

Copyright © 1999, by the author(s).
All rights reserved.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission.

**CAPACITY AND MUTUAL INFORMATION OF
WIDEBAND MULTIPATH FADING CHANNELS**

by

Emre Telatar and David Tse

Memorandum No. UCB/ERL M99/6

15 January 1999

**CAPACITY AND MUTUAL INFORMATION OF
WIDEBAND MULTIPATH FADING CHANNELS**

by

Emre Telatar and David Tse

Memorandum No. UCB/ERL M99/6

15 January 1999

ELECTRONICS RESEARCH LABORATORY

College of Engineering
University of California, Berkeley
94720

Capacity and Mutual Information of Wideband Multipath Fading Channels

Emre Telatar* David Tse†

Abstract

In this paper we will investigate the capacity and mutual information of a broadband fading channel consisting of a finite number of time-varying paths. We will show that the capacity of the channel in the wideband limit is the same as that of a wideband Gaussian channel with the same average received power. However, the input signals needed to achieve the capacity must be “peaky” in time or frequency. In particular, we show that if white-like signals are used instead (as is common in spread-spectrum systems), the mutual information is inversely proportional to the number of *resolvable paths* \tilde{L} with energy spread out, and in fact approaches 0 as the number of paths get large. This is true even when the paths are assumed to be tracked perfectly at the receiver. A critical parameter \tilde{L}_{crit} is defined in terms of system parameters to delineate the threshold on \tilde{L} over which such over-spreading phenomenon occurs.

1 Introduction

Wireless communication takes place over multipath fading channels. Typically the transmitted signal travels to the receiver along a multitude of paths, the delays and gains of which vary with time. One design approach to communication systems for such channels is to separate the channel measurement and data transmission problems: one assumes that the receiver can perfectly track the time varying channel characteristics, and decodes the transmitted signal using this knowledge; one then updates the channel estimate from the knowledge of the transmitted and received signal pair. When the channel is known to the receiver and the noise is additive Gaussian the best input

*Rm. 2C-174, Bell Labs, Lucent Technologies, 600 Mountain Avenue, Murray Hill, NJ 07974, telatar@bell-labs.com.

†Rm. 261M, Cory Hall, Dept. of EECS, U.C. Berkeley, Berkeley, CA 94720, dtse@eeecs.berkeley.edu. This author is partially supported by AFOSR under grant F49620-96-1-0199 and by a NSF CAREER Award under grant NCR-9734090. Part of the work was done when he was visiting Bell Labs.

signals to use are those that look like samples of Gaussian noise. However, it is not clear if the channel characteristics can be reliably estimated when such input signals are used. This issue is particularly pertinent when the signals are spread over a very large bandwidth, as in the proposed third-generation wideband CDMA systems.

To answer these questions, we study in this paper the capacity and mutual information of multipath fading channels *without* the *a priori* assumption of knowledge of the channel at the receiver. We consider a channel having a finite number L of paths and a large transmission bandwidth W . To state the results we introduce the notion of the number of *resolvable paths* \tilde{L} : This is the number of paths one would see if one could only differentiate paths whose delays differ by more than $1/W$. Three results are presented in this paper:

1. With no restriction on the input signal other than an average power constraint, as the bandwidth gets large, one can achieve communication rates over a multipath fading channel equal to the capacity of an infinite bandwidth additive white Gaussian channel of the same SNR without fading. Moreover, this can be achieved by frequency shift keying and non-coherent detection.

2. In contrast, if one uses signals the energy of which are spread evenly over time and frequency, then the mutual information decreases in inverse proportion to the number of resolvable paths \tilde{L} , assuming that the energy is divided more or less equally among all paths and that the path gains are independent. Thus, as the number of paths gets large, the mutual information approaches zero. This result holds even when the receiver can track perfectly the timing of each path and the only uncertainty is in the phases and amplitudes. Observe that the bandwidth does not directly influence the mutual information, but if the underlying number of paths L is very large and the delays of these paths spread out, then \tilde{L} will increase with increasing bandwidth.

3. Without side information about the timing of the paths, if one uses signals that are spread evenly over time and frequency, the mutual information approaches zero with increasing bandwidth even when there is a single fixed gain path with random time varying delay.

The study of the wideband fading channel dates back to the early 60's. Kennedy has shown that the capacity of an infinite bandwidth Rayleigh fading channel is the same as that of an infinite bandwidth AWGN channel with the same average received power (see [1, §8.6], [2]). Our first theorem is a parallel result, applicable to any channel with a finite number of paths.

More recently, Gallager and Medard [3] showed that if the channel is such that the fading processes at different frequencies are independent, then the mutual information achievable over this channel approaches zero with increasing bandwidth if white-like

input signals are used. The assumption of the independence of fades at different frequencies is roughly equivalent to assuming an infinite number of paths. It is not clear *a priori* whether a similar result holds if the number of paths is finite. This issue is important in a wideband system, because any finite set of paths will eventually be resolvable as the bandwidth gets sufficiently large. This in part motivates us to deal directly with a model with finite number of paths.

The above results show that the answer to this question is somewhat subtle. Suppose there are a few dominant paths. If we assume that the receiver has side information on the timing but not the phases and magnitudes of the paths, then the limitation to mutual information comes from the number of *resolvable paths* \tilde{L} rather than the channel bandwidth W . Otherwise, if we assume that no side information is available at receiver about path delays, the limitation comes from the necessity to estimate these delays more and more accurately as bandwidth gets large to be able to decode a white-like transmitted signal. This results in the decay of mutual information to zero with bandwidth. In typical wireless settings, the path delays vary at a much slower time-scale than the path gains (phase and amplitude), so to the first approximation, the first of the scenarios described will hold. The effects predicted for the second scenario (that the mutual information goes to zero with increasing bandwidth even when a only a finite number of paths is present) takes place only at very large bandwidths.

In addition to the above qualitative conclusions, we have also computed explicit upper and lower bounds to the mutual information as a function of key channel parameters. These bounds lead us to define a *critical parameter*:

$$\tilde{L}_{\text{crit}} := \frac{PT_c}{N_0},$$

where P is the average received power constraint, $N_0/2$ is the power spectral density of the additive Gaussian noise, and T_c is the coherence time of the channel. The parameter \tilde{L}_{crit} delineates the regime in which *over-spreading* occurs. If the number of resolvable paths \tilde{L} is much smaller than \tilde{L}_{crit} , then the mutual information achieved by spread-spectrum signal is close to the capacity of the non-fading white Gaussian noise channel. On the other hand, if \tilde{L} is much larger than \tilde{L}_{crit} , the mutual information achieved is negligibly small.

The rest of the paper is organized as follows. In Section 2, we present the fading channel model. Section 3 focuses on how to achieve the capacity of the channel with only average power constraint. In Section 4, we study the mutual information achieved by wideband signals, and derive upper and lower bounds as a function of the number of resolvable paths \tilde{L} and other channel parameters. In Section 5, we turn to the problem of detection of binary orthogonal broadband signals with multipath diversity reception, when the path gains are unknown or imperfectly estimated. We observe the performance deterioration as the number of multipaths grow, in a manner akin to the scaling of mutual information. This provides a more intuitive understanding

of the information theoretic results and an interpretation of the critical parameter \tilde{L}_{crit} in particular. In Section 6, the scenario of single path with time-varying delay is considered. Section 7 contains our conclusions.

A word about notation: unless otherwise stated, the information rates in this paper are in the units of nats per second.

2 Channel Model

We consider a general multi-path fading channel: when the channel input waveform is $x(t)$, the channel output $y(t)$ is given by

$$y(t) = \sum_{\ell=1}^L a_{\ell}(t)x(t - \tau_{\ell}(t)) + z(t), \quad (1)$$

where L is the number of paths, $a_{\ell}(t)$ is the gain of path ℓ at time t , $\tau_{\ell}(t)$ is the delay of the path ℓ at time t , and $z(t)$ is white Gaussian noise with power spectral density $N_0/2$. We will assume that $a_{\ell}(t)$ and $\tau_{\ell}(t)$ are stationary and ergodic stochastic processes, and independent of each other.

We begin by identify a number of key parameters of this channel. The *delay spread* T_d quantifies the uncertainty in the delay of the paths; it satisfies

$$\sup_{\ell, t} \tau_{\ell}(t) - \inf_{\ell, t} \tau_{\ell}(t) \leq T_d. \quad (2)$$

The *coherence time* T_c is the duration of time over which the passband channel remains essentially time invariant; it satisfies

$$\sup_{\substack{\ell, s, t: \\ |s-t| \leq T_c}} f_c [\tau_{\ell}(t) - \tau_{\ell}(s)] \ll 1, \quad (3)$$

where f_c is the carrier frequency of the communication system. We also require that the power spectrum of $a_{\ell}(t)$ is contained within $[-1/T_c, 1/T_c]$. We will assume that the delay spread is much less than the coherence time of the channel. This is the case for most wireless channels where typical delay spreads run in the microseconds whereas typical coherence times are measured in milliseconds [5]. The average transmitter power is constrained to P , and the bandwidth of the input signals is constrained to be W around the carrier frequency.

The channels we are interested in are “narrowband” in the sense that the bandwidth is much smaller than the carrier frequency, but “broadband” in the sense that power per degree of freedom is very small, i.e., we are power limited as opposed to bandwidth limited. That the bandwidth is small compared to the carrier frequency is the reason why we can define the coherence time only with respect to the carrier frequency f_c in (3).

3 Capacity via Frequency-Shift Keying

This section is devoted to proving the following theorem.

Theorem 1. *Under an average power constraint, the capacity of an infinite bandwidth multipath fading channel is the same as that of an infinite bandwidth additive white noise channel with the same received power, and this capacity can be achieved by frequency shift keying modulation.*

Suppose we wish to transmit one of M messages. Let T_s be chosen such that $T_d \ll T_s \ll T_c$. During this interval T_s , the channel can be thought of as a linear time invariant channel at the frequencies of interest. To each message we assign a signal

$$x_m(t) = \begin{cases} \sqrt{\lambda} \exp(j2\pi f_m t) & 0 \leq t \leq T_s \\ 0 & \text{else.} \end{cases}$$

That is, each message is a sinusoid at frequency f_m with amplitude $\sqrt{\lambda}$. We will choose f_m to be an integer multiple of $1/(T_s - 2T_d)$. When x_m is transmitted, the received signal y is given by

$$\sum_{\ell=1}^L a_\ell(t) x_m(t - \tau_\ell(t)) + z(t).$$

Over the interval $[T_d, T_s - T_d]$, we can assume that $a_\ell(t)$ and $f_m \tau_\ell(t)$ are essentially unchanged due to (3) and that $T_s \ll T_c$, and we can write the received signal as

$$\begin{aligned} y(t) &= \sum_{\ell=1}^L a_\ell \sqrt{\lambda} \exp(j2\pi f_m(t - \tau_\ell)) + z(t) \\ &= G \sqrt{\lambda} \exp(j2\pi f_m t) + z(t) \end{aligned}$$

where $G = \sum_{\ell=1}^L a_\ell \exp(j2\pi f_m \tau_\ell)$ is the complex phasor representing the amplitude gain and phase shift during the interval $[T_d, T_s - T_d]$. Without loss of generality, we will assume that $E(|G|^2) = 1$.

At the receiver, the received signal is correlated against all the possible transmitted signals x_l . Namely, the receiver forms:

$$R_l = \frac{1}{\sqrt{N_0(T_s - 2T_d)}} \int_{T_d}^{T_s - T_d} \exp(-j2\pi f_l) y(t) dt$$

for $1 \leq l \leq M$. Note that for $l = m$,

$$R_m = \sqrt{\lambda(T_s - 2T_d)/N_0} G + W_m$$

where W_m is a circularly symmetric complex Gaussian random variable with variance 1. For $l \neq m$, since $(f_l - f_m)$ is an integer multiple of $1/(T_s - 2T_d)$, x_m and x_l are orthogonal on this interval, and the signal component at the output of the correlator vanishes and we are left with

$$R_l = W_l$$

where W_l is again a circularly symmetric complex Gaussian random variable with variance 1. Note that because of the orthogonality of the x_l 's $\{W_l\}$ form a set of independent random variables.

To transmit message m , we will repeat the transmission x_m on N disjoint time intervals to average over the fading of the channel. The receiver will form the correlations $R_{l,n}$ for each possible message $1 \leq l \leq M$ and each interval $1 \leq n \leq N$,

$$R_{l,n} = \delta_{lm} \sqrt{\lambda(T_s - 2T_d)/N_0} G(n) + W_{l,n},$$

where $G(n)$ is the complex gain for time interval n , and $W_{l,n}$ are i.i.d. circularly symmetric complex Gaussian random variables with variance 1. The decoder will form the decision variables

$$S_l = \frac{1}{N} \sum_{n=1}^N |R_{l,n}|^2$$

and use a threshold rule to decide on a message: if exactly one of S_l 's say $S_{\hat{l}}$, exceeds $A = 1 + (1 - \epsilon)\lambda(T_s - 2T_d)/N_0$ then it will declare that \hat{l} was transmitted. Otherwise, it will declare a decoding error. We will fix $\epsilon \in (0, 1)$ and later take it to be arbitrarily small. Observe that this is a non-coherent scheme as we do not need to measure the phase nor the amplitude of the channel gain.

The decision variable for the transmitted message S_m is given by

$$\frac{1}{N} \sum_{n=1}^N |G(n)\lambda(T_s - 2T_d)/N_0 + W_{m,n}|^2.$$

By the ergodicity of the fading process, this time average will exceed the threshold with probability arbitrarily close to 1 for any $\epsilon > 0$ as N gets large.

For any message $l \neq m$, its decision variable is given by

$$\frac{1}{N} \sum_{n=1}^N |W_{l,n}|^2.$$

Note that $|W_{l,n}|^2$ are independent exponentially distributed random variables with mean 1, and we will bound the probability

$$\Pr[S_l \geq A]$$

using a Chernoff bound:

$$\Pr[S_l \leq A] \leq \exp(-NE(A))$$

where

$$\begin{aligned} E(A) &= \sup_r [rA - \log(E[\exp(r|W_{1,1}|^2)])] \\ &= \sup_r [rA + \log(1-r)] \\ &= A - 1 - \log(A). \end{aligned}$$

Using the union bound we see that the probability that one of the decision variables S_l , $l \neq m$, exceeds A is upper bounded by

$$\exp\left(-N\left[E(A) - \frac{1}{N} \log(M)\right]\right)$$

This probability decays to zero exponentially in N as long as

$$\frac{1}{N} \log M < A - 1 - \log(A).$$

Substituting the value for A we can rewrite our condition as

$$R(\lambda) = \frac{1}{NT_s} \log M \leq (1 - \epsilon) \left(1 - 2\frac{T_d}{T_s}\right) \frac{\lambda}{N_0} - \frac{1}{T_s} \log \left[1 + (1 - \epsilon) \left(\frac{\lambda(T_s - 2T_d)}{N_0}\right)\right]$$

We now introduce another parameter θ , which represents the fraction of time we transmit information. During this time, we use the scheme described above with $\lambda = P/\theta$, and the rest of the time the transmitter transmits nothing. This will maintain the average power to be P . The average rate that we achieve is given by:

$$\theta R(P/\theta) = (1 - \epsilon) \left(1 - 2\frac{T_d}{T_s}\right) \frac{P}{N_0} - \frac{\theta}{T_s} \log \left[1 + (1 - \epsilon) \left(\frac{P(T_s - 2T_d)}{\theta N_0}\right)\right].$$

As θ approaches 0, this expression approaches

$$(1 - \epsilon) \left(1 - 2\frac{T_d}{T_s}\right) \frac{P}{N_0}$$

which differs from the capacity of an infinite bandwidth additive white Gaussian noise channel only by the factor $1 - 2T_d/T_s$, after we note that ϵ can be chosen arbitrarily small. Under our assumption that $T_d \ll T_s$, the capacity lost is negligible.

4 Mutual Information For White-Like Signals

There are a number of interesting properties of the capacity-achieving scheme described in the previous section. First, the input signals are “peaky” in frequency. Each occupies a single narrow band. Second, they are peaky in time as well. The parameter θ introduced represents the duty cycle of the transmitted signal, and it approaches zero to get close to capacity. Third, the channel is never explicitly measured at the receiver; the detection is non-coherent.

The above properties of the input signals are quite different than more traditional CDMA waveforms which are broadband and which are transmitted continuously over time. We now turn our attention to mutual information achieved using such signals. The main conclusion we will show, under some simplifying assumptions, is that the mutual information achieved using these signals is inversely proportional to the number of equal-energy resolvable paths and in fact approaches 0 as the number of such paths gets large.

We will make a few further assumptions on the fading process. First, we assume that the complex gain for path ℓ , $A_\ell = a_\ell(t) \exp(j2\pi f_c \tau_\ell(t))$ is constant over a time interval of duration T_c , and jumps to a new independent value at the end of this interval. While typically the channel varies in a more continuous manner, this model greatly simplifies the analysis while capturing the essential idea of channel coherence. Moreover, because f_c is typically very large, we will assume that the gains A_ℓ ’s are circularly symmetric. We will also assume that there is negligible spillover of the input signal across intervals, consistent with our assumption of the delay spread being much less than the coherence time.

Under our assumptions, the channel in different intervals of length T_c are independent, and we can focus on analyzing the mutual information achievable on one such interval. We shift to baseband and sample the continuous-time system (1) at a rate of $1/W$. In the discrete time model we have

$$Y_i = \sqrt{\frac{PT_c}{N_0 K_c}} \sum_{\ell=1}^L A_\ell X_{i-\tau_\ell} + Z_i, \quad i = 1, \dots, K_c \quad (4)$$

where $K_c = \lfloor WT_c \rfloor$, $\tau_\ell = \lfloor W\tau_\ell(t) \rfloor$, and Z_i are the samples of the noise process. The normalization is done such that $E[|Z_i|^2] = 1$ and the X_i ’s satisfy the energy constraint:

$$E \left[\frac{1}{K_c} \sum_{i=1}^{K_c} X_i^2 \right] \leq 1.$$

The sampled delays τ_ℓ ’s are the actual delays sampled at a resolution of $\frac{1}{W}$. There may be more than one path with the same sampled delay. These paths are not resolvable at this sampling rate and from the receiver point of view can be considered as single

paths. Let \tilde{L} be the number of such resolvable paths and let $D_1, \dots, D_{\tilde{L}}$ be the distinct sampled delays of these paths. If we let

$$G_\ell = \sum_{m: \tau_m = D_\ell} A_m$$

be the sum of the gains of the paths with the same (sampled) delay D_ℓ , then we can rewrite eqn. (4) as

$$Y_i = \sqrt{\frac{\mathcal{E}}{K_c}} \sum_{\ell=1}^{\tilde{L}} G_\ell X_{i-D_\ell} + Z_i \quad i = 1, \dots, K_c \quad (5)$$

where $\mathcal{E} = PT_c/N_0$.

At this point, we have a discrete tap model of the channel with a finite number of resolvable paths, each of which may in turn be a sum of a number of paths. The gains of these paths are independent from one interval (of length T_c) to the next. In wireless scenarios, the delays D_ℓ , though random, typically vary at a much slower time-scale than the path gains. This is because the coherence time for the path gains is inversely proportional to the carrier frequency f_c , while the time for the delay of a path to change by one tap is inversely proportional to W . Since typically $W \ll f_c$, the delay of a path is changing at a much slower time-scale than its gain. For example, if we take $W = 10^6 \text{ Hz}$ and $f_c = 10^9 \text{ Hz}$, then for a transmitter moving at 60 mph towards the receiver, it takes about 18 seconds for the direct path to move from one tap to another, while the path gain is rotating at about 55 Hz. Thus, here we make the assumption that the path delays D_ℓ 's can be tracked perfectly at the receiver, i.e. timing acquisition has already been performed. This assumption is consistent with the fact that timing acquisition in spread-spectrum systems is usually much easier than tracking of path gains and phases. We will further make the assumption that the delays D_ℓ 's and the path gains G_ℓ 's are independent. In Section 6, we will consider the situation when path timing is not assumed to be known *a priori*.

4.1 Upper Bound on Mutual Information

In the above scenario, the quantity of interest is $I(X; Y|D)$, which gives the mutual information per T_c . We now present an upper bound on this quantity.

Lemma 1.

$$I(X; Y|D) \leq E_{X, G, D} \log \left(E_H \exp \left[\frac{2\mathcal{E}}{K_c} \Re \left\{ \sum_{i=1}^{K_c} \left(\sum_{\ell=1}^{\tilde{L}} X_{i-D_\ell} G_\ell \right) \left(\sum_{m=1}^{\tilde{L}} X_{i-D_m} H_m \right)^* \right\} \right] \right). \quad (6)$$

where $\{H_\ell\}$ are independent of $\{G_\ell\}$ and $\{D_\ell\}$ and each H_ℓ is identically distributed as G_ℓ .

Proof. See Appendix A. \square

We now further bound the right hand side of eqn. (6). Let $H_\ell = |H_\ell|e^{-j\psi_\ell}$ and $G_\ell = |G_\ell|e^{-j\phi_\ell}$. Circular symmetry implies that ϕ 's and ψ 's are uniform in $[-\pi, \pi]$. For the expectation inside the logarithm, condition on everything else and take the expectation with respect to the ψ 's first. We then get:

$$\begin{aligned} E_H \exp \left[\frac{2\mathcal{E}}{K_c} \Re \left\{ \sum_{i=1}^{K_c} \left(\sum_{\ell=1}^{\bar{L}} X_{i-D_\ell} G_\ell \right) \left(\sum_{m=1}^{\bar{L}} X_{i-D_m} H_m \right)^* \right\} \right] \\ = E_{|H|} E_\psi \exp \left[2\mathcal{E} \sum_{m=1}^{\bar{L}} \Re \left\{ e^{-j\psi_m} |H_m| \left(\sum_{\ell=1}^{\bar{L}} G_\ell C(D_\ell, D_m) \right) \right\} \right] \end{aligned}$$

where $C(m, n) = K_c^{-1} \sum_{i=1}^{K_c} X_{i-m} X_{i-n}^*$ is the empirical auto-correlation function of the input signal. Now $E \exp(\Re(ae^{j\theta})) = I_0(2|a|)$, where I_0 is the 0th order modified Bessel function of the 1st kind. Using the inequality $I_0(x) \leq \exp(x^2/4)$ we get

$$\begin{aligned} E_H \exp \left[\frac{2\mathcal{E}}{K_c} \Re \left\{ \sum_{i=1}^{K_c} \left(\sum_{\ell=1}^{\bar{L}} X_{i-D_\ell} G_\ell \right) \left(\sum_{m=1}^{\bar{L}} X_{i-D_m} H_m \right)^* \right\} \right] \\ \leq E_{|H|} \exp \left[\mathcal{E}^2 \sum_{m=1}^{\bar{L}} |H_m|^2 \left| \sum_{\ell=1}^{\bar{L}} G_\ell C(D_\ell, D_m) \right|^2 \right] \end{aligned}$$

Using Jensen's inequality, our bound on $I(X; Y)$ is thus

$$I(X; Y) \leq \log E_{D, G, |H|} \exp \left[\mathcal{E}^2 \sum_{m=1}^{\bar{L}} |H_m|^2 \left| \sum_{\ell=1}^{\bar{L}} G_\ell C(D_\ell, D_m) \right|^2 \right]$$

Suppose now the input signal is stationary and white with autocorrelation function $\delta(n)$. Assuming the coherence time bandwidth product is large such that the empirical auto-correlation of the input is the same as the auto-correlation function, i.e. $C(m, n) = \delta_{mn}$. Then,

$$\sum_{m=1}^{\bar{L}} |H_m|^2 \left| \sum_{\ell=1}^{\bar{L}} G_\ell C(D_\ell, D_m) \right|^2 = \sum_{m=1}^{\bar{L}} |H_m|^2 |G_m|^2$$

and so

$$I(X; Y|D) \leq \sum_{m=1}^{\bar{L}} \log E_{|H_m|, |G_m|} \exp(\mathcal{E}^2 |H_m|^2 |G_m|^2)$$

This bound can be explicitly computed for specific distribution of the path amplitudes. To exhibit asymptotic behavior as the number of resolvable paths get large, consider

the case when $|H_m|^2$'s are identically distributed for all m , and $E(|H_m|^2) = \frac{1}{\tilde{L}}$, i.e. the resolvable paths have equal amount of energy. Then:

$$I(X; Y|D) \leq \tilde{L} \log E_{|H_1|, |G_1|} \exp(\mathcal{E}^2 |H_1|^2 |G_1|^2) = \tilde{L} \log g(\mathcal{E}^2)$$

where $g(r)$ is the generating function of $|H_1|^2 |G_1|^2$ (assumed exists.) If the number of distinguishable paths \tilde{L} large,

$$g(\mathcal{E}^2) \approx 1 + \mathcal{E}^2 E(|H_1|^2 |G_1|^2) = 1 + \frac{\mathcal{E}^2}{\tilde{L}}$$

and hence the upper bound on $I(X; Y|D)$ is approximately

$$\frac{\mathcal{E}^2}{\tilde{L}}.$$

Thus, for large \tilde{L} , an approximate upper bound on the mutual information per unit time is

$$\frac{P^2 T_c}{N_0^2 \tilde{L}} \quad (7)$$

We observe that this bound is inversely proportional to the number of resolvable paths but do not depend directly on the bandwidth W . As the number of equal energy paths go large, the mutual information goes to zero.

We have made a few assumptions above to simplify the calculations, but they are not really necessary. For instance, it is not difficult to show that if the energy in path i is c_i/\tilde{L} , where c_i 's are bounded and bounded away from 0, the same asymptotic upper bound (7) holds. Thus, it is not necessary that the paths have identical energy as long as their energies are all becoming small at the same rate.

It is also not necessary to assume that the autocorrelation function is $\delta(n)$. It is enough that the random variable

$$U = \sum_{m=1}^{\tilde{L}} |H_m|^2 |F_m|^2$$

with $F_m = \sum_{\ell=1}^{\tilde{L}} G_\ell C(D_\ell, D_m)$ that appears in our bound for the mutual information is in some sense small. For this, it suffices to assume that the empirical autocorrelation function $C(m, n)$ of the input process has some summability properties. For example, assuming that

$$\sum_{\ell} (m - \ell)^2 |C(\ell, m)|^2 < \alpha$$

for all m for some α , we can get a similar upper bound as in (7). The derivation of this is given in Appendix C.

4.2 Lower Bound on Mutual Information

The upper bound (7) shows that the mutual information goes to zero when the number of resolvable paths become large. What happens when the number of paths are bounded even though the bandwidth is large? We address this issue by deriving a lower bound to the mutual information $I(X; Y|D)$ below.

For simplicity, we assume that the input $\{X_i\}$ is i.i.d. complex circular symmetric Gaussian. We begin with the following relationships:

$$I(X; Y|D) = I(Y; X, G|D) - I(Y; G|X, D) \geq I(Y; X|G, D) - I(Y; G|X, D) \quad (8)$$

where the first equality follows from the chain rule. Conditional on the paths gains G and the delays D , X and Y are jointly Gaussian. The first term is then given by:

$$I(Y; X|G, D) = E_{G,D} \log \det(I + \frac{\mathcal{E}}{K_c} AA^*)$$

where A is a K_c by K_c matrix such that $A_{im} = G_\ell$ if $m = i - D_\ell$ and 0 otherwise. A is a Toeplitz matrix. By our assumption, the delay spread T_d is much smaller than the coherence time T_c . Hence $D_\ell \ll K_c$. In this regime, the eigenvalues of AA^* are well approximated by $|C(\frac{k}{K_c})|^2, k = 0 \dots K_c - 1$, where

$$C(f) = \sum_{\ell=1}^L G_\ell \exp(2\pi j D_\ell f)$$

is the Fourier transform of the impulse response of the channel. Hence

$$\begin{aligned} I(Y; X|G, D) &= E_{G,D} \left[\sum_{k=1}^{K_c} \log(1 + \frac{\mathcal{E}}{K_c} |C(\frac{k}{K_c})|^2) \right] \\ &= K_c E_{G,D} \log(1 + \frac{\mathcal{E}}{K_c} |C(0)|^2) \\ &= K_c E_G \log(1 + \frac{\mathcal{E}}{K_c} \sum_{\ell=1}^L |G_\ell|^2) \end{aligned} \quad (9)$$

The second step follows from the fact that $C(f)$ is identically distributed for every f , which in turns follows from the circular symmetry and independence of the G_ℓ 's.

We can upper bound the second term in (8) by making a worst-case assumption that the paths gains G_ℓ 's are circularly symmetric and Gaussian with the same variance.

$$I(Y; G|X, D) \leq E_{X,D} \log \det(I + \frac{\mathcal{E}}{K_c} B \Lambda B^*) \quad (10)$$

where $B_{im} = X_{i-m}$ and $\Lambda = \text{diag}(\sigma_1^2, \dots, \sigma_L^2)$, where $\sigma_\ell^2 = E[|G_\ell|^2]$. Now,

$$\begin{aligned}
& I(Y; G|X, D) \\
& \leq E_{X,D} \log \det(I + \frac{\mathcal{E}}{K_c} B \Lambda B^*) \\
& = E_{X,D} \log \det(I + \frac{\mathcal{E}}{K_c} B^* B \Lambda) \\
& \leq \tilde{L} E_{X,D} \log \left[\frac{1}{\tilde{L}} \text{tr}(I + \frac{\mathcal{E}}{K_c} B^* B \Lambda) \right] \quad (\text{Jensen's inequality}) \\
& = \tilde{L} E_{X,D} \log \left[1 + \frac{\mathcal{E}}{\tilde{L}} \sum_{\ell=1}^{\tilde{L}} \sigma_\ell^2 \left(\frac{1}{K_c} \sum_{m=1}^{K_c} |X_{m-D_\ell}|^2 \right) \right] \\
& \leq \tilde{L} E_D \log \left[1 + \frac{\mathcal{E}}{\tilde{L}} \sum_{\ell=1}^{\tilde{L}} \sigma_\ell^2 E_X \left(\frac{1}{K_c} \sum_{m=1}^{K_c} |X_{m-D_\ell}|^2 \right) \right] \quad (\text{Jensen's inequality}) \\
& \leq \tilde{L} \log(1 + \frac{\mathcal{E}}{\tilde{L}}).
\end{aligned}$$

The last inequality follows from the energy constraint on the input and that $\sum_\ell \sigma_\ell^2 = 1$. Combining this with eqn. (9) yields the following lower bound:

$$I(X; Y|D) \geq K_c E_G \log(1 + \frac{\mathcal{E}}{K_c} |\sum_{\ell=1}^{\tilde{L}} G_\ell|^2) - \tilde{L} \log(1 + \frac{\mathcal{E}}{\tilde{L}}).$$

Let us now examine this bound in the wideband limit. For large W , $K_c = \lfloor WT_c \rfloor$ is large and the first term approaches:

$$K_c E_G \log(1 + \frac{\mathcal{E}}{K_c} |\sum_{\ell=1}^{\tilde{L}} G_\ell|^2) \rightarrow \mathcal{E} E_G |\sum_{\ell=1}^{\tilde{L}} G_\ell|^2 = \mathcal{E}.$$

The quantity $I(X; Y|D)$ is the mutual information per coherence time interval. Thus, in the wideband limit, we have the following lower bound on the mutual information *per unit time*:

$$I(X; Y|D) \geq \frac{P}{N_0} - \frac{\tilde{L}}{PT_c} \log(1 + \frac{PT_c}{N_0 \tilde{L}})$$

Note that the second term is always less than the first term, so that this lower bound is strictly positive. The first term is the capacity of the infinite bandwidth AWGN channel. The second term can therefore be interpreted as an upper bound on the capacity penalty due to channel uncertainty. Observe that this term depends only on

the number of resolvable paths and not on the bandwidth. In particular, if the number of paths is bounded, then the mutual information is bounded away from zero even at infinite bandwidth. This further emphasizes that the fundamental limitation comes from the number of equal-energy resolvable paths.

As $\tilde{L} \rightarrow \infty$, we have the following asymptotic lower bound:

$$\frac{P^2 T_c}{2N_0^2 \tilde{L}}, \quad (11)$$

which approaches zero as $\tilde{L} \rightarrow \infty$. Compared to the asymptotic upper bound in (7), we see that the upper and lower bounds agree to within a factor of 2.

If we let:

$$\tilde{L}_{\text{crit}} := \frac{PT_c}{N_0} \quad (12)$$

and

$$C_{\text{AWGN}} := \frac{P}{N_0},$$

then we can write the lower bound as

$$C_{\text{AWGN}} \left[1 - \frac{\tilde{L}}{\tilde{L}_{\text{crit}}} \log\left(1 + \frac{\tilde{L}_{\text{crit}}}{\tilde{L}}\right) \right]$$

and the upper bound as

$$C_{\text{AWGN}} \frac{\tilde{L}_{\text{crit}}}{\tilde{L}}.$$

Note that the upper bound holds for large \tilde{L} while the lower bound holds for any \tilde{L} . If $\tilde{L} \ll \tilde{L}_{\text{crit}}$, then

$$\frac{\tilde{L}}{\tilde{L}_{\text{crit}}} \log\left(1 + \frac{\tilde{L}_{\text{crit}}}{\tilde{L}}\right) \approx 0,$$

and the mutual information achievable with spread-spectrum signals is close to the capacity of the infinite-bandwidth AWGN channel. On the other hand, if $\tilde{L} \gg \tilde{L}_{\text{crit}}$, then the upper bound says that the mutual information achievable is negligible compared to that of an AWGN channel. Thus, one may view \tilde{L}_{crit} as the *critical parameter* delineating the regime where “over-spreading” occurs. If one thinks of P/N_0 as a nominal information rate, then \tilde{L}_{crit} is smaller for low-rate users and for systems with shorter coherence time.

At carrier frequency of 1 GHz and vehicle speed of 60 mph, the coherence time is of the order of 18 milliseconds. For a voice user with data rate of 9.6 kbits/s, this gives a value of \tilde{L}_{crit} to be 120. On the other hand, at 10 GHz, the coherence time becomes 1.8 milliseconds, and $\tilde{L}_{\text{crit}} = 12$. The upper and lower bounds are plotted for these scenarios in Figs. (1) and (2), as a function of the number of resolvable paths.

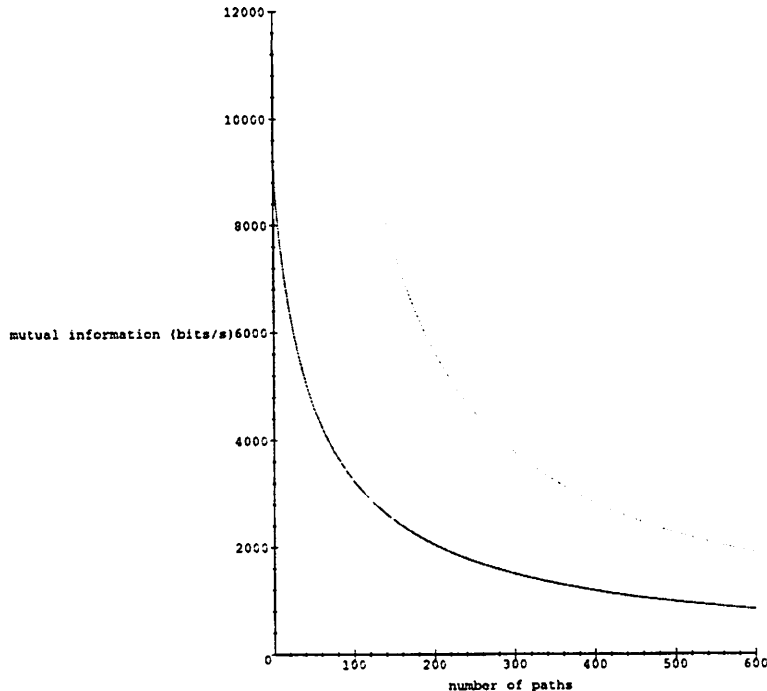


Figure 1: Upper and lower bounds to the achievable mutual information as a function of the number of resolvable paths, for $T_c = 0.018s$. The unit is in bits per second. The upper horizontal line is the capacity of the AWGN channel.

5 Detection of Binary Orthogonal Signals

In the previous sections, we studied the information theoretic properties of broadband multipath channels, focusing on performance scaling when the number of resolvable paths become large. In this section, we will shift our emphasis to the detection error probability of specific binary orthogonal modulation schemes under the same scaling. We will demonstrate performance deterioration as the number of multipaths grow, in a manner akin to the scaling of mutual information. We will also give an intuitive understanding of the critical parameter \tilde{L}_{crit} in terms of estimation errors in the path gains.

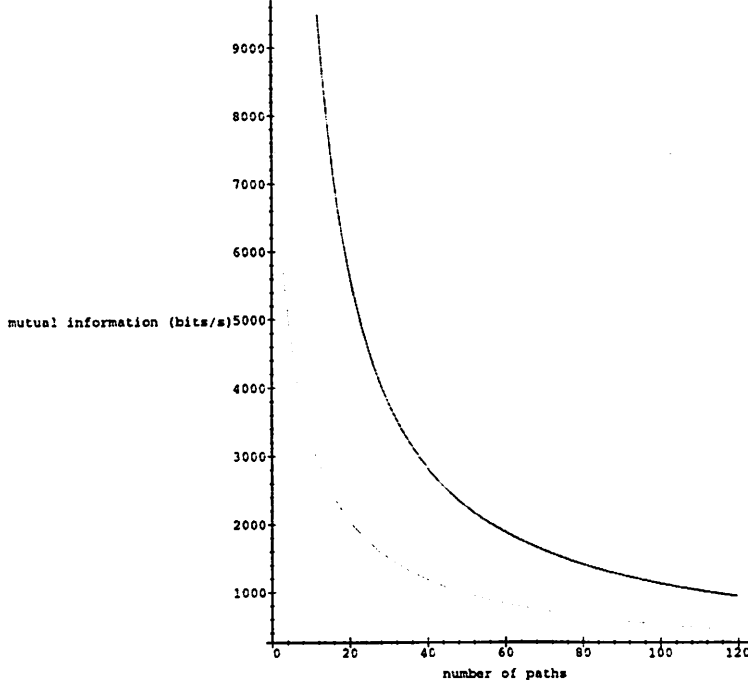


Figure 2: Upper and lower bounds to the achievable mutual information as a function of the number of resolvable paths, for $T_c = 0.0018s$. The unit is in bits per second. The upper horizontal line is the capacity of the AWGN channel.

We use the same channel model as in eqn. (1):

$$y(t) = \sum_{\ell=1}^L a_{\ell}(t)x(t - \tau_{\ell}(t)) + z(t), \quad (13)$$

where each path has independent statistics.

Consider now an uncoded binary modulation scheme when at each symbol time one of two orthogonal waveforms $x_0(\cdot)$ and $x_1(\cdot)$ is transmitted. The symbol duration T_s is chosen such that $T_d \ll T_s \ll T_c$, where T_d and T_c are the delay spread and the coherence time of the channel respectively. The symbol duration much larger than the delay spread means that we can ignore inter-symbol interference. The symbol duration much less than the coherence time means that we can assume that the channel is essentially time-invariant over a symbol duration. The average received energy per bit

is \mathcal{E}_b . The two symbols are assumed to be equiprobable. We compare the performance of narrowband and broadband signaling schemes, under coherent and non-coherent detection.

5.1 Narrowband Signaling

First consider the case when the signals are narrowband FSK signals at frequencies f_0 and f_1 , chosen to be orthogonal. (These are the same as the ones used in the capacity-achieving strategy described in Section 3.) By correlating the received signals with x_0 and x_1 in turn, we obtain two sufficient statistics R_0 and R_1 for detection. Assume without loss of generality that symbol 0 is transmitted. Similar to the development in Section 3, we obtain:

$$R_l = \begin{cases} \sqrt{\mathcal{E}_b}G + W_l & l = 0 \\ W_l & l = 1 \end{cases}$$

where $G = \sum_{\ell=1}^L a_\ell \exp(j2\pi f_0 \tau_\ell)$ and W_0, W_1 are independent circular symmetric complex Gaussian rv's with variance N_0 . (Recall that G is normalized such that $E(|G|^2) = 1$. If G is known to the receiver, then coherent detection can be done, and the error probability, conditional on G , is given by

$$p_e(G) = Q\left(\sqrt{\frac{\mathcal{E}_b}{2N_0}}|G|\right),$$

where $Q(\cdot)$ is the complementary cdf of a $N(0, 1)$ rv. If we now assume that each of the path has uniform phase, magnitude a_ℓ such that $E(a_\ell^2) = \frac{1}{L}$ and Rayleigh distributed, then G is circular-symmetric Gaussian with variance 1, and the probability of error, averaged over G , is given by (see for example [4, eqn. 7.3.8]):

$$p_e = \frac{1}{2} \left[1 - \sqrt{\frac{\frac{\mathcal{E}_b}{N_0}}{2 + \frac{\mathcal{E}_b}{N_0}}} \right].$$

Observe that this expression does not depend on L . If each path is not Rayleigh but still has uniform phase and identically distributed, then this expression holds in the limit when L becomes large, due to the Central Limit Theorem.

If G is not known to the receiver, then non-coherent detection has to be done by comparing the magnitude of R_0 and R_1 (square-law detector). The error probability, conditional on G , is given by [4, eqn. 7.3.11]

$$p_e(G) = \frac{1}{2} \exp\left(-\frac{1}{2} \frac{\mathcal{E}_b}{N_0} |G|^2\right).$$

Assuming that each path is Rayleigh, the average error probability is then [4, eqn. 7.3.12]

$$p_e = \frac{1}{2 + \frac{\mathcal{E}_b}{N_0}}.$$

If each path is not Rayleigh, then this holds only in the limit when L becomes large.

We observe that while, as expected, the performance of non-coherent detection is worse than coherent detection, the performance of the non-coherent detector does not get arbitrarily worse as the number of paths get large. Its limiting performance depends only on the average SNR.

5.2 Wideband Signaling

Let us now consider using spread-spectrum signals, such that x_0 and x_1 are white-like and orthogonal. Without going into the specific details of the structure of the signals, it suffices for our purpose here to assume that the signals have been chosen such that delayed versions are nearly orthogonal to each other. In this case, a reasonable approximation is the standard diversity branch model (see for example [4, Section 7.4]). In this model, the receiver observes \tilde{L} independently faded replicas of the information signal, one for each resolvable path. The additive noise in each branch is white, Gaussian with power spectral density $\frac{N_0}{2}$, and independent between branches. This last assumption ignores the “self-noise” due to interference between delayed versions of the signals, and this is a good approximation if the signals are white-like.

More specifically, suppose that the \tilde{L} resolvable paths are at sampled delays $D_1, \dots, D_{\tilde{L}}$, assumed known to the receiver. Then if symbol 0 is transmitted, the branches at the baseband are given by

$$y_\ell(t) = G_\ell x_0(t - D_\ell) + z_\ell(t), \quad \ell = 1, \dots, \tilde{L}$$

where G_ℓ is the sum of the complex gains of the paths at delay D_ℓ . Match filtering each of the branches with $x_0^*(t - D_\ell)$ and $x_1^*(t - D_\ell)$ gives us the following sufficient statistics for each ℓ :

$$R_{l\ell} = \begin{cases} \sqrt{\mathcal{E}_b} G_\ell + W_{l\ell} & l = 0 \\ W_{l\ell} & l = 1 \end{cases} \quad (14)$$

where $\{W_{l\ell}\}$ are i.i.d. circular symmetric Gaussian random variables with variance N_0 . Note that $G = \sum_{\ell=1}^{\tilde{L}} G_\ell$. For simplicity, we will assume that the gains G_ℓ 's of the resolvable paths are identically distributed, and hence have variance $1/\tilde{L}$, i.e. the energy in the signal is equally spread among the paths. Observe that the narrowband scenario corresponds to $\tilde{L} = 1$.

If the receiver has perfect knowledge of the complex path gains $\{G_\ell\}$, then the optimum detector is to do maximal-ratio combining, weighting each branch by G_ℓ^* and

then adding. This is simply the Rake receiver. Conditional on $\{G_\ell\}$, the probability of error is given by [4, eqn. 7.4.20]

$$p_e(\{G_\ell\}) = Q\left(\sqrt{\frac{1}{2} \frac{\mathcal{E}_b}{N_0}} \sum_{\ell=1}^{\tilde{L}} |G_\ell|^2\right)$$

If we assume that each of the $|G_\ell|$'s has a Rayleigh distribution, the average error probability can be explicitly calculated as [4, eqns. 7.4.15, 7.4.21]:

$$p_e = \left(\frac{1-\mu}{2}\right)^{\tilde{L}} \sum_{\ell=0}^{\tilde{L}-1} \binom{\tilde{L}-1+\ell}{\ell} \left(\frac{1+\mu}{2}\right)^\ell \quad (15)$$

where

$$\mu = \sqrt{\frac{\frac{\mathcal{E}_b}{N_0}}{2\tilde{L} + \frac{\mathcal{E}_b}{N_0}}}.$$

Regardless of whether the path gains are Rayleigh, as \tilde{L} becomes large,

$$\sum_{\ell=1}^{\tilde{L}} |G_\ell|^2 \xrightarrow{\mathcal{P}} 1$$

so that the error probability converges to $Q(\sqrt{\frac{1}{2} \frac{\mathcal{E}_b}{N_0}})$, i.e. the same as that for a non-fading channel with the same received SNR.

The performance of coherent detection as a function of number of resolvable paths is plotted in Figs. 3 and 4 for Rayleigh fading and at different SNRs. The narrowband scenario corresponds to having 1 diversity branch. We see that the performance of the broadband scheme improves monotonically with the number \tilde{L} of resolvable paths. This is the well-known multipath diversity advantage of spread-spectrum schemes.

The picture, however, is different for non-coherent detection. Consider a receiver which does not know the path gains G_ℓ 's and implements a square-law detector, i.e. it computes for $l = 0, 1$,

$$U_l = \sum_{\ell=1}^{\tilde{L}} |R_{l\ell}|^2$$

and makes a decision based on the larger of U_0 and U_1 . The probability of error is

$$\Pr[U_1 > U_0] = \Pr\left[\sum_{\ell=1}^{\tilde{L}} |W_{1\ell}|^2 > \sum_{\ell=1}^{\tilde{L}} |\sqrt{\mathcal{E}_b} G_\ell + W_{0\ell}|^2\right]$$

Let us first examine this error probability in the limit when the number of resolvable paths becomes large. Direct computation shows that

$$E(U_1 - U_0) = -\mathcal{E}_b$$

and hence

$$\lim_{\tilde{L} \rightarrow \infty} E\left(\frac{U_1 - U_0}{\sqrt{\tilde{L}}}\right) = 0.$$

Also,

$$\lim_{\tilde{L} \rightarrow \infty} \text{Var} \left[\frac{U_1 - U_0}{\sqrt{\tilde{L}}} \right] = 2N_0^2.$$

Since U_0 and U_1 are independent and both are a sum of \tilde{L} independent terms, we can apply the Central Limit Theorem and conclude that

$$\frac{U_1 - U_0}{\sqrt{\tilde{L}}} \xrightarrow{D} N(0, 2N_0^2).$$

Hence, the probability of error of the non-coherent scheme approaches 1/2 for a large number of resolvable paths. How large does \tilde{L} have to be for this to happen? A more refined estimate of the error probability yields

$$p_e \approx Q\left(\sqrt{\frac{1}{2\tilde{L}}} \frac{\mathcal{E}_b}{N_0}\right).$$

Hence, when \tilde{L} is compare to the SNR $\frac{\mathcal{E}_b}{N_0}$, then the performance of the non-coherent detector degrades significantly.

For the case when the gain G_ℓ of each branch is Rayleigh, an explicit expression for the error probability can be computed for finite \tilde{L} [4, eqn. 7.4.30]: it is given by formula (15) as in the coherent case, but with μ given instead by

$$\mu = \frac{\frac{\mathcal{E}_b}{N_0}}{2\tilde{L} + \frac{\mathcal{E}_b}{N_0}}.$$

The performance of non-coherent detection is plotted as a function of the number of resolvable paths in Fig. 3 and 4 for different SNR's. We see that for small \tilde{L} , performance of broadband scheme improves over that of the narrowband scheme ($\tilde{L} = 1$) with increasing \tilde{L} . This is due to the effect of multipath diversity. As \tilde{L} is increased further, there is a diminishing return to the benefits from the multipath diversity. On the other hand, the lack of knowledge about the gains of the individual resolvable paths starts to hurt the combining ability of the non-coherent broadband receiver. There is an optimal \tilde{L}^* after which the performance of the non-coherent broadband detector

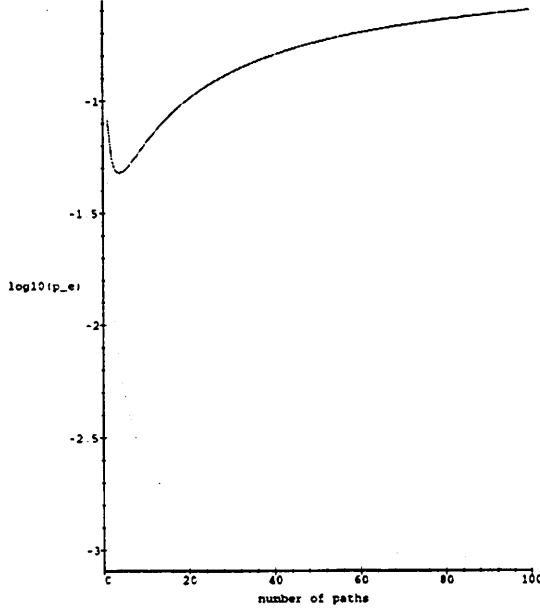


Figure 3: Comparison of error probability under coherent detection (below) and non-coherent detection (above), as a function of the number of paths. $\frac{\mathcal{E}_b}{N_0} = 10dB$.

starts to degrade. As $\tilde{L} \rightarrow \infty$, the non-coherent broadband scheme performs even worse than the non-coherent narrowband scheme and in fact the error probability of the former approaches $1/2$.

Observe the contrast in performance scaling of the coherent and non-coherent broadband schemes. A natural question is whether the poor performance scaling of the non-coherent scheme can be offset to some extent by estimating the path gains and using the estimates in a coherent receiver. To get some insights to this question, let us analyze the performance of a maximal-ratio combiner, using imperfect estimates \hat{G}_ℓ 's instead of G_ℓ . We assume that for each diversity branch $\ell = 1, \dots, \tilde{L}$, the estimate \hat{G}_ℓ is obtained from a set of noisy measurements:

$$S_{\ell k} = \sqrt{\mathcal{E}_p} G_\ell + Z_{\ell k}, \quad k = 1, \dots, K$$

The channel measurements are commonly obtained in two ways: from a pilot signal with known data symbols, or from previously detected symbols. In the former case, \mathcal{E}_p is the energy per bit of the pilot signal, while in the latter case, $\mathcal{E}_p = \mathcal{E}_b$. In either

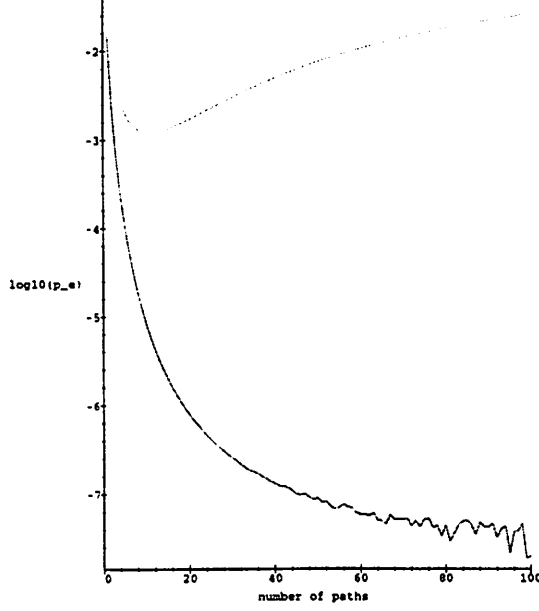


Figure 4: Comparison of error probability under coherent detection (below) and non-coherent detection (above), as a function of the number of paths. $\frac{\mathcal{E}_p}{N_0} = 15dB$.

case, it is reasonable to assume that one can measure over a time interval of length T_c , the coherence time. Hence, the number of measurements K can be taken to be T_c/T_s , where T_s is the symbol duration. The noise $Z_{\ell k}$'s are taken to be i.i.d. circular symmetric random variables with variance N_0 , and also independent of the noise in the interval of the current symbol to be detected.

We employ the LLSE estimate of G_ℓ ; for each ℓ , this is given by:

$$\hat{G}_\ell = \frac{\frac{\mathcal{E}_p}{N_0}}{K \frac{\mathcal{E}_p}{N_0} + \tilde{L}} \sum_{k=1}^K S_{\ell k}.$$

The mean-square error associated with this estimate is:

$$\frac{1}{K \frac{\mathcal{E}_p}{N_0} + \tilde{L}}, \quad (16)$$

same for all branches. The maximal-ratio combiner, using the channel estimates, com-

putes for each $l = 0, 1$

$$V_l := \Re \left\{ \sum_{\ell=1}^{\tilde{L}} \hat{G}_\ell^* R_{l\ell} \right\}$$

where $R_{l\ell}$ is given in eqn. (14), and picks the hypothesis with the larger V_l . The probability of error is

$$p_e = \Pr[V_1 > V_0] = \Pr \left[\Re \left\{ \sum_{\ell=1}^{\tilde{L}} \hat{G}_\ell^* (W_{1\ell} - W_{0\ell}) \right\} > \sqrt{\mathcal{E}_b} \Re \left\{ \sum_{\ell=1}^{\tilde{L}} \hat{G}_\ell^* G_\ell \right\} \right].$$

Direct computation yields:

$$\begin{aligned} E(\hat{G}_\ell^* G_\ell) &= \frac{K \frac{\mathcal{E}_p}{N_0}}{\tilde{L}(\frac{\mathcal{E}_p}{N_0} + \tilde{L})} \\ E(G^*(W_{1\ell} - W_{0\ell})) &= 0 \\ \text{Var}[G^*(W_{1\ell} - W_{0\ell})] &= \frac{2K \mathcal{E}_p}{\tilde{L}(\frac{\mathcal{E}_p}{N_0} + \tilde{L})}. \end{aligned}$$

Applying the Central Limit Theorem, as $\tilde{L} \rightarrow \infty$,

$$\sqrt{\tilde{L}} \Re \left\{ \sum_{\ell=1}^{\tilde{L}} \hat{G}_\ell^* (W_{1\ell} - W_{0\ell}) \right\} \xrightarrow{D} N(0, K \mathcal{E}_p).$$

Also, by a variance computation, one can show that as $\tilde{L} \rightarrow \infty$,

$$\sqrt{\tilde{L}} \sqrt{\mathcal{E}_b} \Re \left\{ \sum_{\ell=1}^{\tilde{L}} \hat{G}_\ell^* G_\ell \right\} \xrightarrow{P} 0.$$

We thus conclude that as the number of resolvable paths grow, the probability of error approaches 1/2 for the coherent scheme using imperfect channel estimates. Using the mean and variance computation done above, a more refined estimate of the error probability for large \tilde{L} is given by

$$p_e \approx Q \left(\sqrt{\frac{1}{2(1 + \frac{\tilde{L} N_0}{K \mathcal{E}_p})} \frac{\mathcal{E}_b}{N_0}} \right).$$

Thus, if $\tilde{L} \ll K \frac{\mathcal{E}_p}{N_0}$, then the performance is very close to that of the coherent receiver with perfect channel estimates. On the other hand, if $\tilde{L} \gg K \frac{\mathcal{E}_p}{N_0}$, then the imperfect channel estimates have a significant impact on performance. An intuitive

explanation can be seen from the expression (16): if $\tilde{L} \gg K \frac{\mathcal{E}_p}{N_0}$, then the mean square error in estimating G_ℓ is approximating $1/\tilde{L}$, the variance of G_ℓ itself. In other words, little information is gained about the G_ℓ 's from the channel measurements. As the number of paths grow large, the receiver meets the same fate as the non-coherent receiver: detection becomes impossible.

The critical parameter

$$\tilde{L}_{\text{crit}} := K \frac{\mathcal{E}_p}{N_0}$$

can be interpreted as the threshold delineating the regime in which the system is “over-spread”: if the number of resolvable paths is significantly larger than \tilde{L}_{crit} , the estimation errors in the paths gains precludes effective combining of the multipaths. Expressing this threshold in terms of system parameters, we find that

$$\tilde{L}_{\text{crit}} = \frac{PT_c}{N_0}$$

where P is the received power of the signal from which channel measurements are obtained. If the measurements are done in a decision-feedback mode, P is the received power of the transmitted signal itself. In this case, the critical parameter defined here for detection coincides with that defined in (12) for the achievable mutual information. If the measurements are done from a pilot, P is the power of the pilot. On the downlink of a CDMA system, it is more economical to have a pilot common to all users; moreover, the power can be larger than the signals for the individual user. This makes coherent combining easier, resulting in a larger \tilde{L}_{crit} . On the uplink, however, it is not possible to have a common pilot, and the channel estimation will have to be done from previously detected symbols or even non-coherently. With a lower received power from the individual users, \tilde{L}_{crit} can be considerably smaller.

In concluding this section, we see that the scaling of the error probability performance of broadband orthogonal modulation schemes mirrors that of the information theoretic properties we derived earlier. As the number of resolvable paths grow large, the performance of such schemes deteriorate arbitrarily badly, whether they try to estimate channel parameters or perform non-coherent detection. Certainly, this is not surprising as the information theoretic results impose fundamental limitation on the performance of *any scheme* given the constraint that spread-spectrum transmitted signals are used. On the other hand, the analysis of specific modulation schemes done here gives a more concrete feeling as to what goes wrong. Basically, as the number of resolvable paths become large *and* their individual energies become corresponding smaller, it is harder to estimate their gains and to combine them effectively. The fact that the threshold \tilde{L}_{crit} identified in both analyses are the same further substantiates this explanation.

6 Timing Uncertainty

In Section 4, we showed that as the number of resolvable paths \tilde{L} with equal energy gets large, the mutual information decreases inversely proportional to \tilde{L} and approaches zero. This holds even when the receiver can track the delay of each path perfectly. In this section, we shall show that if this side information is not *a priori* assumed, the mutual information goes to zero with increasing bandwidth even when there is only one path.

We will focus on a single path channel ($L = 1$), with a fixed gain ($a_1(t) = 1$), but keep the stochastic nature of the delay process, $\tau_1(t)$. We assume that $\tau_1(t)$ remains constant for a time T'_c and jumps to an independent value in the next time-interval of length T'_c . The duration T'_c can be thought of as the *coherence time* for this model, but observe that this is in general different from the coherence time T_c for the path gains considered in Section 4. As explained there, the path delays typically vary much slower than the path gains.

The second assumption is that the delay is uniformly distributed in $[0, T_d]$, where T_d is the delay spread. We will also assume that there is negligible spillover of the input signal across intervals, consistent with our assumption of the delay spread being much less than the coherence time.

Under our assumptions the channel in different intervals of length T'_c are independent, and we can focus on analyzing the mutual information achievable on one such interval. We will start by shifting to baseband, and discretizing time by sampling at a rate of $1/W$ complex samples per second. In this discrete time model we have

$$Y_i = \sqrt{\frac{PT'_c}{N_0K'_c}} X_{i-\tau} + Z_i, \quad i = 1, \dots, K'_c$$

where Y_i are the samples of the received signal, X_i are the scaled samples of the transmitted signal, τ is the random delay in this interval, and Z_i are the samples of the noise process. Note that we have normalized the scaling so that $E[|Z_i|^2] = 1$. The random variable τ takes values in $\{1, \dots, T_d W\}$ and is uniformly distributed on this range. Let $K'_c = WT'_c$, and $\alpha = T_d/T'_c$. The assumption on the delay spread makes sure that $\alpha \ll 1$. Note that the power constraint over $x(t)$ translates into an energy constraint on $\{X_i : i = 1, \dots, K'_c\}$:

$$E \left[\frac{1}{K'_c} \sum_{i=1}^{K'_c} X_i^2 \right] \leq 1.$$

We now present an upper bound to the mutual information which holds for any input distribution.

Lemma 2. Let $\mathcal{E}' = PT'_c/N_0$. Then:

$$I(X; Y) \leq E \left[\frac{1}{\alpha K'_c} \sum_{d=0}^{\alpha K'_c - 1} \log \left(\frac{1}{\alpha K'_c} \sum_{\ell=0}^{\alpha K'_c - 1} \exp \left\{ 2\mathcal{E}' \Re[C'(d, \ell)] \right\} \right) \right].$$

where,

$$C'(m, n) = \frac{1}{K'_c} \sum_{i=1}^{K'_c} X_{i-m} X_{i-n}^*$$

is the empirical autocorrelation function of the input process over the time-interval of length T'_c .

Proof. The proof follows the same lines as that of Lemma 1 . See Appendix B. \square

Suppose now the input signal is stationary and white with autocorrelation function $\delta(n)$. Assuming the coherence time bandwidth product is large such that the empirical auto-correlation of the input is the same as the auto-correlation function, i.e. $C(m, n) = \delta_{mn}$. Substituting into the upper bound above, we get:

$$\begin{aligned} I(X; Y) &\leq \log \left[1 + \frac{1}{\alpha K'_c} (e^{2\mathcal{E}'} - 1) \right] \\ &= \log \left[1 + \frac{1}{WT_d} \left(\exp\left(\frac{2PT'_c}{N_0}\right) - 1 \right) \right] \end{aligned}$$

As the bandwidth W becomes large, the upper bound decays to zero like

$$\frac{1}{WT_d} \left[\exp\left(\frac{2PT'_c}{N_0}\right) - 1 \right].$$

This decay in mutual information is due to the necessity to track the path timing accurately, with the needed resolution increasing linearly with the bandwidth. While such channel measurements are not crucial for communication using narrowband sinusoids, they are when white-like signals are used. As the bandwidth grows, the channel cannot be tracked at the desired accuracy, and communicating reliably is also impossible. However, since T'_c is quite large for typical wireless scenarios, this phenomenon will kick in only when the bandwidth is very large.

7 Conclusions

The main conclusion of this paper is that the mutual information achievable using spread-spectrum signals through a multipath fading channel depends crucially on how the signal energy is divided among the *resolvable paths*. If there are only a few dominant paths, the achievable mutual information is close to the capacity of the AWGN

channel with the channel gains perfectly. If the energy is spread out among many equal-energy resolvable paths, the mutual information achievable is very small, being inversely proportional to the number of resolvable paths \tilde{L} . The limitation comes from the fact that the energy in each path is too small for the gains to be measured accurately enough for effective combining. From a communication theoretic point of view, multipath diversity benefits the system only up to a certain point. When there are too many paths, the uncertainty about the path gain severely limits performance. We have also established a critical parameter \tilde{L} which delineates the threshold on the number of resolvable paths above which this “over-spreading” phenomenon occurs.

Theorem 1 provides a counterpoint to the above result. It shows that the above phenomenon is not intrinsic to the multipath fading channel itself but is rather a consequence of the signaling strategy. Indeed, by using narrowband signals and transmitting at a low duty cycle, capacity of the infinite-bandwidth AWGN channel can be achieved. This is independent of the number of paths.

An interesting point is brought out by these results. Whereas for the infinite-bandwidth AWGN channel, capacity can be achieved using *any* set of orthogonal signals, such is not the case for multipath fading channels. The performance is very much dependent on the *specific* choice of the orthogonal signals. While capacity can be achieved with narrowband sinusoids, the mutual information achievable by spread-spectrum signals can be very small. This is intimately tied to the fact that sinusoids are eigenfunctions of *any* linear time-invariant system, while white-like signals are not.

Acknowledgments

We would like to thank Prof. Shlomo Shamai for interesting discussions.

References

- [1] R. G. Gallager, *Information Theory and Reliable Communication*, New York: John Wiley & Sons, 1968.
- [2] R. S. Kennedy, *Fading Dispersive Communication Channels*, New York: Wiley Interscience, 1969.
- [3] R. G. Gallager and M. Medard, “Bandwidth Scaling for Fading Channels,” *Proc. of the Int. Symp. on Information Theory, Ulm, Germany, 1997*, p. 471.
- [4] J. G. Proakis, *Digital Communications*, 2nd Edition, McGraw Hill, 1989.
- [5] G. Turin, “Spread Spectrum Antimultipath Techniques,” *Proc. IEEE*, vol. 68, pp. 328–353, March, 1980.

A Proof of Lemma 1

Since

$$I(X; Y|D) = H(Y) - H(Y|X, D),$$

we can bound the mutual information by bounding $H(Y|D)$ and $-H(Y|X, D)$ separately. It is easy to upper bound $H(Y)$ via the entropy power inequality:

$$H(Y) \leq k \log \pi e (1 + \mathcal{E}/k).$$

It remains to upper bound $-H(Y|X, D)$. To that end,

$$\begin{aligned} -H(Y|X, D) &= E[\log(p(Y|X, D))] \\ &= E_G E[\log(p(Y|X, D)|G = g)] \\ &= E_G E \log(E_H \pi^{-k} \exp[-\sum_i |\frac{\mathcal{E}}{k} \sum_{\ell} H_{\ell} X_{i-D_{\ell}} + Z_i - \frac{\mathcal{E}}{k} \sum_{\ell} X_{i-D_{\ell}} g_{\ell}|^2]) \\ &= -E_G E \log(E_H \pi^{-k} \exp[-\sum_i |\frac{\mathcal{E}}{k} P_i + Z_i - \frac{\mathcal{E}}{k} Q_i|^2]) \end{aligned}$$

where $P_i = \sum_{\ell} H_{\ell} X_{i-D_{\ell}}$ and $Q_i = \sum_{\ell} X_{i-D_{\ell}} g_{\ell}$. Expanding the square

$$\begin{aligned} -H(Y|X, D) &= -\log \pi - E \sum_i |Z_i|^2 + E_G E \log(E_H \exp(-\frac{\mathcal{E}}{k} \sum_i |P_i - Q_i|^2 - 2\sqrt{\mathcal{E}/k} \sum_i \Re(P_i - Q_i) Z_i^*)) \\ &= -k \log(\pi e) - E_G 2\sqrt{\mathcal{E}/k} E_H \sum_i \Re(P_i Z_i^*) \\ &\quad + E_G E \log(E_H \exp(-\frac{\mathcal{E}}{k} \sum_i |P_i - Q_i|^2 + 2\sqrt{\mathcal{E}/k} \Re(Q_i Z_i^*))) \\ &\leq -k \log(\pi e) + E_G E_{X,D} \log(E_H \exp - \frac{\mathcal{E}}{k} \sum_i |P_i - Q_i|^2 E_Z \exp 2\sqrt{\mathcal{E}/k} \Re(Q_i Z_i^*)) \\ &= -k \log(\pi e) + E_G E_{X,D} \log(E_H \exp - \frac{\mathcal{E}}{k} \sum_i [|P_i - Q_i|^2 - |Q_i|^2]) \end{aligned}$$

where the inequality follows from Jensen's. Thus

$$\begin{aligned} I(X; Y) &\leq k \log(1 + \mathcal{E}/k) \\ &\quad + E_G E_{X,D} \log \left(E_H \exp - \mathcal{E} \frac{1}{k} \sum_{i=1}^k [|P_i - Q_i|^2 - |Q_i|^2] \right) \end{aligned}$$

$$\begin{aligned}
&= \log(1 + \mathcal{E}/k) - E_G E_{X,D} \frac{\mathcal{E}}{k} \sum_{i=1}^k |P_i|^2 \\
&\quad + E_G E_{X,D} \log \left(E_H \exp \mathcal{E} \frac{1}{k} \sum_{i=1}^k \Re(2P_i Q_i^*) \right) \\
&\leq k \log(1 + \mathcal{E}/k) - \mathcal{E} \\
&\quad + E_G E_{X,D} \log \left(E_H \exp \mathcal{E} \frac{1}{k} \sum_{i=1}^k \Re(P_i Q_i^*) \right),
\end{aligned}$$

proving the lemma.

B Proof of Lemma 2

Since

$$I(X; Y) = H(Y) - H(Y|X),$$

we can bound the mutual information by bounding $H(Y)$ and $-H(Y|X)$ separately. It is easy to upper bound $H(Y)$ via the entropy power inequality:

$$H(Y) \leq k \log \pi e (1 + \mathcal{E}/k).$$

It remains to upper bound $-H(Y|X)$. To that end,

$$\begin{aligned}
&-H(Y|X) \\
&= E[\log(p(Y|X))] \\
&= E[E[\log(p(Y|X)) \mid D = d]] \\
&= \frac{1}{\alpha k} \sum_{d=0}^{\alpha k-1} E \left[\log \left(\frac{1}{\alpha k} \sum_{\ell=0}^{\alpha k-1} \pi^{-k} \exp - \sum_{i=1}^k \left| \sqrt{\frac{\mathcal{E}}{k}} X_{i-d} + Z_i - \sqrt{\frac{\mathcal{E}}{k}} X_{i-\ell} \right|^2 \right) \right] \\
&= -k \log \pi - E \left[\sum_{i=1}^k |Z_i|^2 \right] \\
&\quad + \frac{1}{\alpha k} \sum_{d=0}^{\alpha k-1} E \left[\log \left(\frac{1}{k\alpha} \sum_{\ell=0}^{\alpha k-1} \exp \left(-\mathcal{E} \frac{1}{k} \sum_{i=1}^k |X_{i-d} - X_{i-\ell}|^2 \right. \right. \right. \\
&\quad \left. \left. \left. - 2\sqrt{\mathcal{E}/k} \sum_{i=1}^k \Re(X_{i-d} - X_{i-\ell}) Z_i^* \right) \right) \right] \\
&= -k \log(\pi e) - \frac{1}{\alpha k} \sum_{d=0}^{\alpha k-1} 2\sqrt{\mathcal{E}/k} E \left[\sum_{i=1}^k \Re(X_{i-d} Z_i^*) \right]
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{\alpha k} \sum_{d=0}^{\alpha k-1} E \left[\log \left(\frac{1}{\alpha k} \sum_{\ell=0}^{\alpha k-1} \exp \left(-\mathcal{E} \frac{1}{k} \sum_{i=1}^k |X_{i-d} - X_{i-\ell}|^2 \right. \right. \right. \\
& \quad \left. \left. \left. + 2\sqrt{\mathcal{E}/k} \sum_{i=1}^k \Re(X_{i-\ell} Z_i^*) \right) \right) \right] \\
& \leq -k \log(\pi e) + \frac{1}{\alpha k} \sum_{d=0}^{\alpha k-1} E_X \left[\log \left(\frac{1}{\alpha k} \sum_{\ell=0}^{\alpha k-1} \exp -\mathcal{E} \frac{1}{k} \sum_{i=1}^k |X_{i-d} - X_{i-\ell}|^2 \right. \right. \\
& \quad \left. \left. E_Z \exp 2\sqrt{\mathcal{E}/k} \sum_{i=1}^k \Re(X_{i-\ell} Z_i) \right) \right] \\
& = -k \log(\pi e) + \frac{1}{\alpha k} \sum_{d=0}^{\alpha k-1} E_X \left[\log \left(\frac{1}{\alpha k} \sum_{\ell=0}^{\alpha k-1} \exp -\mathcal{E} \frac{1}{k} \sum_{i=1}^k [|X_{i-d} - X_{i-\ell}|^2 - |X_{i-\ell}|^2] \right) \right]
\end{aligned}$$

Thus

$$\begin{aligned}
I(X; Y) & \leq k \log(1 + \mathcal{E}/k) \\
& + \frac{1}{\alpha k} \sum_{d=0}^{\alpha k-1} E_X \log \left(\frac{1}{\alpha k} \sum_{\ell=0}^{\alpha k-1} \exp -\mathcal{E} \frac{1}{k} \sum_{i=1}^k [|X_{i-d} - X_{i-\ell}|^2 - |X_{i-\ell}|^2] \right) \\
& = \log(1 + \mathcal{E}/k) - \frac{1}{\alpha k} \sum_{d=0}^{\alpha k-1} E_X \frac{\mathcal{E}}{k} \sum_{i=1}^k |X_{i-d}|^2 \\
& + \frac{1}{\alpha k} \sum_{d=0}^{\alpha k-1} E_X \log \left(\frac{1}{\alpha k} \sum_{\ell=0}^{\alpha k-1} \exp \mathcal{E} \frac{1}{k} \sum_{i=1}^k \Re(2X_{i-d} X_{i-\ell}^*) \right) \\
& \leq k \log(1 + \mathcal{E}/k) - \mathcal{E} \\
& + \frac{1}{\alpha k} \sum_{d=0}^{\alpha k-1} E_X \log \left(\frac{1}{\alpha k} \sum_{\ell=0}^{\alpha k-1} \exp \mathcal{E} \frac{1}{k} \sum_{i=1}^k \Re(X_{i-d} X_{i-\ell}^*) \right) \tag{17}
\end{aligned}$$

C Upper Bound on Mutual Information for More General Autocorrelation Function

In this appendix, we will show that an upper bound similar to (7) can be derived for the mutual information for signals with empirical auto-correlation function more general than an impulse.

Consider an input signal with empirical autocorrelation function $C(\cdot, \cdot)$ satisfying:

$$\sum_{\ell} (m - \ell)^2 |C(\ell, m)|^2 < \alpha$$

for all m for some α . Then

$$\begin{aligned}
F_m &= \sum_{\ell} C(D_{\ell}, D_m) G_{\ell} \\
&= C(D_m, D_m) G_m + \sum_{\ell \neq m} C(D_{\ell}, D_m) G_{\ell} \\
&= G_m + \sum_{\ell \neq m} (D_{\ell} - D_m) C(D_{\ell}, D_m) \frac{G_{\ell}}{D_{\ell} - D_m}
\end{aligned}$$

The first term G_m , has zero mean and variance $1/\tilde{L}$. The second term is also zero mean and by using the Cauchy-Schwartz inequality its second moment is upper bounded by

$$\frac{1}{\tilde{L}} \left(\sum_{\ell \neq m} (D_{\ell} - D_m)^2 |C(D_{\ell}, D_m)|^2 \right) \left(\sum_{\ell \neq m} \frac{1}{(D_{\ell} - D_m)^2} \right).$$

Since D_{ℓ} 's are distinct integers, we can upper bound the first term by α and the second term by $\sum_{p \neq q} (p - q)^{-2} \leq \pi^2/3$. which we can further upper bound by $\alpha\pi^2/(3\tilde{L})$. Thus F_m has zero mean and its variance is proportional to $1/\tilde{L}$. Since H_m has second moment $1/\tilde{L}$, we see that $E[U]$ is proportional to $1/tL$.

Let us now look at the second moment of U . Expanding out $|U|^2$ we get

$$E[|U|^2] = \frac{1}{\tilde{L}^2} \sum_{m \neq m'} E[|F_m F_{m'}|^2] + \sum_m E[|H_m|^4 |F_m|^4].$$

The second sum is order $1/\tilde{L}^3$. For the first, we will bound each term:

$$\begin{aligned}
F_m F_{m'} &= \left(G_m + C(D_m, D_m) G_m + \sum_{\ell \notin \{m, m'\}} C(D_{\ell}, D_m) G_{\ell} \right) \left(G_{m'} + C(D_{m'}, D_{m'}) G_{m'} + \sum_{\ell' \notin \{m, m'\}} C(D_{\ell'}, D_{m'}) G_{\ell'} \right) \\
&= G_m G_{m'} (1 + |C(D_m, D_m)|^2) + |G_m|^2 C(D_m, D_{m'}) + |G_{m'}|^2 C(D_{m'}, D_m) \\
&\quad + G_m \sum_{\ell \notin \{m, m'\}} C(D_{\ell}, D_{m'}) G_{\ell} \\
&\quad + G_{m'} \sum_{\ell \notin \{m, m'\}} C(D_{\ell}, D_m) G_{\ell} \\
&\quad + C(D_m, D_{m'}) G_m \sum_{\ell \notin \{m, m'\}} C(D_{\ell}, D_m) G_{\ell} \\
&\quad + C(D_{m'}, D_m) G_{m'} \sum_{\ell \notin \{m, m'\}} C(D_{\ell}, D_{m'}) G_{\ell} \\
&\quad + \sum_{\ell \notin \{m, m'\}} \sum_{\ell' \notin \{m, m'\}} C(D_{\ell}, D_m) C(D_{\ell'}, D_{m'}) G_{\ell} G_{\ell'}
\end{aligned}$$

The first three of the terms that appear in the last equality are proportional to $1/\tilde{L}$. For the next term we can apply the Cauchy-Schwartz inequality to get

$$\left| \sum_{\ell \notin \{m, m'\}} C(D_\ell, D_{m'}) G_\ell \right|^2 \leq \sum_{\ell \notin \{m, m'\}} (D_\ell - D_{m'})^2 |C(D_\ell, D_{m'})|^2 \sum_{\ell \notin \{m, m'\}} \frac{|G_\ell|^2}{(D_\ell - D_{m'})^2}$$

and thus bound its expectation by a quantity proportional to $1/\tilde{L}$ just as in bounding the variance of F_m . The same method works for the next three terms also. For the last term

$$\begin{aligned} E \left| \sum_{(\ell, \ell')} C(D_\ell, D_m) C(D_{\ell'}, D_{m'}) G_\ell G_{\ell'} \right|^2 \\ \leq \sum_{(\ell, \ell')} (D_\ell - D_m)^2 (D_{\ell'} - D_{m'})^2 |C(D_\ell, D_m) C(D_{\ell'}, D_{m'})|^2 E \sum_{(\ell, \ell')} \frac{|G_\ell|^2}{(D_\ell - D_m)^2} \frac{|G_{\ell'}|^2}{(D_{\ell'} - D_{m'})^2} \\ \leq \alpha^2 \frac{\pi^4}{9} E[G_1^4] \end{aligned}$$

where the last step again follows from the fact that D_ℓ 's are distinct integers. Since $E[G_1^4]$ is proportional to $1/\tilde{L}^2$ we conclude that $E[|U|^2]$ is proportional to $1/\tilde{L}$. Together with some constraints on H and G , this is sufficient to make

$$E \exp(\mathcal{E}^2 U) \approx 1 + \text{constant} \mathcal{E}^2 / \tilde{L}$$

from which the mutual information again decays to zero.