda_{\underline{\tilde{F}}\underline{S}}$, respectively. The channel has additive Gaussian noise $\underline{N}$. Let $\underline{Y}$ be the output of this channel. Figure II.12 shows our channel.

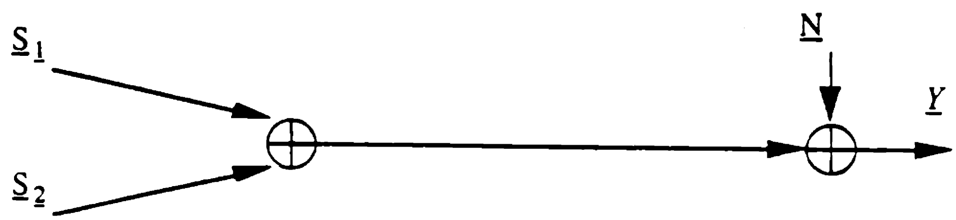


Figure II.12 : Channel for the proof of II-3.[57].

Let us look at the mutual information between the input $\underline{S}_1$ and the output $\underline{Y}$. Since $\underline{S}_1$ and $\underline{S}_2$ are independent, $\underline{S}_2$ acts as noise to $\underline{S}_1$ in the expression $I(\underline{S}_1 ; \underline{Y})$.

$$I(\underline{S}_1 ; \underline{Y}) = \frac{1}{2}\ln\left(\left|\left(\Lambda_{\underline{N}} + \Lambda_{\underline{\tilde{F}}\underline{S}}\right)^{-1}\underline{F}\,\Lambda_{\underline{S}}\,\underline{F}^{T} + \underline{I}\right|\right)$$

II-3.[58].

We may write that

$$I\left(\left(\underline{S}_2, \underline{S}_1\right); \underline{Y}\right) = I\left(\underline{S}_1; \underline{Y}\right) + I\left(\underline{S}_2; \underline{Y} \mid \underline{S}_1\right)$$

II-3.[59].

Since the inputs $\underline{S}_1$ and $\underline{S}_2$ are independent.

$$I\left(\underline{S}_2; \underline{Y} \mid \underline{S}_1\right) = \frac{1}{2}\ln\left(\left|\Lambda_{\underline{N}}^{-1}\Lambda_{\underline{\tilde{F}}\underline{S}} + \underline{I}\right|\right)$$

II-3.[60].

Moreover,

$$I\left(\left(\underline{S}_2, \underline{S}_1\right); \underline{Y}\right) = h\left(\underline{Y}\right) - h\left(\underline{N}\right)$$

II-3.[61].

Since the covariance matrix of $\underline{Y}$ is $\underline{\hat{F}}\,\Lambda_{\underline{S}}\,\underline{\hat{F}}^T + \Lambda_{\underline{\tilde{F}}\underline{S}} + \Lambda_{\underline{N}}$ and $\underline{Y}$ is Gaussian, we may rewrite II-3.[61] as

$$I\left(\left(\underline{S}_2, \underline{S}_1\right); \underline{Y}\right) = \frac{1}{2}\ln\left(\left|\Lambda_{\underline{N}}^{-1}\left(\underline{\hat{F}}\,\Lambda_{\underline{S}}\,\underline{\hat{F}}^T + \Lambda_{\underline{\tilde{F}}\underline{S}}\right) + \underline{I}\right|\right)$$

II-3.[62].

From combining II-3.[58], and II-3.[60] through II-3.[62], we obtain II-3.[57].

Using Hadamard's inequality,

$$\frac{1}{2}\ln\left(\left|\Lambda_{\underline{N}}^{-1}\Lambda_{\underline{\tilde{F}}\underline{S}} + \underline{I}\right|\right)$$

$$\leq \frac{k}{2}\ln\left(1 + \frac{\sum_{i=1}^{\Delta} \mathrm{Var}\left(\underline{\tilde{F}}^{\Delta,\Delta}[i]\,\mathrm{S}\right)}{\sigma_N^2}\right) + \frac{k}{2}\ln\left(1 + \frac{\sum_{i=1}^{\Delta} \mathrm{Var}\left(\underline{\tilde{F}}^{\Delta,\Delta}[i+k]\,\mathrm{S}\right)}{\sigma_N^2}\right)$$

(since $\underline{\tilde{F}}$ and $\underline{S}$ are independent, $\underline{S}$ is stationary and $\underline{\tilde{F}}$ is zero-mean)

$$= \frac{k}{2}\ln\left(1+\frac{\sum_{i=1}^{\Delta}\tilde{\underline{F}}^{\Delta.\Delta}[i]\sigma_S^2}{\sigma_N^2}\right)+\frac{k}{2}\ln\left(1+\frac{\sum_{i=1}^{\Delta}\tilde{\underline{F}}^{\Delta.\Delta}[i+k]\sigma_S^2}{\sigma_N^2}\right)$$

(using II-3.[48])

$$= k\ln\left(1+\frac{\sigma_S^2}{k_0\Delta P}\right)$$

II-3.[63].

For large enough k and k_0, with $k>k_0$, we have that

$$k\ln\left(1+\frac{\sigma_S^2}{k_0\Delta P}\right)\leq\frac{2(k-k_1)\varepsilon}{3}$$

(from II-3.[63])

$$\Rightarrow \frac{1}{2(k-k_1)}\ln\left(\left|\Lambda_{\underline{N}}^{-1}\Lambda_{\tilde{\underline{F}}\underline{S}}+\underline{I}\right|\right)\leq\frac{\varepsilon}{3}$$

II-3.[64].

Combining II-3.[57] and II-3.[64], we obtain II-3.[56] as desired. Hence II-3.[53] holds.Cascading II-3.[49], II-3.[52] and II-3.[53], we obtain II-3.[46].

Q.E.D.

We may easily extend the same analysis to the case where the channel is periodic, *in the sampled space*, with a period p known at the receiver. Given the typically random nature of our channel, such a periodicity is unlikely and we shall not consider it beyond the following corollary. We just perform the same analysis as for the time-invariant case but with a spacing equal to the period in between measurements. We first measure the first tap of the first period by transmitting over one time sample and then being silent for D-1+pD samples and repeating this operation k_0 times. We then measure the second tap of the first period in the same manner, etc... After measuring all the taps for the first period, we measure in the same fashion the taps of the second, third, ... p^{th} periods. Similarly, we may extend the analysis when the channel is a sum of periodic channels, whose different periods are known. Indeed, a sum of periodic channels has itself a period equal to at most the product of the periods of the channels. Therefore, we may state the following corollary.

Corollary II-3.1:

Let $\underline{F}^{2k}$ be a channel with a known finite number of taps Δ, known *apriori* distribution and additive Gaussian noise. Let $\underline{F}^{2k}$ be periodic with finite period p in the sampled time domain. Let $\underline{S}_{2k}$ be a stationary, Gaussian distributed and power limited input to the channel with an arbitrary correlation matrix. For any positive ε, we may choose a probabilistic input $\underline{S}_{2k}'$ whose corresponding output we denote by $\underline{Y}_{2k}'$ such that for all k large enough

$$\left(\frac{I\left(\underline{Y}_{2k}; \underline{S}_{2k} \mid \underline{F}^{2k}\right) - I\left(\underline{Y}_{2k}'; \underline{S}_{2k}'\right)}{2k} \right) \leq \varepsilon$$

II-3.[65].

If the sampling rate is large enough with respect to the Doppler spread, this case may approximate the multipath channel with fixed Doppler shifts and fixed path amplitudes.

We have shown that the fact that the channel is unknown at the receiver does not reduce the mutual information with respect to the case where the channel is perfectly known.if the channel does not change in time. Our argument relied on sounding the channel for an arbitrarily long time. If the channel were to vary in time in a non-periodic manner, this measurement method would not be applicable. Therefore, we suspect that we shall lose some mutual information per symbol to not knowing the channel at the receiver and that that loss of mutual information will be related to the manner in which the channel varies. We also suspect that, as the channel varies more and more "slowly" (we shall later give a precise definition of what we mean by "slowly"), we shall lose less and less mutual information per symbol because of not knowing the channel at the receiver. The following sections shall quantify our intuition.

II-3-3-b- Case of a channel modeled as a stochastic process.

We determined in the last section that, if the channel is time-invariant, we do not lose mutual information owing to not knowing the channel at the receiver. Our purpose in this section is to determine how the channel variability impacts the average loss, in mutual information per transmitted symbol, due to not knowing the channel at the receiver. Throughout this section, we use the notation of II-2.[24] and II-2.[25] with complex numbers, rather than the notation of II-2.[32] and II-2.[33]. Therefore, we have k complex samples rather than 2k real samples. If the channel is close to static, then the previous section leads us to believe that we do not have much loss from not knowing the channel at the receiver. We suspect that the loss may be more significant if the channel experiences dramatic variations. But quantizing the variability of the channel is in itself a difficult task. For instance, a channel may vary rapidly and significantly but in a predictable manner. In that case, our result in corollary II-3.1 on channels which are periodic in the sampled time would indicate that not much mutual information is lost by not knowing the channel at the receiver. If the channel is not easily predictable, it is still difficult to characterize its variability. We use an information theoretic approach in order to present results independently of implementation issues.

Let us assume that the channel has a stationary and ergodic structure such that any row $\underline{F}^{i,i}$ depends on at most μ past rows, i.e. that $\underline{F}^{i,i}$ conditioned on $\underline{F}^{i-\mu,i-1}$ is independent of $\underline{F}^{1,i-\mu-1}$. In steady state, this is equivalent to stating that for any $k > i+m$, $\underline{F}^{i,i}$ conditioned on $\underline{F}^{i+1,i+\mu}$ is independent of $\underline{F}^{i+\mu+1,k}$. Such a model is that of a μ^{th} order Markov chain. We shall in the sequel use the second definition for convenience. Markov models are often used to represent channels ([Gil60], [Fri67], [AFK72]) and in particular mobile channels ([McMan70], [Gol94]). Markov models offer great accuracy of representation, albeit at the expense of added complexity. We may note that our model does not assume that we have a finite state Markov channel, and therefore the existing coding theorems for Markov channels are not applicable. As mentioned before, appendix B gives the appropriate coding theorem for special class of Markov channels.

Our purpose is to obtain limit results for the loss in mutual information, per transmitted symbol, due to not knowing the channel at the receiver. In order to accomplish this, we first express

$$I\left(\underline{Y}_k; \underline{S}_k \mid \underline{F}^k\right) - I\left(\underline{Y}_k; \underline{S}_k\right) = I\left(\underline{F}^k; \underline{S}_k \mid \underline{Y}_k\right)$$

II-3.[5]

as a sum. Our motivation is to obtain a summation of terms, each of which has a limit. Thus, we may obtain an ensemble average limit by simply taking the limit of the components. We assume that the channel is stationary and ergodic.

Let us first establish some properties which we shall use in our proof. The main property we shall use to prove our theorem is that $\underline{F}^i$ is independent of Y[i+1] and S[i+1] given $\underline{F}^{i+1,k}$, $\underline{S}_i$ and $\underline{Y}_i$.

Lemma II-3.1 :

$$h\left(\underline{F}^i \mid \underline{F}^{i+1,k}, Y[i+1], S[i+1], \underline{Y}_i, \underline{S}_i\right) = h\left(\underline{F}^i \mid \underline{F}^{i+1,k}, \underline{Y}_i, \underline{S}_i\right)$$

II-3.[66].

Proof:

In order to see why II-3.[66] is true, we may write the following entropy expression, using repeatedly the fact that h(A, B) = h(B) + h(A|B):

$$h\left(\underline{F}^i \mid \underline{F}^{i+1,k}, Y[i+1], S[i+1], \underline{Y}_i, \underline{S}_i\right)$$

$$= h\left(\underline{F}^i, \underline{F}^{i+1,k}, Y[i+1], S[i+1] \mid \underline{Y}_i, \underline{S}_i\right) - h\left(\underline{F}^{i+1,k} \mid \underline{Y}_i, \underline{S}_i\right)$$

$$- h\left(S[i+1] \mid \underline{F}^{i+1,k}, \underline{Y}_i, \underline{S}_i\right) - h\left(Y[i+1] \mid \underline{F}^{i+1,k}, \underline{Y}_i, \underline{S}_i, S[i+1]\right)$$

$$= h\left(\underline{F}^{i+1,k} \mid \underline{Y}_i, \underline{S}_i\right) + h\left(\underline{F}^i \mid \underline{Y}_i, \underline{S}_i, \underline{F}^{i+1,k}\right)$$

$$+ h\left(S[i+1] \mid \underline{Y}_i, \underline{S}_i, \underline{F}^i, \underline{F}^{i+1,k}\right) + h\left(Y[i+1] \mid \underline{Y}_i, \underline{S}_i, S[i+1], \underline{F}^i, \underline{F}^{i+1,k}\right)$$

$$- h\left(\underline{F}^{i+1,k} \mid \underline{Y}_i, \underline{S}_i\right) - h\left(S[i+1] \mid \underline{F}^{i+1,k}, \underline{Y}_i, \underline{S}_i\right) - h\left(Y[i+1] \mid \underline{F}^{i+1,k}, \underline{Y}_i, \underline{S}_i, S[i+1]\right)$$

II-3.[67].

Let us show that the fourth and last terms in the above expression are in fact equal. Indeed, if we know the channel response at time i+1, the output up to time i and the input up to time i+1, knowing the channel response at time i tells us nothing further about the channel output at time i+1, since the channel is independent of both the noise and the input. More rigorously, we may write that

$$h\left(Y[i+1] \mid \underline{F}^{i+1,k}, \underline{F}^{i}, S[i+1], \underline{Y}_i, \underline{S}_i \right)$$

$$= h\left(\underline{F}^{i+1,i+1} \underline{S}_{i+1} + N[i+1] \mid \underline{F}^{i+1,k}, \underline{F}^{i}, S[i+1], \underline{Y}_i, \underline{S}_i \right)$$

(using independence)

$$= h\left(\underline{F}^{i+1,i+1} \underline{S}_{i+1} + N[i+1] \mid \underline{F}^{i+1,k}, S[i+1], \underline{S}_i, \underline{Y}_i \right)$$

$$= h\left(Y[i+1] \mid \underline{F}^{i+1,k}, S[i+1], \underline{S}_i, \underline{Y}_i \right)$$

II-3.[68].

Note that the conditioning on $\underline{Y}_i$ on the RHS of II-3.[68] could be removed because of our independence assumptions.

We may similarly show that the third and sixth terms in the RHS of II-3.[67] are equal because the input is independent of the channel. Moreover, the first and fifth terms are equal and cancel. Therefore, canceling pairwise the fourth and seventh terms, and the third and sixth termsi in II-3.[67], we obtain II-3.[66].

Q.E.D.

Let us now express the loss in mutual information due to not knowing the channel as a sum. We may write that, from expression II-3.[6],

$$I\left(\underline{Y}_k; \underline{S}_k \mid \underline{F}^k \right) - I\left(\underline{Y}_k; \underline{S}_k \right) = h\left(\underline{F}^k \mid \underline{Y}_k \right) - h\left(\underline{F}^k \mid \underline{S}_k, \underline{Y}_k \right)$$

where, using the fact that h(A, B) = h(B) + h(A|B), we may rewrite the first term as

$$h\left(\underline{F}^{k,k} \mid \underline{Y}_k\right) + h\left(\underline{F}^{k-1} \mid \underline{Y}_k, \underline{F}^{k,k}\right)$$

and from II-3.[66], we may rewrite the second term as

$$- h\left(\underline{F}^{k,k} \mid \underline{S}_k, \underline{Y}_k\right) - h\left(\underline{F}^{k-1} \mid \underline{S}_{k-1}, \underline{Y}_{k-1}, \underline{F}^{k,k}\right).$$

Therefore, we may write that

$$I\left(\underline{Y}_k; \underline{S}_k \mid \underline{F}^k\right) - I(\underline{Y}_k; \underline{S}_k) = h\left(\underline{F}^{k,k} \mid \underline{Y}_k\right) + h\left(\underline{F}^{k-1} \mid \underline{Y}_k, \underline{F}^{k,k}\right)$$
$$- h\left(\underline{F}^{k,k} \mid \underline{S}_k, \underline{Y}_k\right) - h\left(\underline{F}^{k-1} \mid \underline{S}_{k-1}, \underline{Y}_{k-1}, \underline{F}^{k,k}\right)$$

(rearranging the terms and repeating the same manipulation)

$$= h\left(\underline{F}^{k,k} \mid \underline{Y}_k\right) - h\left(\underline{F}^{k,k} \mid \underline{S}_k, \underline{Y}_k\right)$$
$$- h\left(\underline{F}^{k-1,k-1} \mid \underline{Y}_k, \underline{F}^{k,k}\right) + h\left(\underline{F}^{k-1,k-1} \mid \underline{S}_{k-1}, \underline{Y}_{k-1}, \underline{F}^{k,k}\right)$$
$$+ h\left(\underline{F}^{k-2} \mid \underline{Y}_k, \underline{F}^{k-1,k}\right) - h\left(\underline{F}^{k-2} \mid \underline{S}_{k-2}, \underline{Y}_{k-2}, \underline{F}^{k-1,k}\right)$$

(using repeatedly the fact that h(A, B) = h(B) + h(A|B) and II-3.[66])

$$\ldots$$
$$= \sum_{i=\Delta}^{k} \left[h\left(\underline{F}^{i,i} \mid \underline{Y}_k, \underline{F}^{i+1,k}\right) - h\left(\underline{F}^{i,i} \mid \underline{S}_i, \underline{Y}_i, \underline{F}^{i+1,k}\right) \right]$$
$$+ h\left(\underline{F}^{\Delta-1} \mid \underline{Y}_k, \underline{F}^{\Delta,k}\right) - h\left(\underline{F}^{\Delta-1} \mid \underline{S}_{\Delta-1}, \underline{Y}_{\Delta-1}, \underline{F}^{\Delta,k}\right)$$

(using the fact that the channel is μ^{th} order Markov)

$$= \sum_{i=\Delta}^{k} \left[h\left(\underline{F}^{i,i} \mid \underline{Y}_k, \underline{F}^{i+1,k}\right) - h\left(\underline{F}^{i,i} \mid \underline{S}_i, \underline{Y}_i, \underline{F}^{i+1,i+\mu}\right) \right]$$
$$+ h\left(\underline{F}^{\Delta-1} \mid \underline{Y}_k, \underline{F}^{\Delta,k}\right) - h\left(\underline{F}^{\Delta-1} \mid \underline{S}_{\Delta-1}, \underline{Y}_{\Delta-1}, \underline{F}^{\Delta,k}\right)$$

II-3.[69].

II-3.[69] expresses the loss in mutual information due to not knowing the channel as a sum over the number of symbols transmitted. Using such a form, we shall in Theorem II-3.2 find a limit for the information lost per transmitted symbol by finding the limit of each term in the sum. In general, the term in the sum in the above RHS of II-3.[69] may or may not reach a limit. Therefore, we make the assumption of stationarity for the channel and the input. Under this condition, we may state the following theorem.

Theorem II-3.2 : Let $\underline{F}^k$ be an unknown time-varying channel with finite number of taps Δ and a known *apriori* stationary distribution. Let any row $\underline{F}^{i,i}$ conditioned on $\underline{F}^{i+1,i+\mu}$ be independent of $\underline{F}^{i+\mu+1,k}$ and let $h\left(\underline{F}^{i,i} \mid Y[i], \underline{F}^{i+1,i+\mu}, \underline{S}_i, \underline{F}^{i-\mu,i-1}\right) > -\infty$. Let the input $\underline{S}_k$ be stationary. Then

$$\lim_{k\to\infty}\left(\frac{I\left(\underline{Y}_k; \underline{S}_k \mid \underline{F}^k\right) - I\left(\underline{Y}_k; \underline{S}_k\right)}{k}\right)$$

$$= \lim_{i\to\infty}\lim_{k\to\infty}\left(h\left(\underline{F}^{i,i} \mid \underline{Y}_k, \underline{F}^{i+1,k}\right) - h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+\mu}, \underline{S}_i\right)\right)$$

II-3.[70].

Proof:

Let us first consider the first term in the sum of the RHS of II-3.[69]. For any given i<k, let us define $\kappa_{i,k}$ as

$$\kappa_{i,k} = h\left(\underline{F}^{i,i} \mid \underline{Y}_k, \underline{F}^{i+1,k}\right)$$

II-3.[71].

We see that $\kappa_{i,k}$ is decreasing in k from the fact that conditioning decreases entropy. Moreover, the following lemma shows that $\kappa_{i,k}$ is bounded below by a quantity independent of κ.

Lemma II-3.2 :

$$h\left(\underline{F}^{i,i} \mid \underline{Y}_k, \underline{F}^{i+1,k}\right) \geq h\left(\underline{F}^{i,i} \mid Y[i], \underline{F}^{i+1,i+\mu}, \underline{S}_i, \underline{F}^{i-\mu,i-1}\right)$$

II-3.[72].

Proof: The proof is relegated to appendix A.

The RHS of II-3.[72] depends only on i and is not $-\infty$, from our assumption that $h\left(\underline{E}^{i,i} \mid Y[i], \underline{E}^{i+1,i+\mu}, \underline{S}_i, \underline{E}^{i-\mu,i-1}\right) > -\infty$. Therefore, since $\kappa_{i,k}$ is decreasing in κ and bounded below, $\lim_{k\to\infty} \kappa_{i,k}$ exists and is denoted by κ_i.

Since conditioniing decreases entropy, we see from II-3.[71] that the elements $\kappa_{i,i+j}$ are decreasing in i for any fixed positive j. Indeed, from stationarity, $h\left(\underline{E}^{i,i} \mid \underline{E}^{i+1,i+j}, \underline{Y}_{i+j}\right) = h\left(\underline{E}^{i+1,i+1} \mid \underline{E}^{i+2,i+j+1}, \underline{Y}_{2,i+j+1}\right)$. The RHS may be lower bounded by $h\left(\underline{E}^{i+1,i+1} \mid \underline{E}^{i+2,i+j+1}, \underline{Y}_{i+j+1}\right)$, because conditioning decreases entropy. For any i, we may express κ_i as the limit in j of $\kappa_{i,i+j}$. From our discussion, the elements k_i are decreasing in i. From II-3.[71]-[72], we see that for i larger than Δ and μ, we may write that

$$\kappa_i \geq h\left(\underline{E}^{i,i} \mid Y[i], \underline{E}^{i+1,i+\mu}, \underline{S}_i, \underline{E}^{i-\mu,i-1}\right)$$

(using the fact that the input is independent from the channel and the noise and that the channel has finite time spread and the channel is μ^{th} order Markov)

$$= h\left(\underline{E}^{i,i} \mid Y[i], \underline{E}^{i+1,i+\mu}, \underline{S}_{i-\Delta,i}, \underline{E}^{i-\mu,i-1}\right)$$

II-3.[73].

Hence, κ_i is lower bounded and decreasing, and therefore it reaches a limit which we denote by κ.

Let us choose an arbitrarily small positive ε. For j large enough,

$$\left|\kappa_{i,i+j} - \kappa_i\right| \leq \frac{\varepsilon}{4}$$

II-3.[74]

for all i and for all $j>j_0$. Therefore, we may choose an i_0 such that for all $i>i_0$ and $j>j_0$, we have that

$$\left|\kappa_{i,i+j} - \kappa\right| \le \frac{\varepsilon}{2}$$

II-3.[75].

Hence, for all $i>i_0$ and $k>i_0+j_0$,

$$\frac{\sum_{i=1}^{k} \kappa_{i,k}}{k} \le \frac{\sum_{i=1}^{i_0-1} \kappa_{i,k}}{k} + \frac{k - i_0}{k}\left(\frac{\varepsilon}{2} + \kappa\right)$$

II-3.[76].

Therefore, we may write that

$$\lim_{i\to\infty} \lim_{k\to\infty} \left(\frac{\sum_{i=1}^{k} h\left(\underline{F}^{i,i} \mid \underline{Y}_k, \underline{F}^{i+1,k}\right)}{k} \right) = \kappa$$

II-3.[77].

Let us now consider the second term in the sum of the RHS of II-3.[69].We can see, from the stationarity of the channel and input and the fact that entropy is reduced by conditioning, that the elements ζ_i given by

$$\zeta_i = h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+\mu}, \underline{S}_i\right)$$

II-3.[78]

are decreasing in i. Moreover, since conditioning reduces entropy,

$$h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+\mu}, \underline{S}_i\right) \ge h\left(\underline{F}^{i,i} \mid \underline{Y}_k, \underline{F}^{i+1,k}, \underline{S}_i\right)$$

(from II-3.[72])

$$\geq h\left(\underline{F}^{i,i} \mid Y[i], \underline{S}_i, \underline{F}^{i+1,i+\mu}, \underline{F}^{i-\mu,i-1}\right)$$

II-3.[79]

hence ζ_i has a limit, which we denote by ζ. Therefore,

$$\lim_{i \to \infty} \lim_{k \to \infty} \left(\frac{\sum_{i=1}^{k} h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+\mu}, \underline{S}_i\right)}{k} \right) = \zeta$$

II-3.[80].

Combining II-3.[77] and II-3.[80] yields II-3.[70].

Q.E.D.

The RHS of II-3.[70] may not be easy to evaluate. However, since conditioning reduces entropy, we may write that for k large enough

$$h\left(\underline{F}^{i,i} \mid \underline{Y}_k, \underline{F}^{i+1,k}\right) \leq h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+\mu}\right)$$

II-3.[81].

We may substitute II-3.[78] into II-3.[67] to state the following corollary.

Corollary II-3.2: Let $\underline{F}^k$ be an unknown time-varying channel with finite number of taps D and a known *a priori* stationary distribution. Let $\underline{F}^k$ also have the property that any row $\underline{F}^{i,i}$ conditioned on $\underline{F}^{i+1,i+m}$ is independent of $\underline{F}^{i+m+1,k}$ and let $h\left(\underline{F}^{i,i} \mid Y[i], \underline{F}^{i+1,i+\mu}, \underline{S}_i, \underline{F}^{i-\mu,i-1}\right) > -\infty$. Let the input $\underline{S}_k$ have a stationary distribution. Then

$$\lim_{k \to \infty} \left(\frac{I\left(\underline{Y}_k; \underline{S}_k \mid \underline{F}^k\right) - I\left(\underline{Y}_k; \underline{S}_k\right)}{k} \right)$$

$$\leq \lim_{i \to \infty} \left(h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+\mu}\right) - h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+\mu}, \underline{S}_i\right) \right)$$

II-3.[82].

We may also write the RHS of II-3.[82] as $\lim_{i\to\infty}\left(I\left(\underline{F}^{i,i}; \underline{S}_i \mid \underline{Y}_i, \underline{F}^{i+1,i+\mu}\right)\right)$. Let us make some remarks about the bound in II-3.[82].

- First, we may point out that upper bound in II-3.[82] is exact if the S[i]s are IID and the channel has a single tap. Indeed, conditioned on $\underline{Y}_i$ and $\underline{F}^{i+1,i+\mu}$, $\underline{F}^{i,i}$ depends on $\underline{Y}_{i+1,k}$ and $\underline{F}^{i+\mu+1,k}$ only through the estimates of $\underline{S}_i$. Therefore, II-3.[82] holds with equality.

- Second, we may show (see appendix D) that for the bound of II-3.[82] to hold with equality is equivalent to $\lim_{i\to\infty}\lim_{k\to\infty}\left(h\left(\underline{Y}_{i+1,k} \mid \underline{F}^{i+1,k}\right) - h\left(\underline{Y}_{i+1,k} \mid \underline{F}^{i+2,k}\right)\right)$ being 0. The difference between the bound of II-3.[82] and the equality of II-3.[70] can therefore be reduced to the difference one extra initial measurement makes.

- Third, using the expression for the limit of an average, we have that

$$\lim_{i\to\infty}\left(I\left(\underline{F}^{i,i}; \underline{S}_i \mid \underline{Y}_i, \underline{F}^{i+1,i+\mu}\right)\right) = \lim_{i\to\infty}\left(\frac{1}{i}\sum_{j=1}^{i} I\left(\underline{F}^{j,j}; \underline{S}_j \mid \underline{Y}_j, \underline{F}^{j+1,j+\mu}\right)\right)$$

(rewriting the mutual informations as differences of entropies)

$$= \lim_{i\to\infty}\left(\frac{1}{i}\sum_{j=1}^{i} h\left(\underline{F}^{j,j} \mid \underline{Y}_j, \underline{F}^{j+1,j+\mu}\right) - h\left(\underline{F}^{j,j} \mid \underline{S}_j, \underline{Y}_j, \underline{F}^{j+1,j+\mu}\right)\right)$$

(using on the first term of the summation the fact that conditioning decreases entropy)

$$\geq \lim_{i\to\infty}\left(\frac{1}{i}\sum_{j=1}^{i} h\left(\underline{F}^{j,j} \mid \underline{Y}_j, \underline{F}^{j+1,i+\mu}\right) - h\left(\underline{F}^{j,j} \mid \underline{S}_j, \underline{Y}_j, \underline{F}^{j+1,j+\mu}\right)\right)$$

(using the Markov property on the second term of the summation)

$$= \lim_{i \to \infty} \left(\frac{1}{i} \sum_{j=1}^{i} h\left(\underline{E}^{j,j} \mid \underline{Y}_j, \underline{E}^{j+1,i+\mu}\right) - h\left(\underline{E}^{j,j} \mid \underline{S}_j, \underline{Y}_j, \underline{E}^{i+1,j+\mu}\right) \right)$$

(grouping the entropy terms into mutual information terms)

$$= \lim_{i \to \infty} \left(\frac{1}{i} \sum_{j=1}^{i} I\left(\underline{E}^{j,j}; \underline{S}_i \mid \underline{Y}_i, \underline{E}^{j+1,j+\mu}\right) \right)$$

(using the chain rule for entropies)

$$= \lim_{i \to \infty} \left(\frac{1}{i} I\left(\underline{E}^{i}; \underline{S}_i \mid \underline{Y}_i, \underline{E}^{i+1,i+\mu}\right) \right)$$

II-3.[83].

Therefore, our upper bound to the average loss per symbol due to not knowing the channel at the receiver is not stronger than the upper bound given by assuming that we know the final m sample values of the channel. These two remarks simply give a flavor that these limiting arguments for the loss of mutual information per symbol due to not knowing the channel at the receiver are affected by any assumptions of initial or final knowledge of the channel.

Even though we obtain a fairly simple upper bound, computing it in general may be difficult. Out μ^{th} order Markov model allows for many different probabilistic channel descriptions. We shall therefore give an example of a fairly common model which is tractable enough to yield a tight upper bound. In the next section, we apply corollary II-3.2 to obtain some limiting results in the case of a Gauss-Markov channel. The complexity of obtaining results even for this relatively simple model suggests that more complicated μ^{th} order Markov channel models might be best dealt with by using mathematical packages or by approximating them as combination of μ^{th} order Markov channels. We suspect that some of the insight that is gained from the Gauss-Markov model may extend to more general μ^{th} order Markov models, although proving it may be computationally intensive.

II-3-3-c An example of how the coherence time is related to the loss in mutual information due to not knowing the channel *apriori* at the receiver.

Expressions II-3.[70] and II-3.[82] show some sort of dependency between how much the channel changes from sample to sample and how much mutual information is lost per symbol owing to not knowing the channel. As an important example, let us suppose that the channel can be described by a Gauss-Markov model for the individual terms of the channel response matrix. As in II-2.[25], $F[i,j]$ for $1<i\leq k$ and $i\leq j\leq i+\Delta$ is the real part of the response at time i to the real input at time j and $F[i,j]$ for $k+1<i\leq 2k$ and $i-k\leq j\leq i-k+\Delta$ is the complex part of that same response. We shall restrict ourselves to real inputs for simplicity of derivation. We look at the limit as k goes to ∞, and the arguments here are independent of k. The taps evolve as:

$$F[i,j] = \alpha[i-1,j-1]F[i-1,j-1] + \Xi[i-1,j-1]$$

II-3.[84]

where $\{\Xi[i,j]\}$ is a set of zero-mean mutually independent Gaussian random variables independent of the noise. The random variable Ξ is commonly known as the innovation term. The assumption that the $\Xi[i,j]$ are independent implies that the taps of the channel are mutually independent. We see that we satisfy the conditions of Theorem II-3.2 with $\mu = 1$. Since we wish to consider a stationary channel, we assume that, for each j such that $0<j\leq\Delta$ (for the real components of the taps) or $k+1<j\leq k+1+\Delta$ (for the complex components of the taps)$\{\Xi[i,i+j] \mid i\geq 0\}$ is a set of IID random variables. Similarly, we may express the expected power of the innovation in terms of the expected power of the channel by writing:

$$E\left[F[i,j]^2\right] = E\left[\left(\alpha[i-1,j-1]\right)^2 F[i-1,j-1]^2 + \Xi[i-1,j-1]^2\right]$$

(using the fact that the expected power of F is unchanged by the same shift in both i and j because F is stationary)

$$\Leftrightarrow E\left[\Xi[i-1,j-1]^2\right] = \left(1 - \left(\alpha[i-1,j-1]\right)^2\right) E\left[F[i-1,j-1]^2\right]$$

II-3.[85].

Since the expected power of F remains constant after the same shift in both i and j, the expected power of the innovation term is also unchanged by the same shift in both i and j, i.e. our stationarity assumption requires $\alpha[i, i+n]$ to be a function of n only. In the following, for the sake of simplicity, we shall take the different values that $\alpha[i,j]$ takes for different i and j to be constant and denoted by α. We could extend our discussion to the case where a takes different values for different i - j.

The term a indicates how fast the channel decorrelates. Indeed, we may find the term a by writing:

$$\frac{E\left[F[i-1,j-1]F[i,j]\right]}{E\left[F[i,j]^2\right]} = \frac{E\left[\alpha[i-1,j-1]F[i-1,j-1]F[i-1,j-1] + F[i-1,j-1]\,\Xi[i-1,j-1]\right]}{E\left[F[i,j]^2\right]}$$

(using the fact that the term Ξ is zero-mean and independent of past F, and the fact that the expected power of the F term is unchanged by the same shift in both i and j)

$$= \alpha[i-1,j-1]$$

II-3.[86].

The form of the constant a depends on the channel coherence time, $T_{coherence}$. We choose *apriori* a level of decorrelation, ϕ, through which to define our coherence time. We denote by ϕ the value of $\frac{E\left[F[i-m, j-m]F[i,j]\right]}{E\left[F[i,j]^2\right]}$ which we deem sufficient to indicate that the channel has decorrelated in m time units. The time the channel takes to decorrelate, in accordance with the value of ϕ we have chosen, is the coherence time. Therefore, $T_{coherence}$ is m/W. From II-3.[83], we see that

$$\alpha^{T_{coherence} W} = \gamma \Leftrightarrow \ln(\alpha) = \ln(\phi)\frac{1}{T_{coherence} W}$$

II-3.[87].

Therefore, if $T_{coherence}$ goes to ∞, a goes to 1.

There are different ways in which to choose φ. In general, depending on how we define the coherence time, we shall have that $T_{coherence}$ is inversely proportional to $B_{Doppler}$, with a proportionality constant which will depend on the amount of decorrelation we deem necessary in our definition of coherence time. Coherence times are often defined in terms of the envelope correlation coefficient for two received signals separated by Δf in frequency and Δt in time, denoted by $\rho(\Delta f, \Delta t)$. In [Jak], the coherence time is taken to be the time it takes for the received signal envelope correlation coefficient to become 0.5. In the literature, the correlation coefficient ϕ is taken to vary from 0.9 ([CL75]) to 0.37 ([BN63]) for a time separation of $T_{coherence}$.

We consider that the real input sequence $\underline{S}_i$ is composed of IID zero mean random variables. Since the sender does not have any knowledge of the channel, such an assumption is reasonable. If $\underline{S}_i$ had an arbitrary distribution, we might not be able to find a limit for some of our expressions. In order for our IID distribution for the S[i] to hold well, we must have $W >> B_{Doppler}$, otherwise sampling at a higher rate than W_{input} would preclude the samples S[i] from being IID.

Let us first look at the case where we have a single, real tap. Such a model will allow us to generalize to several complex taps later without having to tackle immediately the extra difficulty of having several taps. In the single tap case, $\underline{F}^{i,i}$ may be described simply by F[i,i].

Using the fact that conditioning decreases entropy, the bound in II-3.[82] can be further upper bounded by:

$$h\left(F[i,i] \mid \underline{Y}_i, F[i+1,i+1]\right) - h\left(F[i,i] \mid \underline{Y}_i, F[i+1,i+1], \underline{S}_i\right)$$

$$\leq h\left(F[i,i] \mid F[i+1,i+1]\right) - h\left(F[i,i] \mid \underline{Y}_i, F[i+1,i+1], \underline{S}_i\right)$$

II-3.[88].

Let us consider separately the two terms of the RHS of II-3.[88]. We have already argued that the entropy of a random variable given a second random variable is upper bounded by the entropy of the error of the LLSE estimate of the first random variable from the second. Since all the random variables involved are Gaussian, the first term may be found from the variance of the LLSE estimate of F[i,i] given F[i+1, i+1]. The inverse of the variance of the LLSE estimate of F[i,i] given F[i+1, i+1] is $\frac{1}{\sigma_F^2}+\frac{\alpha^2}{\sigma_\Xi^2}$.

The first term of the RHS of II-3.[82] may be rewritten as:

$$h\left(F[i,i] \mid F[i+1,i+1]\right) = \frac{1}{2}\ln(2\pi e) - \frac{1}{2}\ln\left(\frac{1}{\sigma_F^2}+\frac{\alpha^2}{\sigma_\Xi^2}\right)$$

(using II-3.[85])

$$= \frac{1}{2}\ln\left(2\pi e\left(1-\alpha^2\right)\sigma_F^2\right)$$

II-3.[89]

where σ_F^2 is the variance of the tap.

Let us now consider the second term of II-3.[88]. Assume that we wish to obtain an estimate of F[i,i] from F[i+1, i+1], $\underline{S}_i$ and $\underline{Y}_i$. We shall use a Kalman filter approach to first obtain the estimates of F[1,1], F[2,2], ..., F[i,i] from $\underline{S}_i$ and $\underline{Y}_i$. Next, we shall combine our estimate of F[i,i] obtained from the Kalman filter with our estimate of F[i,i] from F[i+1,i+1]. Equation II-3.[85] gives the Gauss-Markov model for the evolution of the channel. The Kalman filter gives us $\hat{F}[i,i]$, the LSE estimate for F[i,i] given $\underline{S}_i$ and $\underline{Y}_i$. $\hat{F}[i,i]$ is a Gaussian random variable. We can then perform recursive estimation to use the observation of F[i+1,i+1] in order to improve our estimate of F[i,i]. It might at first seem odd that we should use F[i+1,i+1] as just another observation. However, the innovation term X[i,i] is independent from all previous innovation terms and all previous channel noise terms N[1], ..., N[i]. Therefore, performing Kalman filtering on $\underline{Y}_i$ and $\underline{S}_i$ to obtain $\hat{F}[i,i]$ and then using F[i+1,i+1] to refine our estimate yields the LSE estimate of F[i,i] conditioned on $\underline{S}_i$, $\underline{Y}_i$ and F[i+1,i+1]. This estimate is a Gaussian random variable. Since we are dealing with Gaussian random variables, the LSE estimate is also the LLSE estimate. Therefore, we may write that:

$$h(F[i,i] \mid \underline{Y}_i, F[i+1,i+1], \underline{S}_i) = h(F[i,i] - \hat{F}[i,i] \mid \underline{Y}_i, F[i+1,i+1], \underline{S}_i)$$

II-3.[90].

Therefore, since $\hat{F}[i,i]$ is Gaussian, the variance of the error on $\hat{F}[i,i]$ will determine the entropy $h(F[i,i] \mid \underline{Y}_i, F[i+1,i+1], \underline{S}_i)$.

Let us denote by λ_j the variance of the estimate of F[j,j] given $\underline{Y_j}$ and $\underline{S_j}$ using Kalman filtering. We have the following recursive expression for λ_j, which is independent of the values $\underline{y_j}$,

$$\frac{1}{\lambda_1} = \frac{1}{\sigma_F^2 + \sigma_\Xi^2} + \frac{(s[1])^2}{\sigma_N^2}$$

$$\frac{1}{\lambda_{j+1}} = \frac{1}{\alpha^2 \lambda_j + \sigma_\Xi^2} + \frac{(s[j+1])^2}{\sigma_N^2}$$

II-3.[91].

For arbitrary $\underline{s_j}$, we see that it not possible to write a limiting expression for λ_j. Therefore, let us define λ'_j thus:

$$\frac{1}{\lambda'_1} = \frac{1}{\sigma_N^2 + \sigma_\Xi^2} + \frac{\sigma_S^2}{\sigma_N^2}$$

$$\frac{1}{\lambda'_{j+1}} = \frac{1}{\alpha^2 \lambda'_j + \sigma_\Xi^2} + \frac{\sigma_S^2}{\sigma_N^2}$$

II-3.[92].

We have the following lemma, whose proof is relegated to appendix C.

Lemma II-3.3: for the channel model of II-3.[85] with a single tap and $\underline{S}_i$ a sequence of Gaussian IID zero-mean variables, we have that for all $i \geq j \geq 1$

$$\ln(\lambda'_j) \leq E_{\underline{S}_i}\left[\ln(\lambda_j)\right]$$

II-3.[93].

We may find, as $i \to \infty$, the limit λ' for λ'_i from II-3.[92]:

$$\frac{1}{\lambda'} = \frac{1}{\alpha^2 \lambda' + \sigma_\Xi^2} + \frac{\sigma_S^2}{\sigma_N^2}$$

$$\Leftrightarrow \sigma_N^2 \alpha^2 \lambda' + \sigma_\Xi^2 \sigma_N^2 - \sigma_S^2 \alpha^2 \lambda'^2 - \sigma_S^2 \sigma_\Xi^2 \lambda' - \sigma_N^2 \lambda' = 0$$

$$\Leftrightarrow \lambda' = \frac{\sigma_N^2 \alpha^2 - \sigma_N^2 - \sigma_S^2 \sigma_\Xi^2 + \sqrt{\left(\sigma_N^2 \alpha^2 - \sigma_N^2 - \sigma_S^2 \sigma_\Xi^2\right)^2 + 4\sigma_S^2 \sigma_\Xi^2 \sigma_N^2 \alpha^2}}{2\sigma_S^2 \alpha^2}$$

II-3.[94]

since λ'_i is positive.

Let us now look at the variance, denoted by λ''_i, of the estimate of F[i,i] from $\underline{Y}_i$, $\underline{S}_i$ and F[i+1, i+1]. From our previous discussion, we know that λ''_i is given by the variance of an estimate using the observation of F[i+1,i+1] with the previous estimate of F[i,i]

$$\frac{1}{\lambda''_i} = \frac{1}{\lambda_i} + \frac{\alpha^2}{\sigma_\Xi^2}$$

II-3.[95].

Therefore,

$$h\left(F[i,i] \mid \underline{Y}_i, F[i+1,i+1], \underline{S}_i\right) = \frac{1}{2}\ln\left(2\pi e\lambda''_i\right)$$

II-3.[96].

Let us define the following :

$$\frac{1}{\lambda'''_i} = \frac{1}{\lambda'_i} + \frac{\alpha^2}{\sigma_{\Xi}^2}$$

$$= \frac{1}{\lambda'_i} + \frac{\alpha^2}{\left(1-\alpha^2\right)\sigma_F^2}$$

II-3.[97].

We have, from appendix C, the following lemma.

Lemma II-3.4 : for the channel model of II-3.[88] with a single tap and $\underline{S}_i$ a sequence of IID zero-mean variables, we have that for all $i \geq j \geq 1$

$$\ln\left(\lambda'''_j\right) \leq E_{\underline{S}_i}\left[\ln\left(\lambda''_j\right)\right]$$

II-3.[98].

Therefore, from II-3.[96]-[97],

$$h\left(F[i,i] \mid \underline{Y}_i, F[i+1,i+1], \underline{S}_i\right) \leq \frac{1}{2}\ln\left(2\pi e\lambda'''_i\right)$$

II-3.[99].

Hence, we may write the following theorem.

Theorem II-3.3 : for the channel model of II-3.[88] with a single real tap and $\underline{S}_i$ a sequence of real IID zero-mean real variables, we have that

$$\lim_{k\to\infty}\left(\frac{I\left(\underline{Y}_k;\underline{S}_k\mid\underline{F}^k\right)-I\left(\underline{Y}_k;\underline{S}_k\right)}{k}\right)\le\frac{1}{2}\ln\left(\frac{\left(1-\alpha^2\right)\sigma_F^2}{\lambda'}+\alpha^2\right)$$

II-3.[100].

Proof:

Combining Theorem II-3.2, II-3.[89], II-3.[97] and II-3.[99], we obtain

$$\lim_{k\to\infty}\left(\frac{I\left(\underline{Y}_k;\underline{S}_k\mid\underline{F}^k\right)-I\left(\underline{Y}_k;\underline{S}_k\right)}{k}\right)$$

$$\le\lim_{i\to\infty}\left(\frac{1}{2}\ln\left(2\pi e\left(1-\alpha^2\right)\sigma_F^2\right)-\frac{1}{2}\ln\left(2\pi e\left(\frac{1}{\lambda'_i}+\frac{\alpha^2}{\left(1-\alpha^2\right)\sigma_F^2}\right)^{-1}\right)\right)$$

(using the continuity of the functions involved to bring the limit to λ'_i)

$$=\frac{1}{2}\ln\left(2\pi e\left(1-\alpha^2\right)\sigma_F^2\right)-\frac{1}{2}\ln\left(2\pi e\left(\frac{1}{\lambda'}+\frac{\alpha^2}{\left(1-\alpha^2\right)\sigma_F^2}\right)^{-1}\right)$$

II-3.[101].

Expression II-3.[100] follows immediately.

Q.E.D.

Let us look, in the limit as $\alpha\to 1$, at the behavior of $\frac{\left(1-\alpha^2\right)\sigma_F^2}{\lambda'}$. From II-3.[94], we may write

$$\frac{\lambda'}{\left(1-\alpha^2\right)\sigma_F^2} = -\frac{\sigma_N^2}{2\sigma_S^2\alpha^2\sigma_F^2} - \frac{1}{2\alpha^2} + \frac{\sqrt{\left(\sigma_N^2+\sigma_S^2\sigma_F^2\right)^2\left(1-\alpha^2\right) + 4\sigma_S^2\sigma_F^2\sigma_N^2\alpha^2}}{2\sigma_S^2\alpha^2\left(1-\alpha^2\right)\sigma_F^2}$$

II-3.[102]

whose limit is $+\infty$ as $\alpha \to 1$. Therefore, the limit as $\alpha \to 1$ of the RHS II-3.[100] is 0. This limit agrees with our intuition and Theorem II-3.1.

It may seem natural to consider the limit of the RHS of II-3.[100] as $\alpha \to 0$. However, if $\alpha \to 0$, II-3.[87] indicates that $T_{coherence} \to 0$, and therefore our initial assumption that $B_{Doppler} << W$ no longer holds. The RHS of II-3.[100], correspondingly, gives a very weak limit when $\alpha \to 0$. From II-3.[94] we see that the limit may be expressed as

$$\lim_{\alpha \to 0}\left[\frac{1}{2}\ln\left(\left(1-\alpha^2\right)\sigma_F^2\left(\frac{1}{\left(1-\alpha^2\right)\sigma_F^2} + \frac{\sigma_S^2}{\sigma_N^2}\right)\right)\right] = \frac{1}{2}\ln\left(1 + \frac{\sigma_F^2\sigma_S^2}{\sigma_N^2}\right)$$

II-3.[103].

Equivalently, the RHS of II-3.[103] is I(Y[i], ; s[i] | F[i,i]), which is the capacity of an AWGN channel with noise variance σ_N^2 and Gaussian signal with variance $\sigma_F^2\sigma_S^2$. Therefore, as a becomes very small, our bound is quite weak.

Let us now consider what happens when we have several complex taps, i.e. when $\Delta >$ 1 and each tap has a real and a complex component. We still send a real signal. Expression II-3.[88] becomes

$$h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+1}\right) - h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+1}, \underline{S}_i\right)$$

$$\leq h\left(\underline{F}^{i,i} \mid \underline{F}^{i+1,i+1}\right) - h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+1}, \underline{S}_i\right)$$

II-3.[104].

Assuming all the taps are independent (our assumption can be relaxed with appropriate modifications), the first term of the RHS of II-3.[104] is, using the chain rule for entropies,

$$h\left(\underline{F}^{i,i} \mid \underline{F}^{i+1,i+1}\right) = \sum_{j=i}^{i+\Delta} h\left(F[i,j] \mid \underline{F}^{i+1,i+1}, \underline{F}^{i,i}_{i,j-1}\right)$$

(using the fact that conditioning decreases entropy)

$$\leq \sum_{j=i}^{i+\Delta} h(F[i,j] \mid F[i+1,j+1])$$

$$\left(\begin{array}{l}\text{denoting by } \sigma^2_{\Xi_j} \text{ the variance of the innovation on the real part of the } j^{th} \text{ tap,} \\ \text{which is equal to the variance of the innovation on the imaginary part of the } j^{th} \text{ tap} \\ \text{and proceeding as in II-3.[89]}\end{array}\right)$$

$$\leq \frac{1}{2}\sum_{j=0}^{\Delta} \ln\left(2\pi e \sigma^2_{\Xi_j}\right)$$

II-3.[105].

The second term of II-3.[104] cannot be manipulated as the second term of II-3.[88]. Intuitively, the reason is that, for any given tap, the signal not only contributes to reducing the variance of the measurement error in a Kalman filter, but also contributes to the noise by multiplying the other taps. We could also perform Kalman filtering for vectors, but since we can prove a tight bound without having to resort to such tedious algebraic manipulations, we shall not. Therefore, we use the following bound:

$$h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+1}, \underline{S}_i\right) \geq \sum_{j=i}^{i+\Delta} h\left(F[i,j] \mid \underline{Y}_i, \underline{F}^{i+1,i+1}, (F[i,n])_{n\neq j}, \underline{S}_i\right)$$

II-3.[106]

using the chain rule for entropies and the fact that conditioning decreases entropy. The RHS of II-3.[106] corresponds to the case where, for each tap component, we eliminate the ISI. We can lower bound each term in the summation on the RHS of II-3.[106] as we lower bounded the second term in the RHS of II-3.[88]. Therefore, we may bound the limit as $i \to \infty$ of the LHS of II-3.[104] by:

$$\lim_{i\to\infty}\left(h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+1}\right) - h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+1}, \underline{S}_i\right)\right)$$

(using II-3.[104]-[106])

$$\leq \frac{1}{2}\sum_{j=0}^{\Delta} \ln\left(2\pi e\sigma^2_{\Xi_j}\right) - \sum_{j=i}^{i+\Delta} h\left(F[i,j] \mid \underline{Y}_i, \underline{F}^{i+1,i+1}, \{F[i,n]\}_{n\neq j}, \underline{S}_i\right)$$

(using II-3.[99] and the fact that we have eliminated the ISI in the RHS of II-3.[106])

$$\leq \sum_{j=0}^{\Delta}\frac{1}{2}\ln\left(\alpha^2+\left(\frac{-\sigma_N^2}{2\sigma_S^2\alpha^2\sigma_{F_{jR}}^2}-\frac{1}{2\alpha^2}+\frac{\sqrt{\left(\sigma_N^2+\sigma_S^2\sigma_{F_j}^2\right)^2\left(1-\alpha^2\right)+4\sigma_S^2\sigma_{F_j}^2\sigma_N^2\alpha^2}}{2\sigma_S^2\alpha^2\left(1-\alpha^2\right)\sigma_{F_j}^2}\right)^{-1}\right)$$

II-3.[107].

We may now apply our results concerning the loss of capacity due to channel variations to discuss the issue of spreading.

II-4- Spreading in time-varying channels with memory.

The issue of spreading is often associated with time-varying channels and one of its touted advantages is that of extra "diversity". Loosely speaking, the time-varying nature of the channels is responsible for fades which could be catastrophic. Therefore, by transmitting over a wider bandwidth, we attempt to avoid being altogether vulnerable to such fades. Such a concept is indeed sound, and spreading has been shown to improve bit error rate by adding diversity ([FS83]). When we consider capacity results, we are looking at averaging over very long periods of time. The fact that spreading allows us to average over more symbols in less time does not increase capacity.

All of our previous discussion provides us with tools for considering the important issue of spreading. Although spreading is seldom associated in commercial systems with single user channels, we wish to study it for a single user before we add the consideration of multiple access. We define the spreading of a system as the ratio of the bandwidth used to the data rate. We investigate what happens when the spreading is increased while holding the power and the data rate fixed. In the following, we consider the channel to be composed of many small slices of spectrum. We assume that the signal S along any slice of bandwidth is IID with respect to the signals along all the other slices of bandwidth. The signal power along any slice of bandwidth is scaled down with increased spreading to maintain constant overall input power. Note that, since the channel has a non-zero coherence bandwidth, the response of these slices will not be independent of each other. However, we shall later consider capacity results, where we average over an arbitrarily large number of symbols. In that case, it is the average behavior that counts, rather than the instantaneous behavior. The fact that two contiguous slices of bandwidth are correlated in their fading will imply that we have less "diversity". However, over many symbols, the average mutual information will be the same as if they faded independently. We may therefore arrive at the desired results while remembering that our intermediary results do not hold for non-zero coherence bandwidth. Since we do not worry about the effect of the coherence bandwidth, we assume that each slice is narrow enough to be described by a single tap. If the bandwidth of each spectrum slice were larger than the coherence bandwidth (which is inversely proportional to T_{spread}), the taps would be spaced less than T_{spread} apart, and therefore the channel slices could not be described by a single tap.

Three remarks are in order about the spreading scheme which we choose. The spreading scheme is optimal if we have AWGN channels. For fading dispersive channels, we know ([Gal], pp. 431-439, [Ken]), that capacity may be achieved by transmitting very high power pulses with a very low duty factor. However, the bandwidth over which we must be able to transmit to approach capacity may be extremely large ([Tel88]). Such signaling assumes that we have very large bandwidth and creates marked spikes in power. Given the limited availability of bandwidth and the regulatory limitations on peak power, this sort of spreading may not be very practical. The sort of spreading we consider here is over limited bandwidths and also assumes that we do not change the type of signaling as we spread. We merely change the amount of energy per unit of bandwidth through multiplying the signal by the appropriate scaling factor.

The second remark is that, in our white noise signalling scheme, we are using all of the available degrees of freedom. Each symbol sent corresponds to a data symbol. Another way in which we could spread over frequency would be to multiply the data by a spreading sequence, as is the practice in most DS CDMA systems ([Qua91]). Each data symbol is then mapped onto several signal symbols. The distribution of the signal sent would appear to be white to an observer who did not know the spreading code, but we are not sending a white signal. Instead, we are sending repetition codes with impulse-like autocorrelation functions.

The third remark is that, when the channel is not perfectly known at the receiver but not fully described by a scattering function, it may sometimes be preferable to have some sort of repetition code rather than to have independent input signals along all the degrees of freedom. The added problem is that, since we keep the total signal energy constant while increasing the bandwidth over which we transmit, the measurement of any portion of the channel is bound to be worse in the spread case than in the unspread case. We have seen that, for an unknown but time-invariant channel, it is reasonable to send bits which do not contain information but which enable the receiver to measure the channel. Similarly, it may be preferable to repeat a data bit and thus obtain a partial measurement of the channel as well as the transmission of the bit, rather than to send two separate data bits. However, if we send the same data bit repeatedly, we shall obtain good measurements of the channel, but we shall not be able to transmit much information. There must be some trade-off between the number of data bits that are transmittted and the number of bits that are repeated for the sake of channel measurement.

We shall not consider this issue of signalling, but shall instead assume a signalling scheme and consider the trade-off between measurement of the channel and number of data bits transmitted in terms of spreading by using all of the available degrees of freedom. Understanding the case where we signal in this manner yields some insight which is necessary to study the more general case where we consider spreading and signalling schemes together.

<u>II-4-1- The issue of spreading for a single time sample over a channel known with a Gaussian error.</u>

We wish to examine whether spreading in frequency is beneficial in terms of capacity. We denote by S the baseband symbols modulated and sent over a single slice of bandwidth and by Y the corresponding output for that slice of bandwidth. F is, as in our previous discussion, the multiplicative part of the channel over that slice of bandwidth, which may have a fixed component and a probabilistic component, and N is AWGN of variance σ_N^2. Let us assume that F is statistically the same in each slice of bandwidth and that the AWGN noise components on the different slices are IID. We have assumed that the form of the distribution of S is the same in each slice and that the signals along the various slices of bandwidth are IID. Hence, $I(S;(FS+N))$ is solely a function of the SNR on each frequency slice. If we denote by E the overall input energy per signalling interval and by n the number of frequency slices we use, $I(S;(FS+N))$ is solely a function of $\frac{E}{n}$, where $\frac{E}{n}$ is the energy per signalling interval per bandwidth slice. Let us therefore define the function f as

$$f\left(\frac{E}{n}\right) = I(S;(FS+N))$$

II-4.[1].

The sign of the derivative $f'\left(\frac{E}{n}\right)$ is positive. Indeed, let us consider what happens when we increase the signal variance by multiplying S by $\sqrt{\alpha}$, where $\alpha > 1$. The mutual information after the increase is

$$I(S;SF+N) = I\left(\sqrt{\alpha}S;\sqrt{\alpha}SF+N\right)$$

$$= I\left(S;SF+\frac{N}{\sqrt{\alpha}}\right)$$

II-4.[2]

Hence, increasing the variance of S by multiplication is tantamount to reducing the variance of the AWGN. Since adding white Gaussian noise cannot be beneficial (by the Data Processing Theorem)

$$f''\left(\frac{E}{n}\right) \geq 0$$

II-4.[3].

If we have n slices of bandwidth such that we send IID signals with energy $\frac{E}{n}$ over each slice of bandwidth, spreading over m extra slices of bandwidth means that we send, over $n+m$ slices of bandwidth, IID signals with the same distribution as before but multiplied by $\frac{\sqrt{n}}{\sqrt{n+m}}$. To say that spreading over m extra slices of bandwidth is beneficial in terms of capacity is equivalent to saying that the mutual information between input and output for the second system is greater than or equal to the mutual information for the first system. Then, spreading from n to $n+m$ slices is beneficial iff

$$n\, f\left(\frac{E}{n}\right) \leq (n+m)\, f\left(\frac{E}{n+m}\right)$$

II-4.[4].

Note that, if we send a nil signal, the mutual information between input and output is 0. Therefore, if I(S; Y) is concave in $\frac{E}{n}$, then II-4.[4] holds for all positive m and n. Conversely, if I(S; Y) is convex, then spreading is detrimental, i.e. II-4.[4] holds with the opposite inequality.

However, a weaker condition is sufficient for spreading to be beneficial for all values of $\frac{E}{n}$ above some threshold δ.

If

$$\left.\frac{\partial f\left(\frac{E}{n}\right)}{\partial \frac{E}{n}}\right|_{\frac{E}{n}=\delta} \leq \frac{f(\delta)}{\delta}$$

II-4.[5]

and

$$\frac{\partial^2 f\left(\frac{E}{n}\right)}{\left(\partial \frac{E}{n}\right)^2} \leq 0$$

II-4.[6]

for all $\frac{E}{n} \geq \delta$, II-4.[2] holds as long as the variance on each slice of bandwidth exceeds δ. Figure II-4.1 illustrates our conditions. Note that, if we have enough signal power and enough bandwidth, we wish to spread so that we have δ over each slice of bandwidth. For instance, if we are transmitting over n slices of bandwidth, each with signal energy δ, then along each slice of bandwidth we obtain the mutual information shown as i in the figure. If we spread over $\frac{dn}{\delta}$ slices of bandwidth, each with signal energy δ, then each of the $\frac{dn}{\delta}$ slices has mutual information shown as i' in the figure. We see that spreading is beneficial iff $\frac{dni'}{\delta} \geq ni$, which is equivalent to $\frac{di'}{\delta} \geq i$. Figure II.13 illustrates that the latter inequality holds. Moving away from the origin on the $\frac{E}{n}$ axis indicates that we are spreading less, since the total energy E is kept fixed.

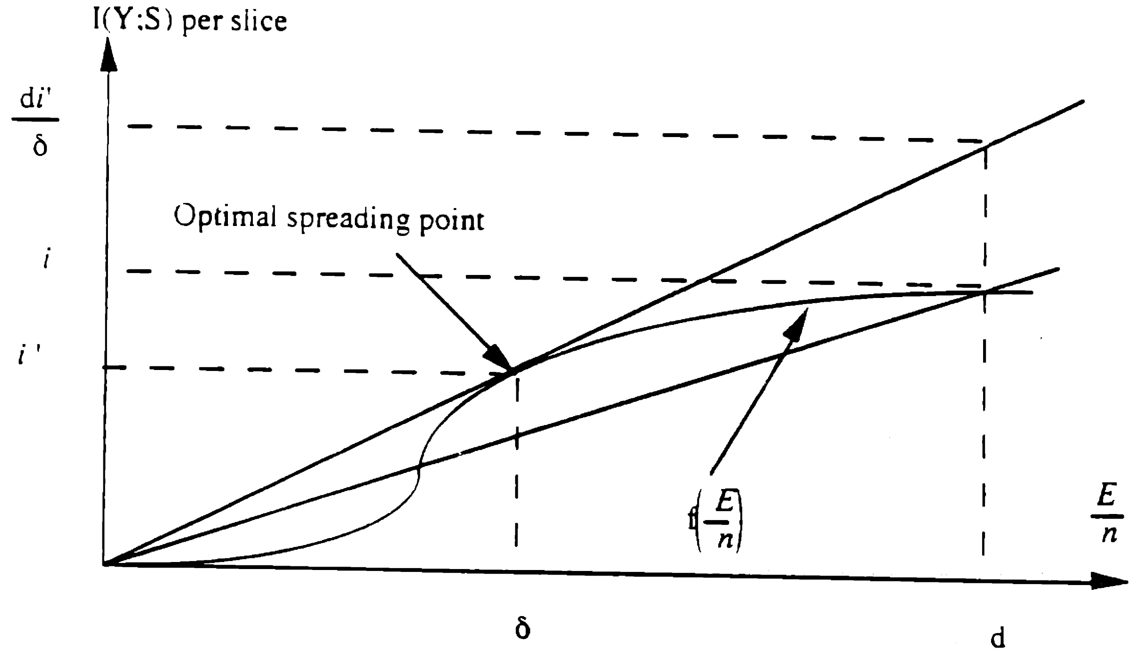


Figure II.13: Illustration of the fact that II-4.[4] holds for a certain range of $\frac{E}{n}$ if conditions II-4.[5] and II-4.[6] hold.

Note that $\frac{f\left(\frac{E}{n}\right)}{\frac{E}{n}}$ is the slope of the line from the origin to the curve $f\left(\frac{E}{n}\right)$. This slope is also $\frac{n\ f\left(\frac{E}{n}\right)}{E}$, which is the total mutual information per signalling interval divided by E. Mutual information is maximized by maximizing this slope, which occurs in the figure at $\frac{E}{n} = \delta$.

Let us assume the random variable F is zero-mean Gaussian, so that we may later extend our results to the case where the channel has a known component and a zero-mean Gaussian error. We may interpret the case where F is zero mean as a Rayleigh channel and the case where F is a known component plus a Gaussian error as a Rician channel. Note that, since we are considering very narrow slices of bandwidth, the fact that we have a Rician channel does not mean that we have a single line of sight path plus a Rayleigh component. Any arbitrary number of known paths will appear at the same tap if we consider narrow enough slices of bandwidth. Therefore, any multipath profile known with a Gaussian error will appear as Rician over a small enough bandwidth.

We may write directly from the formula for mutual information for single variables that

$$I(Y;S) = h(Y) - h(Y|S)$$

(using the formula for the entropy of a Gaussian random variable)

$$= h(Y) - \frac{1}{2} E_S\left[\ln\left(2\pi e\left(S^2\sigma_F^2 + \sigma_N^2\right)\right)\right]$$

II-4.[7].

We may upper bound the first term of the RHS of II-4.[7] by the entropy of a Gaussian random variable with the same variance. Hence, we may upper bound II-4.[7] by

$$I(Y;S) \le \frac{1}{2}\ln\left(\frac{E}{n}\sigma_F^2 + \sigma_N^2\right) - \frac{1}{2}E_S\left[\ln\left(S^2\sigma_F^2 + \sigma_N^2\right)\right]$$

$$= -\frac{1}{2}E_S\left[\ln\left(\frac{S^2\sigma_F^2 + \sigma_N^2}{\frac{E}{n}\sigma_F^2 + \sigma_N^2}\right)\right]$$

$$= -\frac{1}{2}E_S\left[\ln\left(1 + \frac{\left(S^2 - \frac{E}{n}\right)\sigma_F^2}{\frac{E}{n}\sigma_F^2 + \sigma_N^2}\right)\right]$$

II-4.[8].

Let us assume that the distribution of S is such that the fourth central moment of S is upper bounded by a multiple of the square of $\frac{E}{n}$. Such an assumption is not restrictive, since we assume that we have a fixed distribution for S with a scale factor. From the lower bound $-\ln(1+x) \le \frac{x^2}{2}$ for $x \ge 0$, we may upper bound the RHS of II-4.[8] to obtain

$$I(Y{:}S) \le \frac{1}{2}\ln\left(\frac{E}{n}\sigma_F^2 + \sigma_N^2\right) \le \frac{1}{4}E_S\left[\frac{\left(S^2 - \frac{E}{n}\right)^2\sigma_F^4}{\left(\frac{E}{n}s_F^2 + \sigma_N^2\right)^2}\right]$$

II-4.[9].

From our assumption that there exists a C such that $E\left[\left(S^2 - \frac{E}{n}\right)^2\right] \le C\left(\frac{E}{n}\right)^2$, we may further upper bound II-4.[9] by

$$\frac{1}{4}E_S\left[\frac{\left(S^2 - \frac{E}{n}\right)^2\sigma_F^4}{\left(\frac{E}{n}\sigma_F^2 + \sigma_N^2\right)^2}\right] \le \frac{1}{4}\frac{C\left(\frac{E}{n}\right)^2\sigma_F^4}{\left(\frac{E}{n}\sigma_F^2 + \sigma_N^2\right)^2}$$

II-4.[10].

The derivative of the RHS of II-4.[10] with respect to $\frac{E}{n}$ is $\frac{C\sigma_F^4\frac{E}{n}}{2}\left(1 - \frac{\frac{E}{n}\sigma_F^2}{\left(\frac{E}{n}\sigma_F^2 + \sigma_N^2\right)^3}\right)$. Hence, from II-4.[8]-II-4.[10], we see that I(S;Y) cannot be strictly concave in $\frac{E}{n}$, because for $\frac{E}{n}$ close enough to 0, it is upper bounded by a function which approaches 0 and whose first derivative approaches 0 as $\frac{E}{n}$ approaches 0. Figures II.14 and II.15 illustrate the mutual information for a Rayleigh channel and binary signaling. More examples are in appendix E.

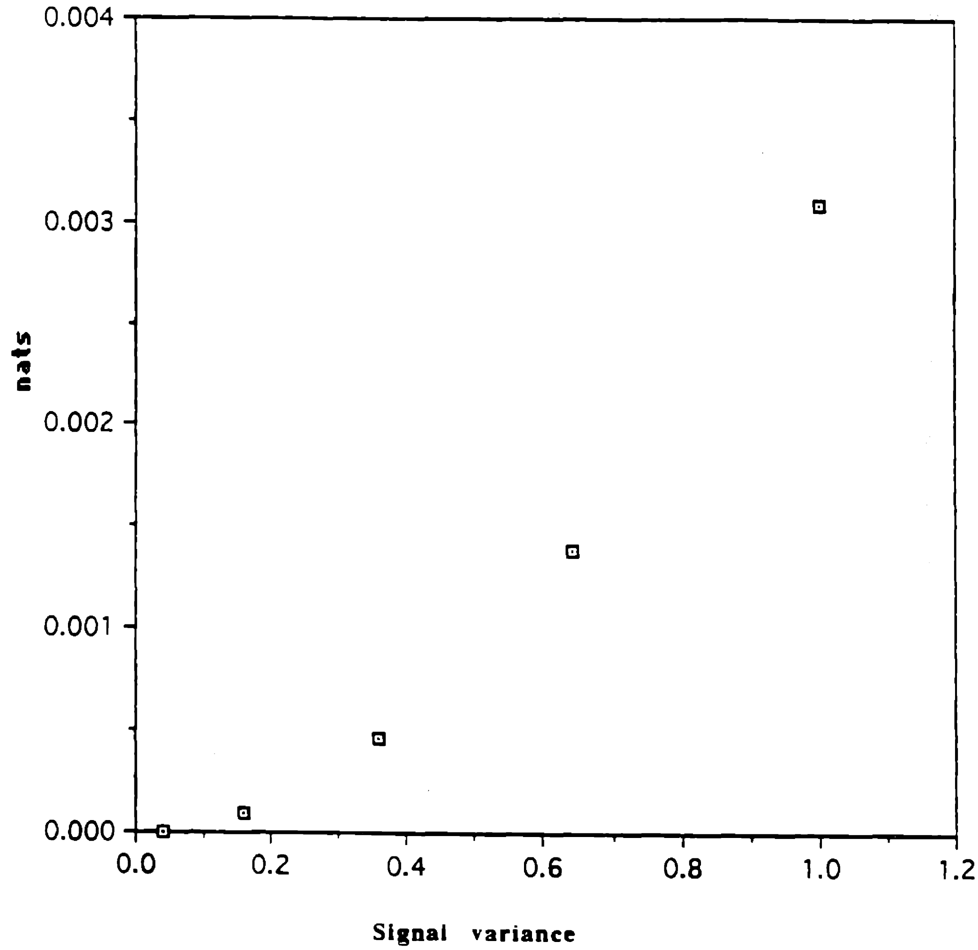


Figure II.14 : Mutual information for 0,1 signaling in a Rayleigh channel with noise variance 4, multiplicative channel variance 1.

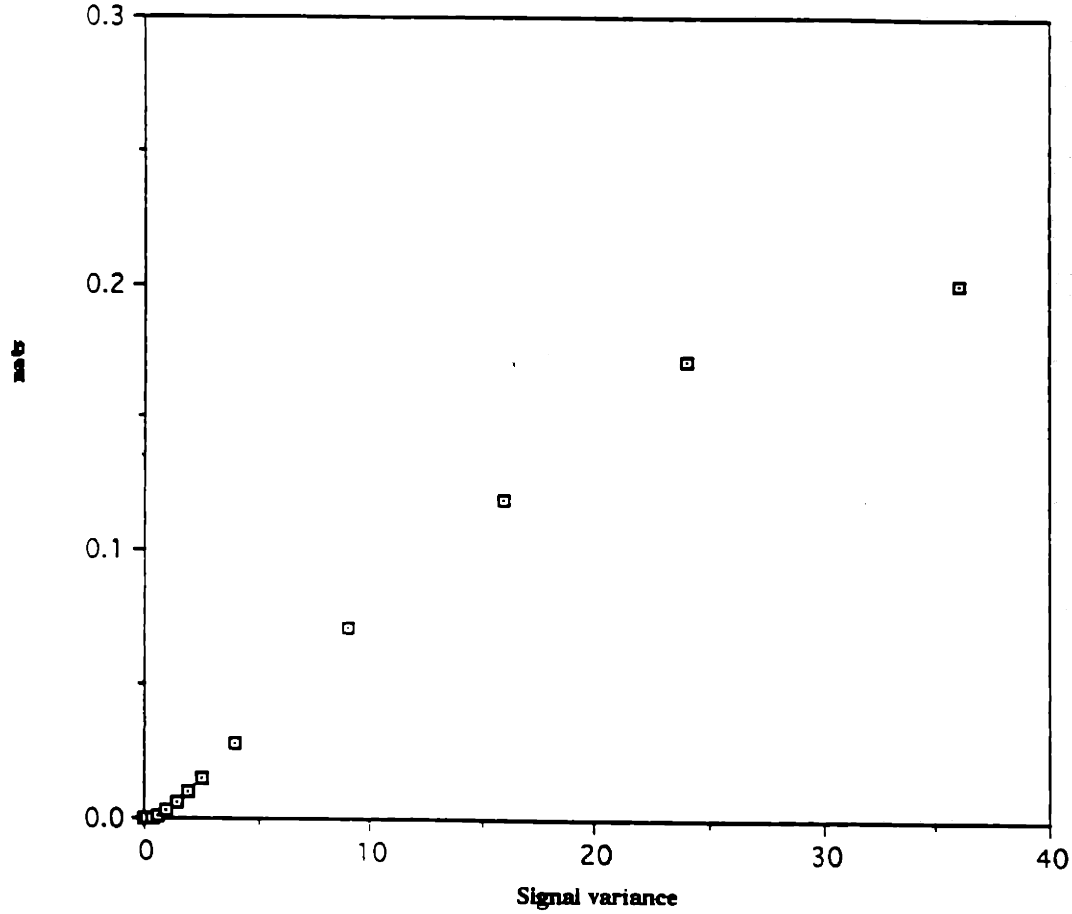


Figure II.15: Mutual information for 0,1 signaling in a Rayleigh channel with noise variance 4, multiplicative channel variance 1.

Let us now consider the case the case where the channel is partially known as a constant with zero-mean Gaussian error of given variance. If the channel is perfectly known, we know that spreading is advantageous. Let us fix the channel energy, i.e. the sum of the error variance and of the square of the known part of the channel.

We denote the energy of the channel by $E[F^2]$. The error variance is $\beta E[F^2]$, where b is in [0,1]. The parameter β represents the accuracy of our knowledge of the channel. Expression II-4.[8] becomes

$$I(Y;S) \le \frac{1}{2}\ln\left(\frac{E}{n}E\left[F^2\right] + \sigma_N^2\right) - \frac{1}{2}E_S\left[\ln\left(S^2 E\left[F^2\right]\beta + \sigma_N^2\right)\right]$$

II-4.[11].

Let us define

$$\tilde{f}\left(\frac{E}{n}E\left[F^2\right]\right) = \frac{1}{2}\ln\left(\frac{E}{n}E\left[F^2\right] + \sigma_N^2\right) - \frac{1}{2}E_S\left[\ln\left(S^2 E\left[F^2\right]\beta + \sigma_N^2\right)\right]$$

$$= -\frac{1}{2}E_S\left[\ln\left(\frac{S^2 E\left[F^2\right]\beta + \sigma_N^2}{\frac{E}{n}E\left[F^2\right] - \sigma_N^2}\right)\right]$$

$$= -\frac{1}{2}E_S\left[\ln\left(\frac{S^2 E\left[F^2\right]\beta - \frac{E}{n}E\left[F^2\right]}{\frac{E}{n}E\left[F^2\right] + \sigma_N^2} + 1\right)\right]$$

II-4.[12],

where S has variance $\frac{E}{n}$. The function defined in II-4.[12] is differentiable and we may use the Taylor series expansion of ln(1+x) to write that :

$$\tilde{f}\left(\frac{E}{n}E\left[F^2\right]\right) = -\frac{1}{2}E_S\left[\frac{S^2 E\left[F^2\right]\beta - \frac{E}{n}E\left[F^2\right]}{\frac{E}{n}E\left[F^2\right] + \sigma_N^2} - \frac{1}{2}\left(\frac{S^2 E\left[F^2\right]\beta - \frac{E}{n}E\left[F^2\right]}{\frac{E}{n}E\left[F^2\right] + \sigma_N^2}\right)^2 + \dots\right]$$

II-4.[13].

The second term of the RHS of II-4.[13] approaches 0 as $\frac{E}{n}$ approaches 0 faster than the first term. As $\frac{E}{n}$ approaches 0, the derivative of the first term with respect to $\frac{E}{n}$ approaches $\frac{(1-\beta)E\left[F^2\right]}{2\sigma_N^2}$. Therefore, we see that $\frac{\partial\tilde{f}\left(\frac{E}{n}E\left[F^2\right]\right)}{\partial\frac{E}{n}}$ approaches $\frac{(1-\beta)E\left[F^2\right]}{2\sigma_N^2}$ as

$\frac{E}{n}$ approaches 0. Therefore, for β close enough to 1, the function f is upper bounded by a function whose value at 0 is 0 and whose first derivative is very small. Therefore, spreading cannot be advantageous for small enough values of $\frac{E}{n}$. Note that our results hold for any distribution on the input.

Let us look at whether, for β small enough, spreading is always beneficial. For a Gaussian distribution on S, we know that for $\beta = 0$, spreading is always beneficial. For both these distributions, the second derivative of I(Y;S) with respect to $\frac{E}{n}$ is continuous in β. If spreading is always beneficial for a perfectly known channel for some signal distribution and if the second derivative of I(Y;S) with respect to $\frac{E}{n}$ is continuous in β, then there is a range of values of β near 0 for which spreading is always advantageous for that type of signaling. Figures II.16 and II.17 show the mutual information for a Rician channel for 0,1 signalling. We see that, as expected, when the channel is poorly known, there is a convex region for low values of $\frac{E}{n}$. When the channel is well known, I(Y;S) is concave in $\frac{E}{n}$ over all values of $\frac{E}{n}$ More examples may be found in appendix E.

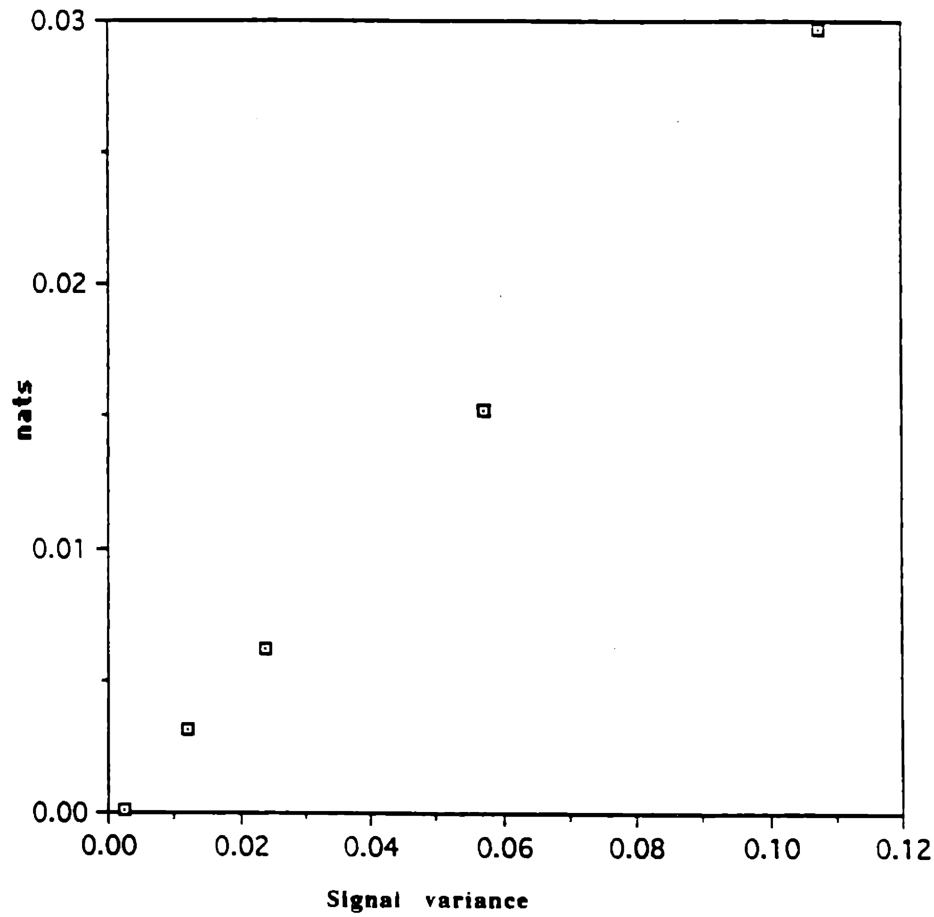


Figure II.16: Mutual information for 0,1 signaling in a Gaussian channel with a known part having 50 % of the channel energy, noise variance 4, multiplicative channel variance 4.

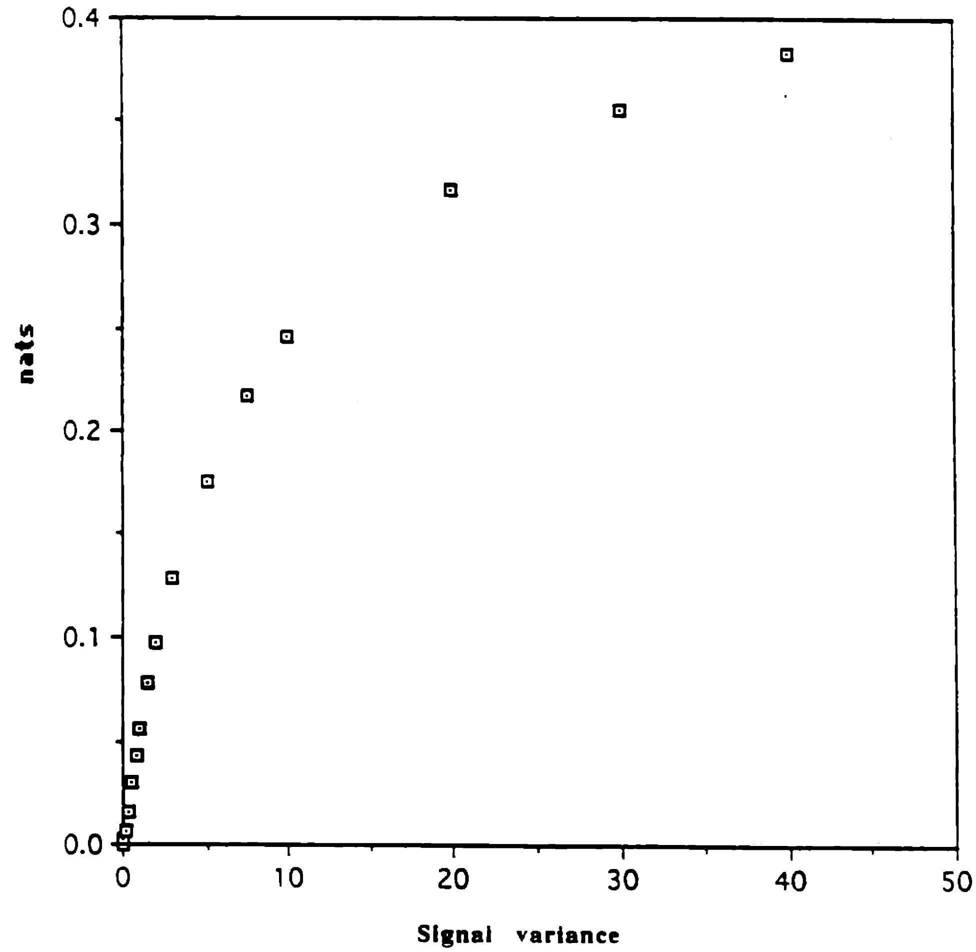


Figure II.17 : Mutual information for 0,1 signaling in a Gaussian channel with a known part having 50 % of the channel energy, noise variance 4, multiplicative channel variance 4.

II-4-2- Spreading for several time samples over an unknown Gaussian channel.

All our previous discussion has dealt with a single time sample, but we would in general wish to examine many time samples so that we may say something about capacity. The main problem in examining several time samples versus just one time sample is that the variance of our error in knowing the channel changes from time to time. Let us first consider the extreme case where each channel sample is independent of the next. Since we sample at $W = W_{input} + B_{Doppler}$, we must be transmitting at a very low rate and $B_{Doppler}$ must be the dominant term in W. In that case, our discussion for the single time sample case carries over. In general, if we have several time samples, at each time sample there is some distribution for the channel based upon our estimate of the channel at that time sample. We may say that as long as $\frac{E}{n} \geq \delta$ for the maximum δ over all possible channel distributions corresponding to the channel estimates at different time samples, then spreading is advantageous. The real difficulty lies in defining what these channel distributions are for the different time samples. How the error on the channel varies in time is difficult to express. If we assume that we decode each symbol correctly, then we may use a Kalman filter to estimate the channel optimally. However, the variance of the error on the channel at any time will depend on the specific symbol sequence up until that time. If we periodically sound the channel with a known sequence, then the error on the channel will be lower bounded by the error variance right after we have sounded the channel plus the variability of the channel over the period between soundings. Overall, predicting the distribution of the channel at each time sample is a difficult task, so we give here some bounds based on the behavior of f.

Let us use the chain rule on mutual information to rewrite the mutual information between an input sequence and an output sequence:

$$I(\underline{S}_i; \underline{Y}_i) = \sum_{j=1}^{i} I(\underline{S}_i; Y[j] \mid \underline{Y}_{j-1})$$

(assuming the inputs symbols are IID and the channel has a single tap)

$$\geq \sum_{j=1}^{i} I(S[j]; Y[j] \mid \underline{Y}_{j-1})$$

[II-4.[14].

Our single tap assumption is not restrictive, since we consider many small contiguous slices of bandwidth and a narrowband single tap model is appropriate for each of these slices of bandwidth. We see that by increasing $\frac{E}{n}$, we have two effects on

$I(S[j]; Y[j] \mid \underline{Y}_{j-1})$. One effect is to improve the measurement of the channel that we obtain from $\underline{Y}_{j-1}$. The other effect is to improve the SNR for time sample j. We may define as in II-4.[4] the functions

$$f_j\left(\frac{E}{n}\right) = I(S[j]; Y[j] \mid \underline{Y}_{j-1}) \qquad \text{II-4.[15]}$$

for all j>1.The threshold at which spreading is advantageous for the different f_j is unknown and it is not clear that it is bounded. We may however make the following remark. Assuming the inputs symbols are IID and the channel has a single tap, we have

$$f\left(\frac{E}{n}\right) \leq f_j\left(\frac{E}{n}\right) \leq \mathcal{f}\left(\frac{E}{n}\right) \qquad \text{II-4.[16]}$$

for f defined as in the previous section, i.e. $f = f_1$, and for $\mathcal{f}$ defined as

$$\mathcal{f}\left(\frac{E}{n}\right) = I(S; (FS + N) \mid F) \qquad \text{II-4.[17].}$$

The function $\mathcal{f}$ we know to be concave in $\frac{E}{n}$. Let us suppose that there exists a δ' such that $\delta \leq \delta'$ and

$$\frac{\delta'}{\delta} f\left(\frac{E}{n}\right)\Bigg|_{\left(\frac{E}{n}\right)=\delta} \geq \mathcal{f}\left(\frac{E}{n}\right)\Bigg|_{\left(\frac{E}{n}\right)=\delta'} \qquad \text{II-4.[18].}$$

Figure II.18 illustrates this condition.

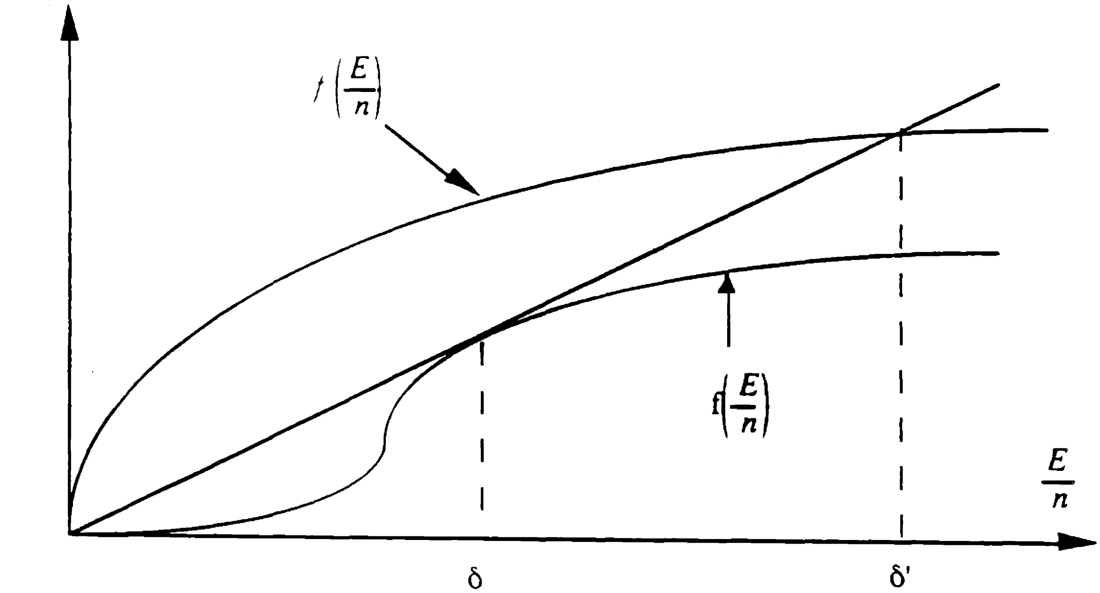


Figure II.18 : Illustration of expression II-4.[18].

If $\frac{E}{n} \geq \delta'$ on a frequency slice, it is preferable in terms of mutual information to spread over enough frequency so that $\frac{E}{n} = \delta$ over each frequency slice and not to know the channel than to not spread and know the channel. Therefore, from II-4.[16], we see that spreading over enough frequency so that $\frac{E}{n} = \delta$ over each frequency slice is advantageous for all mutual informations $I(S[j]; Y[j] \mid \underline{Y}_{j-1})$ if $\frac{E}{n} \geq \delta'$. Therefore, it is always advantageous to spread at least until $\frac{E}{n} = \delta$ over each frequency slice if II-4.[18] holds.

We still do not know how to determine where it is advantageous to spread. Let us use the results of II-3-3-c for Gauss-Markov channels. If we consider to simplify our illustration that we have real taps and send real signals, we have from II-3.[100] that

$$\lim_{j\to\infty}\left(f\left(\frac{E}{n}\right) - f_j\left(\frac{E}{n}\right)\right) \leq \frac{1}{2}\ln\left(\frac{\left(1-\alpha^2\right)\sigma_F^2}{\lambda'} + \alpha^2\right)$$

II-4.[19]

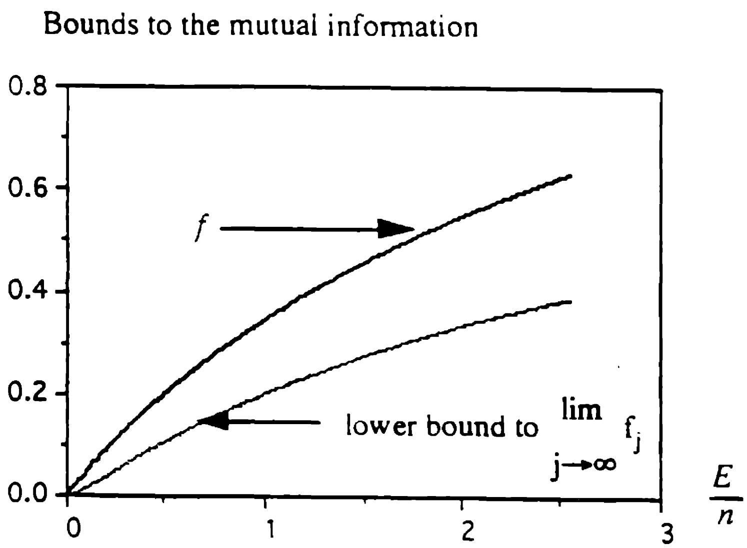
or equivalently

$$f\left(\frac{E}{n}\right) - \frac{1}{2}\ln\left(\frac{\left(1-\alpha^2\right)\sigma_F^2}{\lambda'} + \alpha^2\right) \leq \lim_{j\to\infty} f_j\left(\frac{E}{n}\right)$$

II-4.[20].

Although a changes for the whole channel with W, we do not need to take that change into account because we always look at a slice of bandwidth of fixed size. From II-3.[94], we see that as $\frac{E}{n}$ approaches 0, λ' approaches σ_F^2, therefore the RHS of II-4.[19] approaches 0. We may compute the lower bound to $\lim_{j\to\infty} f_j\left(\frac{E}{n}\right)$ given by the RHS of II-4.[20] and find its maximum slope for small $\frac{E}{n}$, i.e. large n for E fixed. The point at which this maximum slope occurs gives us an upper bound on δ. The intersection of the line through the origin with this slope and f yields an upper bound on δ'.

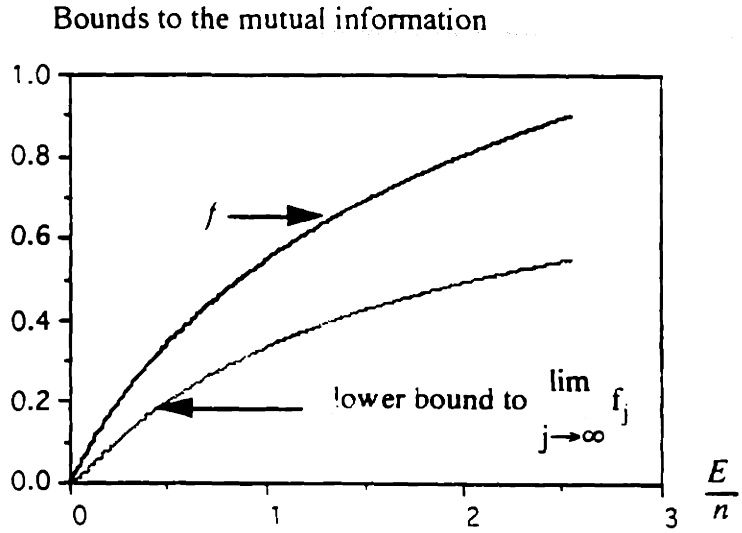
Figures II.19 and II.20 illustrate how our bounds may be used to obtain an upper bound on δ' for two examples where $B_{Doppler}$ = 60 Hz. We send white Gaussian signals. Suppose that we take $\gamma = 0.37$, $T_{coherence} = 1/B_{Doppler}$ and the bandwidth of each slice of spectrum to be 1000 Hz. From II-3.[87], we may compute α to be of the order of 0.9 for $B_{Doppler}$ = 60 Hz and of the order of 0.8 for $B_{Doppler}$ = 200 Hz. Appendix F gives another example for $B_{Doppler}$ = 60 Hz and three examples for $B_{Doppler}$ = 200 Hz.



maximum slope for lower bound is 0.23768772 for $\frac{E}{n} = 0.22$

yielding $\delta' \leq 2.82$

Figure II.19 : Graph of f and $\lim_{j\to\infty} f_j$ versus $\frac{E}{n}$ for $B_{Doppler} = 60$ Hz, spectrum bands of 1000 Hz, γ =.037, channel strength 1 and noise variance 1



maximum slope for lower bound is 0.47534552 for $\frac{E}{n} = 0.11$

yielding $\delta' \leq 1.41$

Figure II.20 : Graph of f and $\lim_{j \to \infty} f_j$ versus $\frac{E}{n}$ for $B_{Doppler} = 60$ Hz, spectrum bands of 1000 Hz, $\gamma = .037$, channel strength 2 and noise variance 1 .

Chapter III - The multiple user case in time-varying channels without feedback.

III-1 - Introduction.

The purpose of Chapter II was to study the effect of time variations on the capacity of a channel without having to take into account the effect of multiple access. In mobile communications, we are interested in having several sources rather than a single source. We now add multiple access considerations to extend the results of Chapter II to several sources. We first extend the model of Chapter II to accommodate several users and then outline the extensions we propose.

As in the previous chapter, each source has its own time-varying channel. It is reasonable to assume that the channels of the different sources are mutually independent. Figure I.1 shows how each source experiences different multipath because it is at a different location than the other sources.

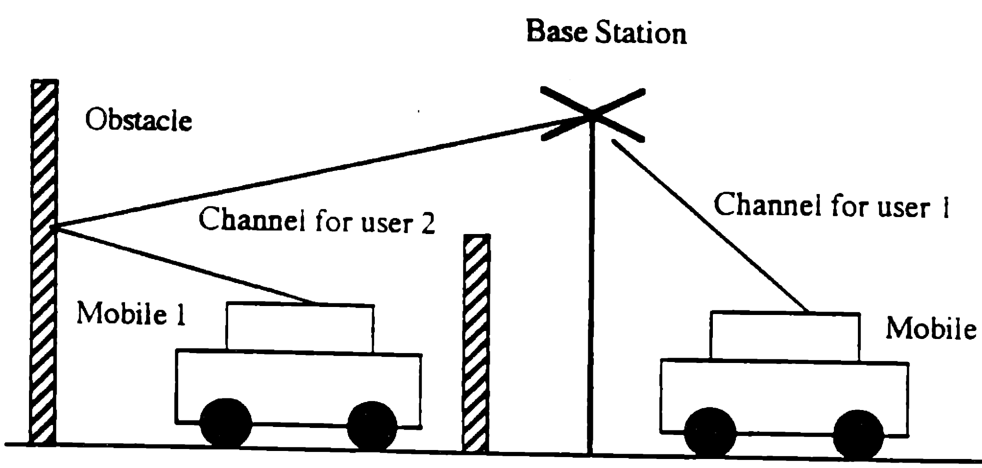


Figure III.1 : Illustration of different channels for different users.

We assume that there is a single portion of spectrum available to all the sources. The users may or may not transmit simultaneously over some or all the bandwidth. We follow the structure of the previous chapter to show how our results for the single-user case extend to the multiple-user case. We first present the time-varying model for a multiple-user multipath channel. We study the multiple access capacity region when the time-varying channels of all the users are perfectly known. When the users cooperate by sending possibly correlated signals, the capacity region may be found by techniques similar to the single user case. When the users are constrained to send independent signals, the problem does not have a simple closed form solution. As in Chapter II, we next consider the case where the channels are not perfectly known. We show how the results of Chapter II for channels with some small Gaussian error and for μ^{th} order Markov channels apply to the multiple user case. In particular, we discuss how our results for the single-user case provide insight into the propagation of the channel measurement error when we perform interference cancellation. Finally, we consider the effectiveness of spreading when we have multiple users in time-varying channels.

III-2- Known channel system.

III-2-1- Model for the multipath channel.

In the following section, we look at the channel response at baseband for a multipath channel. The goal is to develop a continuous time and a tapped-delay line model. In order to derive such models, we first consider the continuous-time response of a known channel and then look at its samples. The input bandwidth, of size W_{input}, is shared by all users. This does not necessarily mean that all users simultaneously transmit over the same portions of the spectrum, although we may choose to have them overlap in time and/or frequency. The samples are chosen so that, using the Nyquist theorem, they are sufficient to reconstitute the continuous-time response. We see from our discussion for the single user case, that it is sufficient to take W to be $W_{input} + \max_{i=1,\ldots,K} \left\{B^i_{Doppler}\right\}$ where $B^i_{Doppler}$ is the Doppler spread associated with the i^{th} channel. Therefore, we choose the sampling interval 1/W to be the smallest of all the sufficient sampling rates for all users.

The continuous-time system which we consider is represented in figure III.2. Discrete data streams from the different users are passed through modulators, which produce continuous signals. The signal from each modulator passes through a different channel. The output of all the channels is added before being received at the receiver. For a given set of channels, modulators, receivers and initial transmission times t_0, ..., t_K for the K users, we may describe a system by figure III.2. The initial transmission times may be incorporated in the channels as arbitrary delays, although in practice they may be considered separately from the channel. Therefore, the model of figure II.1 is adequate for symbol asynchronous systems.

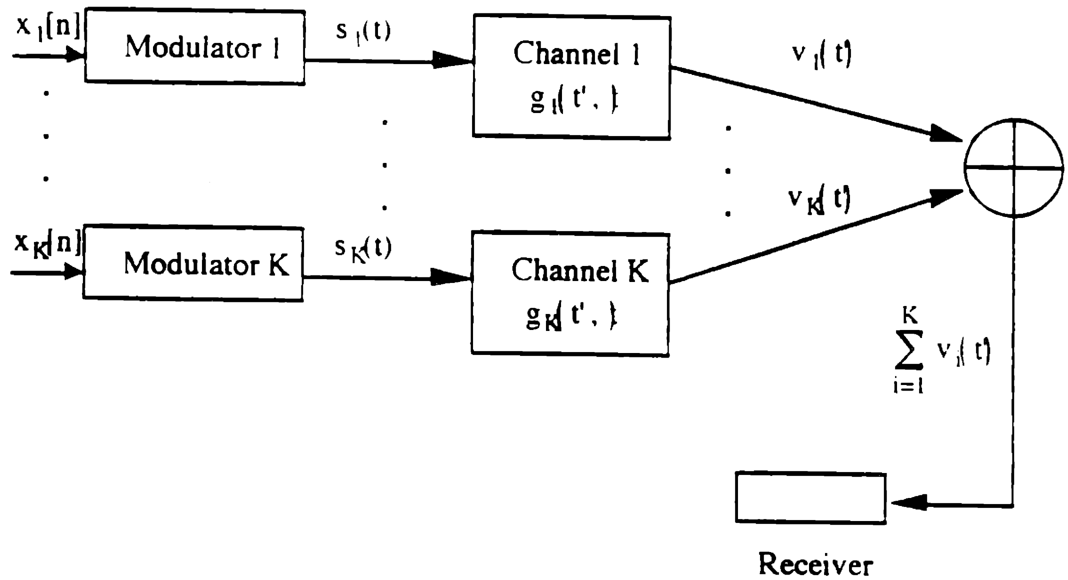


Figure III.2 : Continuous-time multipath system with multiple users.

The impulse response of the multipath channel seen by user i at time t' for a transmission sent at time t'- t is:

$$g_i(t',t) = \sum_{m=1}^{P_i} a_i^m(t')\,\delta\left(\tau_i^m(t') - t\right)$$

III-2.[1]

where P_i is the number of paths in the channel seen by user i, a_i^m is the complex multiplicative constant associated with the m^{th} path of user i (accounting for fading and phase shift) and τ_i^m is the delay associated with the m^{th} path of user i. The issue of Doppler spread was addressed in the previous chapter and the same issues are applicable to each user's channel in the multi-user situation.

Let $s_i(t)$ be the complex signal transmitted by user i. This signal might typically be a data stream modulated by, first, a signature sequence for differentiating users and, second, a modulating waveform for transmission on the channel. For K users, where the channel for user i has P_i paths, the response of the channel is

$$v(t') = \sum_{i=1}^{K} v_i(t')$$

$$= \sum_{i=1}^{K} \int_{t=-\infty}^{+\infty} s_i(t'-t) g_i(t', t)\, dt$$

$$= \sum_{i=1}^{K} \sum_{m=1}^{P_i} \int_{t=-\infty}^{+\infty} s_i(t'-t) a_i^m(t')\, \delta\left(\tau_i^m(t') - t\right) dt$$

III-2.[2].

Let the response be sampled at time intervals of 1/W, where v(t') is bandlimited to [-W/2, +W/2]. As in the previous chapter, we obtain a discrete-time sequence defined as

$$v[k] = v(k/W)$$

III-2.[3]

and we may similarly define for i=1, ..., K

$$s_i[k] = s_i(k/W)$$

III-2.[4].

We wish to determine a discrete channel impulse response to relate v[k] to the K sequences $s_i[k]$. By extending the discussion for the single-user case, we may write that

$$v(t') = \sum_{i=1}^{K} \sum_{m=1}^{P_i} \int_{t=-\infty}^{+\infty} a_i^m(t') \left\{ \sum_{n=-\infty}^{+\infty} s_i[n] \operatorname{sinc}\left\{ \pi W \left(t' - t - \frac{n}{W} \right) \right\} \right\} \delta\left(\tau_i^m(t') - t\right) dt$$

$$= \sum_{i=1}^{K} \sum_{m=1}^{P_i} \sum_{n=-\infty}^{-\infty} s_i[n] \int_{t=-\infty}^{+\infty} a_i^m(t') \operatorname{sinc}\left\{\pi W\left(t' - t - \frac{n}{W}\right)\right\} \delta\left(\tau_i^m(t') - t\right) dt$$

III-2.[5].

Therefore

$$v[k] = \sum_{i=1}^{K} \sum_{m=1}^{P_i} \sum_{n=-\infty}^{+\infty} s_i[n] \int_{t=-\infty}^{+\infty} \operatorname{sinc}\left\{\pi W\left(\frac{k}{W} - t - \frac{n}{W}\right)\right\} a_i^m\left(\frac{k}{W}\right) \delta\left(\tau_i^m\left(\frac{k}{W}\right) - t\right) dt$$

$$= \sum_{i=1}^{K} \sum_{m=1}^{P_i} \sum_{n=-\infty}^{-\infty} s_i[k-n] \int_{t=-\infty}^{+\infty} \operatorname{sinc}\left\{\pi W\left(\frac{n}{W} - t\right)\right\} a_i^m\left(\frac{k}{W}\right) \delta\left(\tau_i^m\left(\frac{k}{W}\right) - t\right) dt$$

III-2.[6] .

Hence,

$$v[k] = \sum_{i=1}^{K} \sum_{m=1}^{P_i} \sum_{n=-\infty}^{+\infty} s_i[k-n] g_i^m[k, n]$$

III-2.[7]

where

$$g_i^m[k, n] = a_i^m\left(\frac{k}{W}\right) \operatorname{sinc}\left\{\pi W\left(\frac{n}{W} - \tau_i^m\left(\frac{k}{W}\right)\right)\right\}$$

III-2.[8].

As in the single user case, our initial assumption that W is large enough that v(t) may be fully reconstituted from v[k] ensures that III-2.[7] gives a complete characterization of the channel response. Figure III.3 shows the discrete-time system, which may be freely substituted for the continuous-time one in figure III.2, with

$$g_i[k,k-n] = \sum_{m=1}^{P_i} g_i^m[k,k-n]$$

III-2.[9].

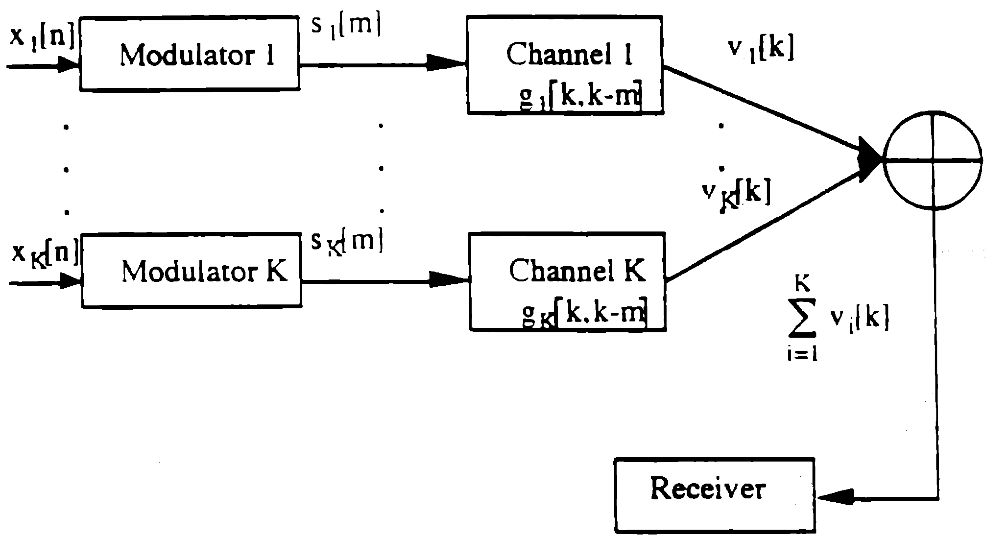


Figure III.3 : Discrete-time multipath system with multiple users.

As in the single-user case, we need to take noise into account and we assume that we have AWGN at the receiver. As in the previous chapter, we sample the bandlimited noise at rate 1/W to have the discrete-time noise variable n[k]. To take the effect of noise into account, we add the term n[k] to v[k] in the discrete-time case to obtain v[k].

As in the single-user case, there may be an infinite number of terms involved in the discrete-time expression. The conditions under which we may approximate the discrete-time case by a finite vector expression may be easily extended from the single-user case. We see from expression III-2.[8] that there are infinitely many terms $g_i^m[k, n]$, whereas in the continuous case there is, for any given t, at most a single t' which yields a non-zero value for $g_i^m(t', t)$. As in the previous chapter, we wish to establish an expression for the discrete-time output of the channel which is a product of finite matrices. We know that the quantities $g_i^m[k, n]$, for $n/W > T_{spread,i}$, where $T_{spread,i}$ is the multipath spread for the channel seen by user i, become vanishingly small as n increases. Figure II.7 in Chapter II holds for every $g_i^m[k, n]$. Moreover, we may include the fact that the channel is causal by ignoring future inputs.

We choose some arbitrary cut-off in n such that $g_i^m[k,n]$ is taken to be zero beyond that point. Therefore, we approximate the sampled channel output by :

$$y[k] \approx \sum_{i=1}^{K} \left[g_i[k,\Delta_i] \ldots g_i[k,0]\right] \begin{bmatrix} s_i[k-\Delta_i] \\ \ldots \\ s_i[k] \end{bmatrix} + n[k]$$

III-2.[10]

where Δ_i is some integer satisfying $\Delta_i > WT_{spread,i}$.

We assume, as an extension of the single user case, that we have chosen $\Delta > \max_{i=1,\ldots,K} \{T^i_{spread}\}$, so that we take $g_i[k,n] = 0$ for $n > \Delta$. We may then write, with the obvious extensions of notation from the single user case, that

$$\underline{y}_k = \sum_{i=1}^{K} \underline{f}_{ik}\, \underline{s}_{ik} + \underline{n}_k$$

III-2.[11].

We shall use this expression later when trying to compute the capacity of the system.

III-2-2- Capacity in the case of a perfectly known channel.

III-2-2-a- Constant single path channels.

When each channel has a constant single path, since we have a memoryless channel with AWGN, we may compute the achievable rate region as found in [Ahl71] and [Lia72]. Let R_i be the rate per second achieved for used i. For K users, we have that, for each subset U of users,

$$\sum_{i\in U} R_i \le \frac{1}{T} I\left(Y ; \{S_i\}_{i\in U} \mid \{S_j\}_{j\in \overline{U}}\right)$$

III-2.[12].

Therefore, when each user has the same average power constraint P, the capacity region is given by the set of inequalities

$$\sum_{i \in U} R_i \leq \frac{\lfloor TW \rfloor}{T} \ln\left(\frac{\left(a^2 P\,|U| + WN_0\right)}{WN_0}\right)$$

III-2.[13]

where U ranges over all the possible subsets of users.

For each of these constraints, the behavior when W increases is the same as in the single-user case.

III-2-2-b- Specular paths.

The general multipath model we have developed for the multiple acces channel has memory. The capacity region for a certain class of multiple-access channels with memory was found in [Ver89]. However, we must carefully explain the meaning of memory. In [Ver89], the memory is assumed to satisfy the constraints that "the outputs (...) are conditionally independent given the inputs, and each output depends on m consecutive inputs of each user, thus encompassing intersymbol interference of finite duration". Thus, this capacity region allows us to find the capacity region of the multiple-user channel when there is perfectly known multipath *if the multipath is constant*. If the multipath is constant, then we do not encounter fading, and therefore we can perform water-filling once and for all. The solution to this multiple-user water filling problem in channels with time-invariant memory is given in [CV93].

Since we are looking at a multiple access situation rather than at a single user channel, we must consider a rate region rather than a single rate. Therefore, the concept of maximization no longer holds in a straightforward manner. We may choose to maximize the sum of all the rates, or the sum of any subset U of the rates, or the sum of all the rates subject to the rates all being equal, etc... The power constraint may also be different - we may consider that each user has the same power constraint, or that the constraint is on the sum of the powers of all the users, etc... The extension from the single user case to the multiple user case depends on what knowledge of the channel we have at the senders and the receiver. The receiver, however, knows the channels of all the users, hence it does not have to consider all the users independently when it decodes. In particular, it can perform some sort of interference cancellation, especially at times when the fading along the different channels makes the users unequal in instantaneous received power. Note that this scheme still assumes that each sender is told by the receiver at what rate it should send. Indeed, only the receiver can compute the capacity region for the system, since only the receiver knows all the channels.

The cases we shall study will assume that each user knows the channels of all the other users. We shall first solve the problem of maximizing the sum of all the rates, $\sum_{j=1}^{K} R_j$, subject to a constraint on the sum, P, of all the powers, since this is the most straightforward extension of the single user case. We do not restrict the users to be independent of each other, and hence the users act as an antenna array. However, such a model would not be applicable to the case where we have independent users. For users to cooperate in their transmission, they must be linked to each other in some fashion. Unless the users are co-located, as would be the case if we considered each antenna in an antenna array to be a user, such links among users do not exist. Therefore, users, would be independent for most cases of interest. Moreover, having a constraint on the total amount of power rather than the power of the individual users is not realistic. Each user has intrinsic limitations (from its hardware, regulatory concerns, or other factors) which are independent of the behavior of other users.

Therefore, after solving the case above, we shall consider a case where each user has its own individual power constraint, its signal is independent of the other users' signals, and we wish to maximize the sum of the rates. Such a case would be reasonable if the users were sending, for instance, data, with very loose delay requirements. Each user is subject to its own power limitations but the different users may be getting different grades of service.

III-2-2-b-a- Maximizing the total rate under a total power constraint over all users and no independence constraint.

We look at the case where all the users know each other's channels and may or may not transmit independently. The problem is similar to the one dimensional problem, except that we now consider that the channel has more degrees of freedom at the input than at the output. We shall use complex notation rather than consider twice the number of dimensions for real vectors. Following the notation of III-2.[36]-[41] in Chapter II, we may write that

$$[Y[1] \ldots Y[2k]]^T = \sum_{i=1}^{K} \tilde{\underline{f}}_{i2k'}^{2k} [\tilde{S}_i[1] \ldots \tilde{S}_i[2k']]^T + [N[1] \ldots N[2k]]^T$$

III-2.[14].

We see that we have M = 2Kk' input degrees of freedom and 2k output degrees of freedom. Indeed, we may rewrite III-2.[3] as

$$[Y[1] \ldots Y[2k]]^T = \hat{\underline{f}}_M^{2k} [\hat{S}[1] \ldots \hat{S}[M]]^T + [N[1] \ldots N[2k]]^T$$

III-2.[15]

where we have defined

$$[\hat{S}[1] \ldots \hat{S}[M]] = [\tilde{S}_1[1] \ldots \tilde{S}_1[2k'], \tilde{S}_2[1] \ldots \tilde{S}_2[2k'], \ldots, \tilde{S}_K[1] \ldots \tilde{S}_K[2k']]$$

III-2.[16]

and

$$\hat{\underline{f}}_M^{2k} = \left[\tilde{\underline{f}}_{1\,2k'}^{\;2k}, \tilde{\underline{f}}_{2\,2k'}^{\;2k}, \ldots, \tilde{\underline{f}}_{K\,2k'}^{\;2k}\right]$$

III-2.[17].

Let $\hat{\lambda}_1, \ldots, \hat{\lambda}_{2k}$ be the eigenvalues of $\hat{\underline{f}}_M^{2k} \hat{\underline{f}}_M^{2k^T}$. These eigenvalues are real and non-negative. Using Theorem 8.4.1 in [Gal], for each positive eigenvalue $\hat{\lambda}_i$ there exists a vector $\hat{\underline{\varphi}}_{i2k}$ such that

$$\hat{\underline{f}}_M^{2k} \hat{\underline{f}}_M^{2k^T} \hat{\underline{\varphi}}_{ik} = \hat{\lambda}_i \hat{\underline{\varphi}}_{i2k}$$

III-2.[18]

and the vectors $\hat{\underline{\varphi}}_{i2k}$ form an orthonormal set. We shall assume for ease of exposition that all $\hat{\lambda}_i$ are positive. Moreover, there exist 2k orthonormal vectors $\hat{\underline{\theta}}_{iM}$ such that

$$\hat{\underline{f}}_M^{2k}\, \hat{\underline{\theta}}_{iM} = \sqrt{\hat{\lambda}_i}\, \hat{\underline{\varphi}}_{i2k} \qquad \text{III-2.[19].}$$

Let us complete the orthonormal basis with M-2k orthonormal vectors which we shall also denote by $\hat{\underline{\theta}}_{iM}$. Let us also complete the set of eigenvalues by assigning nil eigenvalues $\hat{\lambda}_i$ to the vectors $\hat{\underline{\theta}}_{iM}$ which complete the orthonormal basis. We may then express any random input vector $[\hat{S}[1] \ldots \hat{S}[M]]$ as a linear combination of the vectors $\hat{\underline{\theta}}_{iM}$:

$$\hat{\underline{S}}_M = \sum_{i=1}^{M} U_i\, \hat{\underline{\theta}}_{iM} \qquad \text{III-2.[20]}$$

where the coefficients U_i are real random variables. The noise may be expressed as

$$\underline{N}_{2k} = \sum_{i=1}^{2k} v_i\, \hat{\underline{\varphi}}_{i2k} \qquad \text{III-2.[21]}$$

where the coefficients v_i are IID real zero-mean Gaussian random variables with mean σ_N^2. We may express the noiseless output as

$$\hat{\underline{f}}_M^{2k}\hat{\underline{S}}_M = \sum_{i=1}^{2k} U_i \sqrt{\hat{\lambda}_i}\, \hat{\underline{\varphi}}_{i2k} \qquad \text{III-2.[22].}$$

We have decomposed our multiple-access channels into 2k parallel independent channels. The input has M-2k additional degrees of freedom, but those degrees of freedom do not reach the output. The maximization along the active k channels may now be performed using water-filling techniques with the following constraint on input power

$$\operatorname{tr}\left(E\left[\underline{\hat{S}}_M \underline{\hat{S}}_M^T\right]\right) \leq TP\,W$$

III-2.[23]

where T is the duration of the transmission. Since our decomposition of $\hat{S}_M$ in III-2.[9] is along an orthonormal basis, III-2.[12] is equivalent to

$$\sum_{i=1}^{2k} \left(E\left[U_i^2\right]\right) \leq TP\; W$$

III-2.[24].

We may now carry out arguments similar to those in [CT]. We choose

$$E\left[U_i^2\right] = \left(\gamma - \frac{N_0 W}{2\hat{\lambda}_i}\right)^+ \text{ if } \hat{\lambda}_i \neq 0$$

III-2.[25]

where γ satisfies

$$\sum_{i \text{ s.t. } \hat{\lambda}_i \neq 0} \left(\gamma - \frac{WN_0}{2\hat{\lambda}_i}\right)^+ = TP\; W$$

III-2.[26].

We have reduced several channels, each with its own user, to a single channel with a composite user. The result we obtain from water-filling gives us the maximum sum of the rates. Hence, the sum of all the rates is upper bounded by

$$\sum_{j=1}^{K} R_j \leq \frac{1}{T} \sum_{i=1}^{M} \frac{1}{2} \ln\left(1 + \frac{E\left[U_i^2\right]\hat{\lambda}_i}{\frac{N_0 W}{2}}\right)$$

$$= \frac{1}{T} \sum_{i \text{ s.t. } \hat{\lambda}_i \neq 0} \frac{1}{2} \ln\left(1 + \frac{\hat{\lambda}_i \left(\gamma - \frac{N_0 W}{2\hat{\lambda}_i} \right)^+}{\frac{W N_0}{2}} \right)$$

III-2.[27].

In our case, it is not relevant to look at the other inequalities that constitute the rate region, since we are in effect looking only at a single composite user with a single rate denoted as $\sum_{j=1}^{K} R_j$.

III-2-2-b-b- The case of maximizing the total rate sum when the users are constrained to have independent signals.

Let us now consider that each source has an individual power constraint, that the signal of each source is independent of every other source, but that we still seek to maximize $\sum_{j=1}^{K} R_j$. For instance, for the two-user case, we may write the optimization problem as:

$$\max \ \ln\left(\left| \underline{E}_1^{2k} \Delta_{\underline{S}_{12k}} \underline{E}_1^{2k^T} + \underline{E}_2^{2k} \Delta_{\underline{S}_{22k}} \underline{E}_2^{2k^T} + \Delta_{\underline{N}_{2k}} \right| \right)$$

subject to

$\Delta_{\underline{S}_{12k}}$ positive semi-definite

$\Delta_{\underline{S}_{22k}}$ positive semi-definite

$$\mathrm{tr}\left(\Delta_{\underline{S}_{12k}} \right) \leq P\,WT$$

$$\mathrm{tr}\left(\Delta_{\underline{S}_{22k}} \right) \leq P\,WT$$

III-2.[28].

We see that all feasible vectors of inputs are regular (i.e. the vectors formed by the partial derivatives of the constraints form an independent system) and, therefore, we have some hope of being able to apply Lagrangian techniques. We may write necessary and sufficient conditions for local extrema by using the Kuhn-Tucker conditions, but these conditions are tedious. Appendix G shows the explicit solution for the Kuhn-Tucker conditions for a k=k'=1 problem. For larger values of k, k', the expression is complicated enough that a symbolic algebra program (Maple V) cannot solve the system of equations.

Fortunately, the problem is concave and the region over which we optimize is convex. Let us first show that the problem is concave. In order to prove concavity, we first define $\underline{\Delta}_1$ as

$$\underline{E}_1^{2k}\underline{\Delta}_{\underline{S}_{1_{2k}}}\underline{E}_1^{2k^T} + \underline{E}_2^{2k}\underline{\Delta}_{\underline{S}_{2_{2k}}}\underline{E}_2^{2k^T} = \underline{\Delta}_1$$

III-2.[29]

Similarly, let $\underline{\Delta}_{\underline{S}'_{1_{2k}}}$ and $\underline{\Delta}_{\underline{S}'_{2_{2k}}}$ be another pair of covariance matrices that satisfy the conditions of III-2.[28]

$$\underline{E}_1^{2k}\underline{\Delta}_{\underline{S}'_{1_{2k}}}\underline{E}_1^{2k^T} + \underline{E}_2^{2k}\underline{\Delta}_{\underline{S}'_{2_{2k}}}\underline{E}_2^{2k^T} = \underline{\Delta}_2$$

III-2.[30].

Let us denote by $\underline{Z}_{1_{2k}}$ and $\underline{Z}_{2_{2k}}$ two vector random variables with with zero-mean Gaussian distributions and covariance matrices $\underline{\Delta}_1$ and $\underline{\Delta}_2$ respectively. Let us consider an AWGN channel with noise $\underline{N}_{2k}$. Then, from theorem 4.4.2 in [Gal], we know that for any $0 \leq \theta \leq 1$

$$\theta h\left(\underline{Z}_{1_{2k}} + \underline{N}_k\right) + (1-\theta) h\left(\underline{Z}_{2_{2k}} + \underline{N}_{2k}\right) \leq h\left(\underline{Z}_{2k} + \underline{N}_{2k}\right)$$

III-2.[31]

where $\underline{Z}_{2k}$ is a random vector, independent of $\underline{N}_{2k}$, whose p.d.f. is the sum of θ times the p.d.f. of $\underline{Z}_1$ and (1-θ) times the p.d.f. of $\underline{Z}_2$. The variance of $\underline{Z}_{2k}$ + $\underline{N}_{2k}$ is given by $\theta\underline{\Delta}_1 + (1-\theta)\underline{\Delta}_2 + \underline{\Delta}_{\underline{N}_k}$. Since entropy is maximized by a Gaussian distribution for any given covariance matrix,

$$h(\underline{Z} + \underline{N}) \leq \ln\left((2\pi e)^{2k}\left|\theta\underline{\Delta}_1 + (1-\theta)\underline{\Delta}_2 + \underline{\Delta}_{\underline{N}_{2k}}\right|\right)$$

III-2.[32].

Combining III-2.[32] and III-2.[31] yields

$$\theta \ln\left(\left|\theta\underline{\Delta}_1 + \underline{\Delta}_{\underline{N}_k}\right|\right) + (1-\theta)\ln\left(\left|\underline{\Delta}_2 + \underline{\Delta}_{\underline{N}_k}\right|\right) \leq \ln\left(\left|\theta\underline{\Delta}_1 + (1-\theta)\underline{\Delta}_2 + \underline{\Delta}_{\underline{N}_k}\right|\right)$$

III-2.[33].

Therefore, the function in III-2.[28] is concave. The region over which we maximize is convex, because the convex linear combination of positive semi-definite matrices is positive semi-definite and because the trace operator is linear. This means that any kind of hill-climbing technique can be used to perform the maximization. In particular, we may use water filling alternately over the two users. For any given covariance matrix for user 1, we may optimize the distribution for user 2. Then, we may take that distribution for user 2 as being fixed and optimize the distribution for user 1, and so on. A stationary point found by this technique is guaranteed to be a global maximum by our concavity results.

We may note that, when the channel is time-invariant multipath, we have a very particular structure. The complex channel response matrix for user 1 may be written in terms of complex terms as

$$\underline{E}_l^{\,k} = \begin{bmatrix} \underline{E}_l[1,1] & \cdots & \underline{E}_l[1,\Delta] & 0 & \cdots & 0 \\ 0 & \underline{E}_l[1,1] & \cdots & \underline{E}_l[1,\Delta] & 0\cdots & 0 \\ 0 & 0 & \underline{E}_l[1,1] & \cdots & \underline{E}_l[1,\Delta] & 0\cdots 0 \\ 0 & \cdots & 0 & \ddots & \cdots & \vdots \\ \vdots & \ddots & \vdots & 0 & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & 0 & \underline{E}_l[1,1] \end{bmatrix}$$

III-2.[34]

therefore, the rows are all identical to within a shift (except for an end effect of Δ).

Related problems may not exhibit the same convex behavior. We give here some overview of the problems that might be of interest when we wish to maximize capacity for known channels. A natural extension of the case where we wish to maximize the sum of the rates is the problem where each user has its own power constraint and we wish to maximize the rate that may be guaranteed to *every* user. We then have a max-min problem over a convex region. Such a case is germane to the situation where all the users require the same rate and have the same power limitations. For instance, we may wish to give everybody the clearest possible voice channel, but we do not want to give some users very good channels at the expense of poor ones for others. However, we can expect that, statistically, the users will get the same average grade of service even though they may have great differences in their grades of service for prolonged periods of time. Therefore, it is not clear that we need to solve the max min problem explicitly, because over long periods of time the problem in III-2.[28] will probably approach the solution to the max min problem, by symmetry.

III-3- The effect of unknown channels.

In this section, we seek to extend the results concerning channel measurement uncertainty to the multiple user case and consider the effect of spreading when we have several users. The use of broadband signals under multipath channel conditions is often advocated because of the possibility of resolving the multiple paths at the receiver and then combining them in some approximately optimal way [LP87], such as by a rake receiver [PG58]. When we deal with multiple access situations, such as the situation of mobile communications, where the users share bandwidth under some spread-spectrum protocol, the issue of spreading is further complicated by multiple-user interference. The most commonly proposed approach [Qua91] is to decode each user as if all other users were noise. The advantage of spreading under these conditions is to reduce the power spectral density of the users, when the users are all under a maximum power constraint, so that the system becomes limited by the additive Gaussian noise in the channel rather than being interference-limited. Such an approach is, from an information theoretic sense, quite suboptimal (unless we use FDMA [Par81])) when we can measure the channels for the various users [RTM94]. In effect, by considering other users as noise, the multiple-access channel is reduced to an interference channel. For the sake of illustration, figure III.4 shows the achievable rates region for an AWGN channel with noise spectral density N_0. The gray shaded area shows the region of achievable rates for the interference channel.

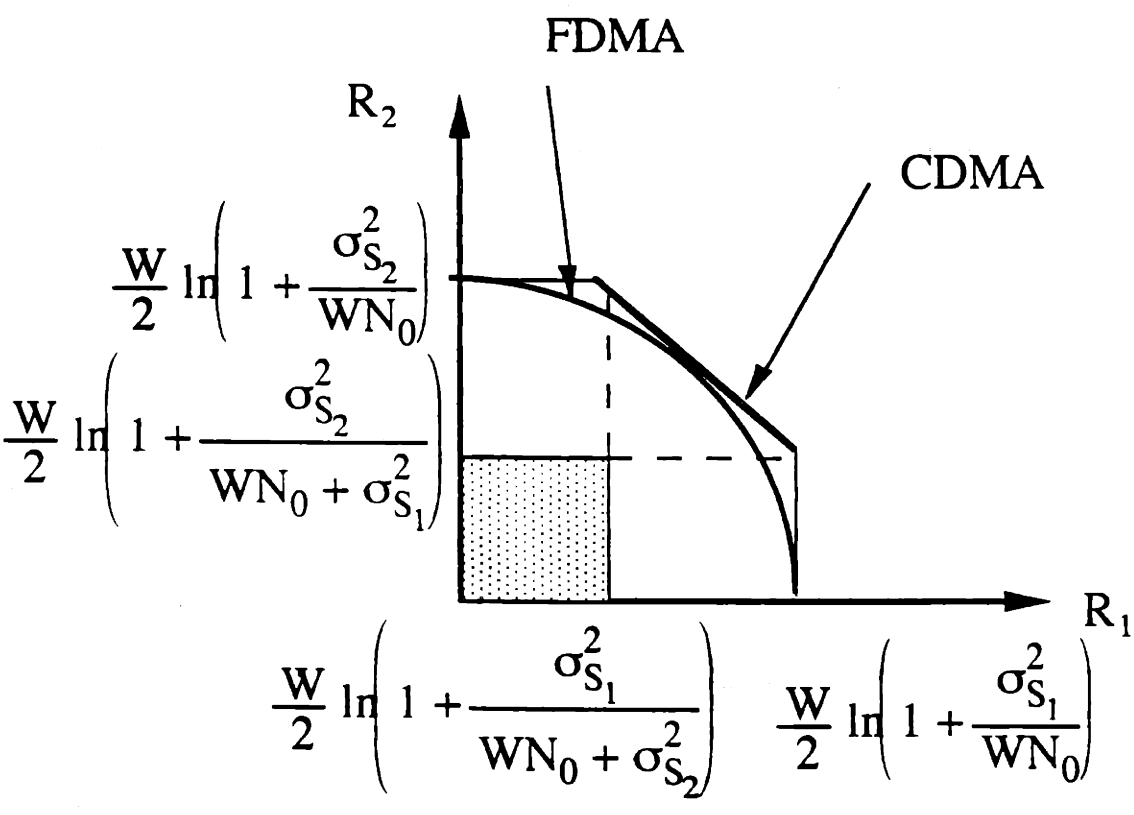


Figure III.4 : Region of feasible rates for two users in an AWGN channel.

We therefore wish to show that interference cancellation is still possible when we do not know the channel as long as we may measure the channel for each user. We know that under certain ergodic conditions, when equalization is possible, interference cancellation is possible with a small added error due to the channel uncertainty ([Gal94]). However, we consider explicitly the loss in capacity due to not being able to perform equalization perfectly. The main purpose of this section is to extend the results of Chapter II to incorporate the effect of channel uncertainty. In particular, we may bound the effect of channel estimation error upon interference cancellation.

In order to show that such bounds in terms of mutual information have some significance for reliable communications, we extend the multiple-access coding theorem to channels with certain decorrelating properties. Appendix B gives this theorem, which leads to an upper bound on the error probability which is exponential in the block lengths. We must be careful in defining memory. Our channels have finite memory in that, conditioned upon the channels only, the output at any time depends on only a finite number of inputs. Similarly, when conditioned solely upon inputs, the output at any time depends on the same finite number of inputs. However, if we condition the output at any one time upon all past outputs and all past inputs, we may not eliminate, in general, the dependency on any of the past inputs. Such a case encompasses the μ^{th} order Markov channels, for instance. If the channel never decorrelates, in a sense which is made precise in appendix B, then a coding theorem may be difficult, because the channel may remain indefinitely in some poor state where communication is virtually impossible.

III-3-1- Channels known with a small error of known variance.

The inequalities we have found for the single user case hold for every inequality which defines that feasible rate region. We shall show the effect of knowing the channels of the users with a certain error of given variance.We shall give our examples for two users, since we may give a graphical interpretation of our bounds. The signals of the users and of the channels are assumed to be mutually independent. The bounds II-3.[21] and II-3.[23] of Chapter II lead to the following upper bounds on the region of feasible mutual informations for the single symbol case (the subscripts denote the users) :

$$I(Y;S_1 \mid S_2) \leq \frac{1}{2}\ln\left(1 + \frac{\overline{F_1}^2 \sigma_{S_1}^2 + \sigma_{S_1}^2 \sigma_{F_1}^2}{\sigma_N^2}\right)$$

III-3.[1]

$$I(Y;S_2 \mid S_1) \le \frac{1}{2}\ln\left(1 + \frac{\overline{F_2}^2\sigma_{S_2}^2 + \sigma_{S_2}^2\sigma_{F_2}^2}{\sigma_N^2}\right)$$

III-3.[2]

$$I(Y;(S_2, S_1)) \le \frac{1}{2}\ln\left(1 + \frac{\overline{F_2}^2\sigma_{S_2}^2 + \sigma_{S_2}^2\sigma_{F_2}^2 + \overline{F_1}^2\sigma_{S_1}^2 + \sigma_{S_1}^2\sigma_{F_1}^2}{\sigma_N^2}\right)$$

III-3.[3].

Each bound of III-3.[1]-[3] may be interpreted as in the single user case. The upper bound may be given by the case where all the received signal components other than the AWGN N are part of the input signal. Indeed, we may write that

$$I(Y;S_1 \mid S_2) \le I(Y;S_1 \mid S_2, F_1, F_2)$$

$$= h(Y \mid S_2, F_1, F_2) - h(Y \mid S_2, S_1, F_1, F_2)$$

$$= h(Y \mid S_2, F_1, F_2) - h(N)$$

(since a Gaussian distribution maximizes entropy for any given variance)

$$\le \frac{1}{2}E_F\left[\ln\left(\sigma_N^2 + \sigma_S^2(F_1)^2\right)\right] - \frac{1}{2}\ln\left(\sigma_N^2\right)$$

(since the ln function is concave)

$$\le \frac{1}{2}\ln\left(1 + \frac{\overline{F_1}^2\sigma_{S_1}^2 + \sigma_{S_1}^2\sigma_{F_1}^2}{\sigma_N^2}\right)$$

III-3.[4].

Therefore, the channel uncertainty on all channels acts as extra input signal power for each combination of users. The inequality in III-3.[3] can be obtained by considering the sum of the signals from the two users as being a single signal.

The lower bounds to the region of achievable rates may be more interesting and are given by:

$$I(Y; S_1 \mid S_2) \geq \frac{1}{2} \ln\left(1 + \frac{\overline{F_1}^2 \sigma_{S_1}^2}{\sigma_N^2 + \sigma_{S_2}^2 \sigma_{F_2}^2 + \sigma_{S_1}^2 \sigma_{F_2}^2}\right)$$

III-3.[5]

$$I(Y; S_2 \mid S_1) \geq \frac{1}{2} \ln\left(1 + \frac{\overline{F_2}^2 \sigma_{S_2}^2}{\sigma_N^2 + \sigma_{S_2}^2 \sigma_{F_2}^2 + \sigma_{S_1}^2 \sigma_{F_2}^2}\right)$$

III-3.[6]

$$I(Y; (S_2, S_1)) \geq \frac{1}{2} \ln\left(1 + \frac{\overline{F_2}^2 \sigma_{S_2}^2 + \overline{F_1}^2 \sigma_{S_1}^2}{\sigma_N^2 + \sigma_{S_2}^2 \sigma_{F_2}^2 + \sigma_{S_1}^2 \sigma_{F_2}^2}\right)$$

III-3.[7].

Again, the bounds in III-3.[5]-[6] bear a similar interpretation to the one we gave for the single user case. The effect of the channel uncertainty can be no worse than to contribute to the AWGN. Each user must contend with the uncertainty of its own channel but also with the uncertainty of the other users' channels. In particular, when we consider the mutual information between user 1 and the output conditioned on user 2, we must take into account the fact that there is a residual effect upon user 1 of the uncertainty on the channel of user 2. Indeed, when we perform interference cancellation, the fact that we are not able to cancel a user exactly because of error in its channel measurement means that there is extra noise beyond the AWGN. Indeed,

$$I(Y;S_1 \mid S_2) = h(S_1 \mid S_2) - h(S_1 \mid Y, S_2)$$

(since S_1 and S_2 are independent)

$$= h(S_1) - h(S_1 \mid Y, S_2)$$

$$= h(S_1) - h\left(S_1 \mid \left(Y - \overline{F_2} S_2\right), S_2\right)$$

(since conditioning reduces entropy)

$$\geq h(S_1) - h\left(S_1 \mid \left(Y - \overline{F_2} S_2\right)\right)$$

III-3.[8].

We may use the LLSE of S_1 from Y as in II-3.[18]-[20] to obtain III-3.[5] from III-3.[8].

Using the lower bounds of III-3.[5]-[7], we may obtain a lower bound to the feasible rate region and represent it as in figure III.4. The perfectly known rate region corresponds to the case where the variance of the channel measurement errors are 0. Figure III.5 shows our lower bound to the feasible rate region when the channels are known with some error of given variance. The effect of the channel error may be bounded as the effect of extra AWGN on the channel. Note that the corners of the feasible rate regions correspond to the rates that are achievable when the users transmit white signals over the same bandwidth at the same time and the receiver performs interference cancellation. When the lower bound is a pentagon rather than a square, we are guaranteed that it is preferable to perform interference cancellation rather than to treat each user as noise to the other user.

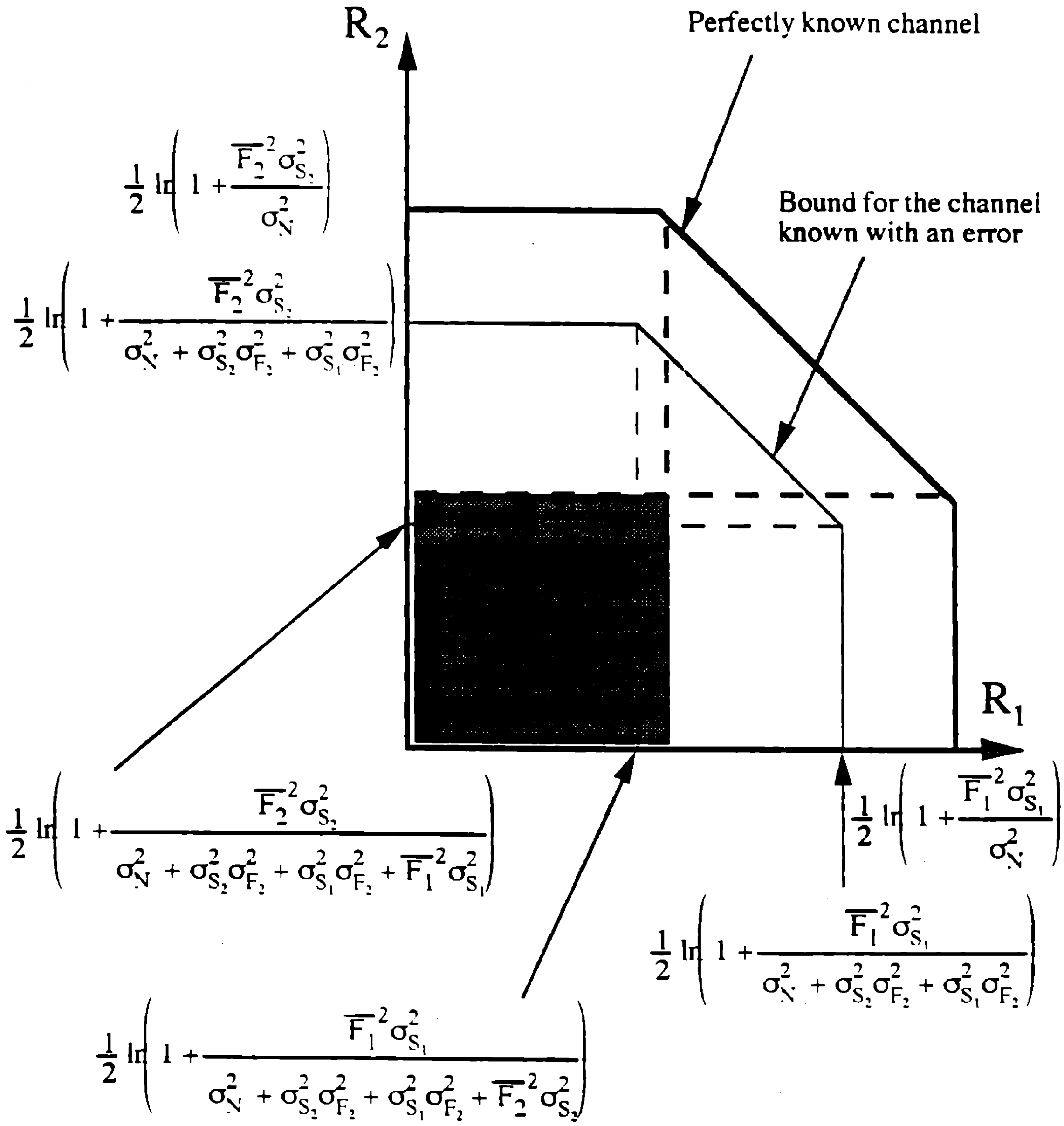


Figure III.5 : Lower bound to the feasible rate region for channels known with a given error variance.

Figure III.5 considers that the channel is known for each user with an error of a given variance. As in the single user case, the actual error variance will depend on the structure of the channel. The next section considers the case of channels with a Gauss-Markov structure.

For the multiple symbol case, the bounds II-3.[43] and II-3.[44] extend in the same way as II-3.[21] and II-3.[23] extend for the single symbol case. Let us simply write the lower bounds:

$$I(\underline{S}_1 ; \underline{Y} | \underline{S}_2) \geq \frac{1}{2} \ln\left(\left| \left(\Lambda_{\underline{\tilde{F}}_1 \underline{S}_1} + \Lambda_{\underline{\tilde{F}}_2 \underline{S}_2} + \Lambda_{\underline{N}} \right)^{-1} \overline{\underline{F}_1} \Lambda_{\underline{S}_1} \overline{\underline{F}_1}^T + \underline{I} \right| \right)$$

III-3.[9]

$$I(\underline{S}_2 ; \underline{Y} | \underline{S}_1) \geq \frac{1}{2} \ln\left(\left| \left(\Lambda_{\underline{\tilde{F}}_1 \underline{S}_1} + \Lambda_{\underline{\tilde{F}}_2 \underline{S}_2} + \Lambda_{\underline{N}} \right)^{-1} \overline{\underline{F}_2} \Lambda_{\underline{S}_2} \overline{\underline{F}_2}^T + \underline{I} \right| \right)$$

III-3.[10]

$$I((\underline{S}_1, \underline{S}_2) ; \underline{Y}) \geq \frac{1}{2} \ln\left(\left| \left(\Lambda_{\underline{\tilde{F}}_1 \underline{S}_1} + \Lambda_{\underline{\tilde{F}}_2 \underline{S}_2} + \Lambda_{\underline{N}} \right)^{-1} \left(\overline{\underline{F}_2} \Lambda_{\underline{S}_2} \overline{\underline{F}_2}^T + \overline{\underline{F}_1} \Lambda_{\underline{S}_1} \overline{\underline{F}_1}^T \right) + \underline{I} \right| \right)$$

III-3.[11].

We do not give a graphical representation of the multiple symbol case, since such an extension would be difficult and would not improve our intuition.

<u>III-3-2- The relation between the rate of change of the channels and loss in terms of mutual information from not knowing the channel a priori at the receiver.</u>

The results of Chapter II extend to the multiple user case with the appropriate modifications to take into account the multiple access. Theorem II-3.1 may be extended as follows:

Theorem III-3.1 : Let $\left\{\underline{F}_i^{\,k}\right\}_{i=1,\dots,K}$ be a set of K unknown time-invariant channels with a known finite number of taps Δ, known *a priori* distributions and additive Gaussian noise. Let $\left\{\underline{S}_{ik}\right\}_{i=1,\dots,K}$ be a set of arbitrary stationary probabilistic, mutually independent and power limited inputs to the channel. For any positive ε, we may choose a set of probabilistic inputs $\left\{\underline{S'}_{ik}\right\}_{i=1,\dots,K}$ whose corresponding output we denote by $\underline{Y}_k'$ such that for all k large enough

$$\left(\frac{ I\left(\left\{\underline{S}_{ik}\right\}_{i\in U} ; \underline{Y}_k \,|\, \underline{F}^k, \left\{\underline{S}_{ik}\right\}_{j\notin U} \right) - I\left(\left\{\underline{S'}_{ik}\right\}_{i\in U} ; \underline{Y}_k' \,|\, \left\{\underline{S'}_{ik}\right\}_{j\notin U} \right) }{k} \right) \leq \varepsilon$$

III-3.[12]

for all subsets U of $\{1 \ldots K\}$.

The main idea in extending the proof is the following chain rule for mutual informations:

$$I\left(\{\underline{S}_{ik}\}_{i=1\ldots K};\underline{Y}_k\right) = I(\underline{S}_{1k};\underline{Y}_k) + I(\underline{S}_{2k};\underline{Y}_k \mid \underline{S}_{1k}) + \ldots + I\left(\underline{S}_{Kk};\underline{Y}_k \mid \{\underline{S}_{ik}\}_{i=1\ldots K-1}\right)$$

III-3.[13].

Each of the terms in the RHS of III-3.[7] may be computed by treating as noise all input signals which are not in the conditioning.

Similarly, theorem II-3.2 may be extended to the multiple access case:

<u>*Theorem III-3.2*</u> : Let $\left\{\underline{E}_i^{k}\right\}_{i=1,\ldots K}$ be a set of K unknown time-invariant channels with a known finite number of taps Δ, known *apriori* distributions and additive Gaussian noise. Let any row $\underline{E}_i^{j,j}$ conditioned on $\underline{E}_i^{j+1,j+\mu}$ be independent of $\underline{E}_i^{j+\mu+1,j+k}$ and let $h\left(\left\{\underline{E}_i^{j,j}\right\}_{i=1\ldots K} \mid Y[i], \left\{\underline{E}_i^{j+1,j+\mu}\right\}_{i=1\ldots K}, \left\{\underline{S}_{ij}\right\}_{i=1\ldots K}, \left\{\underline{E}_i^{j-\mu,j-1}\right\}_{i=1\ldots K}\right) > -\infty$ hold. Let $\left\{\underline{S}_{ik}\right\}_{i=1,\ldots K}$ be a set of arbitrary stationary probabilistic, mutually independent inputs to the channel. Then

$$\lim_{k\to\infty}\left(\frac{I\left(\left\{\underline{S}_{ik}\right\}_{i\in U};\underline{Y}_k \mid \underline{E}^{k}, \left\{\underline{S}_{jk}\right\}_{j\notin U}\right) - I\left(\left\{\underline{S}_{ik}\right\}_{i\in U};\underline{Y}_k \mid \left\{\underline{S}_{jk}\right\}_{j\notin U}\right)}{k}\right)$$

$$= \lim_{j\to\infty}\lim_{k\to\infty}\left(\begin{array}{l} h\left(\left\{\underline{E}_i^{j,j}\right\}_{i=1\ldots K} \mid \underline{Y}_k, \left\{\underline{E}_i^{j+1,k}\right\}_{i=1\ldots K}, \left\{\underline{S}_{ik}\right\}_{j\notin U}\right) \\ -\, h\left(\left\{\underline{E}_i^{j,j}\right\}_{i=1\ldots K} \mid \underline{Y}_j, \left\{\underline{E}_i^{j+1,j+\mu}\right\}_{i=1\ldots K}, \left\{\underline{S}_{ij}\right\}_{i=1\ldots K}\right)\end{array}\right)$$

III-3.[14]

for all subsets U of $\{1 \ldots K\}$.

Expressions III-3.[15] and III-3.[16] allow us to extend the results about Gauss-Markov channels to the multiple access case. Let us give the following extension of theorem II-3.3:

Theorem III-3.3 : For the model of II-3.[88] with a single tap for each user's channel and $\underline{S}_i$ a sequence of IID white Gaussian random, we have that

$$\lim_{k \to \infty} \left(\frac{I\left(\left\{\underline{S}_{ik}\right\}_{i \in U} ; \underline{Y}_k \mid \underline{F}^k, \left\{\underline{S}_{jk}\right\}_{j \notin U}\right) - I\left(\left\{\underline{S}_{ik}\right\}_{i \in U} ; \underline{Y}_k \mid \left\{\underline{S}_{jk}\right\}_{j \notin U}\right)}{k} \right)$$

$$\leq \lim_{k \to \infty} \frac{1}{2} |U| \ln\left(\frac{\left(1-\alpha^2\right) \sigma_F^2}{\lambda} + \alpha^2 \right)$$

III-3.[15]

where

$$\lambda = \frac{\left(\alpha^2 - 1\right)\left(\sigma_N^2 + K\, \sigma_S^2 \sigma_F^2\right) - \sigma_S^2\, \sigma_\Xi^2}{2\sigma_S^2\, \alpha^2} + \frac{\sqrt{\left(\left(-\alpha^2 + 1\right)\left(\sigma_N^2 + K\, \sigma_S^2 \sigma_F^2\right) + \sigma_S^2\, \sigma_\Xi^2\right)^2 + 4\, \sigma_S^2\, \sigma_\Xi^2 \left(\sigma_N^2 + K\, \sigma_S^2 \sigma_F^2\right) \alpha^2}}{2\sigma_S^2\, \alpha^2}$$

III-3.[16].

In order to see why III-3.[15] and III-3.[16] hold, we may first write that

$$\lim_{k \to \infty} \left(\frac{I\left(\left\{\underline{S}_{ik}\right\}_{i \in U} ; \underline{Y}_k \mid \underline{F}^k, \left\{\underline{S}_{jk}\right\}_{j \notin U}\right) - I\left(\left\{\underline{S}_{ik}\right\}_{i \in U} ; \underline{Y}_k \mid \left\{\underline{S}_{jk}\right\}_{j \notin U}\right)}{k} \right)$$

$$= \sum_{p=1}^{|U|} \lim_{k \to \infty} \left(\frac{ I\left(\underline{S}_{i_U(p)_k}; \underline{Y}_k \mid \underline{F}^k, \left\{ \underline{S}_{jk} \right\}_{\left((j \notin U) \vee (j \in U_p)\right)} \right) - I\left(\underline{S}_{i_U(p)_k}; \underline{Y}_k \mid \left\{ \underline{S}_{jk} \right\}_{\left((j \notin U) \vee (j \in U_p)\right)} \right) }{k} \right)$$

III-3.[17]

where $i_U(p)$ is the index of the p^{th} user in set U and U_p is the set of all users with indices $i_U(1)$ through $i_U(p-1)$. For each term in the summation of the RHS of III-3.[9], we see that we must contend with the users in U-U_{p+1} acting as noise and the residual noise from the channel uncertainty of the users not in U. The effect of the users in U-U_{p+1} may be upper bounded by considering that the whole power of their output signal before AWGN goes into being additional AWGN. For the users not in U or in U_p, we may upper bound the effect of their residual noise by considering that the whole power of their output signal before AWGN goes into being additional AWGN. We are therefore taking into account more than just the power of the component due to channel measurement error. Expression III-3.[16] may be therefore be interpreted as being an extension of II-3.[94] using the above discussion to replace σ_N^2 by $\sigma_N^2 + K\,\sigma_S^2\sigma_F^2$.

Note that our extension is slightly less general than the single user case, because we restrict the users to send IID white Gaussian signals. The reason for this restriction is that we wish to be able to consider users as noise to other users in the terms of the sum in the RHS of III-3.[17]. A more general formulation would be much more cumbersome. Figures III.5 and III.6 give some examples of how the Doppler spread affects the region of achievable rates. These figures show, for a given $|U|$, the ratio of the bound in III-3.[15] to the mutual information which would be obtained with perfectly known AWGN channels with the same energy as the Gauss-Markov channels. Figure III.5 illustrates the case for $|U| = 1$ and figure III.6 the case for $|U| = 3$. In appendix H, we give more examples.

Figure III.5.

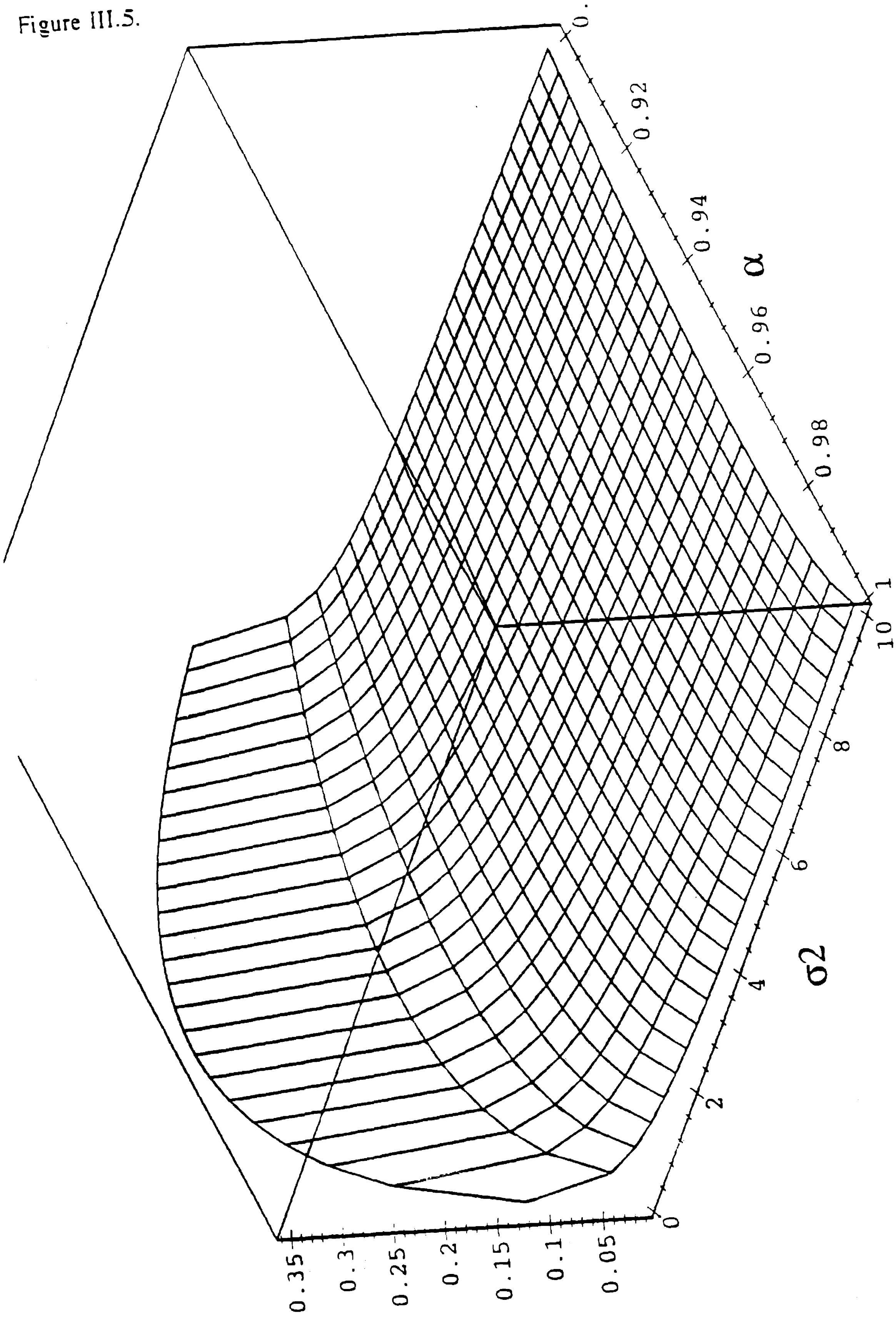
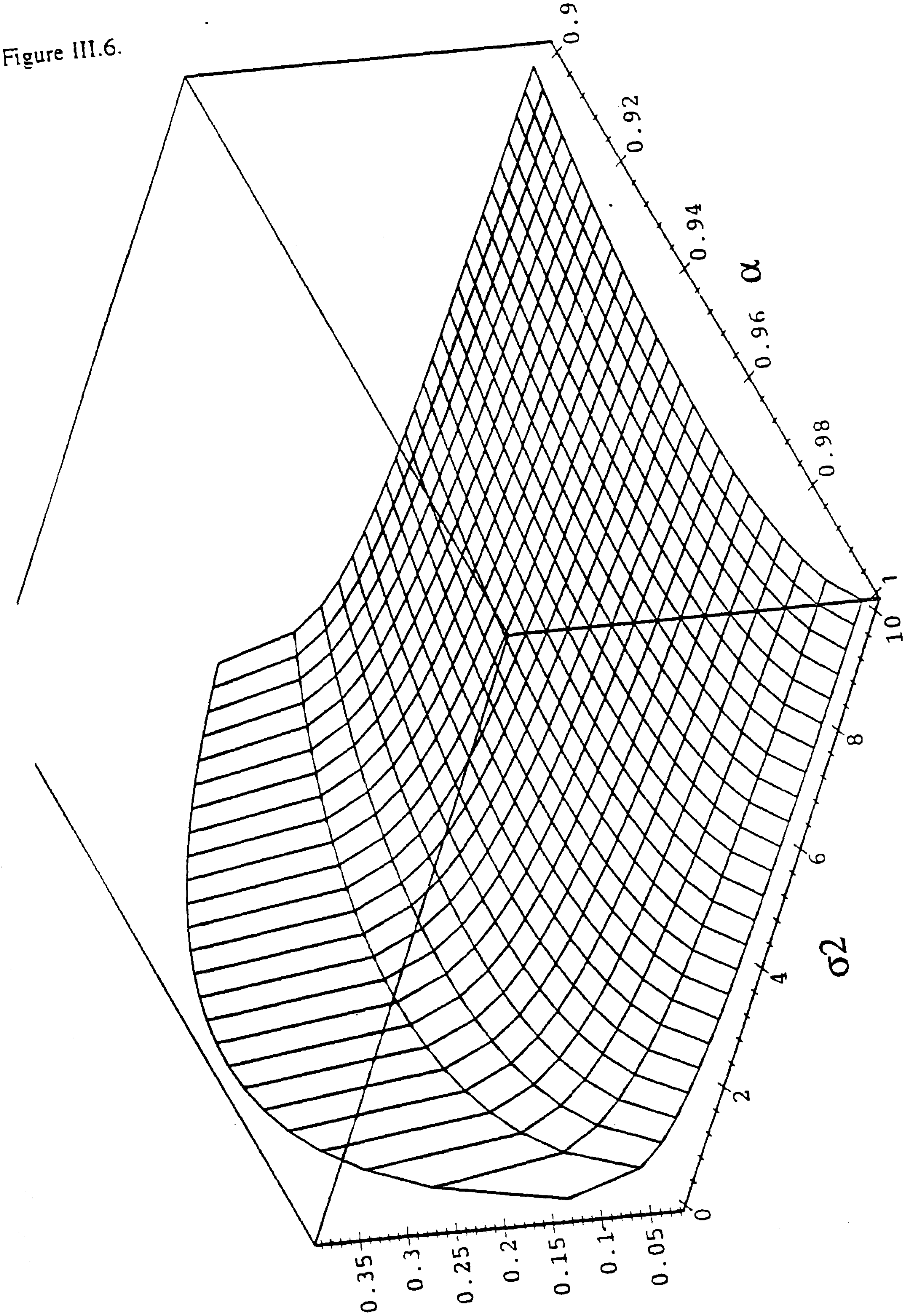


Figure III.6.



III-4- Spreading in time-varying channels with memory.

We have examined in Chapter II the benefits of spreading for a single user. We have seen that there exists a threshold SNR per degree of freedom for which spreading is advantageous. Below that threshold, spreading may be disadvantageous. For multiple users, we may spread users over each other or over separate time and/or frequency intervals. If the users are spread separately, we are back to the single user case for each user. Each user wishes to spread until it reaches its SNR threshold. Let us now consider what happens when the two users are spread over each other, i.e. they transmit at the same time over the same bandwidth. We invoke III-3.[15] once again and examine the terms in its RHS for the two user case:

$$I(\{\underline{S}_{ik}\}_{i=1,2}; \underline{Y}_k) = I(\underline{S}_{1k}; \underline{Y}_k) + I(\underline{S}_{2k}; \underline{Y}_k \mid \underline{S}_{1k})$$

III-4.[1].

In the first term, $I(\underline{S}_{1k}; \underline{Y}_k)$, user 2 acts as noise to user 1. The SNR per degree of freedom for user 1 is then lower than if user 2 were not present. Note that, for some degrees of freedom, the SNR may be affected differently than for others, depending on the distribution of user 2's signal. For the second term, $I(\underline{S}_{2k}; \underline{Y}_k \mid \underline{S}_{1k})$, the product of the channel of user 1 and the known signal of user 1 acts as noise for user 2. Again, the SNR per degree of freedom for user 2 is lower than if user 1 were not present. Therefore, the threshold to which we want to spread the two users jointly is higher than the threshold to which we would spread the users separately.

If the multiplicative part of the channel is perfectly known, then we may send the users spread over each other without any loss in capacity. For instance, if the channel is AWGN and the users are appropriately separated in power so that interference cancellation is possible, then we wish to spread the users jointly as much as possible. We may relax the assumption that the channel is perfectly known and assume that equalization is possible. If we assume that equalization is possible, then we have shown that we are in effect assuming that we do not lose any mutual information by not knowing the channel. We therefore still want to spread the two users as much as possible, even when we are spreading jointly. If we relax the assumption that the channel may be equalized, then spreading jointly behaves differently. The fact that the channel is not perfectly known causes users to act as noise to each other in the measurement of the channel. Therefore, if bandwidth is not limited and the channel is not known, we do not want to spread jointly.

Chapter IV - Conclusions and directions for further research.

IV-1 - Conclusions.

Let us first recapitulate the main results of this thesis. We

- derived a sampled model for single and multiple user time varying multipath channels

- gave capacity results for perfectly known channels at the sender and the receiver for the single and multiple user case

- bounded the effect upon capacity of having a measurement error of known variance

- showed that, for an unknown time-invariant channel, the capacity of the system is not affected by whether or not we know the channel a priori at the receiver

- established exact limit results for the loss in capacity due to not knowing the channel at the receiver for Markov channel models of order m

- computed explicit bounds for the intrinsic effect of time and frequency spreading upon capacity with respect to the perfectly known channel assumption for a Gauss-Markov model

- showed that spreading for an unknown channel, when subject to certain assumptions about the transmitted signal, is beneficial up to a certain SNR per degree of freedom but may be detrimental below and studied the effect of multiple access upon spreading

- established a strong multiple access coding theorem for channels that decorrelate in time and showed that interleaving with block codes can achieve capacity.

The main thrust of this thesis has been to investigate the effect of channel variations upon channel capacity. The effect of channel variations and the ensuing uncertainty about the channel at the receiver are crucial in determining how we wish to transmit. Indeed, the original research interest that sparked this thesis was power control. However, power control is useful only because we have time-varying multipath.

Without understanding the effect of time-varying multipath upon capacity, we cannot devise effective power control techniques. Suppose, for instance, that we do not perform interference cancellation but spread the users simultaneously over the same bandwidth. We then consider the multiple access channel to be an interference channel. Each user is interference to all other users. Our power control will require the users to be close in received power so as to avoid having one user overwhelm the others. But if we were to perform interference cancellation, even with some error in our estimate of the channel at the receiver, then we could allow users to drift away from each other in terms of received power. The strongest user would be decoded first, its estimated noiseless output signal subtracted from the received signal, then the second strongest would be decoded, and so on We have seen that if we know the channel well enough, it is preferable to perform interference cancellation rather than to decode the users independently.

Another aspect of power control which is affected by imperfect knowledge of the channel at the receiver is the fact that mutual information for a multipath channel known with some Gaussian error does not degrade gracefully when the SNR falls below a certain level. Not only do we not want to spread so that the SNR per degree of freedom is too low, but we also must be aware that variations in the channel which could cause us to go below our minimum desirable SNR are much more nefarious than if we knew the channel. Therefore, our power control should seek to avoid allowing low SNRs, even if we are performing interference cancellation and allowing the power of the users to drift away from each other. Our discussion suggests some directions for further research.

IV-2- Directions for further research.

The use of feedback is a natural extension to our research. The study of the use of feedback in information theoretic terms usually assumes total feedback. Since the resources allocated to feedback are no less costly than the resources allocated to data transmission, such an assumption is not satisfactory. The benefit of feedback, as we already mentioned, is due solely to the fact that we have multipath. If a single user channel is memoryless, for instance and AWGN, then feedback cannot help ([Dob58]). If the channel has memory but is time-invariant, we wish to sound the channel and then transmit our measure of the channel to the sender. The feedback channel is then used very briefly, but its use permits large gains in capacity because the sender can tailor its signalling scheme to the channel. It would be reasonable to assume that, if the channel changes slowly enough, periodic sounding and feedback may be beneficial for capacity.

Another problem related to feedback is the delay between the reception at the receiver and the feedback. If feedback were faster than the decorrelating time of the channel, then we could know the value of noise before it decorrelated, therefore we would modulate the data through noise and have an infinite capacity. The bound for the benefit of feedback in [CT] assumes infinite feedback and also implicitly assumes that we do not have instantaneous feedback, although the form of the determinant optimization does not show this assumption. The main reason why the existing theoretical results for channels with feedback do not apply to our channels is that these results attempt to perform estimation on the noise, which decorrelates very rapidly, but we wish to estimate the multiplicative part of the channel, which changes slowly with respect to our data rate.

The fact that delay is important makes data rate arguments difficult. Indeed, we cannot use very long block codes in the feedback channel, because if we get a measurement of the channel long after it was taken, that measurement may be of no use. In particular, the sort of interleaved block code that we have shown achieves capacity for channels which decorrelate as defined in appendix B entails long delays, guaranteeing that the channel has decorrelated. Therefore, capacity arguments where time is allowed to go to infinity are difficult to apply to time-variant channels with feedback. Information stability of the input and output in a channel with feedback is guaranteed by the input signal being stationary and ergodic and the multiplicative part of the channel and the additive noise being weakly mixing when the input and channel are independent. When there is feedback, the input and the channel are no longer independent, since we modify our signal according to our knowledge of the channel.Therefore, capacity arguments are more difficult.

Finally, we must consider multiple access in our feedback. If we have several users, providing each user with feedback about every other user is costly. The results in [Tho87] for Gaussian multiple access channels assume an infinite feedback channel. Again the issue of delay is important. If every use could know immediately what every other user transmits, then the users could cooperate coherently and act as an antenna array. Such cooperation is unlikely to be practical. however. because it would require very rapid communication and a link from every user to every other user.

An approach which might be attractive would be to look at single symbols and consider he deviation from the optimal signalling covaraiance matrix due to the fact that we do not have perfect knowledge of the multiplicative part of the channel at the receiver. Allocating more power/bandwidth to feedback and sounding is beneficial as long as the incremental benefit from reducing the variance of the error on the channel outstrips the benefit from having those resources used for transmission. The sort of bounds that we found in II-3 and III-3 could be useful for considering the use of feedback.

Appendix A.

Proof of Lemma II-3.2.

$$h\left(\underline{F}^{i,i} \mid \underline{Y}_k, \underline{F}^{i+1,k}\right) \geq h\left(\underline{F}^{i,i} \mid \underline{Y}_k, \underline{F}^{i+1,k}, \underline{S}_i\right)$$

(rewriting $\underline{Y}_k$)

$$\geq h\left(\underline{F}^{i,i} \mid \underline{Y}_{i+1,k}, \underline{Y}_{i-1}, Y[i], \underline{F}^{i+1,k}, \underline{S}_i\right)$$

(since conditioning reduces entropy)

$$\geq h\left(\underline{F}^{i,i} \mid \underline{Y}_{i+1,k}, \underline{Y}_{i-1}, Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-1}\right)$$

(using the fact that $\underline{Y}_{i-1}$ is $\underline{F}^{i-1}\underline{S}_{i-1} + \underline{N}_{i-1}$)

$$= h\left(\underline{F}^{i,i} \mid \underline{Y}_{i+1,k}, \underline{N}_{i-1}, Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-1}\right)$$

(using the fact that the noise is white and independent from the other random variables)

$$= h\left(\underline{F}^{i,i} \mid \underline{Y}_{i+1,k}, Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-1}\right)$$

(using the fact that the channel is μ^{th} order Markov)

$$= h\left(\underline{F}^{i,i} \mid \underline{Y}_{i+1,k}, Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-\mu,i-1}\right)$$

(using the fact that conditioning decreases entropy)

$$\geq h\left(\underline{F}^{i,i} \mid \underline{Y}_{i+1,k}, Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-\mu,i-1}, \underline{S}_{i+1,k}, \underline{N}_{i+1,k}\right)$$

(using the fact that $\underline{Y}_i$ is $\underline{F}^i\underline{S}_i + \underline{N}_i$)

$$= h\left(\underline{F}^{i,i} \mid Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-\mu,i-1}, \underline{S}_{i+1,k}, \underline{N}_{i+1,k}\right)$$

(using the fact that the noise is white and independent from the other random variables)

$$= h\left(\underline{F}^{i,i} \mid Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-\mu,i-1}, \underline{S}_{i+1,k}\right)$$

(using the fact that h(A, B) = h(A) + h(B|A))

$$= h\left(\underline{F}^{i,i}, \underline{S}_{i+1,k} \mid Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-\mu,i-1}\right) - h\left(\underline{S}_{i+1,k} \mid Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-\mu,i-1}\right)$$

(using the fact that the input is independent from the channel and the noise)

$$= h\left(\underline{F}^{i,i}, \underline{S}_{i+1,k} \mid Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-\mu,i-1}\right) - h\left(\underline{S}_{i+1,k} \mid \underline{S}_i\right)$$

(using the fact that h(A, B) = h(A) + h(B|A))

$$= h\left(\underline{F}^{i,i} \mid Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-\mu,i-1}\right) + h\left(\underline{S}_{i+1,k} \mid Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-\mu,i-1}, \underline{F}^{i,i}\right) - h\left(\underline{S}_{i+1,k} \mid \underline{S}_i\right)$$

(using the fact that the input is independent from the channel and the noise)

$$= h\left(\underline{F}^{i,i} \mid Y[i], \underline{F}^{i+1,k}, \underline{S}_i, \underline{F}^{i-\mu,i-1}\right) + h\left(\underline{S}_{i+1,k} \mid \underline{S}_i\right) - h\left(\underline{S}_{i+1,k} \mid \underline{S}_i\right)$$

(canceling the last two terms and using the fact that the channel is μ^{th} order Markov)

$$= h\left(\underline{F}^{i,i} \mid Y[i], \underline{F}^{i+1,i+\mu}, \underline{S}_i, \underline{F}^{i-\mu,i-1}\right)$$

[II-3.[59].

Appendix B : Capacity region for a channel with infinite memory on past inputs and outputs and finite memory on past inputs only.

This appendix determines a coding theorem for multiple-access channels without finite memory on past inputs and outputs but with finite memory on inputs only. We first define the sort of channel we are examining. We then establish a source coding theorem for it.

B-I Introduction.

The multipath channel is intrinsically dependent on past states since the input is "echoed" by several paths.The case of several users must take into account the interference from the multipath of different users rather than just that of just one user. We must examine carefully the meaning of memory in the channel. We may always write that:

$$p_{\underline{Y}_i|\underline{S}_i}(\underline{y}_i|\underline{s}_i) = \prod_{j=1}^{i} p_{Y[j]|(\underline{S}, \underline{Y}_{j-1})}(y[j]|\underline{s}_j, \underline{y}_{j-1})$$

B-I.[1].

When the channel is memoryless, we have that for a single user channel B-I.[1] may be rewritten as:

$$p_{\underline{Y}_i|\underline{S}_i}(\underline{y}_i|\underline{s}_i) = \prod_{j=1}^{i} p_{Y[j]|S[j]}(y[j]|s[j])$$

B-I.[2].

One common meaning for the channel having finite memory Δ is that ([Khi56], [Ver89]):

$$p_{\underline{Y}_i|\underline{S}_i}(\underline{y}_i|\underline{s}_i) = p_{\underline{Y}_\Delta|\underline{S}_\Delta}(\underline{y}_\Delta|\underline{s}_\Delta) \prod_{j=\Delta+1}^{i} p_{Y[j]|\underline{S}_{j-\Delta+1,j}}(y[j]|\underline{s}_{j-\Delta+1,j})$$

B-I.[3].

Expression B-I.[3] would apply if the multipath profile were constant and known at the receiver. However. considering that the multipath is constant is not appropriate for a fading channel. Indeed. if the channel is constant. there *is no fading* other than constant multiplicative attenuation. The results are then a special case of Chapter II. section II for known channel for a channel which does not change. Moreover. even if the channel is unknown at the receiver, we have shown in that we do not lose anything in terms of mutual information to not knowing the channel at the receiver. Therefore. B-I.[3] captures neither the fading nor the effect of the unknown channel.

If we condition on knowing the multiplicative part of the channel, then we a modified version of B-I.[3] holds with the appropriate conditioning on the channel:

$$p_{\underline{Y}_i|\underline{S}_i,\underline{F}^i}\left(\underline{y}_i \mid \underline{s}_i, \underline{f}^i\right) = p_{\underline{Y}_\Delta|\underline{S}_\Delta,\underline{F}^\Delta}\left(\underline{y}_\Delta \mid \underline{s}_\Delta, \underline{f}^\Delta\right) \prod_{j=\Delta+1}^{i} p_{Y[j]|\underline{S}_{j-\Delta+1,j},\underline{F}^j}\left(y[j] \mid \underline{s}_{j-\Delta+1,j}, \underline{f}^j\right)$$

B-I.[4].

Expression B-I.[5] simply brings us back to the case of a known channel at the receiver, which was studied in Chapter II, section II. It allows us to consider the effect of fading but assumes pefect knowledge of the channel at the receiver. However, our goal is to consider the effect upon achievable rates of the uncertainty in the channel at the receiver.

Another way of discussing finite memory on both past states of the channel and past outputs is the finite state model found in [Gal]. Finite-state Markov models are often used to represent channels ([Gil60], [Fri67], [AFK72]) and in particular mobile channels ([McMan70], [Gol94]). Markov models offer great accuracy of representation, albeit at the expense of added complexity. The finite state assumption requires us to have a good grasp of the structure of the channel and its accuracy depends heavily on how we create our finite states. It is not compatible with some sort of Gauss-Markov model.

In practice, the memory of the channel is often not truly finite, but simply decaying rapidly enough that it eventually becomes, in some sense, "unimportant". The time threshold after which the memory is "unimportant" depends on what rate we are trying to send at. on the particular structure of the channel. etc... As stated in [Pfa71], "since the memory of most physical channels is not really finite. but only decreasing in some sense with increasing time distance between incoming and outgoing signals. the assumption of a finite-memory channel is not quite satisfactory". We wish to give some idea of how the structure of the channel, especially its time variation. affects the way in which we need to code. In particular, we wish to find some sort of exponential bound on the probability of error in terms of the length of our code. Such a result gives us some sense of a trade-off between delay and probability of error for a given channel.

For finite multipath spread when the multipath profile is not constant. if we condition a single output symbol on the input only, then we could write that

$$p_{Y[i]|\underline{S}_i}\left(y[i]\,|\,\underline{s}_i\right) = p_{Y[i]|\underline{S}_{i-\Delta+1,i}}\left(y[i]\,|\,\underline{s}_{i-\Delta+1,i}\right).$$

B-I.[5]

which is a much weaker condition than B-I.[4]. We see that B-I.[5] alone is unlikely to yield a coding theorem and it does not contain any sense about the channel structure. In B-II, we give some conditions for the decorrelation of the channel which allow us to establish an error bound which decreases exponentially in the length of the block code.

Several coding theorem results have been established for channels with some sort of weak mixing. In [KW72], a coding theorem is given for a channel with a known multiplicative part and a strongly mixing noise component. In B-V, we point out that all we need is weak mixing for both the multiplicative part and the additive noise part of the channel. In [Pfa71], a coding theorem is given with some sort of weak mixing condition on the channel input to output transition probabilities. More generally, [Tin62] shows that information stability is enough to establish a coding theorem. However, it also shows that is necessary. All of these papers make some sort of Strong Law of Large Numbers argument to establish that atypical sequences, without good mutual information properties between the input and output sequences, occur with vanishing probability. We present in B-V how we may use such conditions to construct a coding theorem. However, such coding theorems do not establish how rapidly the error probability vanishes. Therefore, we have a weak coding theorem, which does not give us much of an idea as to how we may wish to code or what the trade-off is between delay and error probability.

Our discussion of the literature has until now dwelt mostly on the aspect of channel memory and decorrelation. We must also contend with multiple access. For discrete-time, memoryless, multi-access channels, the capacity region has been given in [Lia72]. Similar results for just two senders and one receiver have been established by Ahlswede in [Ahl71]. In [Ahl74], it is postulated that the results extend to an arbitrary number of receivers and senders, as long as the senders transmit to all the receivers. This postulate is shown to be true by Ulrey in ([Ulr75]), although there the worst link between a sender and receiver is examined, which is not the sort of problem we wish to consider. The capacity region established in [Lia72] is just a special case of that for the channel with s senders and r receivers considered by Ulrey where r=1. When the channel has finite memory as in the usual sense ([Ver89,1] and[Ver89,2]) establish the capacity region for the multiple access case. The source coding theorem specifically uses B-I.[3] (extended to one output and several inputs) and is therefore not applicable for our purposes. The converse to the coding theorem for multi-user channels with memory which is established in [Ver89,1] does not specifically make use of B-I.[3] and is applicable to our channel model. In the following, we shall assume stationarity and ergodicity. We do not seek to establish results with stationarity and no ergodicity. As for synchronism, asynchronism among the users can be included in the path delays, as mentioned in Chapter III. Therefore, we do not need to consider separately the synchronous (in frame only or in symbol also) case from the asynchronous (in frame only or in symbol also) case. In the memoryless multiple access channel, asynchronism affects the capacity region only by the lack of a convex hull operation ([HH85]). In the channel with memory as in B-I.[3], frame asynchronism, and hence symbol synchronism, reduces the capacity region because the users cannot cooperate in time ([Ver89,2]).

We establish a multiple access coding theorem for users with infinite memory on outputs and finite memory on past inputs only. We proceed in a way similar to that of [Gal68] and [Lia72]. We first establish an exponential bound on the probability of error. We next establish that the bound is vanishingly small for rates in a certain capacity region. Our coding theorem indicates that one way to achieve capacity is to interleave in such a way as to weather out the fades. Such a practice can easily be seen to help combat fading, and engineers have been adopting it without any need for a coding theorem. However, this coding theorem confirms that such a practice is in some sense optimal and also gives some idea as to what delay trade-offs we may expect.

B-II Model and notation.

We use a similar notation and model to that found in [Gal68] and [Lia72], with minor changes due to the nature of the problem. Unless otherwise specified, the subscript associated with the vectors is N. We first consider a discrete-time channel with two users, the case for an arbitrary number of users being considered later as an extension. At any time, we know the probability of the output, given the previous inputs from the two users. The goal of the decoder is to furnish us with a pair of codewords, to reflect the codewords sent by the two users. We perform maximum-likelihood decoding on the output, i.e., we choose as the decoded pair of codewords one such pair which maximises the probability of the observed output given that pair of codewords was the channel input. Since we are trying to show that we may transmit up to rates within a certain capacity region expression, it is sufficient to show that we may achieve such transmission by performing maximum-likelihood decoding. Each user has its own codebook, where each word is mapped to a sequence of N letters. Each user has words selected independently from its codebook with a given probablity mass function, Q, defined below. We denote probability density functions by a small p.

We have the following notation :

M_i : number of code words for the codebook of user i

N: number of channel input letters per user source codebook word

N' : number of output letters per receiver codebook word

R_i : rate of transmission of user i in nats. $R_i = \ln M_i/N$

$\underline{X_i}$: random variable denoting the N-tuple of letters forming the codeword sent by user i

$\underline{Y}$: random variable denoting the N'-tuple of letters forming the output codeword

$\underline{x}_i$: N-tuple of letters: value that may be taken by $\underline{X}_i$

$\underline{y}$: N'-tuple of letters: value that may be taken by $\underline{Y}$

$p(\underline{y} \mid \underline{x}_1, \underline{x}_2)$: simplified notation for $p(\underline{Y} = \underline{y} \mid \underline{X}_1 = \underline{x}_1, \underline{X}_2 = \underline{x}_2)$; maximum likelihood decoding, with this notation, reduces to choosing the pair $(\underline{x}_1, \underline{x}_2)$ which maximises $p(\underline{y} \mid \underline{x}_1, \underline{x}_2)$ for $\underline{y}$ given

$q^{i,N}(\underline{x}_i)$: probability density that the codeword $\underline{x}_i$ was chosen for any particular codeword of user i

m_i : message sent by user i

$\hat{m}_i$: decoded message for user i

Pe_1 (Pe_2) : probability that the decoded codeword for user 1 (2) is different from that sent by user 1 (2) while the decoded codeword for user 2 (1) is the same as the one sent by user 2 (1), i.e.

$m_1 \neq \hat{m}_1$ and $m_2 = \hat{m}_2$ $\qquad \left(m_2 \neq \hat{m}_2 \text{ and } m_1 = \hat{m}_1\right)$

such errors will be denoted as error of type 1 (2)

$Pe_{1,2}$: probability that the decoded codewords for both users 1 and 2 are different from those sent by those users, i.e.

$m_i \neq \hat{m}_i$ for i = 1,2

such error will be denoted as error of type 3

$I_{q^{1,N}, q^{2,N}}(\underline{X}_1; \underline{Y} \mid \underline{X}_2)$

$$= \int_{\underline{y}} \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1,N}(\underline{x}_1)\, q^{2,N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2) \ln\left\{\frac{p(\underline{y} \mid \underline{x}_1, \underline{x}_2)}{p(\underline{y} \mid \underline{x}_2)}\right\} d\underline{x}_2\, d\underline{x}_1\, d\underline{y}$$

B-II.[1]

$$I_{q^{1,N},q^{2,N}}(\underline{X}_2; \underline{Y} | \underline{X}_1)$$

$$= \int_{\underline{y}} \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1,N}(\underline{x}_1)\, q^{2,N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2) \ln\left(\frac{p(\underline{y} \mid \underline{x}_1, \underline{x}_2)}{p(\underline{y} \mid \underline{x}_1)} \right) d\underline{x}_2\, d\underline{x}_1\, d\underline{y}$$

B-II.[2]

$$I_{q^{1,N},q^{2,N}}((\underline{X}_2, \underline{X}_1); \underline{Y})$$

$$= \int_{\underline{y}} \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1,N}(\underline{x}_1)\, q^{2,N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2) \ln\left(\frac{p(\underline{y} \mid \underline{x}_1, \underline{x}_2)}{p(\underline{y})} \right) d\underline{x}_2\, d\underline{x}_1\, d\underline{y}$$

B-II.[3].

Although we shall not state it explicitly, all distributions q^1 and q^2 which we consider are, by assumption, such that the limit of the average over N as $N \rightarrow \infty$ of the RHS of B-II.[1]-[3] exists. The above notation will later be extended when we consider the case of an arbitrary number of users.

In order to establish our coding theorem with an exponentially decreasing upper bound on the probability of error, we define the following conditions :

$$\forall\, \varepsilon > 0, \forall N_0 \in \mathbb{N}^{*+}, \exists\, N_1 \in \mathbb{N}^{*+} \text{ such that } \forall\, k \in \mathbb{N}^{*+}$$

$$\exp(-\varepsilon) \leq \frac{p(\underline{y}_{k+1,\,k+N_0} \mid \underline{x}_{1_{k+1,\,k+N_0}})}{p(\underline{y}_{k+1,\,k+N_0} \mid \underline{x}_{1_{k+1,\,k+N_0}}, \Xi_1, \Xi)} \leq \exp(\varepsilon)$$

for Ξ_1 any subset of $\underline{x}_{1_{k+1-N_1}}$, Ξ any subset of $\underline{y}_{k+1-N_1}$

B-II.[4]

$$\forall\, \varepsilon > 0, \forall N_0 \in \mathbb{N}^{*+}, \exists\, N_1 \in \mathbb{N}^{*+} \text{ such that } \forall\, k \in \mathbb{N}^{*+}$$

$$\exp(-\varepsilon) \leq \frac{p(\underline{y}_{k+1,\,k+N_0} \mid \underline{x}_{2_{k+1,\,k+N_0}})}{p(\underline{y}_{k+1,\,k+N_0} \mid \underline{x}_{2_{k+1,\,k+N_0}}, \Xi_2, \Xi)} \leq \exp(\varepsilon)$$

for Ξ_2 any subset of $\underline{x}_{2_{k+1-N_1}}$, Ξ any subset of $\underline{y}_{k+1-N_1}$

B-II.[5]

$$\forall\, \varepsilon>0,\ \forall N_0 \in \mathbb{N}^{*+},\ \exists\, N_1 \in \mathbb{N}^{*+} \text{ such that } \forall\, k \in \mathbb{N}^{*-}$$

$$\exp(-\varepsilon) \le \frac{p\left(\underline{y}_{k+1,\,k+N_0} \mid \underline{x}_{1\,k+1,\,k+N_0};\ \underline{x}_{2\,k+1,\,k+N_0}\right)}{p\left(\underline{y}_{k+1,\,k+N_0} \mid \underline{x}_{1\,k+1,\,k+N_0};\ \underline{x}_{2\,k+1,\,k+N_0};\ \Xi_1,\ \Xi_2,\ \Xi\right)} \le \exp(\varepsilon)$$

for Ξ_1 any subset of $\underline{x}_{1\,k+1-N_1}$, Ξ_2 any subset of $\underline{x}_{2\,k+1-N_1}$, Ξ any subset of $\underline{y}_{k+1-N}$

B-II.[6]

to which we shall refer as the decorrelating channel conditions. These conditions will later be contrasted with weaker conditions for the decorrelation of channels which yield a weaker form of the coding theorem. Note that N_0 in the context of this appendix has no relation to the spectral density of noise.

B-III A source coding theorem for a two-user channel with decorrelating channels.

We begin by bounding the probability of error with an appropriate exponential argument, whose behavior we later explore. Our arguments up to B-III.[55] follow very closely the ones of [Gal68] and [Lia72]. The main difference is the use of the full $q^{1,N}$ and $q^{2,N}$ pdfs because of the infinite memory in both inputs and outputs. After B-II.[55], we make use of B-II.[4]-[6] and ergodicity.

Let us first consider errors of type 1. Let us denote by Pe_{1,m_1,m_2} the probability that an error of type 1 occurs conditioned on messages m_1 and m_2 being sent. Using the overbar to denote expectation, we have:

$$\overline{Pe}_{1,m_1,m_2} = \int_{\underline{y}} \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1,N}(\underline{x}_1)\, q^{2,N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2) \times P\left(\left\{(\hat{m}_1 \ne m_1) \cap (\hat{m}_2 = m_2)\right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2\right) d\underline{x}_2\, d\underline{x}_1\, d\underline{y}$$

B-III.[1]

where

$$P\left(\left\{(\hat{m}_1 \ne m_1) \cap (\hat{m}_2 = m_2)\right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2\right) = P\left(\bigcup_{m \ne m_1} \left\{(\hat{m}_1 = m) \cap (\hat{m}_2 = m_2)\right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2\right)$$

B-III.[2].

Using the union bound ([Gal68], pg. 136) on the above RHS, we obtain

$$P\left(\bigcup_{m \neq m_1} \left\{ \left(\hat{m}_1 = m\right) \cap \left(\hat{m}_2 = m_2\right) \right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2 \right)$$

$$\leq \left\{ \sum_{m \neq m_1} P\left(\left\{ \left(\hat{m}_1 = m\right) \cap \left(\hat{m}_2 = m_2\right) \right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2 \right) \right\}^{\rho}$$

B-III.[3]

$\forall\, 0 \leq \rho \leq 1$.

Since we are using maximum-likelihood decoding, the event $\left\{ \left(\hat{m}_1 = m\right) \cap \left(\hat{m}_2 = m_2\right) \right\}$ can occur only if $\underline{x}$ is the m^{th} code word for user 1 and if

$$\frac{p\left(\underline{y} \mid \underline{x}, \underline{x}_2\right)}{p\left(\underline{y} \mid \underline{x}_1, \underline{x}_2\right)} \geq 1$$

B-III.[4].

Since we have a random choice of codewords with probability mass function given by the function q, we may write that

$$P\left(\left\{ \left(\hat{m}_1 \neq m_1\right) \cap \left(\hat{m}_2 = m_2\right) \right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2 \right) \leq \int_{\underline{x} \,:\, \frac{p\left(\underline{y} \mid \underline{x}, \underline{x}_2\right)}{p\left(\underline{y} \mid \underline{x}_1, \underline{x}_2\right)} \geq 1} q^{1..N}(\underline{x})\, d\underline{x}$$

B-III.[5].

Therefore, we have the inequality

$$P\left(\left\{ \left(\hat{m}_1 \neq m_1\right) \cap \left(\hat{m}_2 = m_2\right) \right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2 \right) \leq \int_{\underline{x}} q^{1..N}(\underline{x}) \left\{ \frac{p\left(\underline{y} \mid \underline{x}, \underline{x}_2\right)}{p\left(\underline{y} \mid \underline{x}_1, \underline{x}_2\right)} \right\}^{\theta} d\underline{x}$$

B-III.[6]

$\forall\, \theta > 0$.

Using B-III.[6] and the fact that there are (M_1-1) erroneous messages for user 1, we obtain from B-III.[3] and B-III.[2] the inequality

$$\overline{Pe}_{1,m_1,m_2} \le \int_{\underline{y}} \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1,N}(\underline{x}_1)\, q^{2,N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2) \times \left\{ (M_1-1) \int_{\underline{x}} q^{1,N}(\underline{x}) \left(\frac{p(\underline{y} \mid \underline{x}, \underline{x}_2)}{p(\underline{y} \mid \underline{x}_1, \underline{x}_2)} \right)^{\theta} d\underline{x} \right\}^{\rho} d\underline{x}_2\, d\underline{x}_1\, d\underline{y}$$

B-III.[7].

We can take $\theta = 1/(1+\rho)$, yielding

$$\overline{Pe}_{1,m_1,m_2} \le (M_1-1)^{\rho} \int_{\underline{y}} \int_{\underline{x}_2} q^{2,N}(\underline{x}_2) \left(\int_{\underline{x}} q^{1,N}(\underline{x})\, p(\underline{y} \mid \underline{x}, \underline{x}_2)^{\frac{1}{1+\rho}} d\underline{x} \right)^{1+\rho} d\underline{x}_2\, d\underline{y}$$

B-III.[8].

Rewriting B-III.[8] to use the definition of rate, we have the following lemma.

Lemma B-III.1 : for all $\rho \in [0, 1]$, $q^{1,N}$ and $q^{2,N}$ probability density functions for $\underline{X_1}$ and $\underline{X_2}$ respectively, we have

$$\overline{Pe}_{1,m_1,m_2} \le \exp\left(-N\left(-\rho R_1 + E_0^1\left(\rho, q^{1,N}, q^{2,N}\right)\right)\right)$$

B-III.[9],

where we have defined, analogously to [Gal68],

$$E_0^1\left(\rho, q^{1,N}, q^{2,N}\right) = -\frac{1}{N} \ln \left(\int_{\underline{y}} \int_{\underline{x}_2} q^{2,N}(\underline{x}_2) \left(\int_{\underline{x}} q^{1,N}(\underline{x})\, p(\underline{y} \mid \underline{x}, \underline{x}_2)^{\frac{1}{1+\rho}} d\underline{x} \right)^{1+\rho} d\underline{x}_2\, d\underline{y} \right)$$

B-III.[10].

The same reasoning may be applied to obtain

Lemma B-III.2 : for all $\rho \in [0, 1]$, $q^{1,N}$ and $q^{2,N}$ probability density functions for $\underline{X}_1$ and $\underline{X}_2$ we have

$$\overline{Pe}_{2,m_1,m_2} \leq \exp\left(-N\left(-\rho R_2 + E_0^2\left(\rho, q^{1,N}, q^{2,N}\right)\right)\right)$$

B-III.[11],

where

$$E_0^1\left(\rho, q^{1,N}, q^{2,N}\right)$$

$$= -\frac{1}{N}\ln\left\{\int_{\underline{Y}}\int_{\underline{X}_1} q^{1,N}\left(\underline{x}_1\right)\left(\int_{\underline{X}'} q^{2,N}\left(\underline{x}'\right) p\left(\underline{y} \mid \underline{x}_1, \underline{x}'\right)^{\frac{1}{1+\rho}} d\underline{x}'\right)^{1+\rho} d\underline{x}_1\, d\underline{y}\right\}$$

B-III.[12].

We see that it suffices to determine the behavior of the exponents in B-III.[9] and B-III.[11], to determine whether or not the upper bound to error probability becomes vanishingly small. We therefore proceed to show the following lemma, which closely parallels [Gal68], pg. 142 and [Lia72], pg. 108.

Lemma B-III.3: if $I(\underline{Y} ; \underline{X}_1 \mid \underline{X}_2) > 0$ then for all $1 \geq \rho \geq 0$ we have

$$I\left(\underline{X}_1 ; \underline{Y} \mid \underline{X}_2\right) \geq \frac{\partial N E_0^1\left(\rho, q^{1,N}, q^{2,N}\right)}{\partial \rho} > 0$$

B-III.[13]

$$E_0^1\left(\rho, q^{1,N}, q^{2,N}\right) \geq 0$$

B-III.[14]

$$\frac{\partial^2 E_0^1\left(\rho, q^{1,N}, q^{2,N}\right)}{\partial \rho^2} \leq 0$$

B-III.[15].

Proof: We will first compute the first derivative of E_0^1 for any $q^{1,N}$ and $q^{2,N}$ probability density functions for $\underline{X}_1$ and $\underline{X}_2$ respectively.

$$\frac{\partial E_0^1\left(\rho, q^{1,N}, q^{2,N}\right)}{\partial \rho}$$

$$= -\frac{1}{N}\left\{\int_{\underline{y}}\int_{\underline{x}_2} q^{2,N}(\underline{x}_2)\left(\int_{\underline{x}} q^{1,N}(\underline{x})\, p(\underline{y}\mid \underline{x}, \underline{x}_2)^{\frac{1}{1+\rho}}\, d\underline{x}\right)^{1+\rho} d\underline{x}_2\, d\underline{y}\right\}^{-1}$$

$$\times \frac{\partial}{\partial \rho}\left\{\int_{\underline{y}}\int_{\underline{x}_2} q^{2,N}(\underline{x}_2)\left(\int_{\underline{x}} q^{1,N}(\underline{x})\, p(\underline{y}\mid \underline{x}, \underline{x}_2)^{\frac{1}{1+\rho}}\, d\underline{x}\right)^{1+\rho} d\underline{x}_2\, d\underline{y}\right\}$$

B-III.[16].

Let us use the fact that $\frac{\partial}{\partial x}\left(a^{f(x)}\right) = \ln(a)\left(\frac{\partial f(x)}{\partial x}\right) a^{f(x)}$ to rewrite the second term in the RHS of B-III.[16] as

$$\frac{\partial E_0^1\left(\rho, q^{1,N}, q^{2,N}\right)}{\partial \rho}$$

$$= -\frac{1}{N}\left\{\int_{\underline{y}}\int_{\underline{x}_2} q^{2,N}(\underline{x}_2)\left(\int_{\underline{x}} q^{1,N}(\underline{x})\, p(\underline{y}\mid \underline{x}, \underline{x}_2)^{\frac{1}{1+\rho}}\, d\underline{x}\right)^{1+\rho} d\underline{x}_2\, d\underline{y}\right\}^{-1}$$

$$\times \int_{\underline{y}}\int_{\underline{x}_2} q^{2,N}(\underline{x}_2)$$

$$\left\{ \begin{array}{c} (1+\rho)\left(\int_{\underline{x}} q^{1,N}(\underline{x})\, p(\underline{y} \mid \underline{x}, \underline{x}_2)^{\frac{1}{1+\rho}}\, d\underline{x}\right)^{\rho} \\ \times\left\{ \frac{-1}{(1+\rho)^2} \int_{\underline{x}} q^{1,N}(\underline{x})\, p(\underline{y} \mid \underline{x}, \underline{x}_2)^{\frac{1}{1+\rho}} \ln(p(\underline{y} \mid \underline{x}, \underline{x}_2))\, d\underline{x} \right\} \\ + \left(\int_{\underline{x}} q^{1,N}(\underline{x})\, p(\underline{y} \mid \underline{x}, \underline{x}_2)^{\frac{1}{1+\rho}}\, d\underline{x}\right)^{1+\rho} \ln\left(\int_{\underline{x}} q^{1,N}(\underline{x})\, p(\underline{y} \mid \underline{x}, \underline{x}_2)^{\frac{1}{1+\rho}}\, d\underline{x}\right) \end{array} \right\}$$

$d\underline{x}_2\, d\underline{y}$

B-III.[17].

Setting ρ to be 0, we therefore have that

$$\left.\frac{\partial E_0^1\left(\rho, q^{1,N}, q^{2,N}\right)}{\partial \rho}\right|_{\rho=0} = \frac{-1}{N} \int_{\underline{y}} \int_{\underline{x}_2} q^{2,N}(\underline{x}_2) \left\{\int_{\underline{x}_1} q^{1,N}(\underline{x}_1)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2) d\underline{x}_1\right\} d\underline{x}_2\, d\underline{y}$$

$$\times\left\{-\int_{\underline{x}_1} q^{1,N}(\underline{x}_1)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2) \ln(p(\underline{y} \mid \underline{x}_1, \underline{x}_2))\, d\underline{x}_1\right\}$$

$$+\left\{\int_{\underline{x}_1} q^{1,N}(\underline{x}_1) p(\underline{y} \mid \underline{x}_1, \underline{x}_2)\, d\underline{x}_1\right\} \ln\left\{\int_{\underline{x}_1} q^{1,N}(\underline{x}_1) p(\underline{y} \mid \underline{x}_1, \underline{x}_2)\, d\underline{x}_1\right\}$$

$$= \frac{I_{q^{1,N}, q^{2,N}}(\underline{X}_1; \underline{Y} \mid \underline{X}_2)}{N}$$

B-III.[18].

Let us now show that the first derivative is always positive. The function of ρ given by

$$\left(\int_{\underline{x}} q^{1,N}(\underline{x})\, p(\underline{y} \mid \underline{x}, \underline{x}_2)^{\frac{1}{1+\rho}}\, d\underline{x}\right)^{1+\rho}$$

is nondecreasing in ρ, as shown in [Gal68], pg. 523, relation (e). Therefore the same holds for any positive linear combination of such functions. In particular,

$$\int_{\underline{y}} \int_{\underline{x}_2} q^{2,N}(\underline{x}_2) \left(\int_{\underline{x}} q^{1,N}(\underline{x})\, p(\underline{y} \mid \underline{x}, \underline{x}_2)^{\frac{1}{1+\rho}}\, d\underline{x} \right)^{1+\rho} d\underline{x}_2\, d\underline{y}$$

is nondecreasing in ρ. Therefore, by using the fact that

$$E_0^1\left(0, q^{1,N}, q^{2,N}\right) = 0 \qquad \text{B-III.[19]},$$

and using B-III.[13], we have shown B-III.[14].

To show that the second derivative with respect to ρ is negative, we show that E_0 is a convex ∩ function of ρ, i.e.

$$E_0^1\left(\lambda v+(1-\lambda)w, q^{1,N}, q^{2,N}\right) \geq \lambda E_0^1\left(v, q^{1,N}, q^{2,N}\right) + (1-\lambda) E_0^1\left(w, q^{1,N}, q^{2,N}\right) \qquad \text{B-III.[20]}$$

$\forall \lambda \in [0,1]$
$\forall u, v \in [0,1]$

Establishing that E_0 is a convex ∩ function of ρ is equivalent to establishing that E_0 is a convex ∩ function of 1+ρ. Inequality B-III.[20] can therefore be established as follows:

$$\int_{\underline{y}} \int_{\underline{x}_2} q^{2,N}(\underline{x}_2) \left(\int_{\underline{x}_1} q^{1,N}(\underline{x}_1)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{\lambda v+(1-\lambda)w}}\, d\underline{x}_1 \right)^{\lambda v+(1-\lambda)w} d\underline{x}_2\, d\underline{y}$$

(by applying the result of [Gal68], pg. 194, (5.B.7), which is a consequence of Hölder's inequality)

$$\leq \int_{\underline{y}} \int_{\underline{x}_2} q^{2,N}(\underline{x}_2) \left(\int_{\underline{x}_1} q^{1,N}(\underline{x}_1) p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{v}} d\underline{x}_1 \right)^{\lambda v}$$

$$\times \left(\int_{\underline{x}_1} q^{1,N}(\underline{x}_1) p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{w}} d\underline{x}_1 \right)^{(1-\lambda) w} d\underline{x}_2 \, d\underline{y}$$

(from relation (d) in [Gal68], pg. 523)

$$\leq \int_{\underline{y}} \left\{ \int_{\underline{x}_2} q^{2,N}(\underline{x}_2) \left(\int_{\underline{x}_1} q^{1,N}(\underline{x}_1) p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{v}} d\underline{x}_1 \right)^{v} d\underline{x}_2 \right\}^{\lambda}$$

$$\times \int_{\underline{y}} \left\{ \int_{\underline{x}_2} q^{2,N}(\underline{x}_2) \left(\int_{\underline{x}_1} q^{1,N}(\underline{x}_1) p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{w}} d\underline{x}_1 \right)^{w} d\underline{x}_2 \right\}^{(1-\lambda)} d\underline{y}$$

(applying Hölder's inequality)

$$\leq \left\{ \int_{\underline{y}} \int_{\underline{x}_2} q^{2,N}(\underline{x}_2) \left(\int_{\underline{x}_1} q^{1,N}(\underline{x}_1) p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{v}} d\underline{x}_1 \right)^{v} d\underline{x}_2 \, d\underline{y} \right\}^{\lambda}$$

$$\times \left\{ \int_{\underline{y}} \int_{\underline{x}_2} q^{2,N}(\underline{x}_2) \left(\int_{\underline{x}_1} q^{1,N}(\underline{x}_1) p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{w}} d\underline{x}_1 \right)^{w} d\underline{x}_2 \, d\underline{y} \right\}^{(1-\lambda)}$$

B-III.[21].

By taking logarithms in B-III.[21], we obtain

$$-\ln \left\{ \int_{\underline{y}} \int_{\underline{x}_2} q^{2,N}(\underline{x}_2) \left(\int_{\underline{x}_1} q^{1,N}(\underline{x}_1) \, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{\lambda v + (1-\lambda) w}} d\underline{x}_1 \right)^{\lambda v + (1-\lambda) w} d\underline{x}_2 \, d\underline{y} \right\}$$

$$\geq -\lambda \ln \left\{ \int_{\underline{y}} \int_{\underline{x}_2} q^{2,N}(\underline{x}_2) \left(\int_{\underline{x}_1} q^{1,N}(\underline{x}_1) p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{v}} d\underline{x}_1 \right)^{v} d\underline{x}_2 \, d\underline{y} \right\}$$

$$-(1-\lambda)\ln\left\{ \sum_{\underline{y}} \int_{\underline{x}_2} q^{2,N}(\underline{x}_2)\left\{\int_{\underline{x}_1} q^{1,N}(\underline{x}_1)P(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{w}} d\underline{x}_1\right\}^{w} d\underline{x}_2\, d\underline{y} \right\}$$

B-III.[22].

Therefore, we have established B-III.[20], from which we obtain B-III.[15]. Combining B-III.[15] with B-III.[18] and the fact that the first derivative is positive, we obtain B-III.[13].

Q.E.D.

Let $\mathbf{q}^{1,N}$, $\mathbf{q}^{2,N}$ be a pair of probability density functions for the codewords of length N of users 1 and 2. From B-III.[19] , we see that

$$\left\{E_0^1\left(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N}\right) - \rho R_1\right\}\Big|_{\rho=0} = 0$$

B-III.[23].

The first derivative of $\left\{E_0^1\left(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N}\right) - \rho R_1\right\}$ with respect to ρ is given by

$$\frac{\partial E_0^1\left(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N}\right)}{\partial \rho} - R_1$$

and the second derivative is

$$\frac{\partial^2 E_0^1\left(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N}\right)}{\partial \rho^2}.$$

Therefore, from B-III.[15], in order for $\left\{E_0^1\left(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N}\right) - \rho R_1\right\}$ to be strictly positive for some ρ in [0,1], it is necessary and sufficient that

$$\left\{\frac{\partial E_0^1\left(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N}\right)}{\partial \rho} - R_1\right\}\Big|_{\rho=0} > 0$$

B-III.[24].

From B-III.[18], we see that

$$\frac{1}{N} I_{q^{1,N}, q^{2,N}}(\underline{X}_1; \underline{Y} | \underline{X}_2) > R_1 \geq 0 \Rightarrow \left(\frac{\partial E_0^1(\rho, q^{1,N}, q^{2,N})}{\partial \rho} - R_1 \right) \Big|_{\rho=0} > 0$$

B-III.[25].

Therefore, we have the following lemma.

Lemma B-III.4: for all $q^{1,N}$ and $q^{2,N}$ probability density functions for $\underline{X}_1$ and $\underline{X}_2$ respectively, we have

$$\frac{1}{N} I_{q^{1,N}, q^{2,N}}(\underline{X}_1; \underline{Y} | \underline{X}_2) > R_1 \geq 0 \Rightarrow \exists\, \rho \in [0,1] \text{ s.t. } E_0^1(\rho, q^{1,N}, q^{2,N}) - R_1\rho > 0$$

B-III.[26].

In the same manner, we can establish

Lemma B-III.5: for all $q^{1,N}$ and $q^{2,N}$ probability density functions for $\underline{X}_1$ and $\underline{X}_2$ respectively, we have

$$\frac{1}{N} I_{q^{1,N}, q^{2,N}}(\underline{X}_2, \underline{Y} | \underline{X}_1) > R_2 \geq 0 \Rightarrow \exists\, \rho \in [0,1] \text{ s.t. } E_0^2(\rho, q^{1,N}, q^{2,N}) - R_2\rho > 0$$

B-III.[27].

We point out the difference between our results and those for the memoryless case. We need to consider the distribution over all N input symbols because of the memory. When the channel has finite memory as in B-I.[3], then it sufficient to consider $q^{1,\Delta}$ and $q^{2,\Delta}$.

We wish to establish a result analogous to B-III.[26] and B-III.[27] for errors of type 3. The derivation is similar to that for errors of type 1 and 2, with some modifications. We have that

$$Pe_{3,m_1,m_2} = \int_{\underline{y}} \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1,N}(\underline{x}_1)\, q^{2,N}(\underline{x}_2)\, p(\underline{y} | \underline{x}_1, \underline{x}_2)$$
$$\times P\left(\left\{(\hat{m}_1 = m_1) \cap (\hat{m}_2 \neq m_2)\right\} | \underline{y}, \underline{x}_1, \underline{x}_2\right) d\underline{x}_2\, d\underline{x}_1\, d\underline{y}$$

B-III.[28]

where

$$P\left(\left\{\left(\hat{m}_1 \neq m_1\right) \cap \left(\hat{m}_2 \neq m_2\right)\right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2\right)$$
$$= P\left(\bigcup_{m \neq m_1,\, m' \neq m_2} \left\{\left(\hat{m}_1 = m\right) \cap \left(\hat{m}_2 = m'\right)\right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2\right)$$

B-III.[29].

Using the union bound. we obtain

$$P\left(\bigcup_{m \neq m_1,\, m' \neq m_2} \left\{\left(\hat{m}_1 = m\right) \cap \left(\hat{m}_2 = m'\right)\right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2\right)$$
$$\leq \left\{\sum_{m \neq m_1,\, m' \neq m_2} P\left(\left\{\left(\hat{m}_1 = m\right) \cap \left(\hat{m}_2 = m'\right)\right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2\right)\right\}^{\rho}$$

B-III.[30]

$\forall\, 0 \leq \rho \leq 1$.

The event $\left\{\left(\hat{m}_1 = m\right) \cap \left(\hat{m}_2 = m'\right)\right\}$ occurs only if $\underline{x}$ is the m^{th} codeword for user 1 and $\underline{x}'$ is the m'^{th} codeword for user 2 and

$$\frac{p(\underline{y} \mid \underline{x}, \underline{x}')}{p(\underline{y} \mid \underline{x}_1, \underline{x}_2)} \geq 1$$

B-III.[31].

Since we have a random choice of codewords with probability mass function given by the functions q, we may write that

$$P\left(\left\{\left(\hat{m}_1 = m\right) \cap \left(\hat{m}_2 = m'\right)\right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2\right) \leq \int_{\underline{x}, \underline{x}' : \frac{p(\underline{y} \mid \underline{x}, \underline{x}')}{p(\underline{y} \mid \underline{x}_1, \underline{x}_2)} \geq 1} q^{1,N}(\underline{x})\, q^{2,N}(\underline{x}')\, d\underline{x}\; d\underline{x}'$$

B-III.[32].

Hence

$$P\left(\left\{(\hat{m}_1 = m) \cap (\hat{m}_2 = m')\right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2\right) \leq \int_{\underline{x}, \underline{x}'} q^{1,N}(\underline{x})\, q^{2,N}(\underline{x}') \left\{\frac{p(\underline{y} \mid \underline{x}, \underline{x}')}{p(\underline{y} \mid \underline{x}_1, \underline{x}_2)}\right\}^{\theta} d\underline{x}\; d\underline{x}'$$

B-III.[33]

$\forall\, \theta > 0$

From inequality B-III.[28], B-III.[29], B[30] and B-III.[33] and the fact that there are $(M_1-1)(M_2-1)$ erroneous messages for user 1, we obtain

$$Pe_{3.m_1,m_2} \leq \int_{\underline{y}} \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1,N}(\underline{x}_1)\, q^{2,N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2) \times \left\{(M_1-1)(M_2-1) \int_{\underline{x}, \underline{x}'} q^{1,N}(\underline{x})\, q^{2,N}(\underline{x}') \left\{\frac{p(\underline{y} \mid \underline{x}, \underline{x}')}{p(\underline{y} \mid \underline{x}_1, \underline{x}_2)}\right\}^{\theta} d\underline{x}\; d\underline{x}'\right\}^{\rho} d\underline{x}_2\; d\underline{x}_1\; d\underline{y}$$

B-III.[34].

Taking $\theta = 1/(1+\rho)$ yields

$$Pe_{3.m_1,m_2} \leq (M_1-1)^{\rho}(M_2-1)^{\rho} \int_{\underline{y}} \left\{\int_{\underline{x}, \underline{x}'} q^{1,N}(\underline{x})\, q^{2,N}(\underline{x}')\, p(\underline{y} \mid \underline{x}, \underline{x}')^{\frac{1}{1+\rho}} d\underline{x}\; d\underline{x}'\right\}^{1+\rho}$$

B-III.[35].

Using the definitions of R_1 and R_2, we may rewrite B-III.[35] so that for any (m_1, m_2) pair, we have the following lemma

Lemma B-III.6 : for all $\rho \in [0, 1]$, $q^{1,N}$ and $q^{2,N}$ probability density functions for $\underline{X_1}$ and $\underline{X_2}$ we have

$$Pe_{3.m_1,m_2} \leq \exp\left(-N\left(-\rho(R_1+R_2) + E_0^3\left(\rho, q^{1,N}, q^{2,N}\right)\right)\right)$$

B-III.[36]

where

$$E_0^3\left(\rho, q^{1,N}, q^{2,N}\right) = -\frac{1}{N}\ln\left\{\int_{\underline{y}} \left\{\int_{\underline{x}, \underline{x}'} q^{1,N}(\underline{x})\, q^{2,N}(\underline{x}')\, p(\underline{y} \mid \underline{x}, \underline{x}')^{\frac{1}{1+\rho}} d\underline{x}\; d\underline{x}'\right\}^{1+\rho} d\underline{y}\right\}$$

B-III.[37].

We shall therefore try to determine values of R_1+R_2 for which the expression multiplying N in the exponent of B-III.[36] is positive.

Lemma B-III.7: for any $q^{1.N}$ and $q^{2.N}$ probability density functions for $\underline{X}_1$ and $\underline{X}_2$, we have that for all $1\geq\rho\geq 0$

$$E_0^3\left(\rho, q^{1.N}, q^{2.N}\right) \geq 0$$

B-III.[38]

$$I_{q^{1.N}, q^{2.N}}\left(\underline{X}_1, \underline{X}_2; \underline{Y}\right) \geq \frac{\partial N E_0^3\left(\rho, q^{1.N}, q^{2.N}\right)}{\partial \rho} > 0$$

B-III.[39]

$$\frac{\partial^2 E_0^3\left(\rho, q^{1.N}, q^{2.N}\right)}{\partial \rho^2} \leq 0$$

B-III.[40].

Proof: For any $q^{1.N}$ and $q^{2.N}$ probability density functions for $\underline{X}_1$ and $\underline{X}_2$, let us first compute the first derivative of E_0^3 with respect to ρ.

$$\frac{\partial E_0^3\left(\rho, q^{1.N}, q^{2.N}\right)}{\partial \rho}$$

$$= -\frac{1}{N}\left\{\int_{\underline{y}} \left\{\int_{\underline{x}, \underline{x}'} q^{1.N}(\underline{x})\, q^{2.N}(\underline{x}')\, p(\underline{y} \mid \underline{x}, \underline{x}')^{\frac{1}{1+\rho}} d\underline{x}\, d\underline{x}'\right\}^{1+\rho} d\underline{y}\right\}$$

$$\times \frac{\partial}{\partial \rho}\left\{\int_{\underline{y}}\left\{\int_{\underline{x},\,\underline{x}'} q^{1,N}(\underline{x})\, q^{2,N}(\underline{x}')\, p(\underline{y} \mid \underline{x}, \underline{x}')^{\frac{1}{1+\rho}} d\underline{x}\; d\underline{x}'\right\}^{1+\rho} d\underline{y}\right\}^{-1}$$

$$= -\frac{1}{N}\left\{\int_{\underline{y}}\left\{\int_{\underline{x},\,\underline{x}'} q^{1,N}(\underline{x})\, q^{2,N}(\underline{x}')\, p(\underline{y} \mid \underline{x}, \underline{x}')^{\frac{1}{1+\rho}} d\underline{x}\; d\underline{x}'\right\}^{1+\rho} d\underline{y}\right\}$$

B-III.[41].

Therefore

$$\frac{\partial E_0^3\left(\rho,\, q^{1,N},\, q^{2,N}\right)}{\partial \rho}$$

$$= -\frac{1}{N}\left\{\int_{\underline{y}}\left\{\int_{\underline{x},\,\underline{x}'} q^{1,N}(\underline{x})\, q^{2,N}(\underline{x}')\, p(\underline{y} \mid \underline{x}, \underline{x}')^{\frac{1}{1+\rho}} d\underline{x}\; d\underline{x}'\right\}^{1+\rho} d\underline{y}\right\}$$

$$\times \int_{\underline{y}} \left\{ \begin{array}{l} \left\{\int_{\underline{x},\,\underline{x}'} q^{1,N}(\underline{x})\, q^{2,N}(\underline{x}')\, P(\underline{y} \mid \underline{x}, \underline{x}')^{\frac{1}{1+\rho}} d\underline{x}\; d\underline{x}'\right\}^{\rho} \\ \times\left\{-\frac{1}{(1+\rho)^2}\int_{\underline{x},\,\underline{x}'} q^{1,N}(\underline{x})\, q^{2,N}(\underline{x}')\, p(\underline{y} \mid \underline{x}, \underline{x}')^{\frac{1}{1+\rho}} \ln(p(\underline{y} \mid \underline{x}, \underline{x}'))\, d\underline{x}\; d\underline{x}'\right\} \\ +\left\{\int_{\underline{x},\,\underline{x}'} q^{1,N}(\underline{x})\, q^{2,N}(\underline{x}')\, p(\underline{y} \mid \underline{x}, \underline{x}')^{\frac{1}{1+\rho}} d\underline{x}\; d\underline{x}'\right\}^{1+\rho} \\ \times \ln\left\{\int_{\underline{x},\,\underline{x}'} q^{1,N}(\underline{x})\, q^{2,N}(\underline{x}')\, p(\underline{y} \mid \underline{x}, \underline{x}')^{\frac{1}{1+\rho}} d\underline{x}\; d\underline{x}'\right\} \end{array} \right\} d\underline{y}$$

B-III.[42].

Therefore, by setting $\rho = 0$, we obtain

$$\left.\frac{\partial E_0^3\left(\rho, q^{1..N}, q^{2..N}\right)}{\partial \rho}\right|_{\rho=0} = \frac{1}{N} I_{q^{1..N}, q^{2..N}}(\underline{X}_1, \underline{X}_2; \underline{Y})$$

B-III.[43].

Let us show that the first derivative of E_0^3 with respect to ρ is always positive. The function of ρ given by

$$\left(\int_{\underline{x}_1} \int_{\underline{x}_2} q^{1..N}(\underline{x}_1)\, q^{2..N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{1+\rho}}\, d\underline{x}_1\, d\underline{x}_2\right)^{1+\rho}$$

is nondecreasing in ρ, as shown in [Gal68], pg. 523, expression (e). Therefore,

$$\int_{\underline{y}} \left(\int_{\underline{x}_1} \int_{\underline{x}_2} q^{1..N}(\underline{x}_1)\, q^{2..N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{1+\rho}}\, d\underline{x}_1\, d\underline{x}_2\right)^{1+\rho} d\underline{y}$$

is nondecreasing in ρ. Hence, B-III.[38] holds because of B-II.[39] and

$$E_0^3\left(0, Q^{1..N}, Q^{2..N}\right) = 0$$

B-III.[44].

We now wish to show that the second derivative with respect to ρ is negative. To prove that the second derivative of E_0^3 with respect to ρ is negative, we show that E_0^3 is a convex $\cap$ function of $\rho+1$, i.e.

$$E_0^3\left(\lambda v+(1-\lambda)w, Q^{1..N}, Q^{2..N}\right) \geq \lambda E_0^3\left(v, Q^{1..N}, Q^{2..N}\right) + (1-\lambda) E_0^3\left(w, Q^{1..N}, Q^{2..N}\right)$$

B-III.[45]

$\forall\, \lambda \in [0,1]$
$\forall\, u,v \in [0, 1]$.

Showing that E_0^3 is a convex $\cap$ function of ρ is equivalent to showing that E_0^3 is a convex $\cap$ function of $1+\rho$. Inequality B-III.[45] can be established as follows:

$$\int_{\underline{y}} \left\{ \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1..N}(\underline{x}_1)\, q^{2..N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{\lambda v+(1-\lambda)w}}\, d\underline{x}_2\, d\underline{x}_1 \right\}^{\lambda v+(1-\lambda)w} d\underline{y}$$

(by applying the result of [Gal68], pg. 194, (5.B.7), which is a consequence of Hölder's inequality)

$$\leq \int_{\underline{y}} \left\{ \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1..N}(\underline{x}_1)\, q^{2..N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{v}}\, d\underline{x}_2\, d\underline{x}_1 \right\}^{\lambda v}$$
$$\left\{ \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1..N}(\underline{x}_1)\, q^{2..N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{w}}\, d\underline{x}_2\, d\underline{x}_1 \right\}^{(1-\lambda)w} d\underline{y}$$

(applying Hölder's inequality)

$$\leq \left\{ \int_{\underline{y}} \left\{ \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1..N}(\underline{x}_1)\, q^{2..N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{v}}\, d\underline{x}_2\, d\underline{x}_1 \right\}^{v} d\underline{y} \right\}^{\lambda}$$
$$\times \left\{ \int_{\underline{y}} \left\{ \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1..N}(\underline{x}_1)\, q^{2..N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{w}}\, d\underline{x}_2\, d\underline{x}_1 \right\}^{w} d\underline{y} \right\}^{(1-\lambda)}$$

(by taking logarithms)

$$-\ln \left\{ \int_{\underline{y}} \left\{ \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1..N}(\underline{x}_1)\, q^{2..N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{\lambda v+(1-\lambda)w}}\, d\underline{x}_2\, d\underline{x}_1 \right\}^{\lambda v+(1-\lambda)w} d\underline{y} \right\}$$

$$\geq -\lambda \ln \left\{ \int_{\underline{y}} \left\{ \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1..N}(\underline{x}_1)\, q^{2..N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{v}}\, d\underline{x}_2\, d\underline{x}_1 \right\}^{v} d\underline{y} \right\}$$

$$-(1-\lambda) \ln \left\{ \int_{\underline{y}} \left\{ \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1..N}(\underline{x}_1)\, q^{2..N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)^{\frac{1}{w}}\, d\underline{x}_2\, d\underline{x}_1 \right\}^{w} d\underline{y} \right\}$$

B-III.[46].

Therefore, we have established B-III.[40]. Combining B-III.[40], B-III.[43] and the fact that the first derivative of E_0^3 with respect to ρ is positive, we obtain B-III.[39].

Q.E.D.

Let $\mathbf{q}^{1,N}$, $\mathbf{q}^{2,N}$ be two probability density functions for codewords of length N for users 1 and 2. From B-III.[44], we have that

$$E_0^3(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N}) - \rho(R_1+R_2)\Big|_{\rho=0} = 0$$

B-III.[47].

The first derivative of $E_0^3(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N}) - \rho(R_1+R_2)$ with respect to ρ is given by

$$\frac{\partial E_0^3(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N})}{\partial\rho} - R_1 - R_2$$

and the second derivative is

$$\frac{\partial^2 E_0^3(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N})}{\partial\rho^2}.$$

Therefore, from B-III.[40], in order for $E_0^3(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N}) - \rho(R_1+R_2)$ to be strictly positive for some ρ in [0,1], it is necessary and sufficient that

$$\left\{\frac{\partial E_0^3(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N})}{\partial\rho} - R_1 - R_2\right\}\Big|_{\rho=0} > 0$$

B-III.[48].

From B-III.[43], we see that

$$\frac{1}{N}I_{q^{1,N},\, q^{2,N}}(\underline{X}_1, \underline{Y} | \underline{X}_2) > R_1 + R_2 \geq 0 \Rightarrow \left\{\frac{\delta E_0^1(\rho, \mathbf{q}^{1,N}, \mathbf{q}^{2,N})}{\delta\rho} - R_1 - R_2\right\}\Big|_{\rho=0} > 0$$

B-III.[49].

Therefore, we have the following lemma, which parallels lemmas B-III.4 and B-III.5.

Lemma B-III.8: for all $q^{1,N}$ and $q^{2,N}$ probability density functions for $\underline{X_1}$ and $\underline{X_2}$ we have

$$\frac{1}{N} I_{q^{1,N}, q^{2,N}}\left((\underline{X_1}, \underline{X_2}) ; \underline{Y}\right) > R_1 + R_1 \geq 0$$
$$\Rightarrow \exists\, \rho \in [0,1] \text{ s.t. } E_0^3\left(\rho, q^{1,N}, q^{2,N}\right) - (R_1 + R_2)\rho > 0$$

B-III.[50].

Since the probability of error (including any type of error) obeys

$$Pe \leq Pe_1 + Pe_2 + Pe_3$$

B-III.[51].

We may write, using B-III.[9], B-III.[11] and B-III.[36], that for any pair m_1 and m_2 of messsages

$$Pe_{m_1, m_2} \leq 3e^{-NE_{min}}$$

B-III.[52]

where

$$E_{min} = \min\left[\begin{array}{c} \max_{\rho}\left(E_0^1\left(\rho, q^{1,N}, q^{2,N}\right) - R_1\rho\right),\ \max_{\rho}\left(E_0^2\left(\rho, q^{1,N}, q^{2,N}\right) - R_2\rho\right), \\ \max_{\rho}\left(E_0^3\left(\rho, q^{1,N}, q^{2,N}\right) - (R_1 + R_2)\rho\right) \end{array}\right]$$

B-III.[53].

Combining the above equations with lemmas B-III.4, B-III.5 and B-III.8, we have the following theorem.

Theorem B-III.1 : for all $q^{1,N}$ and $q^{2,N}$ probability density functions for $\underline{X_1}$ and $\underline{X_2}$, for any messages m_1 and m_2 of users 1 and 2, we have

$$Pe_{m_1, m_2} \leq 3e^{-NE_{min}}$$

B-III.[54]

and

$$\left\{\frac{1}{N} I_{q^{1,N}, q^{2,N}}(\underline{X}_1; \underline{Y} | \underline{X}_2) > R_1 \geq 0 \right\} \text{ and } \left\{\frac{1}{N} I_{q^{1,N}, q^{2,N}}(\underline{X}_2; \underline{Y} | \underline{X}_1) > R_2 \geq 0 \right\}$$

$$\text{and } \left\{\frac{1}{N} I_{q^{1,N}, q^{2,N}}((\underline{X}_1, \underline{X}_2) ; \underline{Y}) > R_1 + R_2 \geq 0 \right\} \Rightarrow \quad E_{min} > 0$$

B-III.[55].

The above theorem establishes an upper bound to the probability of error. but this is not enough to show that we have reliable communications possible within the rate region of the LHS of B-III.[55]. The term E_{min} varies with N. If it were to decrease faster than N. then we would not have that the probability of error goes to 0 as N goes to infinity. However. if we show that we can obtain a lower bound on E_{min} as N goes to infinity, we would establish a coding theorem. It is now that we make use of our assumption that the channels decorrelate as given in B-II.[4]-[6] and are ergodic.

Let us consider errors of type 3. The channel is ergodic and the probabilities $q^{1,N}$ and $q^{2,N}$ were chosen such that $\lim_{N \to \infty} \frac{1}{N} I_{q^{1,N}, q^{2,N}}((\underline{X}_1, \underline{X}_2) ; \underline{Y})$ exists. Hence. if we have that

$$\lim_{N \to \infty} \frac{1}{N} I_{q^{1,N}, q^{2,N}}((\underline{X}_1, \underline{X}_2) ; \underline{Y}) > R_1 + R_2 \geq 0$$

B-III.[56]

then, for large enough N_0,

$$\frac{1}{N_0} I_{q^{1,N}, q^{2,N}}((\underline{X}_{1N_0}, \underline{X}_{2N_0}) ; \underline{Y}_{N_0}) - (R_1 + R_2) \geq \frac{\varepsilon}{2} \geq 0$$

B-III.[57]

where

$$\lim_{N \to \infty} \left\{ \frac{1}{N} I_{q^{1,N}, q^{2,N}}((\underline{X_1}, \underline{X_2}) ; \underline{Y}) \right\} - (R_1 + R_2) = \varepsilon$$

B-III.[58].

The ability to measure the channel affects the average behavior of the channel in terms of mutual information. If, for instance, the channel varies very rapidly, then we know that we cannot really use past outputs to measure the channel and N_0 is very small. If the channel varies very slowly, then N_0 will be large to take into account that we may use many past outputs to measure the channel.

Combining B-III.[57], B-III.[58] and B-III.[50] of lemma B-III.[8], we obtain that there exists a $\rho_0 \in [0, 1]$ such that

$$E_0^3\left(\rho_0, q^{1,N_0}, q^{2,N_0}\right) - \left(R_1 + R_2\right)\rho_0 = \zeta > 0$$

B-III.[59].

Using B-II.[6], we select an N_1 such that, for all integer k,

$$\exp\left(\frac{-\zeta}{2}\right) \leq \frac{P\left(\underline{y}_{k+1, k+N_0} \mid \underline{x_1}_{k+1, k+N_0}, \underline{x_2}_{k+1, k+N_0}\right)}{P\left(\underline{y}_{k+1, k+N_0} \mid \underline{x_1}_{k+1, k+N_0}, \underline{x_2}_{k+1, k+N_0}, \Xi_1, \Xi_2, \Xi\right)} \leq \exp\left(\frac{\zeta}{2}\right)$$

for Ξ_1 any subset of $\underline{x_1}_{k+1-N_1}$, Ξ_2 any subset of $\underline{x_2}_{k+1-N_1}$, Ξ any subset of $\underline{y}_{k+1-N}$

B-III.[60].

We assume, w.l.o.g., that we choose N_1 to be a multiple of N_0.

Let us denote by $Q^{1,N}$ and $Q^{2,N}$ the input p.d.f.s constructed in the following manner. For every integer i, the sequence $\underline{X_1}_{iN_0+1,\, i(N_0+1)}$ is distributed according to the distribution Q^{1,N_0} and the other elements are set to be nil. Similarly, for every integer i, the sequence $\underline{X_2}_{iN_0+1,\, i(N_0+1)}$ is distributed according to the distribution Q^{2,N_0} and the other elements are set to be nil. We form a block interleaved code in the following manner. Let us take, w.l.o.g., N to be a multiple of N_1. The first block of the code is created as follows. The first block's symbols are the first N_0 symbols, followed by the symbols from sampled time N_1+1 to N_1+N_0, then the symbols from sampled time $2N_1+1$ to $2N_1+N_0$, ... iN_1+1 to iN_1+N_0, ... $\left(\frac{N}{N_1}-1\right)N_1+1$ to $\left(\frac{N}{N_1}-1\right)N_1+N_0$. The second codeword of the first layer of the code is similar but its symbols are shifted by N_0. Thus, there are $\frac{N_1}{N_0}$ blocks, each composed of $\frac{N}{N_1}$ segments of length N_0. Figure B.3 illustrates our coding technique. For the sake of illustration, we show the simple case where $N_0 = 4$, $N_1 = 8$ and $N = 16$.

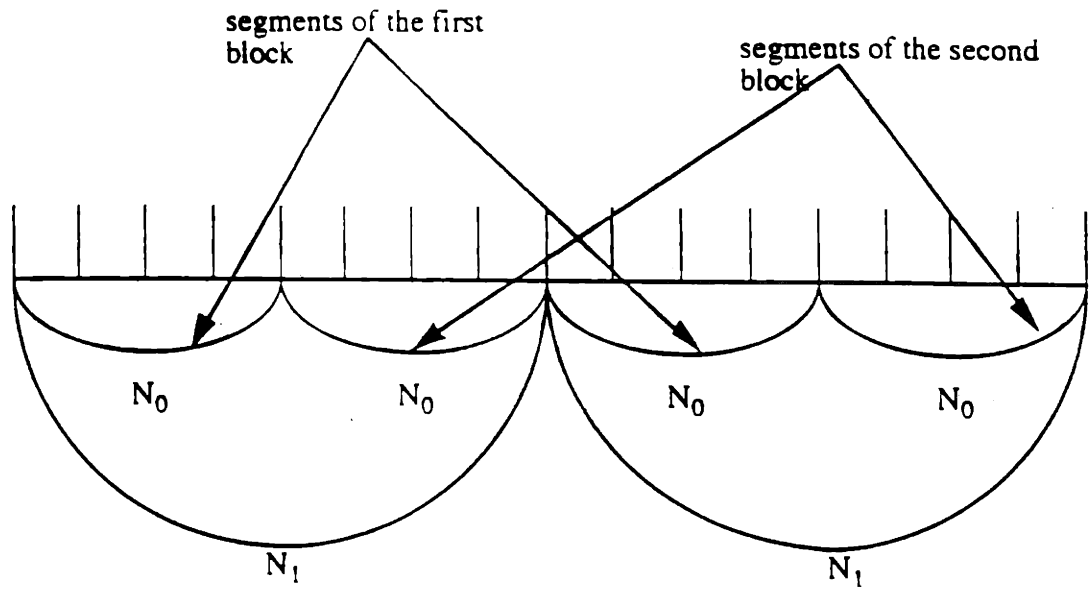


Figure B.3 : Block interleaved coding scheme.

Let us consider the following suboptimal decoding method. We use maximum likelihood decoding for a slightly modified channel in which the coupling among different segements of the code block is broken. We also decode each block independently. Let us define

$$p'(\underline{y} \mid \underline{x}, \underline{x}') = \prod_{i=0}^{\frac{N}{N_1}} p\left(\underline{y}_{iN_1+1,\, iN_1+N_0} \mid \underline{x}_{iN_1+1,\, iN_1+N_0}, \underline{x}'_{iN_1+1,\, iN_1+N_0}\right) \qquad \text{B-III.[61]}$$

and

$$p''(\underline{y} \mid \underline{x}_1, \underline{x}_2) = \prod_{i=0}^{\frac{N}{N_1}} p\left(\underline{y}_{iN_1+1, iN_1+N_0} \mid \underline{x}_{1\, iN_1+1, iN_1+N_0}, \underline{x}_{2\, iN_1+1, iN_1+N_0}\right)$$

B-III.[62].

Let us define the set $\mathcal{E}_3[1]$ of $\underline{x}$ and $\underline{x}'$ for which an error of type 3 occurs with our suboptimal decoding scheme for the first block when $\underline{x}_1$ and $\underline{x}_2$ were sent (the subscript 3 refers to the type of error and the [1] refers to the fact that we are considering the first block):

$$\mathcal{E}_3[1] = \left\{ \begin{array}{c} \underline{x}, \underline{x}' : \dfrac{p'(\underline{y} \mid \underline{x}, \underline{x}')}{p''(\underline{y} \mid \underline{x}_1, \underline{x}_2)} \geq 1 \\ \underline{x} \neq \underline{x}_1, \underline{x}' \neq \underline{x}_2 \end{array} \right\}$$

B-III.[63].

Expression B-III.[32] is replaced by:

$$\int_{\underline{x}, \underline{x}' \in \mathcal{E}_3[1]} q^{1,N}(\underline{x}) q^{2,N}(\underline{x}')\, d\underline{x}\, d\underline{x}' \geq P\left(\left\{(\hat{m}_1 = m) \cap (\hat{m}_2 = m')\right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2\right)$$

B-III.[64]

where we consider messages over the first block.

Hence, for our suboptimal decoding scheme

$$P\left(\left\{(\hat{m}_1 = m) \cap (\hat{m}_2 = m')\right\} \mid \underline{y}, \underline{x}_1, \underline{x}_2\right) \leq \int_{\underline{x}} \int_{\underline{x}'} q^{1,N}(\underline{x}) q^{2,N}(\underline{x}') \left(\frac{p'(\underline{y} \mid \underline{x}, \underline{x}')}{p''(\underline{y} \mid \underline{x}_1, \underline{x}_2)}\right)^{\theta} d\underline{x}\, d\underline{x}'$$

B-III.[65]

$\forall\, \theta > 0$.

Using B-III.[28]-[30] appropriately modified to take into account that we assign $M_1^{\frac{N_0}{N_1}}$ $\left(M_2^{\frac{N_0}{N_1}}\right)$ messages for user 1 (user 2) to the first block, we may write

$$Pe'[1]_{3,\, m_1,\, m_2} \le \int_{\underline{y}} \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1,N}(\underline{x}_1)\, q^{2,N}\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2)$$

$$\left\{ (M_1 - 1)^{\frac{N_0}{N_1}} (M_2 - 1)^{\frac{N_0}{N_1}} \int_{\underline{x}} \int_{\underline{x}'} q^{1,N}(\underline{x}) q^{2,N}(\underline{x}') \left(\frac{p'(\underline{y} \mid \underline{x}, \underline{x}')}{p''(\underline{y} \mid \underline{x}_1, \underline{x}_2)} \right)^{\frac{1}{1+\rho}} d\underline{x}\, d\underline{x}' \right\}^{\rho}$$

$$d\underline{x}_2\, d\underline{x}_1\, d\underline{y}$$

B-III.[66]

where Pe' refers to the probability of error under our suboptimal decoding scheme and the [1] after Pe' refers to the fact that we are considering the first block.

The following lemma, which is similar to B-III.7, holds.

Lemma B-III.9 : for all $\rho \in [0, 1]$, we have

$$Pe'[1]_{3,m_1,\, m_2} \le \exp\left(-N\left(-\rho \left(\frac{N_0}{N_1}\right)(R_1+R_2) + E_0^3\left(\rho,\, q^{1,N},\, q^{2,N}\right)\right)\right)$$

B-III.[67].

Analogously to B-III.[37], we have defined E_0^3 in B-III.[65] as

$$E_0^3\left(\rho,\, q^{1,N},\, q^{2,N}\right)$$

$$= -\frac{1}{N} \ln \left(\int_{\underline{y}} \int_{\underline{x}_1} \int_{\underline{x}_2} q^{1,N}(\underline{x}_1)\, q^{2,N}(\underline{x}_2)\, p(\underline{y} \mid \underline{x}_1, \underline{x}_2) \times \left\{ \int_{\underline{x}} \int_{\underline{x}'} q^{1,N}(\underline{x}) q^{2,N}(\underline{x}') \left(\frac{p'(\underline{y} \mid \underline{x}, \underline{x}')}{p''(\underline{y} \mid \underline{x}_1, \underline{x}_2)} \right)^{\frac{1}{1+\rho}} d\underline{x}\, d\underline{x}' \right\}^{\rho} d\underline{x}_2\, d\underline{x}_1\, d\underline{y} \right)$$

(replacing Q with its product expression and expressing $p(\underline{y} \mid \underline{x}_1, \underline{x}_2)$ as a product)

$$= -\frac{1}{N}\ln\left\{\int_{\underline{y}}\int_{\underline{x}_1}\int_{\underline{x}_2}\prod_{i=0}^{\frac{N}{N_1}} q^{1..N}\left(\underline{x}_{1\,iN_1+1,\,iN_1+N_0}\right) q^{2..N}\left(\underline{x}_{2\,iN_1+1,\,iN_1+N_0}\right) \times p\left(\begin{array}{l}\underline{y}_{iN_1+1,\,iN_1+N_0} \mid \underline{x}_{1\,iN_1+1,\,iN_1+N_0},\\ \underline{x}_{2\,iN_1+1,\,iN_1+N_0},\\ \left\{\underline{y}_{jN_1+1,\,jN_1+N_0}\right\}_{j=0,\ldots,i-1},\\ \left\{\underline{x}_{1\,jN_1+1,\,jN_1+N_0}\right\}_{j=0,\ldots,i-1},\\ \left\{\underline{x}_{2\,jN_1+1,\,jN_1+N_0}\right\}_{j=0,\ldots,i-1}\end{array}\right) \times \left\{\int_{\underline{x}}\int_{\underline{x}'}\prod_{i=0}^{\frac{N}{N_1}} q^{1..N}\left(\underline{x}_{iN_1+1,\,iN_1+N_0}\right) q^{2..N}\left(\underline{x}'_{iN_1+1,\,iN_1+N_0}\right)\left(\frac{p'(\underline{y}\mid\underline{x},\underline{x}')}{p''(\underline{y}\mid\underline{x}_1,\underline{x}_2)}\right)^{\frac{1}{1+\rho}} d\underline{x}\,d\underline{x}'\right\}^{\rho} d\underline{x}_2\,d\underline{x}_1\,d\underline{y}\right\} \qquad \text{B-III.[68]}.$$

We may use B-III.[60] to obtain the following bound from B-III.[68]

$$E_0^3\left(\rho, q^{1..N}, q^{2..N}\right) - \frac{\zeta}{2}\frac{1}{N}\frac{N}{N_1} \leq$$

$$-\frac{1}{N}\ln\left\{\int_{\underline{y}}\int_{\underline{x}_1}\int_{\underline{x}_2}\left(\int_{\underline{x}}\int_{\underline{x}'}\prod_{i=0}^{\frac{N}{N_1}} q^{1,N}\left(\underline{x}_{iN_1+1,\,iN_1+N_0}\right) q^{2,N}\left(\underline{x}'_{iN_1+1,\,iN_1+N_0}\right)\left\{\frac{p'(\underline{y}\mid\underline{x},\underline{x}')}{p''(\underline{y}\mid\underline{x}_1,\underline{x}_2)}\right\}^{\frac{1}{1+\rho}} d\underline{x}\, d\underline{x}'\right)^{\rho+1} d\underline{x}_2\, d\underline{x}_1\, d\underline{y}\right\}$$

$$\leq E_0^3\left(\rho, q^{1,N}, q^{2,N}\right) + \frac{\zeta}{2}\frac{1}{N}\frac{N}{N_1}$$

B-III.[69].

We also have that

$$-\frac{1}{N}\ln\left\{\int_{\underline{y}}\int_{\underline{x}_1}\int_{\underline{x}_2}\left(\int_{\underline{x}}\int_{\underline{x}'}\prod_{i=0}^{\frac{N}{N_1}} q^{1,N}\left(\underline{x}_{iN_1+1,\,iN_1+N_0}\right) q^{2,N}\left(\underline{x}'_{iN_1+1,\,iN_1+N_0}\right)\left\{\frac{p'(\underline{y}\mid\underline{x},\underline{x}')}{p''(\underline{y}\mid\underline{x}_1,\underline{x}_2)}\right\}^{\frac{1}{1+\rho}} d\underline{x}\, d\underline{x}'\right)^{\rho+1} d\underline{x}_2\, d\underline{x}_1\, d\underline{y}\right\}$$

$$=\frac{N_0}{N_1}E_0^3\left(\rho, q^{1,N_0}, q^{1,N_0}\right)$$

B-III.[70].

Using B-III.[69]-[70], we obtain for large enough N that

$$-\frac{\zeta}{2}\frac{N_0}{N_1} \leq E'^{3}_{0}(\rho,\, q^{1..N},\, q^{1..N}) - \frac{N_0}{N_1} E^{3}_{0}(\rho,\, q^{1..N_0},\, q^{1..N_0}) \leq \frac{\zeta}{2}\frac{N_0}{N_1}$$

B-III.[71].

Expressions B-III.[59] and B-III.[71] together yield

$$E^{3}_{0}(\rho_0,\, q^{1..N},\, q^{2..N})\frac{N_0}{N_1} - (R_1 + R_2)\rho_0 \frac{N_0}{N_1} \geq \frac{\zeta}{2}\frac{N_0}{N_1}$$

B-III.[72]

for all large enough N.

We have Pe'[1] ≥ Pe[1]. Moreover, we may write that the probability of error of type 3 over all N is symbols is bounded as

$$Pe_{3,\, m_1,\, m_2} \leq \sum_{i=1}^{\frac{N_1}{N_0}} Pe[i]_{3,\, m_1,\, m_2}$$

B-III.[73].

Combining B-III.[72]-[73], B-III.[59] of lemma B-III.9, B-III.[56] and B-III.[57], we obtain the following lemma.

Lemma B-III.10 : if we have

$$\lim_{N\to\infty} \frac{1}{N} I_{q^{1..N},\, q^{2..N}}((\underline{X}_1, \underline{X}_2)\,;\underline{Y}) > R_1 + R_2 \geq 0$$

B-III.[74]

then there exists $\zeta > 0$ and $z > 0$ such that for all large enough N

$$Pe_{3,\, m_1,\, m_2} \leq z e^{-N\zeta}$$

B-III.[75].

In our derivation, $z = \frac{N_1}{N_0}$ and $\zeta = \frac{N_0}{N_1}\frac{\zeta}{2}$. Similarly, we may derive the following lemmas for errors of type 1 and 2.

Lemma B-III.11 : if we have

$$\lim_{N \to \infty} \frac{1}{N} I_{q^{1,N}, q^{2,N}}(\underline{X_1} ; \underline{Y} \mid \underline{X_2}) > R_1 \geq 0$$

B-III.[76]

then there exists $\zeta' > 0$ and $z' > 0$ such that for all large enough N

$$Pe_{1, m_1, m_2} \leq z' e^{-N\zeta'}$$

B-III.[77].

Lemma B-III.12 : if we have

$$\lim_{N \to \infty} \frac{1}{N} I_{q^{1,N}, q^{2,N}}(\underline{X_2} ; \underline{Y} \mid \underline{X_1}) > R_2 \geq 0$$

B-III.[78]

then there exists $\zeta'' > 0$ and $z'' > 0$ such that for all large enough N

$$Pe_{2, m_1, m_2} \leq z'' e^{-N\zeta''}$$

B-III.[79].

Therefore, if we choose

$$\zeta = \min(\zeta, \zeta', \zeta'')$$

B-III.[80]

and

$z = \max(z, z', z'')$

B-III.[81]

we may write from lemmas B-III.10-12 that

Theorem B-III.2 : if we have

$$\lim_{N \to \infty} \frac{1}{N} I_{q^{1,N}, q^{2,N}}((\underline{X}_1, \underline{X}_2) ; \underline{Y}) > R_1 + R_2 \geq 0$$

B-III.[74]

$$\lim_{N \to \infty} \frac{1}{N} I_{q^{1,N}, q^{2,N}}(\underline{X}_1 ; \underline{Y} \mid \underline{X}_2) > R_1 \geq 0$$

B-III.[76]

$$\lim_{N \to \infty} \frac{1}{N} I_{q^{1,N}, q^{2,N}}(\underline{X}_2 ; \underline{Y} \mid \underline{X}_1) > R_2 \geq 0$$

B-III.[78]

then there exists $\xi > 0$ and $z > 0$ such that for all large enough N

$$Pe \leq z\, e^{-N\xi}$$

B-III.[82].

Remarks :

- We have not used E_{min} directly to prove our coding theorem. However, we can esteablish some properties about E_{min}.

From B-III.[43], we may write that

$$\frac{\partial}{\partial\rho}\left\{E_0^3\left(\rho, q^{1,N}, q^{2,N}\right) - \left(R_1 + R_2\right)\rho\right\}\Big|_{\rho=0} = \frac{1}{N} I_{q^{1,N}, q^{2,N}}\left(\left(\underline{X}_1, \underline{X}_2\right); \underline{Y}\right) - \left(R_1 + R_2\right)$$

B-III.[83].

Moreover,

$$I_{q^{1,N}, q^{2,N}}\left(\left(\underline{X}_1, \underline{X}_2\right); \underline{Y}\right) \geq I_{q^{1,N}, q^{2,N}}\left(\left(\begin{array}{c}\left\{\underline{X}_{1\,iN_0+1,\,(i+1)N_0}\right\}_{i=0}^{\left\lfloor \frac{N}{N_0}\right\rfloor - 1}, \\ \left\{\underline{X}_{2\,iN_0+1,\,(i+1)N_0}\right\}_{i=0}^{\left\lfloor \frac{N}{N_0}\right\rfloor - 1}\end{array}\right); \left\{\underline{Y}_{N_0}\right\}_{i=0}^{\left\lfloor \frac{N}{N_0}\right\rfloor - 1}\right)$$

(rewriting the mutual information as a difference of entropies)

$$= h\left(\left\{\underline{X}_{1\,iN_0+1,\,i(N_0+1)}\right\}_{i=0}^{\left\lfloor \frac{N}{N_0}\right\rfloor - 1}, \left\{\underline{X}_{2\,iN_0+1,\,i(N_0+1)}\right\}_{i=0}^{\left\lfloor \frac{N}{N_0}\right\rfloor - 1}\right)$$

$$- h\left(\left(\begin{array}{c}\left\{\underline{X}_{1\,i(N_0+\Delta)+1,\,i(N_0+\Delta)+N_0}\right\}_{i=0}^{\left\lfloor \frac{N}{N_0+\Delta}\right\rfloor}, \\ \left\{\underline{X}_{2\,i(N_0+\Delta)+1,\,i(N_0+\Delta)+N_0}\right\}_{i=0}^{\left\lfloor \frac{N}{N_0+\Delta}\right\rfloor}\end{array}\right) \Big/ \left\{\underline{Y}_{i(N_0+\Delta)+1,\,(i+1)(N_0+\Delta)}\right\}_{i=0}^{\left\lfloor \frac{N}{N_0+\Delta}\right\rfloor}\right)$$

(rewriting the first term from the definition of q)

$$= \sum_{i=0}^{\left\lfloor \frac{N}{N_0+\Delta} \right\rfloor} h\left(\underline{X}_{1\, i(N_0+\Delta)+1,\, i(N_0+\Delta)+N_0}, \underline{X}_{2\, i(N_0+\Delta)+1,\, i(N_0+\Delta)+N_0}\right)$$

$$- h\left(\left(\begin{matrix} \left(\underline{X}_{1\, i(N_0+\Delta)+1,\, i(N_0+\Delta)+N_0}\right)_{i=0}^{\left\lfloor \frac{N}{N_0+\Delta} \right\rfloor}, \\ \left(\underline{X}_{2\, i(N_0+\Delta)+1,\, i(N_0+\Delta)+N_0}\right)_{i=0}^{\left\lfloor \frac{N}{N_0+\Delta} \right\rfloor} \end{matrix}\right) \middle| \left(\underline{Y}_{i(N_0+\Delta)+1,\, (i+1)(N_0+\Delta)}\right)_{i=0}^{\left\lfloor \frac{N}{N_0+\Delta} \right\rfloor}\right)$$

(using the chain rule for entropy on the second term and using the fact that conditioning reduces entropy)

$$\geq \sum_{i=0}^{\left\lfloor \frac{N}{N_0+\Delta} \right\rfloor} h\left(\underline{X}_{1\, i(N_0+\Delta)+1,\, i(N_0+\Delta)+N_0}, \underline{X}_{2\, i(N_0+\Delta)+1,\, i(N_0+\Delta)+N_0}\right)$$

$$- \sum_{i=0}^{\left\lfloor \frac{N}{N_0+\Delta} \right\rfloor} h\left(\left(\begin{matrix} \underline{X}_{1\, i(N_0+\Delta)+1,\, i(N_0+\Delta)+N_0}, \\ \underline{X}_{2\, i(N_0+\Delta)+1,\, i(N_0+\Delta)+N_0} \end{matrix}\right) \middle| \underline{Y}_{i(N_0+\Delta)+1,\, (i+1)(N_0+\Delta)}\right)$$

(pairing the above entropies into mutual information expressions)

$$= \sum_{i=0}^{\left\lfloor \frac{N}{N_0+\Delta} \right\rfloor} I_{q^{1,N},\, q^{2,N}}\left(\begin{matrix} \left(\underline{X}_{1\, i(N_0+\Delta)+1,\, i(N_0+\Delta)+N_0}, \underline{X}_{2\, i(N_0+\Delta)+1,\, i(N_0+\Delta)+N_0}\right) \\ ; \underline{Y}_{i(N_0+\Delta)+1,\, (i+1)(N_0+\Delta)} \end{matrix}\right)$$

(from the stationarity of the channel)

$$= \left\lfloor \frac{N}{N_0+\Delta} \right\rfloor I_{q^{1,N},\, q^{2,N}}\left(\begin{matrix} \left(\underline{X}_{1\, i(N_0+\Delta)+1,\, i(N_0+\Delta)+N_0}, \underline{X}_{2\, i(N_0+\Delta)+1,\, i(N_0+\Delta)+N_0}\right) \\ ; \underline{Y}_{i(N_0+\Delta)+1,\, (i+1)(N_0+\Delta)} \end{matrix}\right)$$

B-III.[84].

Combining B-III.[83] and B-III.[84], we obtain

$$\frac{\partial}{\partial\rho}\left\{E_0^3\left(\rho, q^{1,N}, q^{2,N}\right) - \left(R_1 + R_2\right)\rho\right\}\Big|_{\rho=0} \geq -\left(R_1 + R_2\right)$$
$$+\frac{1}{N}\left\lfloor\frac{N}{N_0+\Delta}\right\rfloor I_{q^{1,N+\Delta},\, q^{2,N+\Delta}}\left(\left(\underline{X}_{1\, i(N_0+\Delta)+1,\; i(N_0+\Delta)+N_0}, \underline{X}_{2\, i(N_0+\Delta)+1,\; i(N_0+\Delta)+N_0}\right); \underline{Y}_{i(N_0+\Delta)+1,\; (i+1)(N_0+\Delta)}\right)$$

B-III.[85].

For large enough N

$$\frac{\varepsilon}{4} \geq I_{q^{1,N+\Delta},\, q^{2,N+\Delta}}\left(\left(\underline{X}_{1\, i(N_0+\Delta)+1,\; i(N_0+\Delta)+N_0}, \underline{X}_{2\, i(N_0+\Delta)+1,\; i(N_0+\Delta)+N_0}\right); \underline{Y}_{i(N_0+\Delta)+1,\; (i+1)(N_0+\Delta)}\right)$$

$$-\frac{1}{N}\left\lfloor\frac{N}{N_0+\Delta}\right\rfloor I_{q^{1,N+\Delta},\, q^{2,N+\Delta}}\left(\left(\underline{X}_{1\, i(N_0+\Delta)+1,\; i(N_0+\Delta)+N_0}, \underline{X}_{2\, i(N_0+\Delta)+1,\; i(N_0+\Delta)+N_0}\right); \underline{Y}_{i(N_0+\Delta)+1,\; (i+1)(N_0+\Delta)}\right)$$

B-III.[86].

From B-III.[43], we may write that for N large enough

$$\frac{\partial}{\partial\rho}\left\{E_0^3\left(\rho, q^{1,N}, q^{2,N}\right) - \left(R_1 + R_2\right)\rho\right\}\Big|_{\rho=0} \geq \frac{\varepsilon}{4}$$

B-III.[87].

Note that the above is not sufficient to prove our coding theorem. Indeed, the first derivative in the LHS of B-III.[87] might decrease in ρ with N. However, B-III.[87] gives some indication about the behavior of $E_0^3\left(\rho, q^{1,N}, q^{2,N}\right)$. Similar results can be found for $E_0^1\left(\rho, q^{1,N}, q^{2,N}\right)$ and $E_0^2\left(\rho, q^{1,N}, q^{2,N}\right)$.

- We give here another remark about the behavior of $E_0^3\left(\rho, q^{1,N}, q^{2,N}\right)$. The same discussion applies to $E_0^1\left(\rho, q^{1,N}, q^{2,N}\right)$ and $E_0^2\left(\rho, q^{1,N}, q^{2,N}\right)$.

We first state here a lemma which we shall use and whose proof may be found in [Gal68], pages 112-113.

Lemma B.13 : let a_N, for N= 1,2, ...be a sequence such that

$\bar{a} < \infty$

$\underline{a} > -\infty$

where

$\bar{a} = \sup_N a_N$

$\underline{a} = \inf_N a_N$

If for all n≥1 and N≥n

$$a_N \geq \frac{n}{N} a_n + \frac{N-n}{N} a_{N-n}$$

B-III.[88]

then

$$\lim_{N\to\infty} a_N = \bar{a}$$

B-III.[89].

Moreover, if for all n≥1 and N≥n

$$a_N \leq \frac{n}{N} a_n + \frac{N-n}{N} a_{N-n}$$

B-III.[90]

then

$$\lim_{N\to\infty} a_N = \underline{a}$$

B-III.[91].

Let us define

$$F_j^3(\rho) = \max_{q^{1,j},\, q^{2,j}} \left[E_0^3\left(\rho,\, q^{1,j},\, q^{2,j}\right)\right]$$

B-III.[92].

Let us separate the input sequence $\underline{x}_1 = (\underline{x}_1[1], \ldots, \underline{x}_1[N])$ into $\underline{x}'_1 = (\underline{x}_1[1], \ldots, \underline{x}_1[n])$ and $\underline{x}''_1 = (\underline{x}_1[n+1], \ldots, \underline{x}_1[N])$, and the output sequence $\underline{y} = (y_1, \ldots, y_N)$ into $\underline{y}' = (\underline{y}[1], \ldots, \underline{y}[n])$ and $\underline{y}'' = (\underline{y}[n+1], \ldots, \underline{y}[N])$. Let us define the input sequences $\underline{x}'''_1 = (\underline{x}_1[1], \ldots, \underline{x}_1[N-n])$ and $\underline{x}'''_2 = (\underline{x}_2[1], \ldots, \underline{x}_2[N-n])$ and the corresponding output sequence $\underline{y}''' = (\underline{y}[1], \ldots, \underline{y}[N-n]$. Let us consider probability density functions pairs $(q_1^{1,n}\ q_1^{2,n})$ and $(q_2^{1,N-n}\ q_2^{2,N-n})$ for $(\underline{x}'_1, \underline{x}'_2)$ and $(\underline{x}'''_1, \underline{x}'''_2)$ such that they maximize $F_n^3(\rho)$ and $F_{N-n}^3(\rho)$ respectively. Furthermore, let us consider the probability mass function for $\underline{x}$ given by

$$q^{i,N}(\underline{x}) = q_1^{i,n}(\underline{x}_1)\, q_2^{i,N-n}(\underline{x}_2)$$

B-III.[93].

From B-III.[45], we may write that

$$\left(\begin{array}{c} \exp\left(-NE_0^3\left(\rho,\, q^{1,N},\, q^{2,N}\right)\right) \\ -\exp\left(-nE_0^3\left(\rho,\, q^{1,n},\, q^{2,n}\right) - (N-n)E_0^3\left(\rho,\, q^{1,(N-n)},\, q^{2,(N-n)}\right)\right) \end{array}\right)\Bigg|_{\rho=0} = 0$$

B-III.[94].

Let us consider the derivative of the LHS of B-III.[94]:

$$\frac{\partial}{\partial\rho}\left(\begin{array}{c} \exp\left(-NE_0^3\left(\rho,\, q^{1,N},\, q^{2,N}\right)\right) \\ -\exp\left(-nE_0^3\left(\rho,\, q^{1,n},\, q^{2,n}\right) - (N-n)E_0^3\left(\rho,\, q^{1,(N-n)},\, q^{2,(N-n)}\right)\right) \end{array}\right)\Bigg|_{\rho=0}$$

(using B-III.[42] and setting $\rho=0$)

$$-\int_{\underline{y}}\left\{\begin{array}{l}\left(-\int_{\underline{x}_1}\int_{\underline{x}_2} q^{1,N}(\underline{x}_1)\, q^{2,N}(\underline{x}_2)\, P(\underline{y}\mid\underline{x}_1,\underline{x}_2)\ln P(\underline{y}\mid\underline{x}_1,\underline{x}_2)\, d\underline{x}_1\, d\underline{x}_2\right)\\ \quad +\left(\int_{\underline{x}_1}\int_{\underline{x}_2} q^{1,N}(\underline{x}_1)\, q^{2,N}(\underline{x}_2)\, P(\underline{y}\mid\underline{x}_1,\underline{x}_2)\, d\underline{x}_1\, d\underline{x}_2\right)\\ \quad \times\ln\left(\int_{\underline{x}_1}\int_{\underline{x}_2} q^{1,N}(\underline{x}_1)\, q^{2,N}(\underline{x}_2)\, P(\underline{y}\mid\underline{x}_1,\underline{x}_2)\, d\underline{x}_1\, d\underline{x}_2\right)\end{array}\right\} d\underline{y}$$

$$-\int_{\underline{y}}\left\{\begin{array}{l}\left(\begin{array}{l}-\int_{\underline{x}'_1}\int_{\underline{x}'_2}\int_{\underline{x}'''_1}\int_{\underline{x}'''_2} q^{1,n}(\underline{x}'_1)\, q^{1,N-n}(\underline{x}'''_1)\, q^{2,n}(\underline{x}'_2)\, q^{2,N-n}(\underline{x}'''_2)\\ \qquad\times p(\underline{y}'\mid\underline{x}'_1,\underline{x}'_2)\, p(\underline{y}''\mid\underline{x}''_1,\underline{x}''_2)\\ \quad\times\ln\left(p(\underline{y}'\mid\underline{x}'_1,\underline{x}'_2)\, p(\underline{y}'''\mid\underline{x}'''_1,\underline{x}'''_2)\right) d\underline{x}'_1\, d\underline{x}'''_1\, d\underline{x}'_2\, d\underline{x}'''_2\end{array}\right)\\ +\left(\begin{array}{l}\int_{\underline{x}'_1}\int_{\underline{x}'_2}\int_{\underline{x}'''_1}\int_{\underline{x}'''_2} q^{1,n}(\underline{x}'_1)\, q^{1,N-n}(\underline{x}'''_1)\, q^{2,n}(\underline{x}'_2)\, q^{2,N-n}(\underline{x}'''_2)\\ \quad p(\underline{y}'\mid\underline{x}'_1,\underline{x}'_2)\, p(\underline{y}'''\mid\underline{x}'''_1,\underline{x}''_2)\, d\underline{x}'_1\, d\underline{x}'''_1\, d\underline{x}'_2\, d\underline{x}'''_2\end{array}\right)\\ \times\ln\left(\begin{array}{l}\int_{\underline{x}'_1}\int_{\underline{x}'_2}\int_{\underline{x}'''_1}\int_{\underline{x}'''_2} q^{1,n}(\underline{x}'_1)\, q^{1,N-n}(\underline{x}'''_1)\, q^{2,n}(\underline{x}'_2)\, q^{2,N-n}(\underline{x}'''_2)\\ \quad p(\underline{y}'\mid\underline{x}'_1,\underline{x}'_2)\, p(\underline{y}''\mid\underline{x}''_1,\underline{x}''_2)\, d\underline{x}'_1\, d\underline{x}'''_1\, d\underline{x}'_2\, d\underline{x}'''_2\end{array}\right)\end{array}\right\} d\underline{y}$$

(rewriting terms as entropies)

$$= h(\underline{Y}\mid\underline{X}_1,\underline{X}_2) - h(\underline{Y}) - h(\underline{Y}'\mid\underline{X}'_1,\underline{X}'_2) - h(\underline{Y}'''\mid\underline{X}'''_1,\underline{X}'''_2) + h(\underline{Y}') + h(\underline{Y}''')$$

(using the stationarity of the channel to replace $\underline{Y}'''$ by $\underline{Y}''$)

$$= h(\underline{Y}\mid\underline{X}_1,\underline{X}_2) - h(\underline{Y}) - h(\underline{Y}'\mid\underline{X}'_1,\underline{X}'_2) - h(\underline{Y}''\mid\underline{X}''_1,\underline{X}''_2) + h(\underline{Y}') + h(\underline{Y}'')$$

(rewriting $\underline{Y}$ as the concatenation of $\underline{Y}'$ and $\underline{Y}''$ and using the chain rule on the first and second terms)

$$\begin{array}{l}= h(\underline{Y}'\mid\underline{X}'_1,\underline{X}'_2) + h(\underline{Y}''\mid\underline{X}_1,\underline{X}_2,\underline{Y}') - h(\underline{Y}') - h(\underline{Y}''\mid\underline{Y}')\\ \quad - h(\underline{Y}'\mid\underline{X}'_1,\underline{X}'_2) - h(\underline{Y}''\mid\underline{X}''_1,\underline{X}''_2) + h(\underline{Y}') + h(\underline{Y}'')\end{array}$$

(cancelling pairwise the first and fifth terms, and the third and seventh terms)

$$= h\left(\underline{Y''} \mid \underline{X}_1, \underline{X}_2, \underline{Y'}\right) - h\left(\underline{Y''} \mid \underline{Y'}\right) - h\left(\underline{Y''} \mid \underline{X''}_1, \underline{X''}_2\right) + h\left(\underline{Y''}\right)$$

(pairing the first and second terms, and the third and fourth terms into mutual information expressions)

$$= - I\left(\underline{Y''} ; \left(\underline{X}_1, \underline{X}_2\right) \mid \underline{Y'}\right) + I\left(\underline{Y''} ; \left(\underline{X''}_1, \underline{X''}_2\right)\right)$$

(upper bounding the first term by the fact that $I((A,B);C) \geq I(A;C)$)

$$\leq - I\left(\underline{Y''} ; \left(\underline{X''}_1, \underline{X''}_2\right) \mid \underline{Y'}\right) + I\left(\underline{Y''} ; \left(\underline{X''}_1, \underline{X''}_2\right)\right)$$

(rewriting the mutual information terms as entropies)

$$= - h\left(\underline{X''}_1, \underline{X''}_2 \mid \underline{Y'}\right) + h\left(\underline{X''}_1, \underline{X''}_2 \mid \underline{Y'}, \underline{Y''}\right) + h\left(\underline{X''}_1, \underline{X''}_2\right) - h\left(\underline{X''}_1, \underline{X''}_2 \mid \underline{Y''}\right)$$

(cancelling the first and third terms)

$$= h\left(\underline{X''}_1, \underline{X''}_2 \mid \underline{Y'}, \underline{Y''}\right) - h\left(\underline{X''}_1, \underline{X''}_2 \mid \underline{Y''}\right)$$

(rewriting the entropies as a mutual information)

$$= - I\left(\left(\underline{X''}_1, \underline{X''}_2\right); \underline{Y'} \mid \underline{Y''}\right)$$
$$\leq 0$$

B-III.[96].

Therefore, for any positive ε, we can pick j large enough such that there exists a δ for which

$$0 \leq \rho \leq \delta \Rightarrow \overline{F^3}(\rho) - F_j^3(\rho) < \varepsilon$$

B-III.[97].

Similar results can be found for errors of type 1 and type 2. We may note again that B-III.[97] does not by itself prove our coding theorem, because δ may decrease arbitrarily rapidly with N.

B-IV Extension to an arbitrary number of users.

We extend the model used for the two-user case without giving the details of the derivation, since the derivation is a simple extension of the two user case. We need to consider the different error combinations that may arise, rather than just the three cases of the two-user channel. We have the following notation :

K : number of users

U : set $\{1, ..., K\}$.

The following extension to theorem B-III.2 holds:

Theorem B-IV.1 : if we have

$$\lim_{N \to \infty} \frac{1}{N} I_{q^{1,N}, q^{2,N}}\left(\{\underline{X}_i\}_{i \in E} ; \underline{Y} \mid \{\underline{X}_j\}_{j \in U-E}\right) > \sum_{i \in E} R_i \geq 0$$

$$\forall E \subset U$$

B-IV.[1]

then there exists $\xi > 0$ and $z > 0$ such that for all large enough N

$$Pe \leq z\, e^{-N\xi}$$

B-IV.[2].

B-V Extension to weaker decorrelating conditions.

We have discussed in B-I that weaker decorrelating conditions could be given to satisfy a coding theorem without an exponentially decreasing bound of the form of B-IV.[2]. The reasoning is the following. Let the multiplicative part of the channel of the channel be stationary and weakly mixing and the additive noise component of the channel be stationary and weakly mixing. Then, if the input signal is ergodic and all three processes are mutually independent, then the input and the output of the channel are jointly ergodic, from [Pin], pg.74. If the average mutual information per symbol is finite, then the input and output pair is information stable ([Pin], pg. 117). Information stability is defined as the fact that, as the number of samples goes to infinity, the probability that the mutual information between any given sequence of inputs and outputs is different from the expected mutual information goes to 0 ([Pin], pg. 60). Therefore, from [Gal], problem 5.18, we may show that the probability of error goes to 0 as the number of samples increases. However, we do not have an exponential bound, because the definition of information stability does not specify how fast the probability of a sample input-output sequence having the average mutual information approaches 1. Moreover, this sort of argument does not tell us how we would wish to encode.

Conditions B-II.[4]-[6] are not quite satisfied by a Gauss Markov model. However, as long as we restrict the channel not to go above a maximum energy threshold for each tap, then our model holds. Indeed, if we have no threshold, then we could reach a very high energy value for certain taps such that it would take arbitrarily long for the channel to wander off from such a high value. If the channel taps cannot take arbitrarily large energy values, then, after a certain time, the channel will not behave significantly differently whether it started out from a low value or a high value. Such a restriction on the value of the channel taps is not artificial. In a system, arbirtarily high values cannot be registered without eventually running into some sort of saturation. Besides, if the channel is extremely strong, just about any coding will do. When the channel is weaker, transmission of information is more difficult.

Appendix C.

Define

$$z[i] = \frac{z[i-1]}{\alpha^2 + \sigma_\Xi^2 z[i-1]} + \frac{s[i]^2}{\sigma_N^2} \text{ for } i \geq 1$$

$$z[0] = \frac{\alpha^2}{\sigma_F^2}$$

C.[1].

From II-3.[82], we see that, for $i \geq 1$, $z[i]$ is $1/\lambda_i$.

Let us define the function f to be

$$f(z) = \frac{z}{\alpha^2 + \sigma_\Xi^2 z}$$

C.[2].

The first derivative of f is $\frac{\alpha^2}{\left(\alpha^2 + \sigma_\Xi^2 z\right)^2}$, therefore $f'(z)$ is always positive and f is monotonically increasing. The second derivative of f is $\frac{-2\alpha^2 \sigma_\Xi^2}{\left(\alpha^2 + \sigma_\Xi^2 z\right)^3}$, therefore f is concave.

We may rewrite C.[1] as

$$z[i] = f(z[i-1]) + \frac{s[i]^2}{\sigma_N^2}$$

C.[3].

Similarly, we may define

$$z'[i] = f(z'[i-1]) + \frac{\sigma_S^2}{\sigma_N^2} \text{ for } i \geq 1$$

$$z'[0] = \frac{\alpha^2}{\sigma_F^2}$$

C.[4].

From II-3.[83], we see that, for $i \geq 1$, $z'[i]$ is $1/\lambda'_i$. To show Lemma II-3.3, we shall work by induction.

For $i=1$, from C.[1] and C.[3], $z[1] = f\left(\frac{\alpha^2}{\sigma_F^2}\right) + \frac{s[1]^2}{\sigma_N^2}$ and, from C.[4], $z'[1] = f\left(\frac{\alpha^2}{\sigma_F^2}\right) + \frac{\sigma_S^2}{\sigma_N^2}$ Therefore,

$$E_{S[1]}[z[1]] = z'[1]$$

C.[5].

Let us now assume that $E_{\underline{S}_{i-1}}[z[i-1]] \leq z'[i-1]$. In that case,

$$E_{\underline{S}_i}[z[i]] = E_{\underline{S}_{i-1}}[f(z[i-1])] + \frac{\sigma_S^2}{\sigma_N^2}$$

(using the fact that f is concave)

$$\leq f\left(E_{\underline{S}_{i-1}}\left[z[i-1]\right]\right) + \frac{\sigma_S^2}{\sigma_N^2}$$

(using our induction assumption and the fact that f is monotonically increasing)

$$\leq f\left(z'[i-1]\right) + \frac{\sigma_S^2}{\sigma_N^2}$$

C.[6].

Therefore, for all i≥1,

$$E_{\underline{S}_i}\left[z[i]\right] \leq z'[i]$$

C.[7].

Therefore, Lemma II-3.3 and Lemma II-3.4 follow immediately from C.[7].

Appendix D.

First. let us rewrite the argument of the limit in II-3.[67]. We may express the first term as:

$$h\left(\underline{F}^{i,i} \mid \underline{Y}_k, \underline{F}^{i+1,k}\right) = h\left(\underline{F}^{i+1,k}\right) + h\left(\underline{F}^{i,i} \mid \underline{F}^{i+1,k}\right) + h\left(\underline{Y}_k \mid \underline{F}^{i,k}\right) - h\left(\underline{F}^{i+1,k}\right) - h\left(\underline{Y}_k \mid \underline{F}^{i+1,k}\right)$$

(canceling together the first and fourth terms and using the Markov property to rewrite the second term for k large enough with respect to i)

$$= h\left(\underline{F}^{i,i} \mid \underline{F}^{i+1,i+\mu}\right) + h\left(\underline{Y}_k \mid \underline{F}^{i,k}\right) - h\left(\underline{Y}_k \mid \underline{F}^{i+1,k}\right)$$

D.[1]

and the second term as

$$h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+\mu}, \underline{S}_i\right) = h\left(\underline{F}^{i+1,i+\mu}\right) + h\left(\underline{F}^{i,i} \mid \underline{F}^{i+1,i+\mu}\right) + h\left(\underline{S}_i \mid \underline{F}^{i,i+\mu}\right) + h\left(\underline{Y}_i \mid \underline{F}^{i,i+\mu}, \underline{S}_i\right)$$
$$- h\left(\underline{Y}_i \mid \underline{F}^{i+1,i+\mu}\right) - h\left(\underline{F}^{i+1,i+\mu}\right) - h\left(\underline{S}_i \mid \underline{F}^{i+1,i+\mu}\right)$$

(canceling pairwise the third and seventh terms, since the input is independent of the channel, and the first and sixth terms)

$$= h\left(\underline{F}^{i,i} \mid \underline{F}^{i+1,i+\mu}\right) + h\left(\underline{Y}_i \mid \underline{F}^{i,i+\mu}, \underline{S}_i\right) - h\left(\underline{Y}_i \mid \underline{F}^{i+1,i+\mu}\right)$$

D.[2].

Combining D.[1] and D.[2], we obtain:

$$h\left(\underline{F}^{i,i} \mid \underline{Y}_k, \underline{F}^{i+1,k}\right) - h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+\mu}, \underline{S}_i\right)$$
$$= h\left(\underline{F}^{i,i} \mid \underline{F}^{i+1,i+\mu}\right) + h\left(\underline{Y}_k \mid \underline{F}^{i,k}\right) - h\left(\underline{Y}_k \mid \underline{F}^{i+1,k}\right)$$
$$- h\left(\underline{F}^{i,i} \mid \underline{F}^{i+1,i+\mu}\right) + h\left(\underline{Y}_i \mid \underline{F}^{i,i+\mu}, \underline{S}_i\right) - h\left(\underline{Y}_i \mid \underline{F}^{i+1,i+\mu}\right)$$

(canceling the first and fourth terms)

$$=h\left(\underline{Y}_k \mid \underline{F}^{i,k}\right) - h\left(\underline{Y}_k \mid \underline{F}^{i+1,k}\right) + h\left(\underline{Y}_i \mid \underline{F}^{i,i+\mu}, \underline{S}_i\right) - h\left(\underline{Y}_i \mid \underline{F}^{i+1,i+\mu}\right)$$

(rewriting the first and second terms)

$$=h\left(\underline{Y}_i \mid \underline{F}^{i,i}\right) + h\left(\underline{Y}_{i+1,k} \mid \underline{F}^{i+1,k}\right) - h\left(\underline{Y}_i \mid \underline{F}^{i+1,i+1}\right) - h\left(\underline{Y}_{i+1,k} \mid \underline{F}^{i+2,k}\right)$$
$$+ h\left(\underline{Y}_i \mid \underline{F}^{i,i+\mu}, \underline{S}_i\right) - h\left(\underline{Y}_i \mid \underline{F}^{i+1,i+\mu}\right)$$

D.[3].

Now, the RHS of II-3.[79] may be rewritten as:

$$h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,k}\right) - h\left(\underline{F}^{i,i} \mid \underline{Y}_i, \underline{F}^{i+1,i+\mu}, \underline{S}_i\right) = h\left(\underline{F}^{i,i}, \underline{F}^{i+1,i+\mu}, \underline{Y}_i\right) - h\left(\underline{F}^{i+1,i+\mu}, \underline{Y}_i\right)$$
$$- h\left(\underline{F}^{i,i}, \underline{F}^{i+1,i+\mu}, \underline{Y}_i, \underline{S}_i\right) + h\left(\underline{F}^{i+1,i+\mu}, \underline{Y}_i, \underline{S}_i\right)$$

(expanding each term)

$$= h\left(\underline{F}^{i+1,i+\mu}\right) + h\left(\underline{F}^{i,i} \mid \underline{F}^{i+1,i+\mu}\right) + h\left(\underline{Y}_i \mid \underline{F}^{i,i}, \underline{F}^{i+1,i+\mu}\right) - h\left(\underline{Y}_i \mid \underline{F}^{i+1,i+\mu}\right) - h\left(\underline{F}^{i+1,i+\mu}\right)$$
$$- h\left(\underline{F}^{i+1,i+\mu}\right) - h\left(\underline{F}^{i,i} \mid \underline{F}^{i+1,i+\mu}\right) - h(\underline{S}_i) - h\left(\underline{Y}_i \mid \underline{F}^{i,i}, \underline{S}_i\right) + h(\underline{S}_i)$$
$$+ h\left(\underline{F}^{i+1,i+\mu}\right) + h\left(\underline{Y}_i \mid \underline{F}^{i+1,i+\mu}, \underline{S}_i\right)$$

(canceling pairwise the first and fifth terms, the second and seventh terms, the sixth and eleventh terms, the eighth and tenth terms and rewriting the fourth term to eliminate $\underline{F}^{i+1,i+\mu}$)

$$= h\left(\underline{Y}_i \mid \underline{F}^{i,i}\right) - h\left(\underline{Y}_i \mid \underline{F}^{i+1,i+\mu}\right) - h\left(\underline{Y}_i \mid \underline{F}^{i,i}, \underline{S}_i\right) + h\left(\underline{Y}_i \mid \underline{F}^{i+1,i+\mu}, \underline{S}_i\right)$$

D.[4].

Comparing the RHS of D.[3] and D.[4], we see that the bound of II-3.[79] can be obtained from II-3.[67] by upper bounding by 0 the limit $\lim_{i\to\infty} \lim_{k\to\infty}\left(h\left(\underline{Y}_{i+1,k} \mid \underline{F}^{i+1,k}\right) - h\left(\underline{Y}_{i+1,k} \mid \underline{F}^{i+2,k}\right)\right)$. If that limit were indeed 0, then II-3.[79] would be satisfied with equality.

Appendix E.

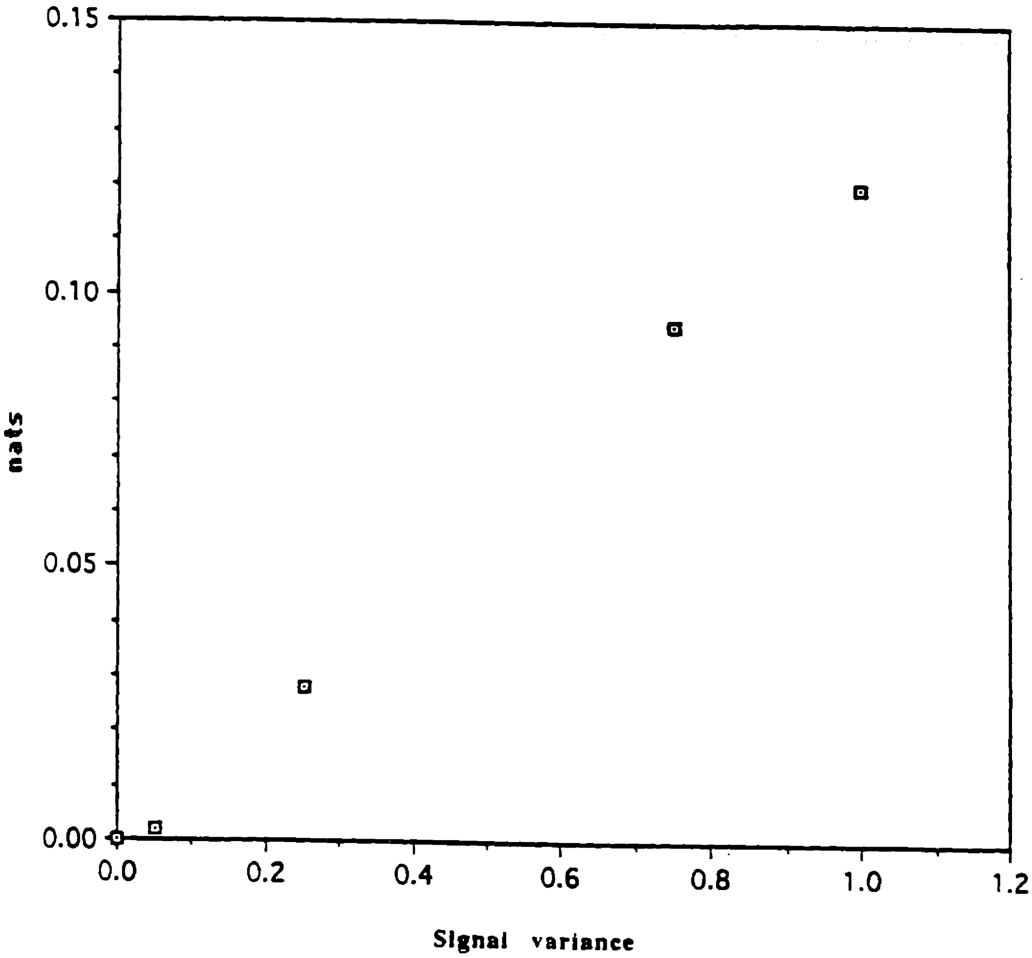


Figure E.1 : Mutual information for 0,1 signaling in a Rayleigh channel with noise variance 1, multiplicative channel variance 4.

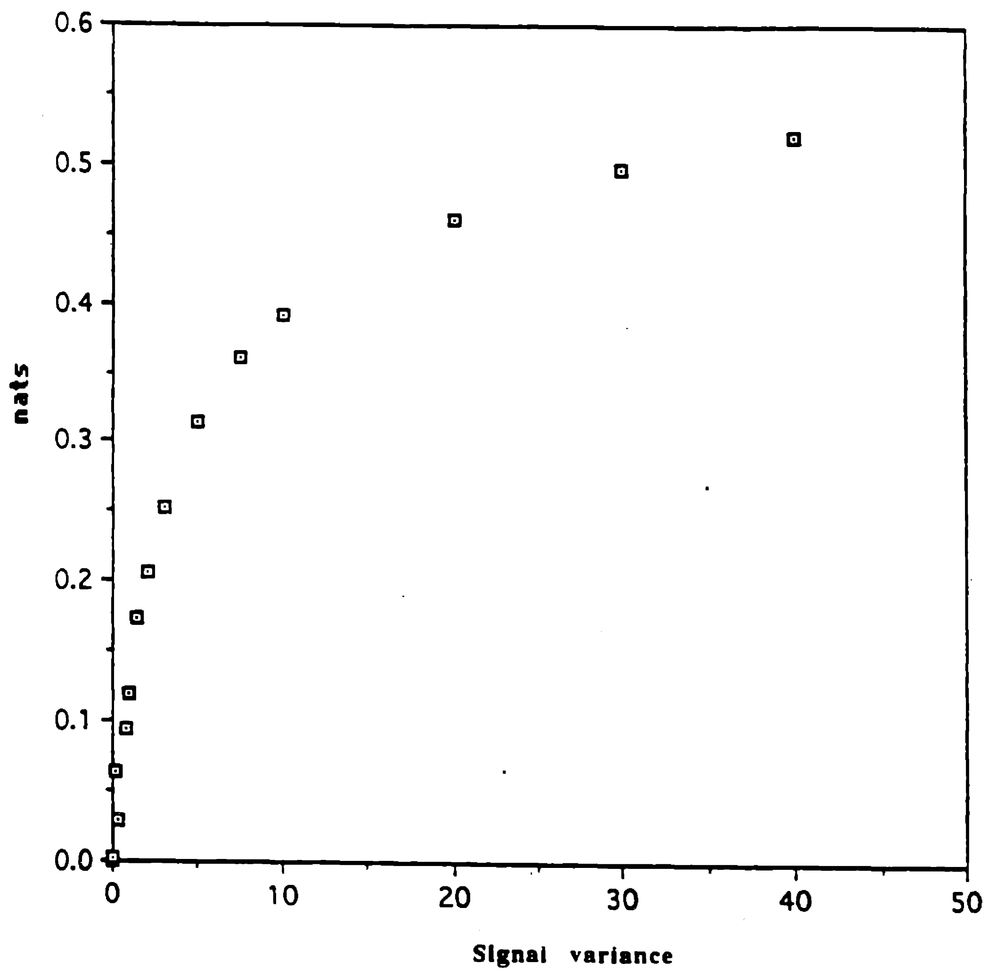


Figure E.2 : Mutual information for 0,1 signaling in a Rayleigh channel with noise variance 1, multiplicative channel variance 4.

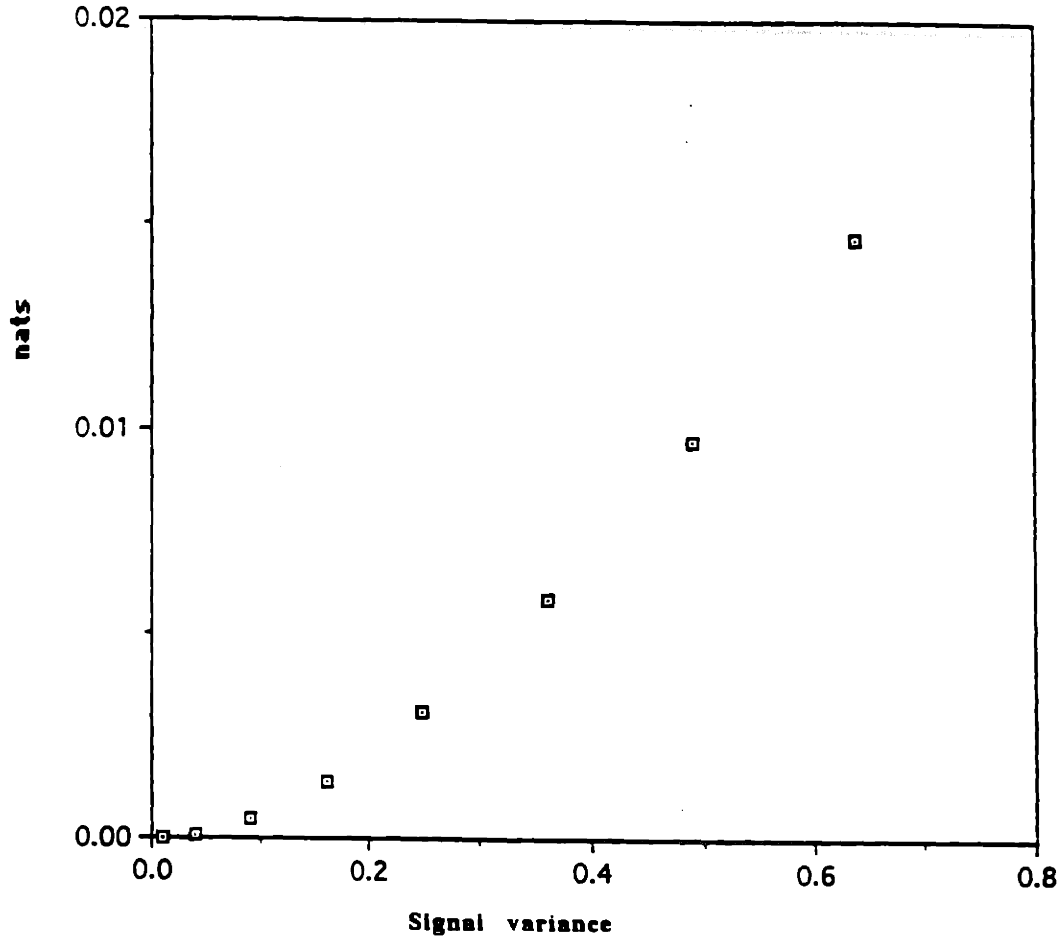


Figure E.3 : Mutual information for 0,1 signaling in a Rayleigh channel with noise variance 4, multiplicative channel variance 4.

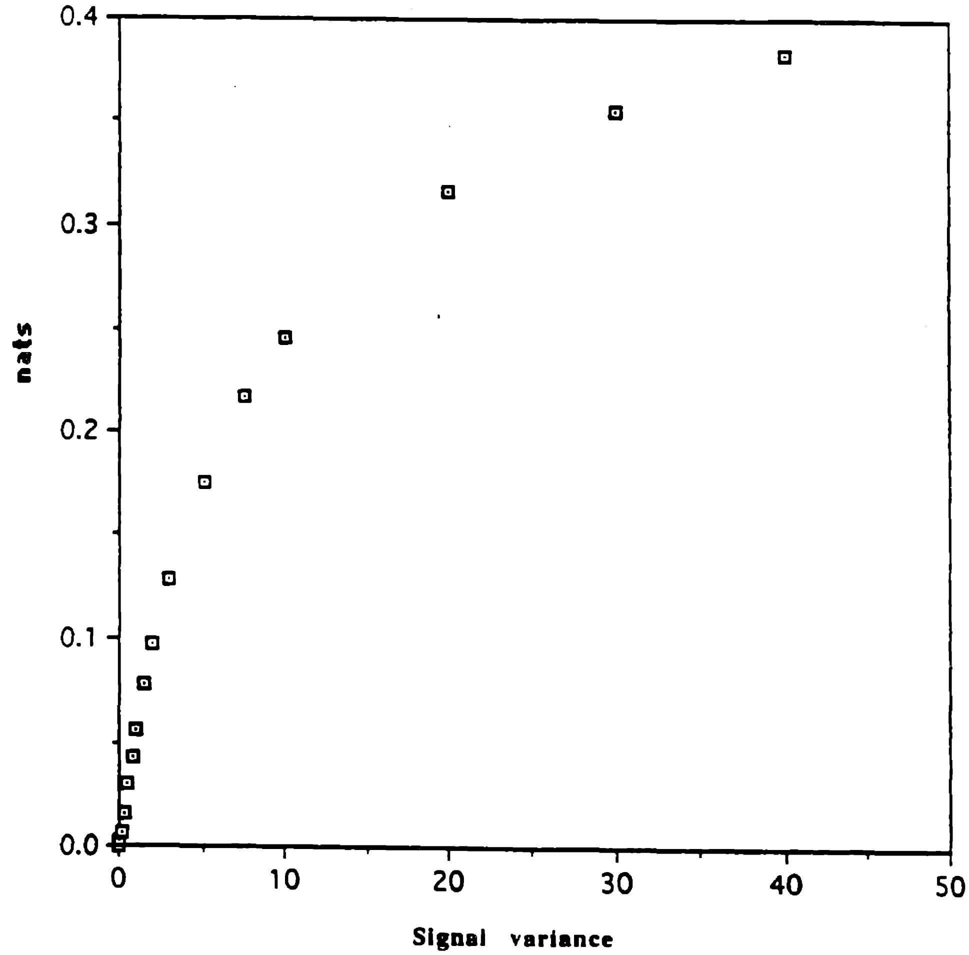


Figure E.4 : Mutual information for 0,1 signaling in a Rayleigh channel with noise variance 4, multiplicative channel variance 4.

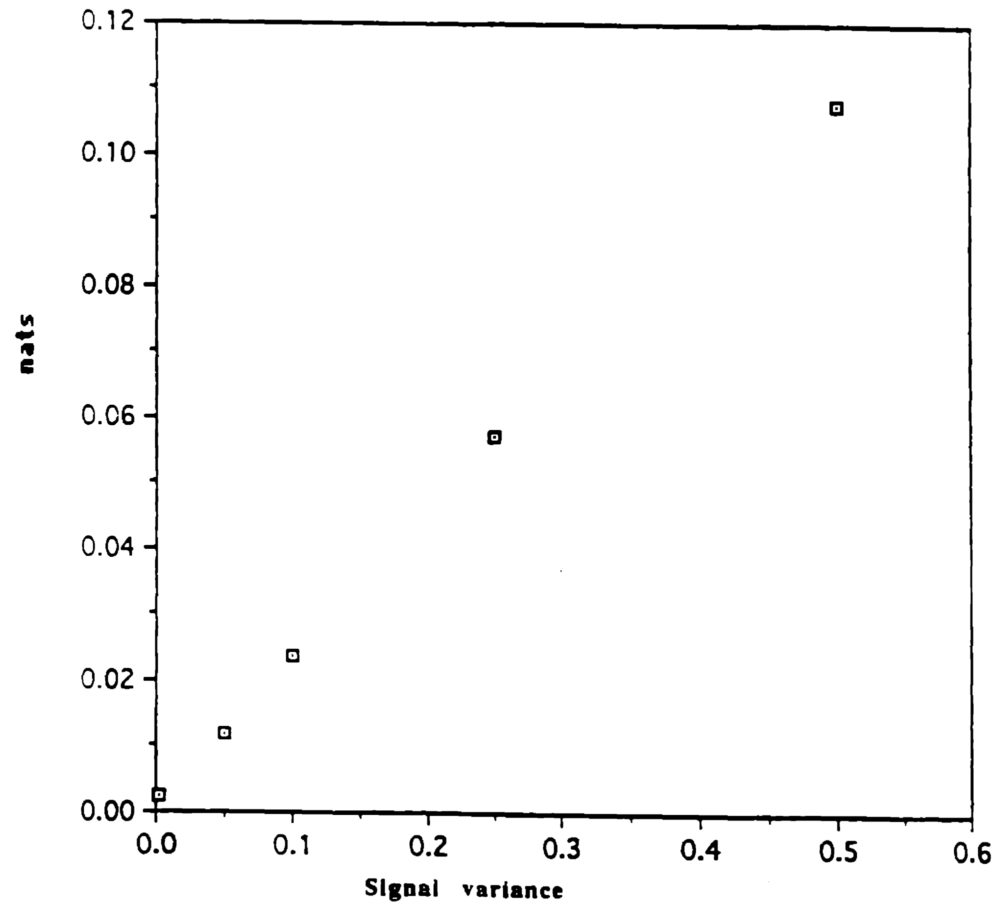


Figure E.5 : Mutual information for 0,1 signaling in a Gaussian channel with a known part having 95 % of the channel energy, noise variance 4, multiplicative channel energy 4.

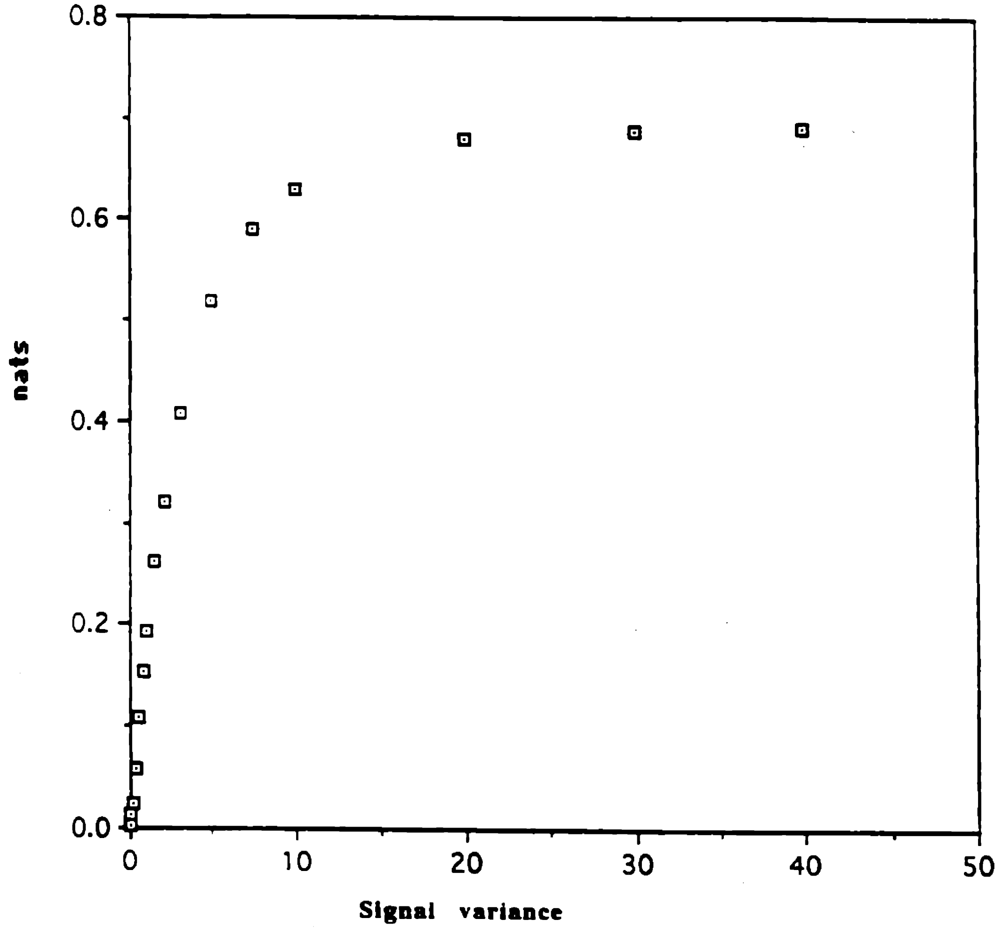
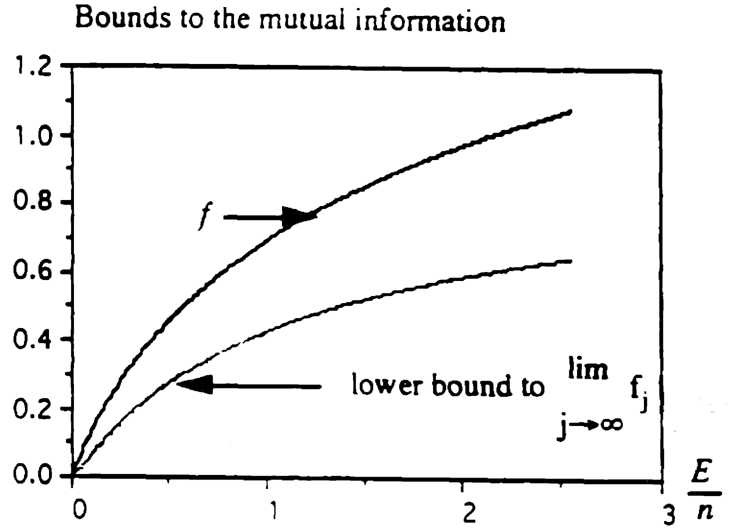


Figure E.6 : Mutual information for 0,1 signaling in a Gaussian channel with a known part having 95 % of the channel energy, noise variance variance 4, multiplicative channel energy 4..

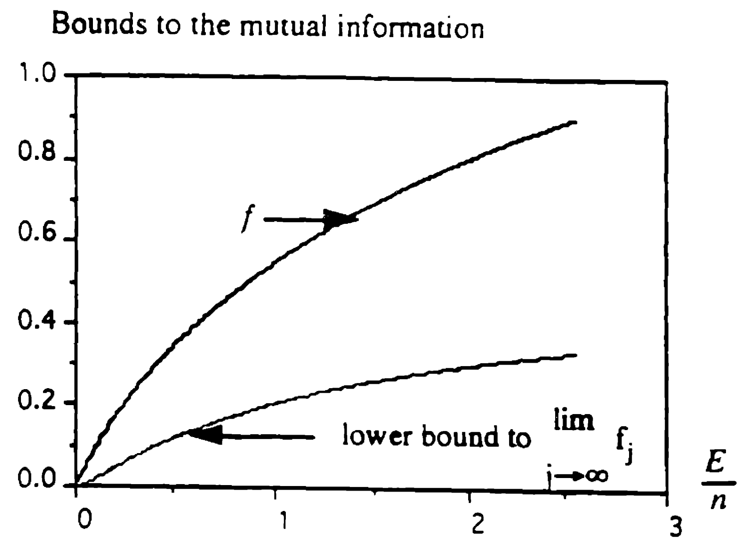
Appendix F.



maximum slope for lower bound is 0.71291933 for $\frac{E}{n} = 0.07$

yielding $\delta' \leq 0.94$

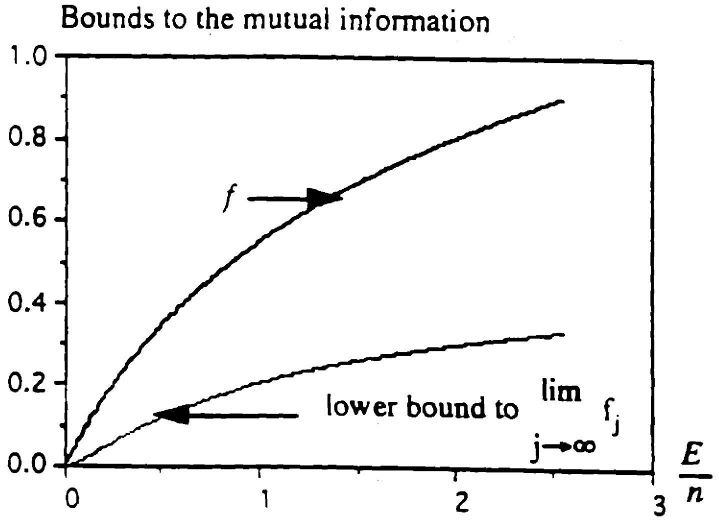
Figure F.1 : Graph of f and $\lim_{j\to\infty} f_j$ versus $\frac{E}{n}$ for $B_{Doppler} = 60$ Hz, spectrum bands of 1000 Hz, $\gamma = .037$, channel strength 3 and noise variance 1



maximum slope for lower bound is 0.1395023 for $\frac{E}{n} = 0.33$

yielding $\delta' \leq 7.81$

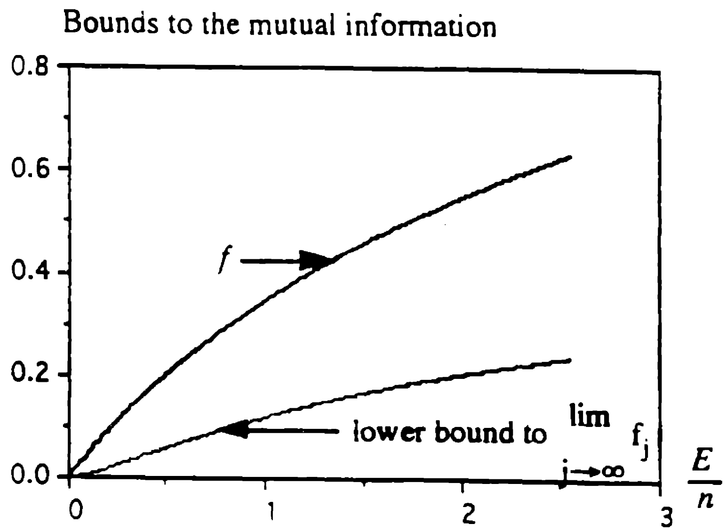
Figure F.2 : Graph of f and $\lim_{j\to\infty} f_j$ versus $\frac{E}{n}$ for $B_{Doppler} = 200$ Hz, spectrum bands of 1000 Hz, $\gamma = .037$, channel strength 1 and noise variance 1



maximum slope for lower bound is 0.27896217 for $\frac{E}{n} = 0.16$

yielding $\delta' \leq 3.9$

Figure F.3 : Graph of f and $\lim_{j \to \infty} f_j$ versus $\frac{E}{n}$ for $B_{Doppler} = 200$ Hz, spectrum bands of 1000 Hz, $\gamma = .037$, channel strength 2 and noise variance 1



maximum slope for lower bound is 0.4184124 for $\frac{E}{n} = 0.11$

yielding $\delta' \leq 2.6$

Figure F.4 : Graph of f and $\lim_{j\to\infty} f_j$ versus $\frac{E}{n}$ for $B_{Doppler} = 200$ Hz, spectrum bands of 1000 Hz, $\gamma = .037$, channel strength 3 and noise variance 1

Appendix G.

The following is a transcript of a Maple V session to find the Kuhn Tucker conditions for two users and k=k'=1.

```
> with(linalg):
Warning: new definition for   norm
Warning: new definition for   trace

[BlockDiagonal, GramSchmidt, JordanBlock, Wronskian,
    add, addcol, addrow, adj, adjoint, angle, augment,
    backsub, band, basis, bezout, blockmatrix, charmat,
    charpoly, col, coldim, colspace, colspan, companion,
    concat, cond, copyinto, crossprod, curl, definite, delcols,
    delrows, det, diag, diverge, dotprod, eigenvals, eigenvects,
    entermatrix, equal, exponential, extend, ffgausselim,
    fibonacci, frobenius, gausselim, gaussjord, genmatrix,
    grad, hadamard, hermite, hessian, hilbert, htranspose,
    ihermite, indexfunc, innerprod, intbasis, inverse, ismith,
    iszero, jacobian, jordan, kernel, laplacian, leastsqrs,
    linsolve, matrix, minor, minpoly, mulcol, mulrow,
    multiply, norm, normalize, nullspace, orthog, permanent,
    pivot, potential, randmatrix, randvector, rank, rsform,
    row, rowdim, rowspace, rowspan, rref, scalarmul,
    singularvals, smith, stack, submatrix, subvector, sumbasis,
    swapcol, swaprow, sylvester, toeplitz, trace, transpose,
    vandermonde, vecpotent, vectdim, vector]
> A1:= matrix([[F111, F112], [F121, F122]]);
```

:

$$A1 := \begin{bmatrix} F111 & F112 \\ F121 & F122 \end{bmatrix}$$

```
> B1 := (s111, s121, s221, s112, s122, s222) -> matrix([[s111, s121], [s121, s221]])
```

$$B1 := (s111, s121, s221, s112, s122, s222) \rightarrow \text{matrix}([[s111, s121], [s121, s221]])$$

```
> A2 := matrix([[F211, F212], [F221, F222]]);
```

$$A2 := \begin{bmatrix} F211 & F212 \\ F221 & F222 \end{bmatrix}$$

```
> B2 := (s111, s121, s221, s112, s122, s222) -> matrix([[s112, s122], [s122, s222]])
  ; evalm(B2(s111, s121, s221, s112, s122, s222));
```

$$B2 := (s111, s121, s221, s112, s122, s222) \rightarrow \text{matrix}([[s112, s122], [s122, s222]])$$

$$\begin{bmatrix} s112 & s122 \\ s122 & s222 \end{bmatrix}$$

```
> N := matrix([[sigma, 0], [0, sigma]]);
```

$$N := \begin{bmatrix} \sigma & 0 \\ 0 & \sigma \end{bmatrix}$$

```
> f := (s111, s121, s221, s112, s122, s222) -> evalm( A1 &* B1(s111, s121, s221, s
  112, s122, s222) &* transpose(A1) + A2 &* B2(s111, s121, s221, s112, s122, s2
  22) &* transpose(A2) + N); evalm(f(s111, s121, s221, s112, s122, s222));
```

$$f := (s111, s121, s221, s112, s122, s222) \rightarrow \text{evalm}(\,(A1 \,\&^*\, \text{B1}(s111, s121, s221, s112, s122, s222)) \,\&^*\, \text{transpose}(A1)) + (\,(A2 \,\&^*\, \text{B2}(s111, s121, s221, s112, s122, s222)) \,\&^*$$

transpose(A2)) + N)

$$[F111^2\, s111 + 2\, F111\, F112\, s121 + F112^2\, s221$$
$$- F211^2\, s112 + 2\, F211\, F212\, s122 + F212^2\, s222 + \sigma,$$
$$F121\, F111\, s111 - F121\, F112\, s121 + F122\, F111\, s121$$
$$+ F122\, F112\, s221 + F221\, F211\, s112 + F221\, F212\, s122$$
$$+ F222\, F211\, s122 + F222\, F212\, s222]$$
$$[F121\, F111\, s111 + F121\, F112\, s121 + F122\, F111\, s121$$
$$+ F122\, F112\, s221 + F221\, F211\, s112 + F221\, F212\, s122$$
$$+ F222\, F211\, s122 + F222\, F212\, s222,\ F121^2\, s111$$
$$+ 2\, F121\, F122\, s121 + F122^2\, s221 + F221^2\, s112$$
$$+ 2\, F221\, F222\, s122 + F222^2\, s222 + \sigma]$$

```
> M := (s111, s121, s221, s112, s122, s222) -> det(f(s111, s121, s221, s112, s122,
  s222)); evalm(M(s111, s121, s221, s112, s122, s222));
```

$$M := (s111, s121, s221, s112, s122, s222) \rightarrow$$
$$\det(f(s111, s121, s221, s112, s122, s222))$$

$$\sigma\, F121^2\, s111 + F112^2\, s221\, \sigma + F211^2\, s112\, \sigma + F111^2\, s111\, \sigma$$
$$+ F111^2\, s111\, F122^2\, s221 + F111^2\, s111\, F221^2\, s112$$
$$+ 2\, F111^2\, s111\, F221\, F222\, s122 + F111^2\, s111\, F222^2\, s222$$
$$- 2\, F111\, F112\, s121\, F121\, F122$$
$$+ 2\, F111\, F112\, s121\, F221^2\, s112$$
$$+ 4\, F111\, F112\, s121\, F221\, F222\, s122$$
$$+ 2\, F111\, F112\, s121\, F222^2\, s222 + F112^2\, s221\, F121^2\, s111$$

$$- F112^2\, s221\, F221^2\, s112 + 2\, F112^2\, s221\, F221\, F222\, s122$$
$$- F112^2\, s221\, F222^2\, s222 + F211^2\, s112\, F121^2\, s111$$
$$- 2\, F211^2\, s112\, F121\, F122\, s121 + F211^2\, s112\, F122^2\, s221$$
$$+ F211^2\, s112\, F222^2\, s222 + \sigma^2$$
$$+ 2\, F211\, F212\, s122\, F121^2\, s111$$
$$+ 4\, F211\, F212\, s122\, F121\, F122\, s121$$
$$+ 2\, F211\, F212\, s122\, F122^2\, s221$$
$$+ 2\, F211\, F212\, s122^2\, F221\, F222$$
$$+ F212^2\, s222\, F121^2\, s111 + 2\, F212^2\, s222\, F121\, F122\, s121$$
$$+ F212^2\, s222\, F122^2\, s221 + F212^2\, s222\, F221^2\, s112$$
$$+ 2\, F111\, F112\, s121\, \sigma + 2\, F211\, F212\, s122\, \sigma$$
$$+ F212^2\, s222\, \sigma + \sigma\, F122^2\, s221 + \sigma\, F221^2\, s112$$
$$+ \sigma\, F222^2\, s222 - F121^2\, F112^2\, s121^2 - F122^2\, F111^2\, s121^2$$
$$- F221^2\, F212^2\, s122^2 - F222^2\, F211^2\, s122^2$$
$$- 2\, \sigma\, F121\, F122\, s121 + 2\, \sigma\, F221\, F222\, s122$$
$$- 2\, F121\, F111\, s111\, F122\, F112\, s221$$
$$- 2\, F121\, F111\, s111\, F221\, F211\, s112$$
$$- 2\, F121\, F111\, s111\, F221\, F212\, s122$$
$$- 2\, F121\, F111\, s111\, F222\, F211\, s122$$
$$- 2\, F121\, F111\, s111\, F222\, F212\, s222$$
$$- 2\, F121\, F112\, s121\, F221\, F211\, s112$$
$$- 2\, F121\, F112\, s121\, F221\, F212\, s122$$

$$- 2\, F121\, F112\, s121\, F222\, F211\, s122$$
$$- 2\, F121\, F112\, s121\, F222\, F212\, s222$$
$$- 2\, F122\, F111\, s121\, F221\, F211\, s112$$
$$- 2\, F122\, F111\, s121\, F221\, F212\, s122$$
$$- 2\, F122\, F111\, s121\, F222\, F211\, s122$$
$$- 2\, F122\, F111\, s121\, F222\, F212\, s222$$
$$- 2\, F122\, F112\, s221\, F221\, F211\, s112$$
$$- 2\, F122\, F112\, s221\, F221\, F212\, s122$$
$$- 2\, F122\, F112\, s221\, F222\, F211\, s122$$
$$- 2\, F122\, F112\, s221\, F222\, F212\, s222$$
$$- 2\, F221\, F211\, s112\, F222\, F212\, s222$$

```
> L := (s111, s121, s221, s112, s122, s222) -> -M(s111, s121, s221, s112, s122, s2
22) + (-s111*s221 + s121^2)*mu1 + (-s112*s222+ s122^2)*mu2 + (s111+s221-P
)*mu3 + (s112+s222 -P)*mu4 - s111*mu5 - s221*mu6 - s112*mu7 - s222*mu8 +(s
121^2 - s111*s221)*mu9 + (s122^2 - s112*s222)*mu10: evalm(L(s111, s121, s22
1, s112, s122, s222));
```

$$L := (s111, s121, s221, s112, s122, s222) \rightarrow$$
$$-M(s111, s121, s221, s112, s122, s222)$$
$$+ (-s111\, s221 + s121^2)\, \mu 1 + (-s112\, s222 + s122^2)\, \mu 2$$
$$+ (s111 + s221 - P)\, \mu 3 + (s112 + s222 - P)\, \mu 4 - s111\, \mu 5$$
$$- s221\, \mu 6 - s112\, \mu 7 - s222\, \mu 8 + (-s111\, s221 + s121^2)\, \mu 9$$
$$+ (-s112\, s222 + s122^2)\, \mu 10$$

$$(-s112\, s222 + s122^2)\, \mu 2 - \sigma\, F121^2\, s111$$
$$+ (-s111\, s221 + s121^2)\, \mu 9 + (-s112\, s222 + s122^2)\, \mu 10$$

$\vdots$

$$- (s112 + s222)\, P\, u4 - (s111\, s221 + s121^2)\, u1$$
$$F112^2\, s221\, \sigma - F211^2\, s112\, \sigma + (s111 + s221)\, P\, u3$$
$$s111\, u5 - s221\, u6 - s112\, u7 - s222\, u8 - F111^2\, s111\, \sigma$$
$$F111^2\, s111\, F122^2\, s221 - F111^2\, s111\, F221^2\, s112$$
$$2\, F111^2\, s111\, F221\, F222\, s122 - F111^2\, s111\, F222^2\, s222$$
$$2\, F111\, F112\, s121^2\, F121\, F122$$
$$2\, F111\, F112\, s121\, F221^2\, s112$$
$$4\, F111\, F112\, s121\, F221\, F222\, s122$$
$$- 2\, F111\, F112\, s121\, F222^2\, s222 - F112^2\, s221\, F121^2\, s111$$
$$- F112^2\, s221\, F221^2\, s112 - 2\, F112^2\, s221\, F221\, F222\, s122$$
$$- F112^2\, s221\, F222^2\, s222 - F211^2\, s112\, F121^2\, s111$$
$$- 2\, F211^2\, s112\, F121\, F122\, s121 - F211^2\, s112\, F122^2\, s221$$
$$- F211^2\, s112\, F222^2\, s222 - \sigma^2$$
$$- 2\, F211\, F212\, s122\, F121^2\, s111$$
$$- 4\, F211\, F212\, s122\, F121\, F122\, s121$$
$$- 2\, F211\, F212\, s122\, F122^2\, s221$$
$$- 2\, F211\, F212\, s122^2\, F221\, F222 - F212^2\, s222\, F121^2\, s111$$
$$- 2\, F212^2\, s222\, F121\, F122\, s121 - F212^2\, s222\, F122^2\, s221$$
$$- F212^2\, s222\, F221^2\, s112 - 2\, F111\, F112\, s121\, \sigma$$
$$- 2\, F211\, F212\, s122\, \sigma - F212^2\, s222\, \sigma - \sigma\, F122^2\, s221$$
$$- \sigma\, F221^2\, s112 - \sigma\, F222^2\, s222 + F121^2\, F112^2\, s121^2$$
$$+ F122^2\, F111^2\, s121^2 + F221^2\, F212^2\, s122^2$$

$$
\begin{aligned}
&- F222^2\, F211^2\, s122^2 + 2\,\sigma\, F121\, F122\, s121 \\
&+ 2\,\sigma\, F221\, F222\, s122 \\
&- 2\, F121\, F111\, s111\, F122\, F112\, s221 \\
&- 2\, F121\, F111\, s111\, F221\, F211\, s112 \\
&+ 2\, F121\, F111\, s111\, F221\, F212\, s122 \\
&- 2\, F121\, F111\, s111\, F222\, F211\, s122 \\
&+ 2\, F121\, F111\, s111\, F222\, F212\, s222 \\
&+ 2\, F121\, F112\, s121\, F221\, F211\, s112 \\
&+ 2\, F121\, F112\, s121\, F221\, F212\, s122 \\
&+ 2\, F121\, F112\, s121\, F222\, F211\, s122 \\
&+ 2\, F121\, F112\, s121\, F222\, F212\, s222 \\
&+ 2\, F122\, F111\, s121\, F221\, F211\, s112 \\
&+ 2\, F122\, F111\, s121\, F221\, F212\, s122 \\
&+ 2\, F122\, F111\, s121\, F222\, F211\, s122 \\
&+ 2\, F122\, F111\, s121\, F222\, F212\, s222 \\
&+ 2\, F122\, F112\, s221\, F221\, F211\, s112 \\
&+ 2\, F122\, F112\, s221\, F221\, F212\, s122 \\
&+ 2\, F122\, F112\, s221\, F222\, F211\, s122 \\
&+ 2\, F122\, F112\, s221\, F222\, F212\, s222 \\
&+ 2\, F221\, F211\, s112\, F222\, F212\, s222
\end{aligned}
$$

```
>
> eq1 := diff(L(s111, s121, s221, s112, s122, s222), s111) = 0:
```

$$
\begin{aligned}
eq1 := & -\sigma F121^2 - s221\,\mu9 - s221\,\mu1 + \mu3 - \mu5 - F111^2\,\sigma \\
& - F111^2\,F122^2\,s221 - F111^2\,F221^2\,s112 \\
& - 2\,F111^2\,F221\,F222\,s122 - F111^2\,F222^2\,s222 \\
& - F112^2\,s221\,F121^2 - F211^2\,s112\,F121^2 \\
& - 2\,F211\,F212\,s122\,F121^2 - F212^2\,s222\,F121^2 \\
& + 2\,F121\,F111\,F122\,F112\,s221 \\
& + 2\,F121\,F111\,F221\,F211\,s112 \\
& + 2\,F121\,F111\,F221\,F212\,s122 \\
& + 2\,F121\,F111\,F222\,F211\,s122 \\
& + 2\,F121\,F111\,F222\,F212\,s222 = 0
\end{aligned}
$$

```
> eq2 := diff(L(s111, s121, s221, s112, s122, s222), s121) = 0;
```

$$
\begin{aligned}
eq2 := & \; 2\,s121\,\mu9 + 2\,s121\,\mu1 - 4\,F111\,F112\,s121\,F121\,F122 \\
& - 2\,F111\,F112\,F221^2\,s112 \\
& - 4\,F111\,F112\,F221\,F222\,s122 \\
& - 2\,F111\,F112\,F222^2\,s222 - 2\,F211^2\,s112\,F121\,F122 \\
& - 4\,F211\,F212\,s122\,F121\,F122 \\
& - 2\,F212^2\,s222\,F121\,F122 - 2\,F111\,F112\,\sigma \\
& + 2\,F121^2\,F112^2\,s121 + 2\,F122^2\,F111^2\,s121 \\
& - 2\,\sigma\,F121\,F122 + 2\,F121\,F112\,F221\,F211\,s112 \\
& + 2\,F121\,F112\,F221\,F212\,s122 \\
& + 2\,F121\,F112\,F222\,F211\,s122 \\
& + 2\,F121\,F112\,F222\,F212\,s222
\end{aligned}
$$

$$- 2\, F122\, F111\, F221\, F211\, s112$$
$$- 2\, F122\, F111\, F221\, F212\, s122$$
$$- 2\, F122\, F111\, F222\, F211\, s122$$
$$- 2\, F122\, F111\, F222\, F212\, s222 = 0$$

```
> eq3 := diff(L(s111, s121, s221, s112, s122, s222), s221) = 0;
```

$$eq3 := -s111\, \mu 9 - s111\, \mu 1 - F112^2\, \sigma + \mu 3 - \mu 6$$
$$- F111^2\, s111\, F122^2 - F112^2\, F121^2\, s111$$
$$- F112^2\, F221^2\, s112 - 2\, F112^2\, F221\, F222\, s122$$
$$- F112^2\, F222^2\, s222 - F211^2\, s112\, F122^2$$
$$- 2\, F211\, F212\, s122\, F122^2 - F212^2\, s222\, F122^2 - \sigma\, F122^2$$
$$+ 2\, F121\, F111\, s111\, F122\, F112$$
$$- 2\, F122\, F112\, F221\, F211\, s112$$
$$+ 2\, F122\, F112\, F221\, F212\, s122$$
$$+ 2\, F122\, F112\, F222\, F211\, s122$$
$$+ 2\, F122\, F112\, F222\, F212\, s222 = 0$$

```
> eq4 := diff(L(s111, s121, s221, s112, s122, s222), s112) = 0;
```

$$eq4 := -s222\, \mu 2 - s222\, \mu 10 + \mu 4 - F211^2\, \sigma - \mu 7$$
$$- F111^2\, s111\, F221^2 - 2\, F111\, F112\, s121\, F221^2$$
$$- F112^2\, s221\, F221^2 - F211^2\, F121^2\, s111$$
$$- 2\, F211^2\, F121\, F122\, s121 - F211^2\, F122^2\, s221$$
$$- F211^2\, F222^2\, s222 - F212^2\, s222\, F221^2 - \sigma\, F221^2$$
$$+ 2\, F121\, F111\, s111\, F221\, F211$$

$$- 2\, F121\, F112\, s121\, F221\, F211$$
$$- 2\, F122\, F111\, s121\, F221\, F211$$
$$- 2\, F122\, F112\, s221\, F221\, F211$$
$$- 2\, F221\, F211\, F222\, F212\, s222 = 0$$

`> eq5 := diff(L(s111, s121, s221, s112, s122, s222), s122) = 0:`

$$eq5 := 2\, s122\, \mu2 + 2\, s122\, \mu10 - 2\, F111^2\, s111\, F221\, F222$$
$$- 4\, F111\, F112\, s121\, F221\, F222$$
$$- 2\, F112^2\, s221\, F221\, F222 - 2\, F211\, F212\, F121^2\, s111$$
$$- 4\, F211\, F212\, F121\, F122\, s121$$
$$- 2\, F211\, F212\, F122^2\, s221$$
$$- 4\, F211\, F212\, s122\, F221\, F222 - 2\, F211\, F212\, \sigma$$
$$+ 2\, F221^2\, F212^2\, s122 + 2\, F222^2\, F211^2\, s122$$
$$- 2\, \sigma\, F221\, F222 + 2\, F121\, F111\, s111\, F221\, F212$$
$$+ 2\, F121\, F111\, s111\, F222\, F211$$
$$+ 2\, F121\, F112\, s121\, F221\, F212$$
$$+ 2\, F121\, F112\, s121\, F222\, F211$$
$$+ 2\, F122\, F111\, s121\, F221\, F212$$
$$+ 2\, F122\, F111\, s121\, F222\, F211$$
$$- 2\, F122\, F112\, s221\, F221\, F212$$
$$+ 2\, F122\, F112\, s221\, F222\, F211 = 0$$

`> eq6 := diff(L(s111, s121, s221, s112, s122, s222), s222) = 0:`

$$eq6 := -s112\, \mu2 - s112\, \mu10 + \mu4 - \mu8 - F111^2\, s111\, F222^2$$

$$- 2\, F111\, F112\, s121\, F222^2 - F112^2\, s221\, F222^2$$
$$F211^2\, s112\, F222^2 - F212^2\, F121^2\, s111$$
$$- 2\, F212^2\, F121\, F122\, s121 - F212^2\, F122^2\, s221$$
$$- F212^2\, F221^2\, s112 - F212^2\, \sigma - \sigma\, F222^2$$
$$+ 2\, F121\, F111\, s111\, F222\, F212$$
$$+ 2\, F121\, F112\, s121\, F222\, F212$$
$$+ 2\, F122\, F111\, s121\, F222\, F212$$
$$- 2\, F122\, F112\, s221\, F222\, F212$$
$$+ 2\, F221\, F211\, s112\, F222\, F212 = 0$$

```
> sols := solve({eq1, eq2, eq3, eq4, eq5, eq6}, {s111, s121, s221, s112, s122, s22
2});
```

$$sols := \{\, s121 = -(-F122\, F121\, F212^2\, \mu 4$$
$$+ F111\, F221^2\, F112\, \mu 8 - F121\, F112\, F222\, F212\, \mu 7$$
$$- \mu 10\, F111\, F112\, \sigma - F121\, F211\, F112\, F221\, \mu 8$$
$$- F222^2\, F111\, F112\, \mu 4 - u2\, F111\, F112\, \sigma$$
$$+ F222^2\, F111\, F112\, \mu 7 - F111\, F221^2\, F112\, \mu 4$$
$$- F122\, F121\, \mu 10\, \sigma + F122\, F211\, F111\, F221\, \mu 4$$
$$+ F122\, F111\, F222\, F212\, \mu 4 - F122\, F111\, F222\, F212\, \mu 7$$
$$+ F121\, F112\, F222\, F212\, \mu 4 - F122\, F211\, F111\, F221\, \mu 8$$
$$+ F121\, F211\, F112\, F221\, \mu 4 + F122\, F121\, F212^2\, \mu 7$$
$$+ F122\, F121\, F211^2\, \mu 8 - F122\, F121\, F211^2\, \mu 4$$
$$- F122\, F121\, u2\, \sigma)/(\%1),\ s222 = -(u9\, \sigma\, F221^2$$
$$+ u9\, F211^2\, \sigma + F112^2\, F121^2\, \mu 7 - F112^2\, F121^2\, \mu 4$$

$- F111^2\, F122^2\, \mu7 - \mu9\, \mu4 - F111^2\, F122^2\, \mu4 + \mu1\, \sigma\, F221^2$
$- \mu1\, F211^2\, \sigma - F211^2\, F122^2\, \mu5 + \mu9\, \mu7$
$- F111^2\, F221^2\, \mu3 + F112^2\, F221^2\, \mu3 - F111^2\, F221^2\, \mu6$
$+ F211^2\, F122^2\, \mu3 - 2\, F121\, F111\, F122\, F112\, \mu7$
$- 2\, F121\, F111\, F122\, F112\, \mu4 - F211^2\, F121^2\, \mu6$
$+ F211^2\, F121^2\, \mu3 - F112^2\, F221^2\, \mu5 - \mu1\, \mu4 + \mu1\, \mu7$
$+ 2\, F121\, F111\, F221\, F211\, \mu6$
$- 2\, F121\, F111\, F221\, F211\, \mu3$
$- 2\, F122\, F112\, F221\, F211\, \mu3$
$+ 2\, F122\, F112\, F221\, F211\, \mu5) / (\%1),\ s221 = ($
$-\mu5\, F221^2\, F212^2 - \mu5\, F222^2\, F211^2 + \mu3\, F221^2\, F212^2$
$+ \mu3\, F222^2\, F211^2 - \mu4\, F111^2\, F221^2 - \mu4\, F121^2\, F212^2$
$- \mu4\, F111^2\, F222^2 - \mu4\, F211^2\, F121^2 - \sigma\, \mu2\, F111^2$
$- \sigma\, \mu2\, F121^2 - \sigma\, \mu10\, F121^2 - \sigma\, \mu10\, F111^2$
$+ \mu8\, F211^2\, F121^2 + \mu8\, F111^2\, F221^2 + \mu7\, F121^2\, F212^2$
$+ \mu7\, F111^2\, F222^2 + \mu3\, \mu2 + \mu3\, \mu10 - \mu5\, \mu2 - \mu5\, \mu10$
$- 2\, \mu7\, F121\, F111\, F212\, F222$
$- 2\, \mu8\, F121\, F111\, F221\, F211$
$- 2\, \mu4\, F121\, F111\, F221\, F211$
$- 2\, \mu4\, F121\, F111\, F212\, F222$
$- 2\, \mu3\, F211\, F212\, F221\, F222$
$+ 2\, \mu5\, F211\, F212\, F221\, F222) / (\%1),\ s122 = - ($

$-F121^2\ F211\ F212\ \mu3 - F121^2\ F211\ F212\ \mu6$

$-F121\ F211\ F222\ F111\ \mu6 + F121\ F211\ F222\ F111\ \mu3$

$-F121\ F212\ \mu6\ F111\ F221 + F121\ F212\ \mu3\ F111\ F221$

$-F211\ F222\ F122\ F112\ \mu5 + F211\ F222\ F122\ F112\ \mu3$

$-F211\ F212\ \mu1\ \sigma + F211\ F212\ F122^2\ \mu5$

$-F211\ F212\ \mu9\ \sigma - F211\ F212\ F122^2\ \mu3$

$+ F222\ F111^2\ F221\ \mu6 - F222\ F111^2\ F221\ \mu3$

$-F222\ \sigma\ F221\ \mu9 + F222\ F112^2\ F221\ \mu5$

$-F222\ \sigma\ F221\ \mu1 - F222\ F112^2\ F221\ \mu3$

$+ F122\ F112\ F221\ F212\ \mu3 - F122\ F112\ F221\ F212\ \mu5)$

$/(\%1),\ s111 = (\mu3\ F221^2\ F212^2 + \mu3\ F222^2\ F211^2$

$-F221^2\ F212^2\ \mu6 + F222^2\ F112^2\ \mu7 - F222^2\ F112^2\ \mu4$

$-F211^2\ F222^2\ \mu6 + F211^2\ F122^2\ \mu8 - F211^2\ F122^2\ \mu4$

$-\sigma\ \mu10\ F112^2 - \sigma\ \mu2\ F112^2 + F122^2\ \mu7\ F212^2$

$-F112^2\ F221^2\ \mu8 - F112^2\ F221^2\ \mu4 - F122^2\ \sigma\ \mu10$

$-F122^2\ \sigma\ \mu2 - F122^2\ \mu4\ F212^2 + \mu3\ \mu2 + \mu3\ \mu10$

$-2\ \mu3\ F211\ F212\ F221\ F222 - \mu2\ \mu6 - \mu10\ \mu6$

$-2\ F122\ F112\ F222\ F212\ \mu7$

$+ 2\ F122\ F112\ F222\ F212\ \mu4$

$+ 2\ F211\ F122\ F112\ F221\ \mu4$

$-2\ F211\ F122\ F112\ F221\ \mu8$

$+ 2\ F211\ F212\ F221\ F222\ \mu6)/(\%1),\ s112 = (-\mu9\ \mu8$

$$
\begin{aligned}
& F112^2\, F222^2\, \mu3 - F212^2\, F121^2\, \mu3 + F212^2\, F121^2\, \mu6 \\
& - F212^2\, \mu1\, \sigma - F212^2\, \mu9\, \sigma + F112^2\, F121^2\, \mu4 + \mu9\, \mu4 \\
& - F111^2\, F122^2\, \mu4 - F122^2\, F212^2\, \mu3 - F122^2\, F111^2\, \mu8 \\
& - F122^2\, F212^2\, \mu5 - \sigma\, F222^2\, \mu9 - \sigma\, F222^2\, \mu1 \\
& - F111^2\, F222^2\, \mu6 - F112^2\, F121^2\, \mu8 - F111^2\, \mu3\, F222^2 \\
& - F112^2\, F222^2\, \mu5 - 2\, F212\, F121\, F222\, F111\, \mu6 \\
& + 2\, F212\, F121\, F222\, F111\, \mu3 - \mu1\, \mu8 \\
& - 2\, F121\, F111\, F122\, F112\, \mu4 \\
& - 2\, F122\, F112\, F222\, F212\, \mu5 \\
& + 2\, F122\, F121\, F111\, F112\, \mu8 \\
& - 2\, F122\, F112\, F222\, F212\, \mu3 + \mu1\, \mu4)/(\%1))
\end{aligned}
$$

$$
\begin{aligned}
\%1 := {} & F122^2\, \mu2\, F111^2 + F122^2\, \mu10\, F111^2 \\
& - 2\, F122\, F121\, \mu2\, F111\, F112 \\
& - 2\, F122\, F121\, \mu10\, F111\, F112 + F121^2\, \mu2\, F112^2 \\
& + F121^2\, \mu10\, F112^2 + F211^2\, F222^2\, \mu9 + F211^2\, F222^2\, \mu1 \\
& - 2\, F211\, F212\, F221\, F222\, \mu9 \\
& - 2\, F211\, F212\, F221\, F222\, \mu1 + \mu2\, \mu9 + F221^2\, F212^2\, \mu1 \\
& + \mu2\, \mu1 + \mu10\, \mu9 + F221^2\, F212^2\, \mu9 + \mu10\, \mu1
\end{aligned}
$$

Appendix H.

The figures in this appendix show, for a given $|U|$, the ratio of the bound in III-3.[17] to the mutual information which would be obtained with perfectly known AWGN channels with the same energy as the Gauss-Markov channels. Figure H.1 illustrates the case for $|U| = 2$. Figure H.2 the case for $|U| = 10$ and Figure H.3 the case for $|U| = 50$.

Figure H.1

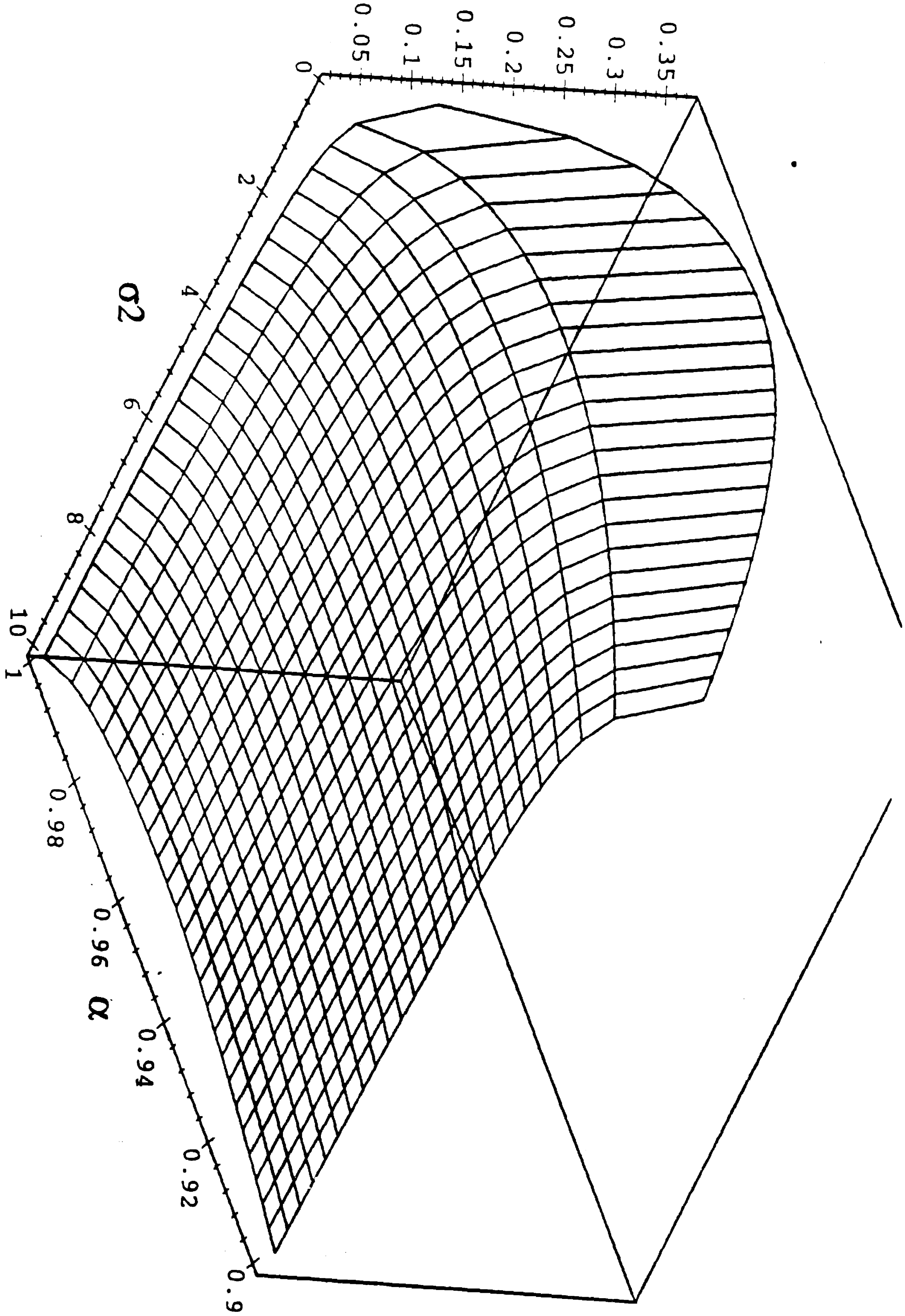


Figure H.2

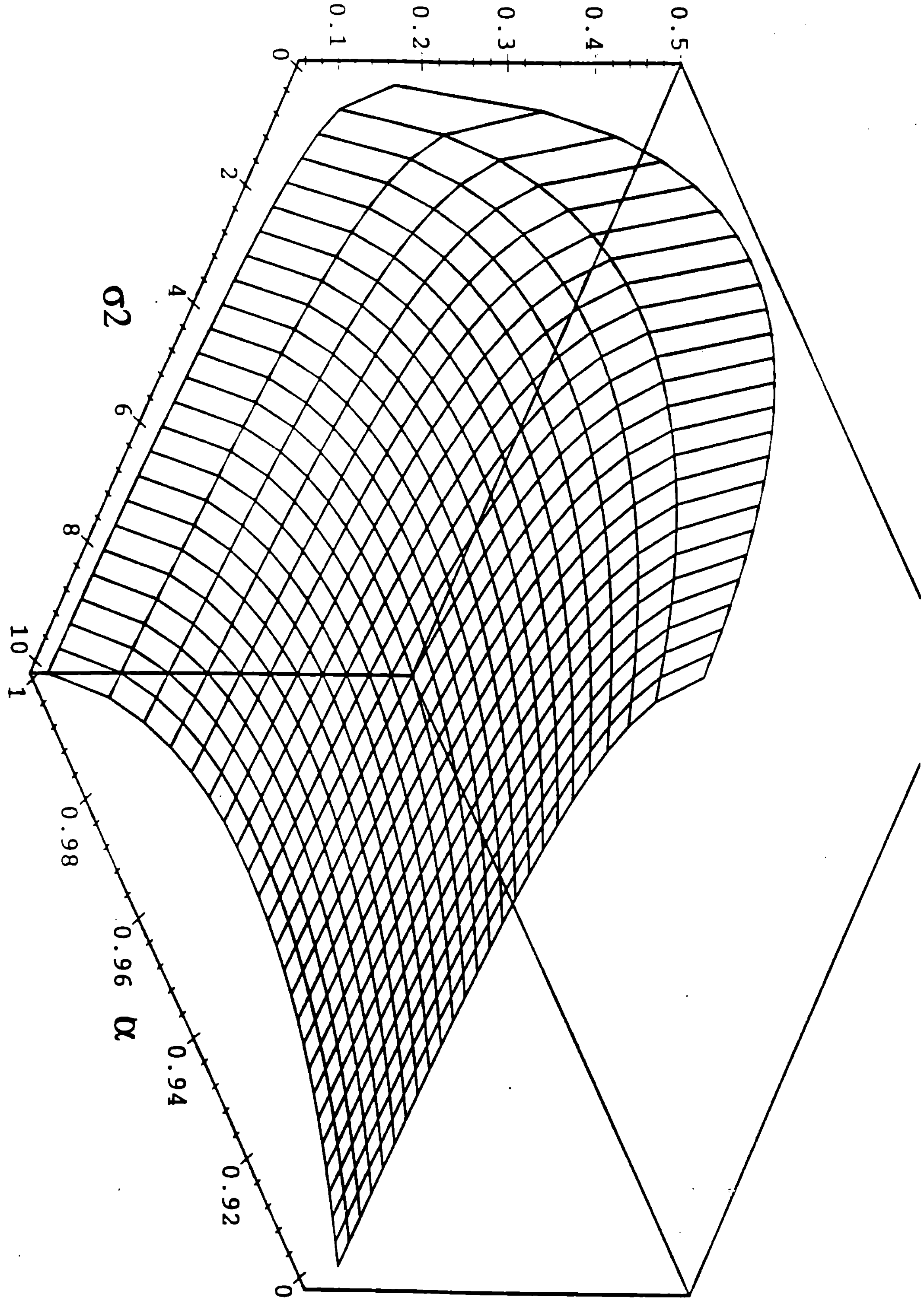
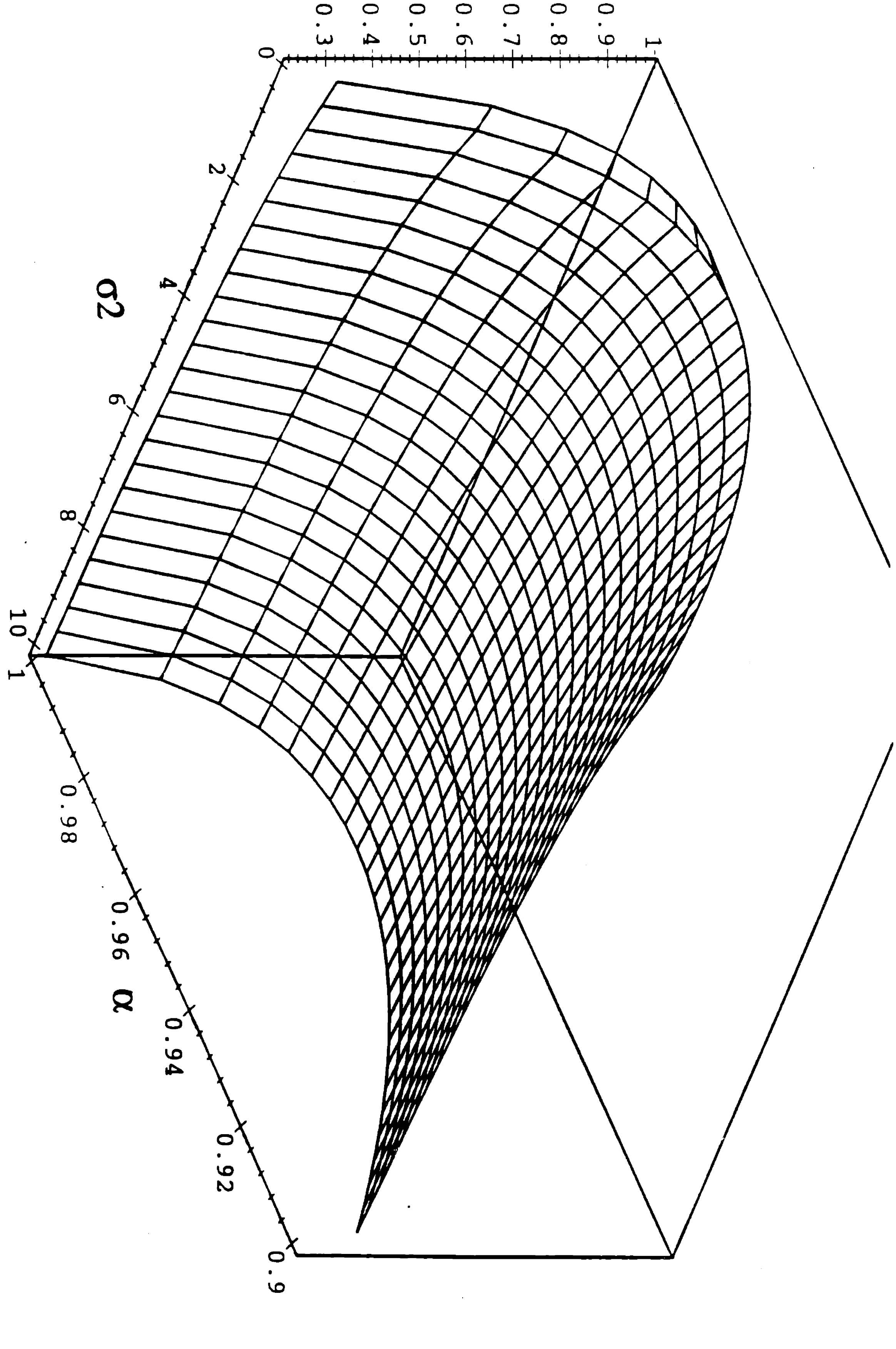


Figure H.3.



Notation.

t, t' : times

g(t', t) : known impulse response of a multipath channel seen at time t' for a transmission sent at t'-t

$g^m(t', t)$: known impulse response of a the m^{th} path of the channel seen at time t' for a transmission sent at t'-t

$a^m(t')$: complex multiplicative factor associated with the m^{th} path

$\tau^m(t')$: delay associated with the m^{th} path

x[n] : discrete input data stream

s(t) : modulated input signal

v(t) : output of the channel with AWGN

f_0 : carrier frequency

$B_{Doppler}$: Doppler spread

T_{spread} : multipath time spread

ϖ : velocity at which sender moves away from the receiver

t_S : time at which a transmission is sent

t_R : time at which a transmission is received

f_S : frequency of sender's transmitted sinusoid

c : speed of light

B^m : Doppler shift for path m

ϖ^m : speed at which the length of path m between a mobile and a base station changes

τ^m : original time shift for the m^{th} path

W_{input} : bandwidth to which the input signal may be bandlimited

W : bandwidth to whih the output signall may be bandlimited

$G^{mFT}_{2}(t', f)$: Fourier transform of g^m with respect to the second variable

$G^{mFT}_{2,1}(f', f)$: Fourier transform of $G^{mFT}_{2}(t', f)$ with respect to t'

f, f' : frequencies

n(t') : AWGN component

$R_N[k]$: autocorrelation function of the filtered discrete time noise n[k]

σ_N^2 : variance of the real or complex component of n[k], equal to $N_0 W/2$

N_0 : spectral density of n(t')

y(t) : output of the channel with AWGN

k, n, i, j : sampled times

Δ : range of non-zero taps

$\underline{v}_{j,k}$: notation for $\left[v[j] \dots v[k] \right]^T$

$\underline{v}_k$: notation for $\left[v[1] \dots v[k] \right]^T$

$\underline{v}_{j,k}^{j',k'}$: matrix with columns j' through k'. rows j through k

$\underline{v}_{k}^{k'}$: matrix with columns 1 through k'. rows 1 through k

$\underline{v}^{k}$: matrix with columns 1 through k. rows 1 through k

$g^m[k, i]$: sampled channel response for path m

$g[k, i]$: sampled channel response

f^k : transfer matrix for the noiseless channel

$\sigma(t, f)$: scattering function

$\underline{S}_k$, $\underline{N}_k$, $\underline{Y}_k$: random vectors corresponding to $\underline{s}_k$, $\underline{n}_k$, $\underline{y}_k$

$\underline{S}'_{2k}$, $\underline{N}'_{2k}$, $\underline{Y}'_{2k}$: randomvectors whose first k and last k components are the real and imaginary parts

P : input power constraint for the real part and for the complex part

T : duration of transmission

$\Lambda_{\underline{v}}$: covaraince matrix of $\underline{v}$

f'^{2k} : transfer matrix for the real representation of the system

S(t), V(t'), N(t') : randomvariables corresponding to s(t), v(t'), n(t')

$\tilde{S}[n]$: sampled version of S(t) sampled at rate W_{input}

$\tilde{g}^m[i, n]$: transfer function of the m^{th} path for an input sampled at rate W_{input} and an output sampled at rate W

$\tilde{\Delta}$: range of non-zero the $\tilde{g}^m[i, n]$ for all m

$\underline{f}_k^{k'}$: transfer matrix constructed from $\sum_m \tilde{g}^m[i, n]$

$\underline{\tilde{f}}_{2k}^{2k'}$: transfer matrix for the real representation of $\underline{f}_k^{k'}$

$\underline{\varphi}_{i2k'}$: eigenvector of $\underline{\tilde{f}}_{2k}^{2k'T} \underline{\tilde{f}}_{2k}^{2k'}$

λ_i : eigenvalue of $\underline{\tilde{f}}_{2k}^{2k'T} \underline{\tilde{f}}_{2k}^{2k'}$

$\underline{\theta}_{i2k}$: output of $\frac{\underline{\varphi}_{i2k'}}{\sqrt{\lambda_i}}$ passed through $\underline{\tilde{f}}_{2k}^{2k'}$

U_i : random variable coefficient of $\underline{\tilde{S}}'_{2k'}$ in terms of $\underline{\varphi}_{i2k'}$

ν_i : random variable coefficient of $\underline{N}'_{2k}$ in terms of $\underline{\theta}_{i2k}$

γ : coefficient for Kuhn-Tucker conditions

$\underline{F}^k$: random variable matrix corresponding to $\underline{f}^k$

$\overline{F}$: known part of a single tap

$\tilde{\overline{F}}$: known part of a single tap

α, $\underline{\alpha}$: multiplicative constant scalar, matrix

$\underline{E}$: error vector forthe LLSE of $\underline{S}$ from $\underline{Y}$

$\underline{\hat{N}}$: zero-mean Gaussian noise with covariance matrix $\Lambda_{\underline{\tilde{F}S}}+\Lambda_{\underline{\Sigma}}$

$\underline{\hat{Y}}$: output to channel $\underline{\overline{F}}$ with AGN $\underline{\hat{N}}$

$\underline{\hat{E}}$:error vector for LLSE of $\underline{S}$ from $\underline{\hat{Y}}$

$\underline{S'}_k$: input to time-invaraint channel, including an initial sounding sequence

$\underline{Y'}_k$: output corresponding to $\underline{S'}_k$

k_0 : number of times a sounding pattern is repeated

k_1 : total duration $k_0\Delta$ of the sounding sequence

$\underline{\hat{F}}^{\Delta,\Delta}[i]$:estimate of the real part of the i^{th} channel tap for a time-invariant channel

$\underline{\hat{F}}^{k+\Delta,k+\Delta}[i]$:estimate of the imaginary part of the i^{th} channel tap

p : period for a channel which is periodic in sampled time

$\kappa_{i,k}$: series given by II-3.[71]

$\kappa_i : \lim_{j \to \infty} \kappa_{i,i+j}$

$\kappa : \lim_{i \to \infty} \kappa_i$

i_0, j_0 : threshold for i, j

ζ_i : series defined by II-3.[78]

$\zeta : \lim_{i \to \infty} \zeta_i$

$\alpha[i, j]$, α : multiplicative constant for Gauss-Markov model

$\underline{\Xi}_i$: innovation vector for Gauss-Markov model

ϕ : level of decorrelation to define coherence time

$\hat{F}[i,i]$ = LSE estimate of $F[i, i]$ given $\underline{S}_i$ and $\underline{Y}_i$

λ_j : variance of Kalman filter

λ'_j : approximation to λ_j

λ' : $\lim_{i \to \infty} \lambda'_i$

λ''_j : variance of Kalman filter with a known final value

λ'''_j : approximation to λ''_j

E : overall input energy

n , m : number of frequency slices over which we spread

$f\left(\frac{E}{n}\right)$: function defined by II-4.[1]

d : value of $\frac{E}{n}$

δ : threshold in $\frac{E}{n}$ for which spreading is desirable for a single time sample

i , i' : mutual information values

C : positive constant

β : proportion of the channel energy which corresponds to the measurement error

$\tilde{f}\left(\frac{E}{n} E\left[F^2\right]\right)$: function given by II-4.[12]

$f_i\left(\frac{E}{n}\right)$: function given by II-4.[15]

$f\left(\frac{E}{n}\right)$: function given by II-4.[17]

δ' : threshold in $\frac{E}{n}$ for which spreading is desirable for several time samples

K : number of users

$B^i_{Doppler}$: Doppler spread associated with the i^{th} channel

$g_i(t, t')$: reponse of the multipath channel seen by user i

P_i : number of paths seen by user i

$a_i^m(t')$: complex factor associated with the m^{th} path of user i

$\tau_i^m(t')$: delay associated with the m^{th} path of user i

$s_i(t)$: complex signal transmitted by user i

$x_i[n]$: data stream for user i

$g_i^m[k, n]$: sampled time response for the m^{th} path of user i

$g_i[k, n]$: sampled time response for user i

R_i : rate for user i

U : subset of users

M : 2Kk' input degrees of freedom

$\tilde{f}^{2k}_{i2k'}$: transfer matrix for the real representation of the channel of user i

$\hat{\underline{f}}_{M}^{2k}$: transfer matrix for the channel which treats all K users as a composite user

$\hat{\lambda}_i$: eigenvalue of $\hat{\underline{f}}_{M}^{2k}\hat{\underline{f}}_{M}^{2k^T}$

$\hat{\underline{\varphi}}_{i2k}$: eigenvector of $\hat{\underline{f}}_{M}^{2k}\hat{\underline{f}}_{M}^{2k^T}$

$\hat{\underline{\theta}}_{iM}$: outputs corresponding to the inputs $\hat{\underline{\varphi}}_{i2k}$ and vectors completing the basis

[illegible] : approximation, given in III-3.[18], to the variance of a Kalman filter for several users

Bibliography.

[AFK72] J.-P. A. Adoul, B.D. Fritchman, L.N. Kanal, *'A Critical Statistic for Channels with Memory'*, IEEE Transactions on Information Theory, vol. IT-18, pp. 133-141, January 1972.

[AM] B.D.O. Anderson, J.B.Moore, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, NJ, 1979

[And73] L.G. Anderson, *'A Simulation Study of Some Dynamic Channel Assignment in a High Capacity Mobile Telecommunications System'*, IEEE Transactions on Communications, vol. COM-21, no. 11, November 1973.

[AN82] H. Alavi, R.W. Nettleton, *'Downstream Power Control for a Spread Spectrum Cellular Mobile Radio System'*, IEEE Global Telecommunications Conference, 1982.

[APO92] B. Aazhang, P. Paris, G.C. Orsak, *'Neural Networks for Multiuser Detection in Code-Division Multiple-Access Communications'*, IEEE Transactions on Communications, vol. 40, no. 7, July 1992.

[Bel63] P. Bello, *'Characterization of Randomly Time-Invariant Channels'*, IEEE Transactions on Communication Systems, December 1963

[Bla62] N.M. Blachman, *'On the Effect of Statistically Dependent Interference upon Channel Capacity'*, IEEE Transactions on Information Theory, September 1962

[BN63] P.A. Bello, B.D. Nelin, *'The Effect of Frequency Selective Fading on the Binary Error probabilities of Incoherent and Differentially Coherent Matched Filter Receivers'*, IEEE Transactions, vol. CS-11, no. 2, June 1963

[Car75] A.B. Carleial, *'A Case Where Interference Does Not Reduce Capacity'*, IEEE Transactions on Information Theory, September 1975.

[Car78] A.B. Carleial, *'Interference Channels'*, IEEE Transactions on Information Theory, Vol. IT-24, no. 1, January 1978.

[CGS80] T.M. Cover, A. El Gamal, M. Salehi, *'Multiple Access Channels with Arbitrarily Correlated Sources'*, IEEE Transactions on Information Theory, vol. IT-26, no. 6, November 1980.

[Cha79] U. Charash, *'Reception Through Nakagami Fading Multipath Channels with Random Delays'*, IEEE Transactions on Communications, Vol. COM-27, No. 4, April 1979.

[Chu89] F.J.P. Chu, *'Improvement of Rake Modem Performance to Reduce Self-Noise in Multipath Channels'*, S.M. Thesis, MIT, February 1989.

[CL75] D.C. Cox, R.P. Leck, *'Correlation Bandwidth and Delay Spread Multipath Propagation Statistics for 910 MHz Urban Mobile Radio Channnels'*, IEEE Transactions, vol. COM-23, no. 11, November 1975

[CMP81] T.M. Cover, R.J. McEliece, E.C. Posner, *'Asynchronous Multiple-Access Channel Capacity'*, IEEE Transactions on Information Theory, vol. IT-27, no. 4, July 1981.

[CNW94] M.C. Chuah, S. Nanda, W.S. Wong, *'Squeezing Out Capacity with Joint Power-Control and Channel Assignment'*, in Wireless and Mobile Communications, edited by J.M. Holtzman, D.J. Goodman, Kluwer Academic Publishers, Boston, 1994.

[CT] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, John Wiley and Sons, New York, 1991.

[CV93] R.S. Cheng, S. Verdú, *'Gaussian Multiaccess Channels with Capacity Region and Multiuser Water-Filling'*, IEEE Transactions on Information Theory, vol. 39, no. 3, May 1993.

[DKR93] M.Duque-Antón, D. Kunz, B. Rüber, *'Channel Assignment for Cellular Radio Using Simulated Annealing'*, IEEE Transactions on Vehicular Technology, vol. 42, no. 1, February 1993.

[Dob58] R.L. Dobrushin, *'Information Transmission in a Channel with Feedback'*, Theory of Probability and its Applications, Volume III, Number 4, 1958.

[Due92,1] A. Duel-Hallen, *'On Suboptimal Detection for Asynchronous Code-Division Multiple-Access Channels'*, AT&T Bell Laboratories, 1992.

[Due92,2] A. Duel-Hallen, *'Decorrelating Decision-Feedback Multiuser Detector for Synchronous Code-Division Multiple-Access Channel'*, AT&T Bell Laboratories, 1992.

[DV94] D.D. Dimitrijevic, J.F. Vucetic, *'Design and Performance Analysis of Algorithms for Channel Allocation in Cellular Networks'*, in Wireless and Mobile Communications, edited by J.M. Holtzman, D.J. Goodman, Kluwer Academic Publishers, Boston, 1994.

[Ekh86] B. Ekhlund, *'Channel Utilization and Blocking Probability in a Cellular Mobile Telephone System with Directed Retry'*, IEEE Transactions on Communications, vol. COM-34, no. 4, April 1986.

[For72] G.D. Forney, *'Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference'*, IEEE Transactions on Information Theory, vol. IT-18, no. 3, May 1972.

[Fri67] B.D. Fritchman, *'A Binary Channel Characterization Using Partitioned Markov Chains'*, IEEE Transactions on Information Theory, vol. IT-13, pp. 221-227, April 1967.

[FS83] G.J. Foschini, J. Salz, *'Digitsal Communications over Fading Radio Channels'*, The Bell System Technical Journal, vol. 62, no. 2, Part 2, February 1983

[Gal64] R.G. Gallager, *Characterization and Measurement of Time- and Frequency-spread Channels*, Technical Report 352, 30 April 1964, MIT Lincoln Laboratories, Lexington, MA.

[Gal85] R.G. Gallager, '*A Perspective on Multiaccess Channels*', IEEE Transactions on Information Theory, vol. IT-31, no. 2, March 1985.

[Gal94] R.G. Gallager, '*An Inequality on the Capacity Region of Multiaccess Multipath Channels*', MIT, 1994.

[Gal] R.G. Gallager '*Information Theory and Reliable Communication*', John Wiley & Sons, 1968.

[GB81] A.J. Giger, W.T. Barnett, '*Effects of Multipath Propagation on Digital Radio*', IEEE Transactions on Communications, Vol. COM-29, September 1981.

[GC80] A. El Gamal, T.M. Cover, '*Multiple User Information Theory*', Proceedings of the IEEE, Vol. 68, no. 12, December 1980.

[Ger90] E. Gerianotis, '*Multiple-Access Capability of Frequency-Hopped Spread-Spectrum Revisited : An Analysis of the Effect of Unequal Power Levels*', IEEE Transactions on Communications, vol. 38, no. 7, July 1990.

[GG92] A. Ganz, Y. Gao, '*Algorithms for SS/TDMA Scheduling*', IEEE Transactions on Communications, vol. 40, no. 8, August 1992.

[GGV93] D.J. Goodman, S.A. Grandhi, R. Viyajan, '*Distributed Dynamic Channel Assignment Schemes*', 43rd IEEE Vehicular Technology Conference, Secaucus, New Jersey, May 1993.

[GH92] D. Greenwood and L. Hanzo, '*Characterization of Mobile Radio Channels*', in *Mobile Radio Communications*, edited by R. Steele, IEEE Press, Piscataway, NJ, 1992.

[Gil60] E.N. Gilbert, '*Capacity of a Burst-Noise Channel*', Bell System Technical Journal, vol. 39, pp. 123-1265, September 1960.

[Gol94] A.J. Goldsmith, '*Design and Performance of High-Speed Communication Systems over Time-Varying Radio Channels*', Ph. D. Dissertation, U.C. Berkeley, 1994.

[Goo91] D.J. Goodman, '*Trends in Cellular and Cordless Communications*', IEEE Communications Magazine, June 1991.

[GP82] E.A. Gerianotis, M.B. Pursley, '*Error Probability for Direct-Sequence Spread-Spectrum Multiple-Access Communications - Part II : Approximations*' , IEEE Transactions on Communications, vol. COM-30, no. 5, May 1982.

[GP82,1] E.A. Gerianotis, M.B. Pursley, '*Error Probabilities for Slow-Frequency-Hopped Spread-Spectrum Multiple-Access Communications Over Fading Channels*', IEEE Transactions on Communications, vol. COM-30, no. 5, May 1982.

[Gra92] S.D. Gray '*Multipath Reduction Using Constant Modulus Gradient Techniques*', IEEE Journal on Selected Areas in Communications, vol. 10, no. 8, October 1992.

[GS76] T. Gonzalez, S. Sahni, *'Open Shop Scheduling to Minimize Finish Time'*, Journal of the Association for Computing Machinery, vol. 23, no. 4, October 1976.

[Hal80] W.K.Hale, *'Frequency Assignment: Theory and Applications'*, Proceedings of the IEEE, vol. 68, no. 12, December 1980.

[Han93] S. H. Hanly, *'Information Capacity of Radio Networks'*, Dissertation Submitted for the Degree of Doctor of Philosophy of the University of Cambridge, August 1993.

[HH85] J.Y.N. Hui, P.A. Humblet, *'The Capacity Region of the Totally Asynchronous Multiple-Access Channel'*, IEEE Transactions on Information Theory, vol. IT-31, no. 2, March 1985.

[HS90] M.V. Hegde, W.E. Stark, *'Capacity of Frequency-Hop Spread-Spectrum Multiple-Access Communication Systems'*, IEEE Transactions on Communications, vol. 38, no. 7, July 1990.

[Jak] W.C. Jakes, *Microwave Mobile Communications*, IEEE Press, Piscataway, NJ, 1993

[Kai59], T. Kailath, *'Sampling Models for Linear Time-variant Filters'*, Technical Report 352, MIT RLE, Cambridge, MA, May 25, 1959

[Kai61] T. Kailath, *'Optimum Receivers for Randomly Varying Channels'*, 'Information Theory', Fourth London Symposium, Butterworths, London, 1961.

[KE89] J. Karlsson, B. Ekhlund, *'A Cellular Mobile Telephone System with Load Sharing - An Enhancement of Directed Retry'*, IEEE Transactions on Communications, vol. 37, no. 5, May 1989.

[Ken] R.S. Kennedy, *Fading Dispersive Communications Channels*, Wiley-Interscience, 1969.

[Khi56] A. Ya. Khinchin, *'Fundamental Theorems of Information Theory'*, Transactions on Mathematical Sciences, vol. XI, no. 1, January-February 1956

[KP93] Z. Kosti´c, G. Pavlovi´c, *'Resolving Sub-chip Spaced Multipath Components in CDMA Communications Systems'* , 43rd IEEE Vehicular Technology Conference, 1993.

[Kun88] D. Kunz, *'Channel Assignment for Cellular Radio Using Neural Networks'* , IEEE Transactions on Vehicular Technology, vol. 40, no. 1, February 1988.

[KW72] T.T. Kadota, A.D. Wyner, *'Coding Theorem for Stationary, Asymptotically Memoryless, Continuous-Time Channels'*, The Annals of Mathematical Statistics, Vol. 43, No. 5, 1972.

[Lee] W.C.Y. Lee, *'Mobile Communications Engineering'*, New York, McGraw -Hill Book Company.

[Lee86] W.C.Y. Lee, *'Elements of Cellular Mobile Radio Systems'*, IEEE Transactions on Vehicular Technology, vol. VT-35, no. 2, May 1986.

[Lee87] W.C.Y Lee *'New Cellular Schemes for Spectral Efficiency'*, IEEE Transactions on Vehicular Technology, vol. VT-36, no. 4, November 1987.

[Lee89] W.C.Y. Lee, *'Spectrum Efficiency in Cellular'*, IEEE Transactions on Vehicular Technology, vol. 38, no. 2, May 1989.

[Lee91] W.C.Y. Lee, *'Overview of Cellular CDMA'*, IEEE Transactions on Vehicular Technology, vol. 40, no. 2, May 1991.

[Lia72] H.H.-J. Liao, *'Multiple-Access Channels'*, Ph. D. Thesis, University of Hawaii, September 1972.

[LM] E.A. Lee, D.G. Messerschmitt, *Digital Communication*, Kluwer Academic Publishers, Boston, 1992

[LP87] J.S. Lehnert, M.B. Pursley, *'Multipath Diversity Reception of Spread-Spectrum Multiple-Access Communications'*, IEEE Transactions on Communications, Vol. COM-35, No. 11, November 1987

[LV89] R. Lupas, S. Verdú, *'Linear Multiuser Detectors for Synchronous Code-Division Multiple-Access Channels'*, IEEE Transactions on Information Theory, vol. 35, no. 1, January 1989.

[LV90] R. Lupas, S. Verdú, *'Near-Far Resistance of Multiuser Detectors in Asynchronous Channels'*, IEEE Transactions on Communications, vol. 38, no. 4, April 1990.

[McGar77] T. P. McGarty, *'Adaptive Decoding for a Specular Multipath Channel'*, NTC'77 Conference Record, vol. 2, IEEE, 1977.

[McMan70] P. McManamon, *'HF Markov Chain Models and Measured Error Averages'*, IEEE Transactions on Communications Technology, vol. COM-18, pp. 201-208, June 1970.

[Meu77] E.C. van der Meulen, *'A Survey of Multi-Way Channels in Information Theory: 1961-1976'*, IEEE Transactions on Information Theory, vol. IT-23, no. 1, January 1977.

[MG92] M.A. Mokhter, S.C. Gupta, *'Power Control Considerations for DS/CDMA Personal Communication Systems'*, IEEE Transactions on Vehicular Technology, November 1992.

[MP90] U. Madhow, M.B. Pursley, *'Limiting Performance of Frequency -Hop Random Access'*, IEEE Transactions on Information Theory, vol. 36, no. 2, March 1990.

[MS83] R.J. McEliece, W.E. Stark, *'Channels with Block Interference'*, IEEE Transactions on Information Theory, vol. IT-30, no. 1, January 1984.

[MS94], R.J. McEliece, K.N. Sivarajan, *'Performance Limits for Channelized Cellular Telephone Systems'*, IEEE Transactions on Information Theory, vol.40, no. 1, January 1994.

[NA83] R.W. Nettleton, H. Alavi, *'Power Control for a Spread Spectrum Cellular Mobile Radio System'*, 33rd IEEE Vehicular Technology Conference, May 1983.

[Nak58] M. Nakagami, *'The m-distribution - A General Formula of Intensity Distribution of Rapid Fading'*, in *Statistical Methods of Radio Wave Propagation*, Proceedings of a Symposium held at UCLA, June 18-20, 1958, Pergamon Press, 1960.

[Net80] R.W. Nettleton, *'Traffic Theory and Interference Management for a Spread Spectrum Cellular Mobile Radio System'*, ICC '80 Conference Record, vol.1.

[Par81] J.H. Park, *'Two-User Dispersive Communication Channels'*, IEEE Transactions on Information Theory, Vol. IT-27, No. 4, July 1981.

[Pfa71] E. Pfaffelhuber, *'Channels with Asymptotically Decreasing Memory and Anticipation'*, IEEE Transactions on Information Theory, Vol. IT-17, No. 4, July 1971.

[PG58] R. Price, P.E. Green, *'Communication Technique for Multipath Channels'*, Proceedings of the IRE, Vol. 46 March 1958.

[Pin] M. S. Pinsker, *'Information and Information Stanility of Random Variables and Processes'*, translated by A. Feinstein, Holden-Day, San Francisco, 1964.

[PMS91] R.L. Pickholts, L.B. Milstein, D.L. Schilling, *'Spread Spectrum for Communications'*, IEEE Transactions on Vehicular Technology, vol. 40. no. 2, May 1991.

[Pri56] R. Price, *'Optimum Detection of Signals in Noise, with Application to Scatter-Multipath Communication, I'*, IRE Transactions on Information Theory, Vol. IT-2, 1956.

[Pro] J.G. Proakis, *Digital Communications*, McGraw-Hill Book Co., New York, 1983

[PSM82] R.L. Pickholtz, D.L. Schilling, L.B. Milstein, *'Theory of Spread-Spectrum Communications - A Tutorial'*, IEEE Transactions on Communications, vol. COM-30, no. 5, May 1982.

[PSS82] M.B. Pursley, D.V. Sarwate, W.E. Stark, *'Error Probability for Direct-Sequence Spread-Spectrum Multiple-Access Communications - Part I: Upper and Lower Bounds'*, IEEE Transactions on Communications, vol. COM-30, no. 5, May 1982.

[PV88] H.V. Poor, S. Verdú, *'Single-User Detectors for Multiuser Channels'*, IEEE Transactions on Communications, vol. 36, no. 1, January 1988.

[Qua91] Qualcomm, *'An Overview of the Application of Code Division Multiple Access to Digital Cellular Systems and Personal Communications Networks'*, Qualcomm, Inc., January 1991

[Qur85] S.U.H. Qureshi, *'Adaptive Equalization'*, Proceedings of the IEEE, vol. 73, no. 9, September 1985

[Ric67] J. S. Richters, *'Communication over Fading Dispersive Channels'*, Ph. D. Thesis. Massachusetts Institute of Technology, September 1967.

[RTM94] M. Rupf, F. Tarköy, J.L. Massey, *'User-Separating Demodulation for Code-Division Multiple-Access Systems'*, IEEE Transactions on Selected Areas in Communications. Vol. 12, No. 5, June 1994.

[Sat77] H. Sato, *'Two-User Communication Channels'*, IEEE Transactions on Information Theory, vol. IT-23, no. 3, May 1977.

[Sch94] R. Schneiderman, *Wireless Persnal Communications*, IEEE Press, New York, 1994.

[SG93], M. Serizawa, D.J. Goodman, *'Instability and Deadlock of Distributed Dynamic Channel Allocation'*, 43rd IEEE Vehicular Technology Conference, Secaucus, New Jersey, May 1993.

[Spi65] J.J. Spilker, Jr. ,*'On the Characterization and Measurement of Randomly Time Varying Filters'*, IEEE Transactions on Circuit Theory, May 1965.

[Ste87] S. Stein, *'Fading Channels in System Engineering'*. IEEE Journal on Selected Areas in Communications, vol. SAC-5, N. 2, February 1987.

[Str] G. Strang, *Linear Algebra and its Applications*, 2nd Edition, Academic Press, New York, 1980

[SW73] D. Slepian, J.K. Wolf, *'A Coding Theorem for Multiple Access Channels With Correlated Sources'*, The Bell System Technical Journal, vol. 52, no. 7, September 1973.

[Tel88] I.E. Telatar, *'Coding and Multiaccess for the Energy Limited Rayleigh Fading Channel'*, M.S. Thesis, MIT, May 1988.

[Tho87] J.A. Thomas, *'Feedback Can at Most Double Gaussian Multiple Access Channel Capacity'*, IEEE Transactions on Information Theory, vol. IT-33, no. 5, September 1987.

[Tin62] H.K. Ting, *'On the Information Stability of a Sequence of Channels'*, Theory of Probability and its Applications, Volume VII, Number 3, 1962.

[Tsa58] I.P Tsaregradskii, *'A Note on the Capacity of a Stationary Channel with Finite Memory'*, Theory of Probability and its Applications, Volume III, Number 1, 1958.

[Tur80] G.L. Turin, *'Introduction to Spread-Spectrum Antimultipath Techniques and Their Application to Urban Digital Radio'*, Proceedings of the IEEE, vol. 68, no. 3, March 1980.

[Tur84] G.L. Turin, *'The Effects of Multipath and Fading on the Performance of Direct-Sequence CDMA Systems'*, IEEE Journal on Selected Areas in Communications, vol. SAC-2, no. 4, July 1984.

[VA90] M.K. Varanasi, B. Aazhang, '*Multistage Detection in Asynchronous Code-Division Multiple-Access Communications*', IEEE Transactions on Communications, vol. 38, no. 4, April 1990.

[Ver89,1] S. Verdú, '*Multiple-Access Channels with Memory with and without Frame Synchronism*', IEEE Transactions on Information Theory, vol. 35, no. 3, May 1989.

[Ver89,2] S. Verdú, '*The Capacity Region of the Symbol-Asynchronous Gaussian Multiple-Access Channel*', IEEE Transactions on Information Theory, vol. 35, no. 4, July 1989.

[Ver86,1] S. Verdú, '*Minimum Probability of Error for Asynchronous Gaussian Multiple-Access Channels*', IEEE Transactions on Information Theory, vol. IT-32, no. 1, January 1986.

[Ver86,2] S. Verdú, '*Optimum Multiuser Asymptotic Efficiency*', IEEE Transactions on Communications, vol. COM-34, no. 9, September 1986.

[VH94] AS, Verdú, T.S. Han, '*A General Formula for Channel Capacity*', IEEE Transactions on Information Theory, vol. 40, no. 4, July 1994

[VMS 84] D. Verhulst, M. Mouly, J. Szpirglas, '*Slow Frequency Hopping Multiple Access for Digital Cellular Radiotelephone*', IEEE Journal on Selected Areas in Communications, vol. SAC-2, no. 4, July 1984.

[WHB80] C.L. Weber, G.K. Huth, B.H. Batson, '*Performance of Code Division Multiple Access Systems*', ICC '80 Conference Record, vol.1.

[Wyn94] A.D. Wyner, '*Shannon-Theoretic Approach to a Gaussian Cellular Multiple Access Channel*', IEEE Transactions on Information Theory, vol. 40, no. 6, November 1994.

[Zan92] J. Zander, '*Performance of Optimum Transmitter Power Control in Cellular Radio Systems*', IEEE Transactions on Vehicular Technology, February 1992.

[ZY89] Zhang, T.K.P. Yum, '*Comparison of Channel-Assignment Strategies in Cellular Mobile Telephone Systems*', IEEE Transactions on Vehicular Technology, vol. 39, no. 4, November 1989.