Capacity and Mutual Information of Wideband Multipath Fading Channels

I. Emre Telatar, *Member, IEEE,* and David N. C. Tse, *Member, IEEE*

***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 gets large. This is true even when the paths are assumed to be tracked perfectly at the receiver. A critical parameter $\tilde{L}_{\mathrm{crit}}$ is defined in terms of system parameters to delineate the threshold on $\tilde{L}$ over which such over-spreading phenomenon occurs.**

***Index Terms*—Capacity, CDMA, fading channel, noncoherent communications, wideband, wireless communications.**

I. 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 only and the noise is additive white Gaussian the best input signals to use are those that look like samples of white Gaussian noise. However, it is not clear if the channel characteristics can be reliably estimated when such input signals are used, or if accurate estimation of the channel is essential to reliable communication. These issues are particularly pertinent when the signals are spread over a very large bandwidth, as in the proposed third-generation wideband code-division multiple-access (CDMA) systems.

Manuscript received September 13, 1998; revised February 25, 2000. The work of D. N. C. Tse was supported in part by AFOSR under Grant F49620-96-1-0199 and by an NSF CAREER Award under Grant NCR-9734090. Part of the work was done when he was visiting Bell Labs.

I. E. Telatar was with Lucent Technologies, Murray Hill, NJ 07974 USA. He is now with Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland (e-mail: Telatar@epfl.ch).

D. N. C. Tse is with the Department of Electrical Engineering and Computer Sciences, Cory Hall, University of California, Berkeley, CA 94720 USA (e-mail: dtse@eecs.berkeley.edu).

Communicated by S. Shamai, Associate Editor for Shannon Theory.

Publisher Item Identifier S 0018-9448(00)05011-2.

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 $1/W$. The following 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 signal-to-noise ratio (SNR) without fading. Moreover, this can be achieved by frequency shift keying and noncoherent detection.
2) In contrast, if one uses "spread-spectrum" white-like signals (the definition of which will be made precise in Section III), then the mutual information is inversely proportional to the number of resolvable paths $\tilde{L}$, assuming that the energy is divided more or less equally among all resolvable paths and that the path gains are independent. Thus if the number of resolvable paths is large, the mutual information is close to 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 spread-spectrum signals, the mutual information approaches zero with increasing bandwidth even when there is a only a single fixed gain path with random time-varying delay.

The study of the wideband fading channel dates back to the early 1960's. Kennedy has shown that the capacity of an infinite-bandwidth Rayleigh fading channel is the same as that of an infinite bandwidth additive white Gaussian noise (AWGN) channel with the same average received power (see [2, Sec. 8.6], [7]). Our first theorem is a parallel result, applicable to any channel with a finite number of paths.

More recently, Gallager and Médard [4] 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 spread-spectrum input signals are used. Their definition of spread-spectrum signals is that for each fixed frequency band, the energy and fourth moment of the signal in that band scale with $1/W$ and $1/W^2$, respectively, as the bandwidth W increases. 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 a 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 much slower 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 only a finite number of paths is present) takes place only at very large bandwidths.

In addition to the above qualitative conclusions, we also compute 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 $\tilde{L}_{\text{crit}}$ delineates the regime in which *overspreading* occurs. If the number of resolvable paths $\tilde{L}$ is much smaller than $\tilde{L}_{\text{crit}}$, then the mutual information achieved by spread-spectrum signal is close to the capacity of the nonfading white Gaussian noise channel. On the other hand, if $\tilde{L}$ is much larger than $\tilde{L}_{\text{crit}}$, the mutual information achieved is negligibly small.

In other related works, Médard [8] has derived bounds on the impact of channel estimation errors on the achievable mutual information of fading wireless channels. Using a similar channel model as in [4], Médard [9] has obtained explicit upper bound on the achievable mutual information for direct-sequence CDMA signals and showed the similar conclusion that the mutual information goes to zero as the bandwidth gets large. Subsequent to the conference version of this work [11], Hajek and Subramanian [5] have obtained more recent results by applying the theory of capacity and reliability function per unit cost [3], [13] to related problems. By using a certain "fourthegy" of the signal as a cost measure (related to the fourth moment), they showed that the mutual information achieved by spread-spectrum signals in fading channels is small because their fourthegy is small. For a comprehensive survey of other results on fading channels, see [1].

The remainder of the paper is organized as follows. In Section II-B, we introduce a fading channel model and focus on the question of how to achieve the capacity of the channel with only an average power constraint. In Section III, we study the mutual information achieved by wideband spread-spectrum signals, and derive upper and lower bounds as a function of the number of resolvable paths $\tilde{L}$ and other channel parameters. In Section IV, 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 diversity branches grows, 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 $\tilde{L}_{\text{crit}}$ in particular. In Section V, the scenario of single path with time-varying delay is considered. Section VI contains our conclusions.

Unless otherwise stated, the information rates in this paper are in units of nats per second.

II. Capacity of a Multipath Fading Channel

In this section, we will introduce a continuous-time multipath fading channel model, and compute its capacity under an average power constraint.

A. Channel Model

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

$$y(t) = \sum_{\ell=1}^{L} a_\ell(t)x(t - d_\ell(t)) + z(t) \tag{1}$$

where L is the number of paths, $a_\ell(t)$ is the gain of path ℓ at time t, $d_\ell(t)$ 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 begin by identifying a number of key parameters defining the statistics of this channel.

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

$$\sup_{\substack{\ell,s,t:\\ |s-t|\le T_c}} f_c[d_\ell(t) - d_\ell(s)] \ll 1 \tag{2}$$

where f_c is the carrier frequency of the communication system. For simplicity, we will assume that the channel processes $\{a_\ell(t)\}$ and $\{d_\ell(t)\}$ are piecewise-constant, with their values remaining fixed on time intervals $[nT_c, (n+1)T_c)$, $n \in Z$. While typically the channel varies in a more continuous manner, this model greatly simplifies the analysis while capturing the essential idea of channel coherence. We will further assume that $\{a_\ell(nT_c)\}$ and $\{d_\ell(nT_c)\}$ are stationary and ergodic discrete-time stochastic processes, and independent of each other.

The *delay spread* T_d quantifies the uncertainty in the delay of the paths; it satisfies

$$\sup_{\ell,m,t}[d_\ell(t) - d_m(t)] \le T_d. \tag{3}$$

In this paper, we focus on the regime 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 [12].

The average received 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 the SNR 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 [2].

B. Capacity via Frequency-Shift Keying

This section is devoted to proving the following theorem. The proof is based on [2, Sec. 8.6], which proves the analogous result for a Rayleigh fading channel, except that we use a threshold decoding rule which allows us to prove a more general result about the multipath fading channel introduced above.

Theorem 1: The capacity of the multipath fading channel without bandwidth constraint is at least

$$\left(1 - 2\frac{T_d}{T_c}\right)\frac{P}{N_0}.$$

It is known that the capacity of the infinite bandwidth fading channel with perfect channel side information at the receiver (but no channel information at the transmitter) is P/N_0, the capacity of an infinite bandwidth AWGN channel with the same average received SNR. The capacity of the infinite bandwidth fading channel *without* channel side information cannot exceed this. Combining this observation with Theorem 1, it can be concluded that in the regime where $T_d \ll T_c$, the capacity of the multipath fading channel is close to P/N_0.

Proof: Suppose we wish to transmit one of M messages. Let T_s be chosen such that $2T_d < T_s \le T_c$. During this interval T_s, the channel is a linear time-invariant channel at the frequencies of interest. To each message we assign a signal

$$x_m(t) = \begin{cases} \sqrt{\lambda}\exp(\bar{\jmath}2\pi f_m t), & 0 \le t \le T_s \\ 0, & \text{else.} \end{cases}$$

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

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

Over the interval $[T_d, T_s - T_d]$, the processes $\{a_\ell(t)\}$ and $\{d_\ell(t)\}$ are constant by the modeling assumption and that $T_s \le T_c$, and we can write the received signal as

$$\begin{aligned} y(t) &= \sum_{\ell=1}^{L} a_\ell\sqrt{\lambda}\exp(\bar{\jmath}2\pi f_m(t - d_\ell)) + z(t) \\ &= G\sqrt{\lambda}\exp(\bar{\jmath}2\pi f_m t) + z(t) \end{aligned}$$

where

$$G = \sum_{\ell=1}^{L} a_\ell \exp(-\bar{\jmath}2\pi f_m d_\ell)$$

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_k. Namely, the receiver forms

$$R_k = \frac{1}{\sqrt{N_0(T_s - 2T_d)}}\int_{T_d}^{T_s - T_d} \exp(-\bar{\jmath}2\pi f_k t)y(t)\,dt$$

for $1 \le k \le M$. Note that for $k = m$

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

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

$$R_k = W_k$$

where W_k is again a circularly symmetric complex Gaussian random variable with variance 1. Note that because of the orthogonality of the x_k's $\{W_k\}$ 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 $R_{k,n}$ for each possible message $1 \le k \le M$ and each interval $1 \le n \le N$

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

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

$$S_k = \frac{1}{N}\sum_{n=1}^{N} |R_{k,n}|^2$$

and use a threshold rule to decide on a message: if S_k exceeds $A = 1 + (1 - \epsilon)\lambda(T_s - 2T_d)/N_0$ for exactly one value of k, say $\hat{k}$, then it will declare that $\hat{k}$ 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 noncoherent 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 $k\neq m$, its decision variable is given by

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

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

$$\Pr\left[S_k\geq A\right]$$

using a Chernoff bound

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

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_k, $k\neq m$, 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

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

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

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

As θ approaches 0, this expression approaches

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

The desired result follows after noting that ϵ can be chosen arbitrarily small and the symbol duration T_s can be chosen as large as the coherence time T_c. □

In the scheme above, transmission of a symbol is repeated over different time slots to obtain diversity against fading. This diversity can alternatively be obtained by frequency hopping, i.e., repeating the transmission at different frequencies that fade independently. In a slow fading environment, frequency hopping may be more preferable than interleaving over time as less delay needs to be incurred.

III. Mutual Information for White-Like Signals

There are a number of interesting properties of the capacity-achieving scheme described in the previous section. First, at a time scale shorter than the channel coherence time, the input signals are "peaky" in frequency. Transmission is done only on a narrow band. Second, on a longer time scale, the transmitted signal is 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 noncoherent.

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 the mutual information achieved using such signals. We first present a discrete-time memoryless approximation of the continuous-time fading channel model in Section II-A, which we will analyze in this section. Then we give two alternative definitions which attempt to capture the imprecise notion of "white-like spread-spectrum" signals. The main conclusion we will show is that under both of these definitions, 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.

A. Discrete-Time Channel Model

Recall that under the continuous-time model used in Section II-A, the complex gain for path ℓ

$$A_\ell(t)=a_\ell(t)\exp\left(-\jmath 2\pi f_c d_\ell(t)\right)$$

is assumed to be constant over each time interval $[nT_c,(n+1)T_c)$. Here, we will make the further simplifying assumption that each gain takes on independent values in different intervals. 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 now shift to baseband and sample the continuous-time system (1) at a rate of $1/W$. Focusing on the interval $[0,T_c)$, we have the following discrete-time input/output relationship:

$$Y_i=\sqrt{\frac{PT_c}{N_0K_c}}\sum_{\ell=1}^{L}A_\ell X_{i-\tau_\ell}+Z_i,\qquad i=0,\cdots K_c-1 \quad (4)$$

where $K_c=\lfloor WT_c\rfloor$, $\tau_\ell=\lfloor Wd_\ell(t)\rfloor$, 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

$$E\left[\frac{1}{K_c}\sum_{i=0}^{K_c-1}X_i^2\right]\leq 1.$$

Although the channel gains are assumed to be independent in different intervals of length T_c, there is spillover of the input signal across intervals due to the nonzero delay spread. Hence, strictly speaking, the channel is not independent in different intervals. However, in the regime where the delay spread is much smaller than the coherence time, this spillover is negligible. To incorporate this assumption into our model, let us then replace the model (4) by the "circular" approximation

$$Y_i = \sqrt{\frac{PT_c}{N_0 K_c}} \sum_{\ell=1}^{L} A_\ell X_{(i-\tau_\ell)} + Z_i, \qquad i = 0, \cdots K_c - 1 \quad (5)$$

where $(n) \equiv n \bmod K_c$. In this modified model, the output in a coherent interval depends only on the input in the same interval. Note that this approximation is similar in spirit to the N-circular Gaussian channel used in [6].

The sampled delays τ_ℓ's are the actual delays sampled at a resolution of $1/W$. 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 $D_1, \cdots, D_{\tilde{L}}$ 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 (4) as

$$Y_i = \sqrt{\frac{\mathcal{E}}{K_c}} \sum_{\ell=1}^{\tilde{L}} G_\ell X_{(i-D_\ell)} + Z_i, \qquad i = 0, \cdots K_c - 1 \quad (6)$$

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

At this point, we have a discrete-time 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 $W = 10^6$ Hz and $f_c = 10^9$ Hz, then for a transmitter moving at 60 mi/h toward the receiver, it takes about 18 s 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 V, we will consider the situation when path timing is not assumed to be known *a priori*.

We have invoked a few simplifying assumptions and made some approximations to derive this discrete-time model from the original continuous-time model. However, it is now precisely defined and the analysis to be presented in this section will be based solely on this discrete-time model. In this model, the channel, conditional on the path delays, is memoryless from one coherence interval to the next. We will therefore focus on analyzing the achievable mutual information $I(X;Y|D)$ on one such interval. Here, $X := (X_0, \cdots, X_{K_c-1})$, $Y := (Y_0, \cdots, Y_{K_c-1})$, and $D = (D_1, \cdots, D_{\tilde{L}})$.

B. Upper Bound on Mutual Information

We would like to investigate the mutual information $I(X;Y|D)$ achievable by "white-like" spread-spectrum signals $\{X_i\}$. First, we derive an upper bound on the mutual information in terms of the number of resolvable paths $\tilde{L}$.

We first need to make the notion of "white-like" signals precise. One possible candidate is to model $\{X_i\}$ as a sequence of independent and identicaly distributed (i.i.d.) random variables. Thus the input signals look like white noise, and this would be a good approximation in the situation where the transmitted information is spread onto a wide bandwidth by means of error-correction coding alone. In fact, this is a way to approach the capacity of the AWGN channel.

An alternative way of spreading, as is common in direct-sequence CDMA (DS-CDMA) systems, is to modulate the information symbols onto a pseudonoise sequence which has pseudorandom properties but is nevertheless known to the receiver. Typically, one information symbol spans many "chips" of the sequence, and so there is actually strong dependency between consecutive transmitted symbols (the chip values). A more appropriate definition for this class of spread-spectrum signals is to instead impose conditions on the *empirical autocorrelation function* of the signal

$$C(m,n) := \frac{1}{K_c} \sum_{i=0}^{K_c-1} X_{(i-m)} X_{(i-n)},$$

requiring that $C(m,n)$ be close to $\delta(n-m)$. (See, for example, [14] for such properties of pseudonoise sequences.)

The upper bound we present below on the mutual information applies to both these definitions of spread-spectrum signals, and it holds for large spreading bandwidths. To avoid technicalities,[1] the result is proved under the assumption that the gains G_ℓ are uniformly bounded by some constant b.

Theorem 2: Assume that the input process $\{X_i\}$ satisfies either:

1) $\{X_i\}$ is an i.i.d. complex-valued process, such that $E(X_i) = E(X_i^2) = 0$ and $E(|X_i|^4) < \infty$;
   or
2) there exists a universal constant d, not dependent on the bandwidth, such that for any realization of the input process $\{X_i\}$

$$|C(m,n) - \delta(n-m)| \le \frac{d}{K_c} \quad (7)$$

for all m, n.

[1]This assumption can probably be relaxed by a more involved truncation argument.

Then as the bandwidth $W \to \infty$, the following asymptotic upper bound holds:

$$I(X;Y \mid D) \le \sum_{l=1}^{\tilde{L}} \log E \exp(\mathcal{E}^2 |H_\ell|^2 |G_\ell|^2) + O\left(\frac{1}{\sqrt{W}}\right)$$

where H_ℓ has the same distribution as and is independent of G_ℓ.

The essential property of "white-like" signals which we use to prove Theorem 2 is that in both cases, the empirical autocorrelation function $C(m,n) \approx \delta(n-m)$ when the coherence time–bandwidth product is large. The following lemma first provides an upper bound in terms of the empirical autocorrelation function of $\{X_i\}$ for any input process.

Lemma 1: For the channel model (6)

$$\begin{aligned} I(X;Y|D) \le & E_{X,G,D} \\ & \log\left(E_H \exp\left[\frac{2\mathcal{E}}{K_c}\mathfrak{Re}\left\{\sum_{i=0}^{K_c-1}\left(\sum_{\ell=1}^{\tilde{L}} X_{(i-D_\ell)}G_\ell\right)\right.\right.\right. \\ & \left.\left.\left. \times \left(\sum_{m=1}^{\tilde{L}} X_{(i-D_m)}H_m\right)^*\right\}\right]\right) \\ = & E_{X,G,D} \log\left(E_H \exp\left[2\mathcal{E}\sum_{m=1}^{\tilde{L}} \right.\right. \\ & \left.\left. \mathfrak{Re}\left\{H_m\left(\sum_{\ell=1}^{\tilde{L}} G_\ell C(D_\ell, D_m)\right)\right\}\right]\right) \end{aligned} \tag{8}$$

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. □

Proof of Theorem 2: Starting with the upper bound in Lemma 1, we can further bound the mutual information by

$$\begin{aligned} & I(X;Y \mid D) \\ & \le E_{X,G,D} \log\left(E_H \exp\left[2\mathcal{E}\mathfrak{Re}\left\{\sum_{\ell=1}^{\tilde{L}} H_\ell G_\ell\right\}\right.\right. \\ & \quad \left.\left. + \left|\sum_{m=1}^{\tilde{L}}\left\{H_m\left(\sum_{\ell=1}^{\tilde{L}} G_\ell\left[C(D_\ell,D_m) - \delta_{\ell m}\right]\right)\right\}\right|\right]\right) \\ & \le E_G \log\left(E_H \exp\left[2\mathcal{E}\mathfrak{Re}\left\{\sum_{\ell=1}^{\tilde{L}} H_\ell G_\ell\right\}\right]\right) \\ & \quad + 2\mathcal{E}b^2 E_D\left[\sum_{\ell=1}^{\tilde{L}}\sum_{l=1}^{\tilde{L}} E_X\left|C(D_\ell,D_m) - \delta_{\ell m}\right|\right] \end{aligned} \tag{9}$$

using the assumption that $|G_\ell|$ (and hence $|H_\ell|$) is uniformly bounded by the constant b. Consider the first term in (9). Let $H_\ell = |H_\ell| e^{-\jmath\psi_\ell}$ and $G_\ell = |G_\ell| e^{-\jmath\phi_\ell}$. Circular symmetry implies that ϕ's and ψ's are uniform in $[-\pi, \pi]$. For the expectation inside the logarithm in the first term in (9), condition on everything else and take the expectation with respect to the ψ's first. We then get

$$\begin{aligned} & \log E_H \exp\left[2\mathcal{E}\mathfrak{Re}\left\{\sum_{\ell=1}^{\tilde{L}} H_\ell G_\ell\right\}\right] \\ & \qquad = \log E_{|H|} E_\psi \exp\left[2\mathcal{E}\mathfrak{Re}\left\{\sum_{l=1}^{\tilde{L}} e^{-\jmath\psi_\ell}|H_\ell|G_\ell\right\}\right]. \end{aligned}$$

Now

$$E_{\psi_\ell} \exp(\mathfrak{Re}(a e^{\jmath\psi_\ell})) = I_0(2|a|)$$

where I_0 is the 0th-order modified Bessel function of the first kind. Using the inequality $I_0(x) \le \exp(x^2/4)$ we get

$$\begin{aligned} & E_H \exp\left[2\mathcal{E}\mathfrak{Re}\left\{\sum_{\ell=1}^{\tilde{L}} H_\ell G_\ell\right\}\right] \\ & \qquad \le \sum_{\ell=1}^{\tilde{L}} \log E_{|H_\ell|} \exp[\mathcal{E}^2 |H_\ell|^2 |G_\ell|^2]. \end{aligned}$$

Using Jensen's inequality, a bound on the first term in (9) is thus

$$\begin{aligned} & E_G \log\left(E_H \exp\left[2\mathcal{E}\,\mathfrak{Re}\left\{\sum_{\ell=1}^{\tilde{L}} H_\ell G_\ell\right\}\right]\right) \\ & \qquad \le \sum_{\ell=1}^{\tilde{L}} \log E_{|H_\ell|,|G_\ell|} \exp[\mathcal{E}^2 |H_\ell|^2 |G_\ell|^2]. \end{aligned} \tag{10}$$

We now focus on the second term in (9).

Under Assumption 1) on the input sequence $\{X_i\}$

$$E_X|C(k,k)-1| \le \sqrt{E_X[(C(k,k)-1)^2]} = \sqrt{\frac{E[|X_1|^4]-1}{K_c}}$$

and for $m > 0$

$$E_X|C(k,k+m)| \le \sqrt{E_X[|C(k,k+m)|^2]} = \sqrt{\frac{E[|X_1|^4]}{K_c}}$$

and hence the second term in (9) is $O(1/\sqrt{W})$.

It is also easy to see that under Assumption 2), the second term is $O(1/W)$.

Combining this with (10) now yields the theorem. □

The bound in Theorem 2 can be explicitly computed for specific distributions of the path amplitudes. To get more insight, consider the special case when the gain amplitudes $|G_\ell|^2$'s are identically distributed such that $E(|G_\ell|^2) = 1/\tilde{L}$ for all ℓ. We are assuming here that the resolvable paths have equal amount of energy. Then the bound becomes

$$\tilde{L} \log E_{|H_1|,|G_1|} \exp\left(\frac{\mathcal{E}^2}{\tilde{L}^2}|H_1|^2|G_1|^2\right) = \tilde{L} \log g(\mathcal{E}^2)$$

where $g(r)$ is the generating function of $|H_1|^2|G_1|^2$. If the number of distinguishable paths $\tilde{L}$ is large, then the squared amplitude of the gain of each path is small, of order $1/\tilde{L}$. The generating function g is approximately linear, and

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

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

$$\frac{P^2T_c}{N_0^2\tilde{L}}. \tag{11}$$

We observe that this bound is inversely proportional to the number of resolvable paths, and holds when the bandwidth W is large. As the number of equal energy paths gets large, the mutual information goes to zero.

C. Lower Bound on Mutual Information

The upper bound (11) shows that the mutual information goes to zero when the number of resolvable paths becomes large. What happens when the number of resolvable paths is small even though the bandwidth is large? We address this issue by presenting an asymptotic lower bound to the mutual information $I(X;Y|D)$ for a specific white input.

Theorem 3: If the input $\{X_i\}$ is i.i.d. complex circular symmetric Gaussian, then

$$I(X;Y|D) \geq \mathcal{E} - \tilde{L}\log\left(1+\frac{\mathcal{E}}{\tilde{L}}\right) + O\left(\frac{1}{W}\right).$$

Proof: We begin with the following relationships:

$$\begin{aligned} I(X;Y\,|\,D) &= I(Y;X,G\,|\,D) - I(Y;G\,|\,X,D) \\ &\geq I(Y;X\,|\,G,D) - I(Y;G\,|\,X,D) \end{aligned} \tag{12}$$

where the first equality follows from the chain rule. Conditional on the path gains G and the delays D, it can be seen that X and Y are jointly Gaussian. The first term is then given by

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

where A is a K_c by K_c matrix such that $A_{im}=G_\ell$ if $m=(i-D_\ell)$ and 0 otherwise. By the circular convolution properties of the discrete Fourier transform (DFT), the eigenvalues of AA^* are precisely $|F(k/K_c)|^2, k=0\cdots K_c-1$, where

$$F(f) = \sum_{\ell=1}^{\tilde{L}} G_\ell \exp\left(2\pi \jmath D_\ell f\right)$$

is the DFT of the impulse response of the channel. Hence

$$\begin{aligned} I(Y;X\,|\,G,D) &= E_{G,D}\left[\sum_{k=0}^{K_c-1}\log\left(1+\frac{\mathcal{E}}{K_c}\left|F\left(\frac{k}{K_c}\right)\right|^2\right)\right] \\ &= K_cE_{G,D}\log\left(1+\frac{\mathcal{E}}{K_c}|F(0)|^2\right) \\ &= K_cE_G\log\left(1+\frac{\mathcal{E}}{K_c}\left|\sum_{\ell=1}^{\tilde{L}}G_\ell\right|^2\right). \end{aligned} \tag{13}$$

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

We can upper-bound the second term in (12) by making a worst case assumption that the paths gains G_ℓ's are circularly symmetric and Gaussian with the same variance

$$I(Y;G\,|\,X,D) \leq E_{X,D}\log\det\left(I+\frac{\mathcal{E}}{K_c}B\Lambda B^*\right) \tag{14}$$

where $B_{im}=X_{(i-m)}$ and $\Lambda=\mathrm{diag}\,(\sigma_1^2,\cdots,\sigma_{\tilde{L}}^2)$, with $\sigma_\ell^2 = E[|G_\ell|^2]$. Now

$$\begin{aligned} &I(Y;G|X,D) \\ &\leq E_{X,D}\log\det\left(I+\frac{\mathcal{E}}{K_c}B\Lambda B^*\right) \\ &= E_{X,D}\log\det\left(I+\frac{\mathcal{E}}{K_c}B^*B\Lambda\right) \\ &\leq \tilde{L}E_{X,D}\log\left[\frac{1}{\tilde{L}}\,\mathrm{tr}\left(I+\frac{\mathcal{E}}{K_c}B^*B\Lambda\right)\right] \text{(Jensen's inequality)} \\ &= \tilde{L}E_{X,D}\log\left[1+\frac{\mathcal{E}}{\tilde{L}}\sum_{\ell=1}^{\tilde{L}}\sigma_\ell^2\left(\frac{1}{K_c}\sum_{m=0}^{K_c-1}|X_{(m-D_\ell)}|^2\right)\right] \\ &\leq \tilde{L}E_D\log\left[1+\frac{\mathcal{E}}{\tilde{L}}\sum_{\ell=1}^{\tilde{L}}\sigma_\ell^2E_X\left(\frac{1}{K_c}\sum_{m=0}^{K_c-1}|X_{(m-D_\ell)}|^2\right)\right] \\ &\qquad \text{(Jensen's inequality)} \\ &\leq \tilde{L}\log\left(1+\frac{\mathcal{E}}{\tilde{L}}\right). \end{aligned}$$

The last inequality follows from the energy constraint on the input and that $\sum_\ell \sigma_l^2 = 1$. Combining this with (13) yields

$$\begin{aligned} &I(X;Y\,|\,D) \\ &\geq K_cE_G\log\left(1+\frac{\mathcal{E}}{K_c}\left|\sum_{\ell=1}^{\tilde{L}}G_\ell\right|^2\right) - \tilde{L}\log\left(1+\frac{\mathcal{E}}{\tilde{L}}\right) \\ &\geq \mathcal{E}E_G\left|\sum_{\ell=1}^{\tilde{L}}G_\ell\right|^2 - \frac{\mathcal{E}^2}{2K_c}\mathcal{E}_G\left|\sum_{\ell=1}^{\tilde{L}}G_\ell\right|^4 - \tilde{L}\log\left(1+\frac{\mathcal{E}}{\tilde{L}}\right) \\ &\geq \mathcal{E} - \tilde{L}\log\left(1+\frac{\mathcal{E}}{\tilde{L}}\right) - \frac{\mathcal{E}^2}{2K_c}\mathcal{E}_G\left|\sum_{\ell=1}^{\tilde{L}}G_\ell\right|^4 \end{aligned}$$

where the second inequality follows from $\log(1+x)\geq x-(x^2/2)$ for $x\geq 0$.

As $W\to\infty$, $K_c=\lfloor WT_c\rfloor\to\infty$, and the third term in the above approaches 0 like $O(1/W)$. The theorem folllows. □

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*:

$$\frac{P}{N_0} - \frac{\tilde{L}}{T_c}\log\left(1+\frac{PT_c}{N_0\tilde{L}}\right).$$

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

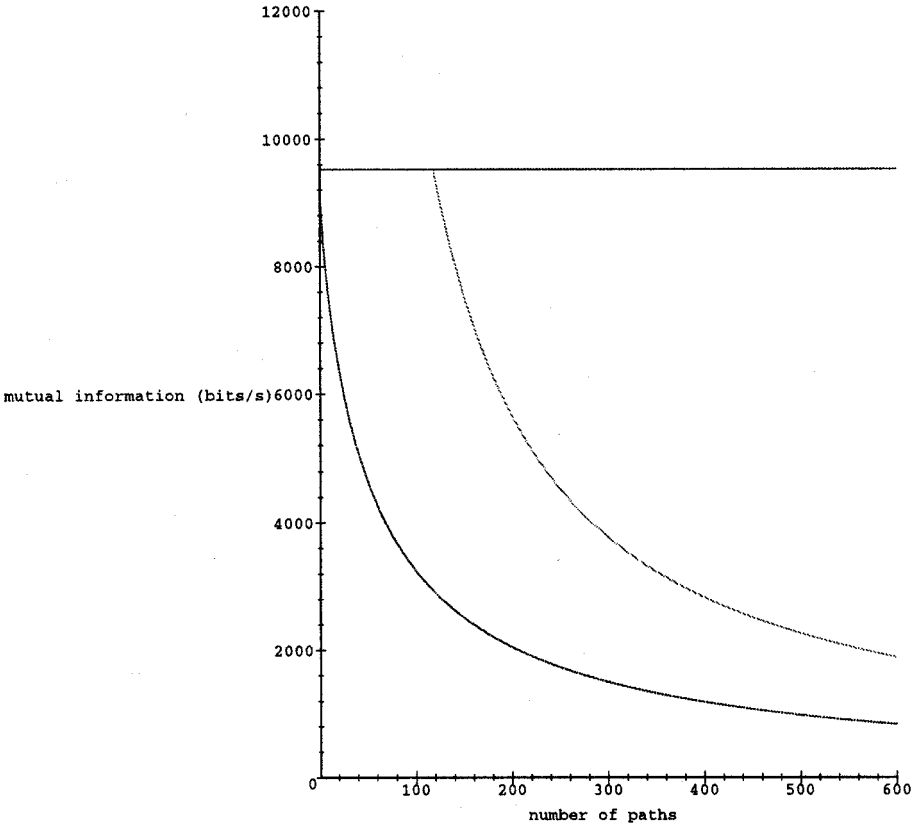

Fig. 1. Upper and lower bounds to the achievable mutual information as a function of the number of resolvable paths, for $T_c = 0.018s$. The unit is in bits per second. The upper horizontal line is the capacity of the AWGN channel.

limitation comes from the number of equal-energy resolvable paths.

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

$$\frac{P^2 T_c}{2N_0^2 \tilde{L}} \tag{15}$$

which approaches zero as $\tilde{L} \to \infty$. Compared to the asymptotic upper bound in (11), we see that the upper and lower bounds agree to within a factor of 2.

If we let

$$\tilde{L}_{\text{crit}} := \frac{PT_c}{N_0} \tag{16}$$

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 $\tilde{L}_{\text{crit}}$ as the *critical parameter* delineating the regime where "over-spreading" occurs. If one thinks of P/N_0 as a nominal information rate, then $\tilde{L}_{\text{crit}}$ is smaller for low-rate users and for systems with shorter coherence time.

At a carrier frequency of 1 GHz and vehicle speed of 60 mi/h, the coherence time is of the order of 18 ms. For a voice user with a nominal AWGN capacity of 9.6 kbits/s, this gives a value of $\tilde{L}_{\text{crit}}$ to be 120. On the other hand, at 10 GHz , the coherence time becomes 1.8 ms, 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.

IV. Detection of Binary Orthogonal Signals

In the previous sections, we studied the information-theoretic properties of broadband multipath channels, focusing on upper and lower bounds on the achievable mutual information in terms

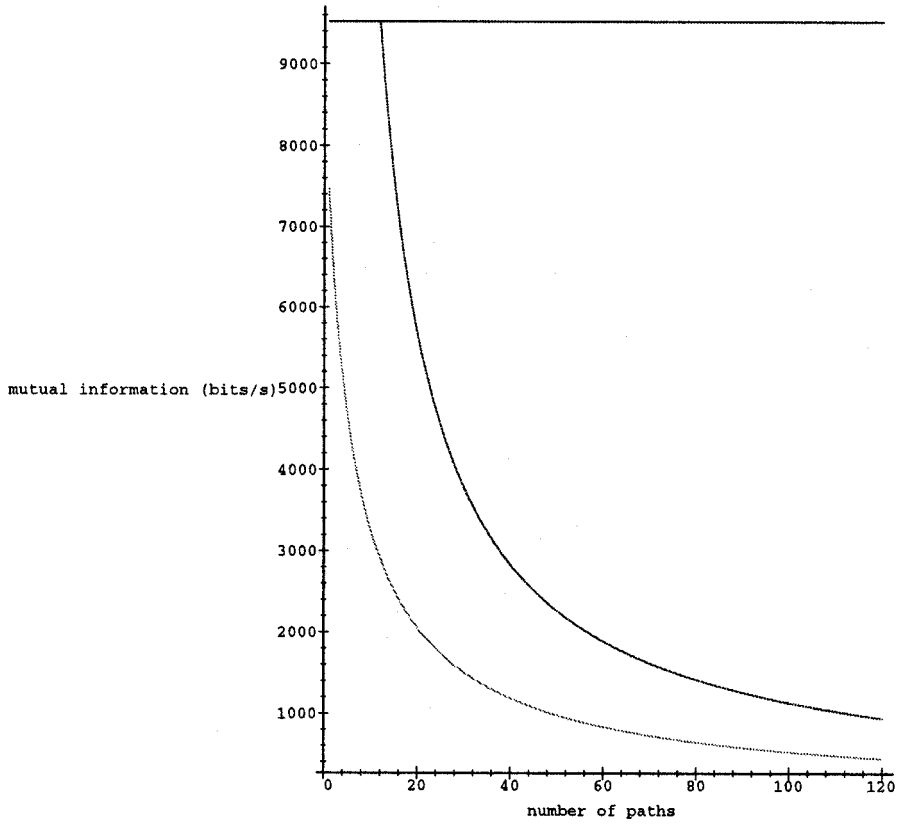

Fig. 2. Upper and lower bounds to the achievable mutual information as a function of the number of resolvable paths, for $T_c = 0.0018s$. The unit is in bits per second. The upper horizontal line is the capacity of the AWGN channel.

of the number of resolvable paths. 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 $\tilde{L}_{\text{crit}}$ in terms of estimation errors in the path gains.

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

$$y(t) = \sum_{\ell=1}^{L} a_\ell(t)x(t - d_\ell(t)) + z(t) \tag{17}$$

where each path has independent statistics.

Consider now an uncoded binary modulation scheme for which 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. A symbol duration much larger than the delay spread means that we can ignore intersymbol interference. A 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 noncoherent detection.

A. 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 II-B.) 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 II-B, we obtain

$$R_k = \begin{cases} \sqrt{\dfrac{\mathcal{E}_b}{N_0}}G + W_0, & k = 0 \\ W_1, & k = 1 \end{cases}$$

where $G = \sum_{\ell=1}^{L} a_\ell \exp(\jmath 2\pi f_0 d_\ell)$ and W_0, W_1 are independent circular symmetric complex Gaussian random variables (rv's) with variance 1. (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 cumulative distribution function (cdf) of an $N(0,1)$ random variable. If we now assume that each path has uniform phase, magnitude a_ℓ such that $E(a_\ell^2) =$

$1/L$, and a Rayleigh distribution, then G is circular symmetric Gaussian with variance 1, and the probability of error, averaged over G, is given by (see, for example, [10, eq. (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 is 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 noncoherent 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 [10, eq. (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 [10, eq. (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 noncoherent detection is worse than that of coherent detection, the performance of the noncoherent detector does not get arbitrarily worse as the number of paths gets large. Its limiting performance depends only on the average SNR.

B. 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, [10, Sec. 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 $N_0/2$, 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 $D_1, \cdots, D_{\tilde{L}}$, assumed known to the receiver. Then if symbol 0 is transmitted and the branches at the baseband are given by

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

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:

$$R_{k\ell} = \begin{cases} \sqrt{\frac{\mathcal{E}_b}{N_0}}G_\ell + W_{0\ell}, & k = 0 \\ W_{1\ell}, & k = 1 \end{cases} \tag{18}$$

where $\{W_{k\ell}\}$ are i.i.d. circular symmetric Gaussian random variables with variance 1. 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, weighing 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 [10, eq. (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 [10, eqs. (7.4.15), (7.4.21)]

$$p_e = \left(\frac{1-\mu}{2}\right)^{\tilde{L}}\sum_{\ell=0}^{\tilde{L}-1}\binom{\tilde{L}-1+\ell}{\ell}\left(\frac{1+\mu}{2}\right)^{\ell} \tag{19}$$

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 $Q\left(\sqrt{\frac{\mathcal{E}_b}{2N_0}}\right)$, i.e., the same as that for a nonfading 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 SNR's. The narrowband scenario corresponds to having one 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 noncoherent 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 $k = 0, 1$

$$U_k = \sum_{\ell=1}^{\tilde{L}}|R_{k\ell}|^2$$

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

$$\Pr[U_1 > U_0] = \Pr\left[\sum_{\ell=1}^{\tilde{L}}|W_{1\ell}|^2 > \sum_{\ell=1}^{\tilde{L}}\left|\sqrt{\frac{\mathcal{E}_b}{N_0}}G_\ell + W_{0\ell}\right|^2\right].$$

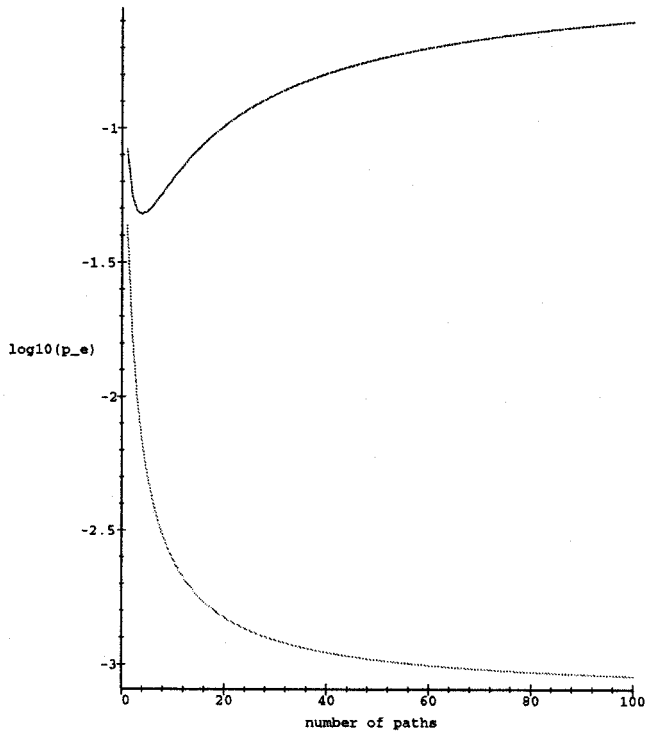

Fig. 3. Comparison of error probability under coherent detection (below) and noncoherent detection (above), as a function of the number of paths. $\mathcal{E}_b/N_0 = 10$ dB.

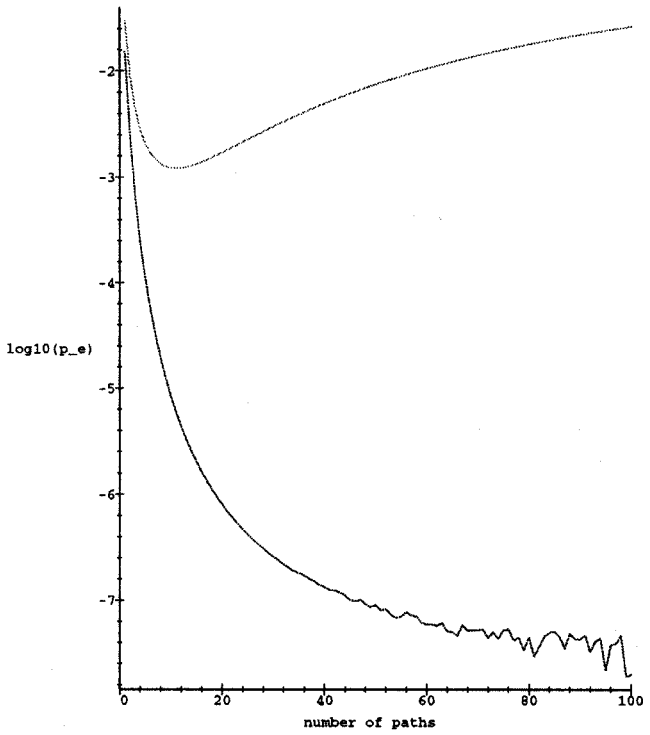

Fig. 4. Comparison of error probability under coherent detection (below) and noncoherent detection (above), as a function of the number of paths. $\mathcal{E}_b/N_0 = 15$ dB.

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

$$E(U_1 - U_0) = -\frac{\mathcal{E}_b}{N_0}$$

and hence

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

Also,

$$\lim_{\tilde{L}\to\infty} \mathrm{Var}\left[\frac{U_1 - U_0}{\sqrt{\tilde{L}}}\right] = 2.$$

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

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

Hence, the probability of error of the noncoherent 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 comparable to the SNR $\mathcal{E}_b/N_0$, then the performance of the noncoherent 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 $\tilde{L}$ [10, eq. (7.4.30)]: it is given by (19) 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 noncoherent detection is plotted as a function of the number of resolvable paths in Figs. 3 and 4 for different SNR's. We see that for small $\tilde{L}$, performance of the 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 noncoherent broadband receiver. There is an optimal $\tilde{L}^*$ after which the performance of the noncoherent broadband detector starts to degrade. As $\tilde{L} \to \infty$, the noncoherent broadband scheme performs even worse than the noncoherent narrowband scheme and in fact the error probability of the former approaches $1/2$.

Observe the contrast in performance scaling of the coherent and noncoherent broadband schemes. A natural question is whether the poor performance scaling of the noncoherent 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 $\hat{G}_\ell$'s instead of G_ℓ. We assume that for each diversity branch $\ell = 1, \cdots, \tilde{L}$, the estimate $\hat{G}_\ell$ is obtained from a set of noisy measurements

$$S_{\ell j} = \sqrt{\frac{\mathcal{E}_p}{N_0}} G_\ell + Z_{\ell j}, \qquad j = 1, \cdots J.$$

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 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 J can be taken to be T_c/T_s, where T_s is the symbol duration. The noise $Z_{\ell j}$'s are taken to be i.i.d. circular symmetric random variables, normalized such that the variance is 1, and also independent of the noise in the interval of the current symbol to be detected.

We employ the linear least-square estimate of G_ℓ; for each ℓ, this is given by

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

The mean-square error associated with this estimate is

$$\frac{1}{J\frac{\mathcal{E}_p}{N_0} + \tilde{L}} \tag{20}$$

same for all branches. The maximal-ratio combiner, using the channel estimates, computes for each $k = 0, 1$

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

where $R_{k\ell}$ is given in (18), and picks the hypothesis with the larger V_l. The probability of error is

$$p_e = \Pr[V_1 > V_0] = \Pr\left[\mathfrak{Re}\left\{ \sum_{\ell=1}^{\tilde{L}} \hat{G}_\ell^* (W_{1\ell} - W_{0\ell}) \right\} > \sqrt{\frac{\mathcal{E}_b}{N_0}} \mathfrak{Re}\left\{ \sum_{\ell=1}^{\tilde{L}} \hat{G}_\ell^* G_\ell \right\} \right].$$

Direct computation yields

$$E(\hat{G}_\ell^* G_\ell) = \frac{J\frac{\mathcal{E}_p}{N_0}}{\tilde{L}\left(\frac{\mathcal{E}_p}{N_0} + \tilde{L}\right)}$$

$$E(G^*(W_{1\ell} - W_{0\ell})) = 0$$

$$\mathrm{Var}[G^*(W_{1\ell} - W_{0\ell})] = \frac{2J\frac{\mathcal{E}_p}{N_0}}{\tilde{L}\left(\frac{\mathcal{E}_p}{N_0} + \tilde{L}\right)}.$$

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

$$\sqrt{\tilde{L}}\, \mathfrak{Re}\left\{ \sum_{\ell=1}^{\tilde{L}} \hat{G}_\ell^* (W_{1\ell} - W_{0\ell}) \right\} \xrightarrow{\mathcal{D}} N(0, J\mathcal{E}_p/N_0).$$

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

$$\sqrt{\tilde{L}} \sqrt{\frac{\mathcal{E}_b}{N_0}}\, \mathfrak{Re}\left\{ \sum_{\ell=1}^{\tilde{L}} \hat{G}_\ell^* G_\ell \right\} \xrightarrow{\mathcal{P}} 0.$$

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

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

Thus if $\tilde{L} \ll J(\mathcal{E}_p/N_0)$, then the performance is very close to that of the coherent receiver with perfect channel estimates. On the other hand, if $\tilde{L} \gg J(\mathcal{E}_p/N_0)$, then the imperfect channel estimates have a significant impact on performance. An intuitive explanation can be seen from (20): if $\tilde{L} \gg J(\mathcal{E}_p/N_0)$, 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 noncoherent receiver: detection becomes impossible.

The critical parameter

$$\tilde{L}_{\mathrm{crit}} := J\frac{\mathcal{E}_p}{N_0}$$

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

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

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 (16) 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 $\tilde{L}_{\mathrm{crit}}$. On the uplink, however, it is not possible to have a common pilot, and the channel estimation will have to be done with a weaker pilot alloted to the individual user, or even noncoherently. With a lower received power from the individual users, $\tilde{L}_{\mathrm{crit}}$ 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 noncoherent 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 correspondingly smaller, it is harder to estimate their gains and to combine them effectively. The fact that the thresh-

olds $\tilde{L}_{\text{crit}}$ identified in both analyses are the same further substantiates this explanation.

V. Timing Uncertainty

In Section III, we showed that as the number of resolvable paths $\tilde{L}$ with equal energy gets large, the mutual information decreases in inverse proportion 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 start with the continuous-time model described in Section II-A, but specialize to the case of single path ($L = 1$) with a fixed gain ($a_1(t) = 1$). We keep the stochastic nature of the delay process $d_1(t)$. We assume that $d_1(t)$ 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 III. 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. Analogous to (4), we will then consider the following baseband discrete-time model for this channel on one coherence interval:

$$Y_i = \sqrt{\frac{PT_c'}{N_0 K_c'}} X_{(i-\tau)} + Z_i, \qquad i = 0, \cdots, K_c' - 1$$

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. Here, $K_c' = WT_c'$. Moreover, the channel is independent from one coherence interval to the next. As in Section III-A, we have normalized the scaling so that $E[|Z_i|^2] = 1$. The random variable τ takes values in $\{1, \cdots, T_d W\}$ and is uniformly distributed on this range. Let $\alpha = T_d/T_c'$. 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 $\{X_i : i = 1, \cdots, K_c'\}$

$$E\left[\frac{1}{K_c'} \sum_{i=0}^{K_c'-1} X_i^2\right] \leq 1.$$

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

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

where

$$C'(m,n) = \frac{1}{K_c'} \sum_{i=0}^{K_c'-1} X_{(i-m)} X_{(i-n)}^*$$

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. □

White input signals are essentially those whose empirical autocorrelation function $C'(m,n) \approx \delta(n-m)$ when the coherence time bandwidth product is large. For these signals, Lemma 2 leads to upper bounds on the achievable mutual information which goes to zero as the bandwidth goes to infinity. The following is a concrete result for the case when $\{X_i\}$ is an i.i.d. process. An analogous result can be obtained for DS-CDMA signals satisfying condition (7).

Theorem 4: If $\{X_i\}$ is an i.i.d. complex-valued process such that $E(X_i) = E(X_i^2) = 0$ and $E(|X_i|^{12}) < \infty$,[2] then as the bandwidth $W \to \infty$, the following asymptotic upper bound holds:

$$I(X;Y) \leq \frac{1}{WT_d}\left[\exp\left(\frac{2PT_c'}{N_0}\right) - 1\right] + O\left(\frac{1}{W^2}\right).$$

Proof: See Appendix C. □

As the bandwidth W becomes large, the upper bound decays to zero like $1/W$. 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.

VI. Conclusion

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 known. 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 $\tilde{L}_{\text{crit}} = PT_c/N_0$ which delineates the threshold on the number of resolvable paths above which this "overspreading" phenomenon occurs: basically, when the SNR per path over a coherence time interval is small.

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 sig-

[2]The latter is a technical condition that can probably be removed.

naling 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.

Appendix A
Proof of Lemma 1

Since

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

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

$$H(Y\,|\,D) \le H(Y) \le k\log\pi e(1+\mathcal{E}/k)$$

since the variance of each of the Y_i's is no greater than $1+\mathcal{E}/k$. It remains to upper-bound $-H(Y\,|\,X,D)$. To simplify notation, let $k = K_c$.

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

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

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

where the inequality follows from Jensen's. Thus we have the expression at the bottom of this page, proving the lemma.

Appendix B
Proof of Lemma 2

To simplify notation, let $k = K_c'$.
Since

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

$$\begin{aligned}
I(X;Y\,|\,D) &\le k\log(1+\mathcal{E}/k) + E_G E_{X,D}\log\Bigg(E_H\exp -\mathcal{E}\frac{1}{k}\sum_{i=0}^{k-1}[|P_i-Q_i|^2 - |Q_i|^2]\Bigg)\\
&= \log(1+\mathcal{E}/k) - E_G E_{X,D}\frac{\mathcal{E}}{k}\sum_{i=0}^{k-1}|P_i|^2 + E_G E_{X,D}\log\Bigg(E_H\exp\mathcal{E}\frac{1}{k}\sum_{i=0}^{k-1}\mathfrak{Re}\,(2P_iQ_i^*)\Bigg)\\
&\le k\log(1+\mathcal{E}/k) - \mathcal{E} + E_G E_{X,D}\log\Bigg(E_H\exp\mathcal{E}\frac{1}{k}\sum_{i=0}^{k-1}\mathfrak{Re}\,(P_iQ_i^*)\Bigg)
\end{aligned}$$

we can bound the mutual information by bounding $H(Y)$ and $-H(Y\,|\,X)$ separately. It is easy to upper-bound $H(Y)$

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

since the variance of each of the Y_i's is no greater than $1+\mathcal{E}/k$. It remains to upper-bound $-H(Y\,|\,X)$. To that end see the expression at the bottom of this page. Thus

$$\begin{aligned} I(X;Y) &\le k\log(1+\mathcal{E}/k) \\ &\quad + \frac{1}{\alpha k}\sum_{d=0}^{\alpha k-1} E_X \log\left(\frac{1}{\alpha k}\sum_{\ell=0}^{\alpha k-1}\exp\right. \\ &\quad \left. - \mathcal{E}\frac{1}{k}\sum_{i=0}^{k-1}\left[\left|X_{(i-d)}-X_{(i-\ell)}\right|^2 - \left|X_{(i-\ell)}\right|^2\right]\right) \\ &= \log(1+\mathcal{E}/k) - \frac{1}{\alpha k}\sum_{d=0}^{\alpha k-1} E_X \frac{\mathcal{E}}{k}\sum_{i=0}^{k-1}\left|X_{(i-d)}\right|^2 \\ &\quad + \frac{1}{\alpha k}\sum_{d=0}^{\alpha k-1} E_X \log\left(\frac{1}{\alpha k}\sum_{\ell=0}^{\alpha k-1}\exp \mathcal{E}\frac{1}{k}\right. \\ &\qquad \left.\times \sum_{i=0}^{k-1}\mathfrak{Re}\left(2X_{(i-d)}X^*_{(i-\ell)}\right)\right) \\ &\le k\log(1+\mathcal{E}/k) - \mathcal{E} \\ &\quad + \frac{1}{\alpha k}\sum_{d=0}^{\alpha k-1} E_X \log\left(\frac{1}{\alpha k}\sum_{\ell=0}^{\alpha k-1}\exp \mathcal{E}\frac{1}{k}\right. \\ &\qquad \left.\times \sum_{i=0}^{k-1}\mathfrak{Re}\left(X_{(i-d)}X^*_{(i-\ell)}\right)\right). \quad (21) \end{aligned}$$

Appendix C
Proof of Theorem 4

Under the i.i.d. assumption on $\{X_i\}$, the empirical autocorrelation function $C'(m,n)$ depends only on the difference $n-m$, and the upper bound in Lemma 2 reduces to

$$E\left[\log\left(\frac{1}{\alpha K'_c}\sum_{\ell=0}^{\alpha K'_c-1}\exp\left\{\frac{\mathcal{E}}{K'_c}\sum_{i=0}^{K'_c-1}\mathfrak{Re}\left[2X_iX^*_{(i-\ell)}\right]\right\}\right)\right].$$

To simplify notation, let $k=K'_c$ and $n=\alpha K'_c$ and

$$S_\ell = \frac{1}{k}\sum_{i=1}^{k}\mathfrak{Re}\left(2X_iX^*_{(i-\ell)}\right), \qquad \ell=0,\cdots,n.$$

Notice that $E[S_0]=2$, and for $\ell>0$, $E[S_\ell]=0$. The upper bound on the mutual information can now be written as

$$(k\log(1+\mathcal{E}/k)-\mathcal{E}) + E\log\left(\frac{1}{n}\sum_{\ell=0}^{n-1}e^{\mathcal{E}S_\ell}\right).$$

The first term for large k behaves like $-\mathcal{E}^2/(2k)$ and we now concentrate on the expectation

$$\begin{aligned} &E\log\left(\frac{1}{n}\sum_{\ell=0}^{n-1}e^{\mathcal{E}S_\ell}\right) \\ &\qquad = \int_{\mathbb{R}^n}\log\left(\frac{1}{n}\sum_{\ell=0}^{n-1}e^{\mathcal{E}s_\ell}\right)dF_k(s_0,\cdots,s_{n-1}) \end{aligned}$$

where F_k is the cumulative distribution function of the random vector $(S_0,\cdots,S_{n-1})$. Let

$$G=\{(s_0,\cdots,s_{n-1}) : \max\{s_0-2,s_1,\cdots,s_{n-1}\}\le 1/\mathcal{E}\}$$

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

and write the integration over $\mathbb{R}^n$ as a sum of two integrals, one over G, the other over $\mathbb{R}^n \setminus G$. We will bound each term separately. For the first, we will use the inequality $e^u \leq 1+u+u^2$ for $u \leq 1$ to get

$$\begin{aligned}
&\int_G \log\left(\frac{1}{n}\sum_{\ell=0}^{n-1} e^{\mathcal{E}s_\ell}\right) dF_k(s_0,\cdots,s_{n-1}) \\
&\leq \int_G \log\left(e^{2\mathcal{E}}\frac{1}{n}(1+\mathcal{E}(s_0-2)+\mathcal{E}^2(s_0-2)^2)\right. \\
&\qquad \left.+\frac{1}{n}\sum_{\ell=1}^{n-1}\left(1+\mathcal{E}s_\ell+\mathcal{E}^2 s_\ell^2\right)\right) dF_k(s_0,\cdots,s_{n-1}) \\
&= \int_G \log\left(1+\frac{1}{n}(e^{2\mathcal{E}}-1+\mathcal{E}e^{2\mathcal{E}}(s_0-2)+e^{2\mathcal{E}}\mathcal{E}^2(s_0-2)^2)\right. \\
&\qquad \left.+\sum_{\ell=1}^{n-1}\mathcal{E}s_\ell+\mathcal{E}^2 s_l^2\right) dF_k(s_0,\cdots,s_{n-1}) \\
&\leq \int_G \frac{1}{n}\left(e^{2\mathcal{E}}-1+e^{2\mathcal{E}}\mathcal{E}(s_0-2)+e^{2\mathcal{E}}\mathcal{E}^2(s_0-2)^2\right. \\
&\qquad \left.+\sum_{\ell=1}^{n-1}\mathcal{E}s_\ell+\mathcal{E}^2 s_l^2\right) dF_k(s_0,\cdots,s_{n-1}) \\
&= \int_G \frac{1}{n}\left(e^{2\mathcal{E}}-1+e^{2\mathcal{E}}\mathcal{E}^2(s_0-2)^2+\sum_{\ell=1}^{n-1}\mathcal{E}^2 s_l^2\right) dF_k \\
&\quad + \int_G \frac{1}{n}\left(e^{2\mathcal{E}}\mathcal{E}(s_0-2)+\sum_{\ell=1}^{n-1}\mathcal{E}s_\ell\right) dF_k.
\end{aligned}$$

In the first integral we can extend the range of integration to $\mathbb{R}^n$ to get an upper bound

$$\frac{1}{n}\left(e^{2\mathcal{E}}-1+e^{2\mathcal{E}}\mathcal{E}^2 E[(S_0-2)^2]+\sum_{\ell=1}^{n-1}\mathcal{E}^2 E\left[S_l^2\right]\right). \quad (22)$$

The second integral equals

$$-\int_{\mathbb{R}^n\setminus G}\frac{1}{n}\left(e^{\mathcal{E}}\mathcal{E}(s_0-2)+\sum_{\ell=1}^{n-1}\mathcal{E}s_\ell\right) dF_k$$

since the means of the random variables $(S_0-2), S_1,\cdots,S_{n-1}$ are zero. We will treat this term shortly. Turning our attention to (22), we evaluate

$$\begin{aligned}
E[(S_0-2)^2] &= \frac{4}{k}(E[|X_1|^4]-1) \\
E\left[S_\ell^2\right] &= \frac{2}{k}, \qquad \ell = 1,\cdots,n-1
\end{aligned}$$

to get an upper bound on (22) of

$$\frac{1}{\alpha k}(e^{2\mathcal{E}}-1)+\frac{\mathcal{E}^2}{k}+O(1/k^2).$$

We will now upper-bound the remaining terms to show that they decay like $O(1/k^2)$. The remaining term is

$$\int_{\mathbb{R}^n\setminus G}\log\left(\frac{1}{n}\sum_{\ell=0}^{n-1}e^{\mathcal{E}S_\ell}\right) dF_k - \int_{\mathbb{R}^n\setminus G}\frac{1}{n}\left(e^{2\mathcal{E}}\mathcal{E}(s_0-2)+\sum_{\ell=1}^{n-1}\mathcal{E}s_\ell\right) dF_k.$$

For the first integral, let us first define $u = \max\{s_0-2, s_1,\cdots,s_{n-1}\}$. and the corresponding random variable $U = \max\{S_0-2, S_1,\cdots,S_{n-1}\}$ with cumulative distribution F_U. Then, the integration region $\mathbb{R}^n \setminus G$ is precisely that for which $u > 1/\mathcal{E}$. Furthermore, the integrand is upper-bounded by $2\mathcal{E}+\mathcal{E}u$. Thus the first integral can be upper-bounded by

$$2\mathcal{E}\Pr(U>1/\mathcal{E})+\mathcal{E}\int_{1/\mathcal{E}}^{\infty} u\, dF_U(u).$$

Integrating by parts, the integral above equals

$$P(U>1/\mathcal{E})+\mathcal{E}\int_{1/\mathcal{E}}^{\infty}\Pr(U>u)\, du.$$

Using the union bound, the probability $\Pr(U>u)$ can be bounded by

$$\Pr(U>u) \leq \Pr(S_0-2>u)+\sum_{\ell=1}^{n-1}\Pr(S_\ell>u).$$

Now note that

$$\Pr(S_0-2>u) \leq E[(S_0-2)^6]/u^6$$
$$\text{and } \Pr(S_\ell>u) \leq E\left[S_\ell^6\right]/u^6.$$

We compute

$$\begin{aligned}
E[(S_0-2)^6] &= \frac{64}{k^6}(kE[(|X_1|^2-1)^6] \\
&\quad +15k(k-1)E[(|X_1|^2-1)^4]E[(|X_1|^2-1)^2] \\
&\quad +20k(k-1)E[(|X_1|^2-1)^3]^2 \\
&\quad +90k(k-1)(k-2)E[(|X_1|^2-1)^2]^3) \\
&\leq 64E[(|X_1|^2-1)^6](90k^{-3}+35k^{-4}+k^{-5}) \\
&\leq 64(1+E[|X_1|^{12}]^{1/6})^6(90k^{-3}+35k^{-4}+k^{-5})
\end{aligned}$$

and, similarly,

$$E\left[S_\ell^6\right] \leq 64E[|X_1|^{12}](90k^{-3}+35k^{-4}+k^{-5}).$$

We thus conclude that

$$P(U>u) \leq \frac{cn}{k^3u^6} = \frac{\alpha c}{k^2u^6}$$

with the constant c depending only on $E[|X_1|^{12}]$. Substituting this into our upper bound, we see that

$$\int_{\mathbb{R}^n\setminus G}\log\left(\frac{1}{n}\sum_{\ell=0}^{n-1}e^{\mathcal{E}S_\ell}\right) dF_k \leq \frac{\alpha c}{k^2}(11\mathcal{E}^7/5+\mathcal{E}^6).$$

We will now show that a similar bound applies to

$$B = -\int_{\mathbb{R}^n\setminus G}\frac{1}{n}\left(e^{\mathcal{E}}\mathcal{E}(s_0-2)+\sum_{\ell=1}^{n-1}\mathcal{E}s_\ell\right) dF_k.$$

Interchanging the sum and the integral, we then bound each term of the sum. For the term

$$\int_{\mathbb{R}^n\setminus G}(-s_\ell)\, dF_k$$

we first bound $-s_\ell$ by $|s_\ell|$, then extend the range of integration to the set for which $v = |\max\{S_0-1, S_1,\cdots,S_{n-1}\}| > 1/\mathcal{E}$.We then see that each term is less than

$$\int_{v>1/\mathcal{E}} v\, dF_k.$$

The term B is thus bounded by $\mathcal{E}e^{2\mathcal{E}}$ times the integral above, and by the same argument as above, this term decays to zero at least as fast as $1/k^2$.

ACKNOWLEDGMENT

The authors wish to thank Prof. Shlomo Shamai and Prof. Robert Gallager for interesting discussions. Comments from the reviewers were very helpful in improving an earlier version.

REFERENCES

[1] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-Theoretic and communications aspects," *IEEE Trans. Information Theory*, vol. 44, pp. 2619–2692, Oct. 1998.

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

[3] ——, "Energy Limited Channels: Coding, Multiaccess and Spread Spectrum," MIT, Cambridge, MA, MIT LIDS Rep. LIDS-P-1714, Nov. 1987.

[4] R. G. Gallager and M. Médard, "Bandwidth scaling for fading channels," in *Proc. Int. Symp. Information Theory*, Ulm, Germany, 1997, p. 471.

[5] B. Hajek and V. G. Subramanian, "Capacity and reliability function per fourth moment cost for WSSUS fading channels," in *Proc. Information Theory Workshop*, Kruger Nat. Park, South Africa, June 1999.

[6] W. Hirt and J. L. Massey, "Capacity of the discrete-time gaussian channel with intersymbol interference," *IEEE Trans. Inform. Theory*, vol. 34, pp. 380–388, May 1988.

[7] R. S. Kennedy, *Fading Dispersive Communication Channels*. New York: Wiley–Interscience, 1969.

[8] M. Médard, "The capacity of time varying multiple user channels in wireless communications," Sc.D. dissertation, MIT, Cambridge, Sept. 1995.

[9] ——, "Bound on mutual information for DS-CDMA spreading over independent fading channels," in *Proc. Asilomar Conf. Signals, Systems and Computers*, Nov. 1997.

[10] J. G. Proakis, *Digital Communications*, 2nd ed. New York: McGraw-Hill, 1989.

[11] I. E. Telatar and D. N. C. Tse, "Capacity and mutual information of broadband multipath channels," in *Proc. Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 395.

[12] G. Turin, "Spread spectrum antimultipath techniques," *Proc. IEEE*, vol. 68, pp. 328–353, Mar. 1980.

[13] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inform. Theory*, vol. 36, pp. 1019–1030, Sept. 1990.

[14] A. J. Viterbi, *CDMA: Principles of Spread Spectrum Communications*. Reading, MA: Addison-Wesley, 1995.