Bandwidth Scaling for Fading Multipath Channels

M. Médard and R. Gallager

Abstract— **We show that very large bandwidths on fading multipath channels cannot be effectively utilized by spread spectrum systems that (in a particular sense) spread the available power uniformly over both time and frequency. The approach is to express the input process as an expansion in an orthonormal set of functions each localized in time and frequency. The fourth moment of each coefficient in this expansion is then uniformly constrained. We show that such a constraint forces the mutual information to 0 inversely with increasing bandwidth. Simply constraining the second moment of these coefficients does not achieve this effect. The results suggest strongly that conventional direct sequence CDMA systems do not scale well to extremely large bandwidths. To illustrate how the interplay between channel estimation and symbol detection affects capacity, we present results for a specific channel and CDMA signaling scheme.**

Keywords— **Fading multipath, Channel capacity, Broadband communication, CDMA, Wireless.**

I. Introduction

The objective of this paper is to help understand the effect of increasing the available bandwidth for channels subject to both additive white Gaussian noise (WGN) and multipath fading. We describe our model for fading multipath channels precisely later, but in essence we are considering a classical scattering model (i.e., a channel with no specular component and with finite time and frequency coherence). We also assume no feedback and no side information about the channel state. For WGN channels without fading, it is well known that, with a power constraint S and with noise spectral density $N_0/2$, the capacity, in natural units per second, as a function of available bandwidth $\mathcal{W}$, is $\mathcal{W}\ln[\infty + \mathcal{S}/(\mathcal{N}\!,\mathcal{W})]$. This increases with $\mathcal{W}$ to the limit S/N_0. This infinite bandwidth capacity can be approached arbitrarily closely by a set of orthogonal equal energy waveforms, and it makes no difference what set of orthogonal waveforms are used. A set of orthogonal time limited sine waves, a set of non-overlapping pulses, or a set of orthogonal pseudo-noise waveforms are all equivalent in terms of probability of decoding error.

For WGN fading multipath channels, there is an old, rather surprising, result due to Kennedy (see, for example, [14], [26], [8], [32], [1]) saying that the infinite bandwidth capacity of the channel is the same as the infinite bandwidth capacity of the non-fading WGN channel of the same average received power. This result differs from the non-fading result in two important ways. First, in the fading case, the infinite bandwidth result is approached impractically slowly with increasing $\mathcal{W}$. Second, although the infinite bandwidth result can be approached with orthogonal codewords for the fading case, the results appear to depend critically on the particular choice of orthogonal set. For example, orthogonal sinusoids of increasingly high power and low duty cycle (so as to remain inside the average power constraint) work, but sinusoids with constant average power do not.

Fading multipath channels filter the input with a response that varies slowly both with time and frequency shifts of the input. Because of these shifts, it is insightful to use an expansion for the signal space in which the available bandwidth is separated into fixed slices of bandwidth, using the sampling theorem to represent the baseband representation of each slice by an orthonormal expansion (with complex coefficients) of normalized sinc functions. The relationship between the slices in this expansion is explained in section V. Representing waveforms by such an expansion, the channel becomes a discrete time channel where each discrete time input corresponds

Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, medard@mit.edu. gallager@mit.edu

to a given time/frequency slot. Note that using these expansions does not constrain the choice of signaling waveforms except for the overall bandwidth constraint.

The capacity of a fading WGN channel is equal to the maximum average mutual information per unit time over the above discrete time channel, modeling the bandwidth constraint by the number of frequency slices available. The major result of this paper is to show that if a particular type of fourth moment constraint is placed on the input variables for this channel, then the maximum mutual information is significantly degraded for large $\mathcal{W}$, in fact approaching 0 at least as $\mathcal{W}^{-\infty}$ as $\mathcal{W} \to \infty$. Coding theorems and converses ([13], [23]) apply to these mutual informations in much the same way as with more conventional channels, so, in what follows, we deal exclusively with mutual informations.

With a bandwidth $\mathcal{W}$, there are $\mathcal{W}$ complex input random variables per second. With a power constraint S, the average second moment constraint on these input variables (which need not be uniformly applied) is $S/\mathcal{W}$. The fourth moment constraint above is then, for any finite constant α, to constrain the fourth moment of each complex input variable to be at most $\alpha S^2/\mathcal{W}^{\in}$. With such a constraint, we show that the average mutual information per unit time approaches 0 as $\mathcal{W}^{-\infty}$. Note that if the input variables are IID Gaussian, with independent real and imaginary parts, then the fourth moment of each complex variable is $2S^2/\mathcal{W}^{\in}$, so the above result applies with $\alpha = 2$. If we want to maintain or increase mutual information with increasing $\mathcal{W}$, it is necessary for the input random variables to either become increasingly non-uniform or to become increasingly "peaky" in distribution.

The results here were first presented in [10]. Some related results for a memoryless fading model were presented in [18], [19], using WGN signals and DS-CDMA signals, respectively. Later related results appear in [33], [29], [30], [31]. In [30], the results in this paper are considered using an output fourth order constraint and a capacity per unit cost approach ([34]). In [31], a peak signal constraint going to 0 is considered. In [33], the type of results in this paper are shown to depend strongly on having no specular component in the channel multipath, and the effect of specular components is analyzed. These results also depend critically on our assumption of finite frequency coherence, which rules out flat fading models as treated in [16]. Finally, for a perspective and survey of known results for a wide variety of models of fading channels, see [5].

The main reason for interest in this result lies in its application to direct sequence code division multiple access (CDMA) systems. The CDMA standard (IS-95) is one of the major systems currently deployed for commercial wireless telephony, and CDMA is particularly attractive for future integrated systems. CDMA is also being deployed in ultrawideband systems as overlays over licensed spectrum. As shown in [35], [11], [27], [15], CDMA has many system strengths for cellular systems, taking advantage of the idle periods in voice and data, and being robust to out of cell interference and noise other than WGN. Also, as shown in [22], [7], [28], [24], CDMA has many system strengths for military systems, including anti-jam and low-probability-of-intercept capabilities. CDMA signals, however, closely resemble WGN over the available spectrum. Thus, for the orthonormal expansion above, the input variables are reasonably modelled as IID Gaussian. The above results then imply that the mutual information per unit time approaches 0 with increasing bandwidth. The bandwidths of current commercial systems are sufficiently small that the above limiting regime is not entered. The results explain, however, why the bandwidth of a pure CDMA system cannot be expanded arbitrarily over time varying channels. This result is particularly relevant in light of some recent developments in the area of ultrawideband systems using CDMA signaling, indicating that such systems should be used only on channels which vary very slowly.

Some intuition about why signals resembling white noise are not very effective at increasingly large bandwidths can be seen by considering a rake receiver for a fading channel. The rake receiver both measures the channel and makes data decisions. The data decisions are made on the basis of the current channel estimate, and then the current decision is used to update the channel estimate. As the bandwidth increases, the power available in any given bandwidth slice decreases. Thus, the accuracy of the channel measurement in that bandwidth slice degrades and also the signal to noise ratio degrades. Because of the combined effect, the mutual information per degree of freedom decreases inversely with the square of the overall bandwidth. We will see in what follows that this effect is independent of the particular receiver structure and depends only on the fourth moment constraint α above. Our results depend on having no side information, so that the channel state estimation depends only on input and output. [25] analyzes a situation with a different assumption about estimate and estimation error.

II. A model for fading multipath channels

First we will look at a system in which the input is bandlimited to some fixed bandwidth W'. Because of Doppler spread, the output bandwidth W will be slightly larger than the input bandwidth W'. We represent the output $\mathcal{Y}(\sqcup)$ as a complex baseband process of bandwidth $W/2$, and using the sampling theorem, we represent it as a complex sampled time process $\ldots Y_{-1}, Y_0, Y_1, \ldots$ with samples at rate W. In particular,

$$
\begin{aligned}
Y_i &= \int_{-\infty}^{\infty} \mathcal{Y}(\sqcup) \frac{\sin[\pi \mathcal{W}(\sqcup - \rangle/\mathcal{W})]}{\pi\sqrt{\mathcal{W}}(\sqcup - \rangle/\mathcal{W})} \lceil \sqcup \\
\mathcal{Y}(\sqcup) &= \sum_i Y_i \frac{\sin[\pi W(t - i/W)]}{\pi\sqrt{W}(t - i/W)} \qquad (1)
\end{aligned}
$$

Even though the input is bandlimited to a smaller band than W, we can still represent the input by the corresponding complex sampled time process $\ldots X_{-1}, X_0, X_1, \ldots$ at rate W. After analyzing this system of fixed bandwidth, we then look at bandwidths that are integer multiples of W'. We will show how to view these broader band systems as combinations made up of multiple slices each of input bandwidth W'.

The channel multipath fading is represented by a randomly time varying linear filter whose impulse response is limited to some multipath time spread T_0. The effect of this filter on the input over the given band can be represented as a complex, time varying, tapped delay line filter with L complex taps at intervals of $1/W$. L must be at least WT_0 because of the effective bandlimiting of the filter impulse response, but the exact value of L is non-critical in the arguments to follow. Let $F_{i,j}$ be the j^{th} tap of the filter at discrete output time i. Thus, the signal, corrupted by the multipath fading but before the addition of noise, is given at time i by $U_i = \sum_{m=0}^{L-1} X_{i-m} F_{i,m}$.

We denote $(F_{i,0}, F_{i,1}, \ldots, F_{i,L-1})^{\mathrm{T}}$ as a random vector $\vec{F}_i$. The sample value of this vector is called the channel state at time i. We assume that the vector stochastic process $\ldots, \vec{F}_{-1}, \vec{F}_0, \vec{F}_1, \ldots$ is zero mean, stationary, and complex Gaussian. We also assume that this process is statistically independent of the input process $\ldots X_{-1}, X_0, X_1, \ldots$. This assumption implies that the source uses no side information about the state of the channel, and thus, for example, assumes that power control is not used. Power control, as used in practice, decreases the average information rate, since it increases power when the channel is badly faded. Even if power control were used to increase the rate (with all the attendant system problems), it would be ineffective if the channel state could not be well estimated at the receiver. We conjecture, for this reason, that CDMA, with the above fourth moment constraint, still breaks down at very high bandwidths even when power control is used.

As a notational convenience, define $\vec{X}_i = (X_i, X_{i-1}, \ldots, X_{i-L+1})^{\mathrm{T}}$. Thus the faded signal at time i is given by $U_i = \vec{X}_i^{\mathrm{T}} \vec{F}_i$. The additive noise, over the bandwidth W, is represented at baseband by a discrete time, zero mean, complex Gaussian process $\ldots Z_{-1}, Z_0, Z_1, \ldots$. The complex random variables Z_i are independent and identically distributed (IID), and are independent of the input process and multipath fading process. The output, at each discrete time i, is given by

$$
Y_i = U_i + Z_i = \vec{X}_i^{\mathrm{T}} \vec{F}_i + Z_i \qquad (2)
$$

A set of complex random variables $Z_1, Z_2, \ldots, Z_n$, and the corresponding random vector $\vec{Z}^n = (Z_1, \ldots, Z_n)^{\mathrm{T}}$, is said to be *circularly sym-*

metric if, for any angle ϕ, the variables $e^{j\phi}Z_1, e^{j\phi}Z_2, \ldots, e^{j\phi}Z_n$ have the same joint distribution as $Z_1, Z_2, \ldots, Z_n$. A complex stochastic process (or set of complex stochastic processes) is *circularly symmetric* (or *jointly circularly symmetric*) if each finite set of complex random variables within the process (or processes) is circularly symmetric. It is physically almost inevitable to assume that the noise process, $\ldots Z_{-1}, Z_0, Z_1, \ldots$ is circularly symmetric and that the multipath fading process $\ldots, \vec{F}_{-1}, \vec{F}_0, \vec{F}_1, \ldots$ is circularly symmetric. Since these processes are independent, they are also jointly circularly symmetric. Finally, conditional on any given input $\ldots, X_{-1}{=}x_{-1}, X_0{=}x_0, X_1{=}x_1, \ldots,$ we see that, for any i, $\vec{x}_i^{\,\mathrm{T}}\vec{F}_i$ is circularly symmetric and thus Y_i is circularly symmetric. More generally, conditional on a given input sequence, the output process $\{Y_i\}$ is circularly symmetric and $\{Y_i\}$, $\{Z_i\}$, and $\{\vec{F}_i\}$ are jointly circularly symmetric.

For some large but finite sequence length n, let $\vec{X}^{\,n} = (X_{2-L}, \ldots, X_n)^{\mathrm{T}}$ and $\vec{Y}^{\,n} = (Y_1, \ldots, Y_n)^{\mathrm{T}}$. Our first objective is to find a useful upper bound to the average mutual information, $I(\vec{X}^{\,n}; \vec{Y}^{\,n})$, over the given band. The reason for including $X_{2-L}, \cdots, X_0$ will become apparent later, but has little effect for large n. Define the number β by,

$$\sum_{j=-\infty}^{\infty} \left\{ \sum_{k=0}^{L-1} \sum_{m=0}^{L-1} \left| \mathsf{E}\left[F_{0,k}F_{j,m}^{*}\right] \right| \right\}^2 = \beta \qquad (3)$$

The term inside the braces is a form of correlation between time 0 and time j, and thus β, suitably normalized, is proportional to the the time coherence. We assume β is finite, as a precise characterization of our assumption of finite time coherence. Also assume a finite fourth moment constraint γ such that

$$\mathsf{E}\left[|X_i|^4\right] \le \gamma \quad ; \qquad 2-L \le i \le n \qquad (4)$$

We then develop the following upper bound on $I(\vec{X}^{\,n}; \vec{Y}^{\,n})$:

Theorem 1: Let a discrete time multipath fading channel have output $Y_i = \vec{X}_i^{\,\mathrm{T}}\vec{F}_i + Z_i$ for the input process $\{X_i\}$, fading process $\{\vec{F}_i\}$, and noise process $\{Z_i\}$ defined above. Then, for any positive integer n such that (4) is satisfied,

$$\frac{1}{n} I(\vec{X}^{\,n}; \vec{Y}^{\,n}) \le \frac{\gamma\beta}{2\sigma_Z^4} \qquad (5)$$

where $\sigma_Z^2 = \mathsf{E}\left[|Z_i|^2\right]$.

This theorem will be proven in the next section. It is valid for all distributions on the input, subject to the constraints above. Note that the theorem contains no explicit constraint on $\mathsf{E}\left[|X_i|^2\right]$, although (4) implicitly implies that $\mathsf{E}\left[|X_i|^2\right] \le \sqrt{\gamma}$. To understand what theorem 1 is saying more clearly, define the *Kurtosis*, α, of a zero mean random variable X to be $\alpha = \mathsf{E}\left[|X^4|\right]/[\mathsf{E}\left[|X^2|\right]]^2$. If X is antipodal, the Kurtosis is 1 and, if complex Gaussian, it is 2. As a more insightful example, if X is 0 with probability $1-p$ and 1 or -1 with probability $p/2$ each, then $\alpha = 1/p$. Thus a random variable with large Kurtosis has a "peaky" distribution. If we constrain each of the inputs in (5) to have a Kurtosis of at most some arbitrary number α and a mean square value at most some number ϵ, then γ can be expressed as $\alpha\epsilon^2$.

We now express this in more familiar signal to noise ratio terms. The channel above has W complex degrees of freedom per second, and thus a power constraint S_W can be met by the constraint $\mathsf{E}\left[|X_i|^2\right] \le S_W/W$. Thus, taking $\epsilon = S_W/W$,

$$\gamma = \frac{\alpha S_W^2}{W^2} \qquad (6)$$

Finally, let $N_0/2$ be the spectral density of the noise. Then $\sigma_Z^2 = N_0$. Substituting this plus (6) in (5), we get

$$\frac{1}{n} I(\vec{X}^{\,n}; \vec{Y}^{\,n}) \le \frac{\alpha S_W^2 \beta}{2W^2 N_0^2} \qquad (7)$$

In the above argument, we have used an energy constraint on each degree of freedom to motivate the relation in (6) between γ and S_W. However, as stated before, the theorem is valid whether or not there is an explicit constraint on signal to noise ratio.

Suppose we view a broadband system with power constraint S as some number b of frequency slices, each with power constraint $S_W =$

S/b. If we assume for the moment that each frequency slice is independent and satisfies (7), then the average mutual information per unit time per slice goes down as $1/b^2$, and the aggregate mutual information over the entire band then approaches 0 as $1/b$.

As will be explained in section V, the slicing interpretation above is over simplified, and we must take into account both the Doppler shift and the correlation in fading between different frequency slices. The problem caused by the Doppler shift is that adjacent frequency slices at the input give rise to overlapping slices at the output. The problem with correlated frequency slices is that the aggregate of average mutual information over several slices might be greater than the sum of the average mutual informations over the individual slices. However, after being careful about these issues, we shall still find that the average mutual information rate goes to zero with increasing bandwidth $\mathcal{W} = \lfloor \mathcal{W}'$ if the fourth moments are bounded as above.

The problems caused by statistical dependence between the fading on different frequency slices are quite tricky and depend critically on the model of multipath fading. The model we have adopted here is a classical scattering model, corresponding to a continuum of infinitesimal paths. A different model, using a finite number m of time varying paths, has been investigated by Telatar and Tse [33]. When they assume that the delay of each path is known (but the amplitude and phase is not), then the mutual information does not approach 0 with increasing $\mathcal{W}$, but rather is inversely proportional to the number of resolvable paths (which are upper bounded by m). They also consider the case in which path delay is unknown. Here they show that the mutual information approaches 0 as $1/\mathcal{W}$ but the bound becomes meaningful only at extremely large $\mathcal{W}$. Neither classical scattering models nor finite path models are completely satisfactory for modelling reflecting surfaces and other such physical multipath mechanisms. All of these models, however, are close enough to physical wireless media to provide some guidance on wideband future systems.

The analysis in the present paper relies heavily on the particular way the input is scaled with increasing $\mathcal{W}$. This type of scaling does not apply to frequency hopping, since, as the set of available frequencies for hopping increases, the fraction of time that a frequency is used decreases. Consequently, as in the example above, the Kurtosis increases as $1/\mathcal{W}$. It also does not apply to the increasingly "peaky" type of distribution used to achieve capacity on fading channels with no bandwidth constraint. This scaling does apply to CDMA type systems, and helps explain why very broadband systems tend to use a combination of frequency hopping and CDMA rather than CDMA alone.

III. Mutual information for a fixed frequency band; Proof of theorem 1

We begin the proof of theorem 1 with some standard relations between expected mutual information and differential entropy. We will then establish a couple of lemmas, and finally complete the proof of the theorem. First, note that

$$I(\vec{X}^{\,n};\vec{Y}^{\,n}) = I(\vec{Y}^{\,n};\vec{X}^{\,n}) = \sum_{i=1}^{n} I(Y_i;\vec{X}^{\,n} \mid \vec{Y}^{\,i-1})$$

$$I(Y_i;\vec{X}^{\,n} \mid \vec{Y}^{\,i-1}) = h(Y_i \mid \vec{Y}^{\,i-1}) - h(Y_i \mid \vec{Y}^{\,i-1},\vec{X}^{\,n}) \tag{8}$$

where I denotes expected mutual information and h denotes differential entropy. Information and entropy for complex random variables and vectors are, by definition, the information and entropy for the joint real and imaginary parts of those complex variables and vectors.

First look at the differential conditional entropy $h(Y_i \mid \vec{Y}^{\,i-1},\vec{X}^{\,n})$ for given sample values $\vec{Y}^{\,i-1}{=}\vec{y}^{\,i-1}$ and $\vec{X}^{\,n}{=}\vec{x}^{\,n}$, i.e., we look at $h(Y_i \mid \vec{y}^{\,i-1},\vec{x}^{\,n})$. Conditional on $\vec{x}^{\,n}$, the random vectors $\vec{Y}^{\,n},\vec{F}_1,\vec{F}_2,\ldots,\vec{F}_n$ are jointly Gaussian, and, as explained earlier, jointly circularly symmetric. The covariance matrix of a zero mean complex random vector $\vec{F}$ is defined to be $\mathrm{K}_{\vec{\mathrm{F}}} = \mathsf{E}\left[\vec{\mathrm{F}}\vec{\mathrm{F}}^{\,*\mathrm{T}}\right]$. A useful property of arbitrary zero mean jointly Gaussian, circularly symmetric random vectors, say $\vec{Y}$ and $\vec{F}$, is that, conditional on some given value $\vec{y}$ for $\vec{Y}$, the conditional distribution for $\vec{F}$ has a mean value $\widehat{F}$ given by

$$\widehat{F}(\vec{y}) = \mathsf{E}\left[\vec{F}\vec{Y}^{*\mathrm{T}}\right] \mathrm{K}_{\vec{\mathrm{Y}}}^{-1}\vec{\mathrm{y}} \tag{9}$$

The conditional fluctuation, $\widetilde{F} = \vec{F} - \widehat{F}$, given $\vec{Y} = \vec{y}$, is zero mean, Gaussian, and circularly symmetric. Its covariance function, $\mathsf{E}\left[\widetilde{F}\widetilde{F}^{*\mathrm{T}}|\vec{y}\right]$, is not a function of $\vec{y}$ and is given by

$$\begin{aligned} &\mathsf{E}\left[\widetilde{F}\widetilde{F}^{*\mathrm{T}}|\vec{y}\right] \\ =\ &\mathsf{E}\left[\vec{F}\vec{F}^{*\mathrm{T}}\right] - \mathsf{E}\left[\vec{F}\vec{Y}^{*\mathrm{T}}\right]\mathrm{K}_{\vec{\mathrm{Y}}}^{-1}\,\mathsf{E}\left[\vec{\mathrm{Y}}\vec{\mathrm{F}}^{*\mathrm{T}}\right] \\ =\ &\mathsf{E}\left[\vec{F}\vec{F}^{*\mathrm{T}}\right] - \mathsf{E}\left[[\hat{F}(\vec{Y})\hat{F}^{*\mathrm{T}}(\vec{Y})]^2\right] \end{aligned} \tag{10}$$

$\widehat{F}$ is the minimum mean square error (MMSE) estimate of $\vec{F}$ given $\vec{Y}=\vec{y}$, and $\widetilde{F}$ is the negative of the estimation error; (9) - (10) are well known formulas of elementary estimation theory.

For the application here, we use $\vec{F}_i$ for $\vec{F}$ and $\vec{Y}^{\,i-1}$ for $\vec{Y}$, with additional conditioning on the input, $\vec{X}^{\,n} = \vec{x}^{\,n}$. With this conditioning, $\vec{F}_i$ and $\vec{Y}^{\,i-1}$ are zero mean, jointly Gaussian, and circularly symmetric, so the conditional mean and covariance of $\vec{F}_i$, given both $\vec{X}^{\,n} = \vec{x}^{\,n}$ and $\vec{Y}^{\,i-1} = \vec{y}^{\,i-1}$ are

$$\widehat{F}_i(\vec{x}^{\,n}, \vec{y}^{\,i-1}) = \mathsf{E}\left[\vec{F}_i\vec{Y}^{(i-1)*\mathrm{T}}|\vec{x}^{\,n}\right] K_{\vec{\mathrm{Y}}|\vec{x}}^{-1}\,\vec{y}^{\,i-1} \tag{11}$$

$$\begin{aligned} \mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}}) = &\ \mathsf{E}\left[\vec{\mathrm{F}}_{\mathrm{i}}\vec{\mathrm{F}}_{\mathrm{i}}^{*\mathrm{T}}\right] - \\ &\mathsf{E}\left[\vec{F}_i\vec{Y}^{(i-1)*\mathrm{T}}|\vec{x}^{\,n}\right]\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}}^{-1}\,\mathsf{E}\left[\vec{\mathrm{Y}}^{\,\mathrm{i}-1}\vec{\mathrm{F}}_{\mathrm{i}}^{*\mathrm{T}}|\vec{\mathrm{x}}^{\mathrm{n}}\right] \end{aligned} \tag{12}$$

where $\mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}}) = \mathsf{E}\left[\widetilde{\mathrm{F}}_{\mathrm{i}}\widetilde{\mathrm{F}}_{\mathrm{i}}^{*\mathrm{T}}|\vec{\mathrm{x}}^{\mathrm{n}}, \vec{\mathrm{y}}^{\mathrm{i}-1}\right]$, $\widetilde{F}_i = \vec{F}_i - \widehat{F}_i$, and $\mathrm{K}_{\vec{\mathrm{Y}}^{\mathrm{i}-1}|\vec{\mathrm{x}}^{\mathrm{n}}}$ is abbreviated $\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}}$. The first term on the right of (12) is not conditioned on $\vec{x}^n$ since $\vec{F}_i$ and $\vec{X}^{\,n}$ are independent. In what follows, we call $\hat{F}_i$ and $\mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}})$ the *idealized* estimate and error covariance, since $\vec{x}^n$ is unknown at the receiver and thus $\hat{F}_i$ can not be measured there.

Lemma 1: Let $\widehat{F}_i(\vec{x}^{\,n}, \vec{y}^{\,i-1})$ and $\mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}})$ be the idealized estimate and covariance of $\vec{F}_i$ as given in (11, 12). Then

$$h(Y_i \mid \vec{x}^{\,n}, \vec{Y}^{\,i-1}) = \ln\left[\pi e\left(\sigma_Z^2 + \vec{x}_i^{\mathrm{T}}\mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}})\,\vec{\mathrm{x}}_{\mathrm{i}}^*\right)\right] \tag{13}$$

$$\begin{aligned} I(Y_i; X^n \mid \vec{Y}^{i-1}) \ \leq\ &\frac{\mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\widehat{F}_i(\vec{X}^{\,n}, \vec{Y}^{i-1})\right|^2\right]^6}{\sigma_Z^2} \\ +\ &\frac{\mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\,\mathsf{E}\left[\vec{F}_i\vec{F}_i^{*\mathrm{T}}\right]\vec{X}_i^*\right|^2\right]}{2\sigma_Z^4} \end{aligned} \tag{14}$$

Proof: Let $\widehat{Y}_i = \mathsf{E}\left[Y_i \mid \vec{x}^{\,n}, \vec{y}^{\,i-1})\right]$ and $\widetilde{Y}_i = Y_i - \widehat{Y}_i$. Since $Y_i = \vec{x}_i^{\mathrm{T}}\vec{F}_i + Z_i$, we have $\widehat{Y}_i = \vec{x}_i^{\mathrm{T}}\widehat{F}_i$ and thus $\widetilde{Y}_i = \vec{x}_i^{\mathrm{T}}\widetilde{F}_i + Z_i$. We can then calculate the conditional variance of Y_i directly in terms of the idealized conditional covariance matrix, $\mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}})$, of $\widetilde{F}_i$. In particular,

$$\mathsf{E}\left[\widetilde{Y}_i\widetilde{Y}_i^* \mid \vec{y}^{\,i-1}, \vec{x}^{\,n}\right] = \vec{x}_i^{\mathrm{T}}\mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}})\vec{\mathrm{x}}_{\mathrm{i}}^* + \sigma_{\mathrm{Z}}^2 \tag{15}$$

Consider the differential conditional entropy, $h(Y_i \mid \vec{y}^{\,i-1}, \vec{x}^{\,n})$. Since differential entropy is invariant to translation, this is equal to $h(\widetilde{Y}_i \mid \vec{y}^{\,i-1}, \vec{x}^{\,n})$. We have seen that $\widetilde{Y}_i$, conditional on $\vec{X}^{\,n}=\vec{x}^{\,n}$ and $\vec{Y}^{\,i-1}=\vec{y}^{\,i-1}$, is Gaussian and circularly symmetric. Because of the circular symmetry, the real and imaginary parts of $\widetilde{Y}_i$ are independent, Gaussian, and equally distributed. Thus the combined entropy in natural units, (taking account of real and imaginary parts) is

$$h(Y_i \mid \vec{x}^{\,n}, \vec{y}^{\,i-1}) = \ln\left[\pi e\ \mathsf{E}\left[\widetilde{Y}_i\widetilde{Y}_i^* \mid \vec{x}^{\,n}, \vec{y}^{\,i-1}\right]\right] \tag{16}$$

Substituting (15) into (16), and recognizing (from the RHS of (15) and (12)) that the covariance in (15) does not depend on $\vec{y}^{\,i-1}$, we have (13), proving the first part of the lemma. Note that this conditional entropy is expressed directly in terms of the idealized error covariance matrix, $\mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}})$.

Next, we find a lower bound to $h(Y_i \mid \vec{x}^{\,n}, \vec{Y}^{\,n-1})$ that can be easily averaged over $\vec{X}^{\,n}$. From (13),

$$\begin{aligned} &h(Y_i \mid \vec{x}^{\,n}, \vec{Y}^{i-1}) \\ =\ &\ln\left[\pi e\sigma_Z^2\right] + \ln\left[1 + \frac{\vec{x}_i^{\mathrm{T}}\mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}})\,\vec{\mathrm{x}}_{\mathrm{i}}^*}{\sigma_Z^2}\right] \qquad (17) \\ \geq\ &\ln\left[\pi e\sigma_Z^2\right] + \frac{\vec{x}_i^{\mathrm{T}}\mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}})\,\vec{\mathrm{x}}_{\mathrm{i}}^*}{\sigma_{\mathrm{Z}}^2} - \frac{[\vec{x}_i^{\mathrm{T}}\mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}})\,\vec{\mathrm{x}}_{\mathrm{i}}^*]^2}{2\sigma_Z^4} \end{aligned}$$

We used the inequality $\ln(1 + u) \geq u - u^2/2$ for

$u \geq 0$ above. We can now take the expected value over $\vec{X}^n$.

$$\begin{aligned} & h(Y_i \mid \vec{X}^n, \vec{Y}^{i-1}) \\ \geq \;& \ln\left[\pi e \sigma_Z^2\right] + \frac{\mathsf{E}\left[\vec{X}_i^{\mathrm{T}} \mathrm{B_i}(\vec{\mathrm{X}}^{\mathrm{n}})\, \vec{\mathrm{X}}_{\mathrm{i}}^{*}\right]}{\sigma_Z^2} \\ - \;& \frac{\mathsf{E}\left[\left(\vec{X}_i^{\mathrm{T}} \mathrm{B_i}(\vec{\mathrm{X}}^{\mathrm{n}})\, \vec{\mathrm{X}}_{\mathrm{i}}^{*}\right)^2\right]}{2\sigma_Z^4} \qquad (18) \\ \geq \;& \ln\left[\pi e \left(\sigma_Z^2 + \mathsf{E}\left[\vec{X}_i^{\mathrm{T}} \mathrm{B_i}(\vec{\mathrm{X}}^{\mathrm{n}})\, \vec{\mathrm{X}}_{\mathrm{i}}^{*}\right]\right)\right] \\ - \;& \frac{\mathsf{E}\left[\left(\vec{X}_i^{\mathrm{T}} \mathrm{B_i}(\vec{\mathrm{X}}^{\mathrm{n}})\, \vec{\mathrm{X}}_{\mathrm{i}}^{*}\right)^2\right]}{2\sigma_Z^4} \qquad (19) \end{aligned}$$

where we used the inequality $u \geq \ln(1+u)$ on the middle term of (18).

Next we need to upper bound $h(Y_i \mid \vec{Y}^{i-1})$ in (8). Breaking Y_i into its real and imaginary parts, we have

$$h(Y_i \mid \vec{Y}^{i-1}) \leq h(Y_i) \leq h(\Re(Y_i)) + h(\Im(Y_i))$$

We have seen that Y_i, conditional on $\vec{X}_i = \vec{x}_i$, is circularly symmetric, and it is then not hard to see that Y_i, unconditionally, is also circularly symmetric (although typically not Gaussian). It follows that $h(\Re(Y_i))$ and $h(\Im(Y_i))$ are equal, and each is upper bounded by the Gaussian entropy of the same variance. Thus

$$h(Y_i \mid \vec{Y}^{i-1}) \leq \ln\left\{\pi e \;\mathsf{E}\left[|Y_i|^2\right]\right\} \qquad (20)$$

We finally bound $\mathsf{E}\left[|Y_i|^2\right]$. Conditional on $\vec{X}^n{=}\vec{x}^n$ and $\vec{Y}^{i-1}{=}\vec{y}^{i-1}$, we have seen that Y_i is Gaussian with mean $\vec{x}_i^{\mathrm{T}} \widehat{F}_i(\vec{x}^n, \vec{y}^{i-1})$ and with variance given by (15). It follows that

$$\begin{aligned} & \mathsf{E}\left[Y_i Y_i^* \mid \vec{x}^n, \vec{y}^{i-1}\right] = \\ & \left|\vec{x}_i^{\mathrm{T}} \widehat{F}_i(\vec{x}^n, \vec{y}^{i-1})\right|^2 + \vec{x}_i^{\mathrm{T}} \mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}})\, \vec{\mathrm{x}}_{\mathrm{i}}^{*} + \sigma_{\mathrm{Z}}^2 \qquad (21) \end{aligned}$$

Taking the expected value over $\vec{X}^n$ and $\vec{Y}^{i-1}$,

$$\begin{aligned} & \mathsf{E}\left[Y_i Y_i^*\right] = \\ & \mathsf{E}\left[\,|\vec{X}_i^{\mathrm{T}} \widehat{F}_i(\vec{X}^n, \vec{Y}^{i-1})|^2 + \vec{X}_i^{\mathrm{T}} \mathrm{B_i}(\vec{\mathrm{X}}^{\mathrm{n}})\, \vec{\mathrm{X}}_{\mathrm{i}}^{*} + \sigma_{\mathrm{Z}}^2\right] \qquad (22) \end{aligned}$$

Substituting this into (20), we get

$$\begin{aligned} & h(Y_i \mid \vec{Y}^{i-1}) \\ \leq \;& \ln\left\{\pi e \;\mathsf{E}\left[|\vec{X}_i^{\mathrm{T}} \widehat{F}_i(\vec{X}^n, \vec{Y}^{i-1})|^2\right.\right. \\ + \;& \left.\left.\vec{X}_i^{\mathrm{T}} \mathrm{B_i}(\vec{\mathrm{X}}^{\mathrm{n}})\, \vec{\mathrm{X}}_{\mathrm{i}}^{*} + \sigma_{\mathrm{Z}}^2\right]\right\} \qquad (24) \end{aligned}$$

Substituting this and (19) into (8), we get

$$\begin{aligned} & I(Y_i; \vec{X}^n \mid \vec{Y}^{i-1}) \\ \leq \;& \ln\left[1 + \frac{\mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}} \widehat{F}_i(\vec{X}^n, \vec{Y}^{i-1})\right|^2\right]}{\sigma_Z^2 + \mathsf{E}\left[\vec{X}_i^{\mathrm{T}} \mathrm{B_i}(\vec{\mathrm{X}}^{\mathrm{n}})\, \vec{\mathrm{X}}_{\mathrm{i}}^{*}\right]}\right] \\ + \;& \frac{\mathsf{E}\left[\left(\vec{X}_i^{\mathrm{T}} \mathrm{B_i}(\vec{\mathrm{X}}^{\mathrm{n}})\, \vec{\mathrm{X}}_{\mathrm{i}}^{*}\right)^2\right]}{2\sigma_Z^4} \qquad (25) \end{aligned}$$

The first term on the right side of (25) can be interpreted as the mutual information of a WGN channel for which first, the channel strength is the idealized estimate of the channel, and for which, second, the noise is the original additive noise plus the idealized covariance. This interpretation helps to support our intuition that the mutual information is intimately tied to channel measurement (even though the idealized estimate can not be determined by the receiver).

To complete the proof of the lemma, use the upper bound $\ln(1+v) \leq v$ in the first term of (25) and drop the extra term in the denominator,

$$\begin{aligned} I(Y_i; \vec{X}^n \mid \vec{Y}^{i-1}) \;\leq\; & \frac{\mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}} \widehat{F}_i(\vec{X}^n, \vec{Y}^{i-1})\right|^2\right]}{\sigma_Z^2} \\ + \;& \frac{\mathsf{E}\left[\left(\vec{X}_i^{\mathrm{T}} \mathrm{B_i}(\vec{\mathrm{X}}^{\mathrm{n}})\, \vec{\mathrm{X}}_{\mathrm{i}}^{*}\right)^2\right]}{2\sigma_Z^4} \qquad (26) \end{aligned}$$

To upper bound the numerator of the final term in (26), note, from (12), that $\mathsf{E}\left[\vec{F}_i \vec{F}_i^{*\mathrm{T}}\right] - \mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}})$ is non-negative definite for each $\vec{x}^n$. Thus,

$$\vec{x}_i^{\mathrm{T}} \mathrm{B_i}(\vec{\mathrm{x}}^{\mathrm{n}})\, \vec{\mathrm{x}}_{\mathrm{i}}^{*} \leq \vec{\mathrm{x}}_{\mathrm{i}}^{\mathrm{T}} \;\mathsf{E}\left[\vec{\mathrm{F}}_{\mathrm{i}} \vec{\mathrm{F}}_{\mathrm{i}}^{*\mathrm{T}}\right] \vec{\mathrm{x}}_{\mathrm{i}}^{*}$$

$$\mathsf{E}\left[\left(\vec{X}_i^{\mathrm{T}}\mathrm{B_i}(\vec{\mathrm{X}}^{\,\mathrm{n}})\,\vec{\mathrm{X}}_\mathrm{i}^*\right)^2\right] \le \ \mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_i^{*\mathrm{T}}\right]\vec{X}_i^*\right|^2\right] \tag{27}$$

Note that $\mathsf{E}\left[\vec{F}_i\vec{F}_i^{*\mathrm{T}}\right]$ is a (non-random) matrix here, and the outer expectation in (27) is thus over $\vec{X}_i$. Substituting (27) into (26) gives us (14), completing the proof of the lemma.

The following lemma is of interest in its own right, since (30) bounds the mutual information in terms of the fourth moment of the pre-noise output $U_i = X_i^{\mathrm{T}}\vec{F}_i$. (29) is slightly stronger, and is needed to complete the proof of theorem 1.

Lemma 2: Given the conditions of theorem 1,

$$I(\vec{X}^{\,n};\vec{Y}^{\,n}) \le \tag{28}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\mathsf{E}\left[\left|{}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_j^{*\mathrm{T}}\right]\vec{X}_j^*\right|^2\right]}{2\sigma_Z^4} \tag{29}$$

$$I(\vec{X}^n;\vec{Y}^n) \le \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\mathsf{E}\left[|U_iU_j^*|^2\right]}{2\sigma_Z^4} \tag{30}$$

Proof: First, we upper bound the numerator of the first term on the right side of (14). We first look at a fixed sample value, $\vec{X}^{\,n} = \vec{x}^{\,n}$ for the input, and bound $\mathsf{E}\left[|\vec{x}_i^{\mathrm{T}}\widehat{F}_i(\vec{x}^n,\vec{Y}^{i-1})|^2\right]$, where the expectation is over the output $\vec{Y}^{i-1}$, conditional on $\vec{x}^n$. For this $\vec{x}^n$, abbreviate $\hat{\vec{Y}}^{\,i-1}$ by $\vec{Y}$, and $\widehat{F}_i(\vec{x}^{\,n},\vec{Y}^{\,i-1})$ by $\widehat{F}_i$. Thus, we want to upper bound

$$\mathsf{E}\left[\left|\vec{x}_i^{\mathrm{T}}\widehat{F}_i\right|^2\right] = \vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\widehat{F}_i\widehat{F}_i^{*\mathrm{T}}\right]\vec{x}_i^* \tag{31}$$

From (11), $\widehat{F}_i$ is given by $\mathsf{E}\left[\vec{F}_i\vec{Y}^{*\mathrm{T}} \mid \vec{x}^n\right]\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}}^{-1}\vec{\mathrm{Y}}$, where $\vec{Y}$ is an abbreviation for $\vec{Y}^{i-1}$. We then have

$$\begin{aligned}
&\mathsf{E}\left[\widehat{F}_i\widehat{F}_i^{*\mathrm{T}}\right] = \\
&\mathsf{E}\left[\vec{F}_i\vec{Y}^{\,*\mathrm{T}} \mid \vec{x}^n\right]\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}}^{-1}\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}}\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}}^{-1}\ \mathsf{E}\left[\vec{\mathrm{Y}}\vec{\mathrm{F}}_\mathrm{i}^{*\mathrm{T}}\right] \\
&= \ \mathsf{E}\left[\vec{F}_i\vec{Y}^{*\mathrm{T}} \mid \vec{x}^n\right]\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}}^{-1}\ \mathsf{E}\left[\vec{\mathrm{Y}}\vec{\mathrm{F}}_\mathrm{i}^{*\mathrm{T}}\right]
\end{aligned} \tag{32}$$

$$\begin{aligned}
&\vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\widehat{F}_i\widehat{F}_i^{*\mathrm{T}}\right]\vec{x}_i^* \\
= \ &\vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{Y}^{*\mathrm{T}}\right]\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}}^{-1}\ \mathsf{E}\left[\vec{\mathrm{Y}}\vec{\mathrm{F}}_\mathrm{i}^{*\mathrm{T}}\right]\vec{\mathrm{x}}_\mathrm{i}^*
\end{aligned} \tag{33}$$

Recall that $Y_j = U_j + Z_j$ where $U_j = \vec{x}_j^{\,\mathrm{T}}\vec{F}_j$.[8] Thus $\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}} = \mathrm{K}_{\vec{\mathrm{U}}} + \sigma_\mathrm{Z}^2\mathrm{I}_{\mathrm{i}-1}$, where $\mathrm{I}_{\mathrm{i}-1}$ is the $i-1$ dimensional identity matrix. It follows that $\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}} - \sigma_\mathrm{Z}^2\mathrm{I}_{\mathrm{i}-1}$ is non-negative definite for each input $\vec{x}^n$. From this, we see that each eigenvalue of $\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}}$ must be greater than or equal to σ_Z^2, and thus, each eigenvalue of $\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}}^{-1}$ must be less than or equal to σ_Z^{-2}. This in turn means that $\sigma_Z^{-2}\mathrm{I}_{\mathrm{i}-1} - \mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}}^{-1}$ must be non-negative definite. It follows that for any complex vector $\vec{v}$,

$$\vec{v}^{\mathrm{T}}\mathrm{K}_{\vec{\mathrm{Y}}|\vec{\mathrm{x}}}^{-1}\vec{\mathrm{v}}^* \le \sigma_\mathrm{Z}^{-2}\,\|\vec{\mathrm{v}}\|^2$$

Taking $\vec{v}^{\,\mathrm{T}}$ as $\vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{Y}^{*\mathrm{T}} \mid \vec{x}^n\right]$, (33) is bounded by

$$\vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\widehat{F}_i^{\mathrm{T}}\widehat{F}_i^{*\mathrm{T}}\right]\vec{x}_i^* \le \sigma_Z^{-2}\left\|\vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{Y}^{\,*\mathrm{T}} \mid \vec{x}^n\right]\right\|^2 \tag{34}$$

Observe that $\vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{Y}^{\,*\mathrm{T}} \mid \vec{x}^n\right]$ is an $i-1$ dimensional row vector whose j^{th} component is given by

$$\begin{aligned}
\vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_iY_j^* \mid \vec{x}^n\right] \ &= \ \vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i(\vec{x}_j^{\,\mathrm{T}}\vec{F}_j + Z_j)^*\right] \\
&= \ \vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_j^{*\mathrm{T}}\right]\vec{x}_j^*
\end{aligned} \tag{35}$$

Thus it follows that

$$\left\|\vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{Y}^{\,*\mathrm{T}}\right]\right\|^2 \ = \ \sum_{j=1}^{i-1}\left|\vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_j^{*\mathrm{T}}\right]\vec{x}_j^*\right|^2 \tag{36}$$

Combining (34) and (36) and taking the expected value over $\vec{X}^n$,

$$\begin{aligned}
&\mathsf{E}\left[\vec{X}_i^{\mathrm{T}}\ \mathsf{E}\left[\widehat{F}_i^{\mathrm{T}}\widehat{F}_i^{*\mathrm{T}}\right]\vec{X}_i^*\right] \\
\le \ &\sigma_Z^{-2}\sum_{j=1}^{i-1}\ \mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_j^{*\mathrm{T}}\right]\vec{X}_j^*\right|^2\right]
\end{aligned} \tag{37}$$

Combining (37) with (31), and substituting this plus (27) into (26),

$$I(Y_i;\vec{X}^{\,n} \mid \vec{Y}^{\,i-1}) \le$$

$$\sum_{j=1}^{i-1}\left[\frac{\mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_j^{*\mathrm{T}}\right]\vec{X}_j^*\right|^2\right]}{\sigma_Z^4}\right]$$

$$+ \quad \frac{\mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_i^{*\mathrm{T}}\right]\vec{X}_i^*\right|^2\right]}{2\sigma_Z^4} \tag{38}$$

Summing over i, we get

$$\begin{aligned} & I(\vec{X}^n;\vec{Y}^n) \\ \leq \ & \sum_{i=1}^{n}\left\{\sum_{j=1}^{i-1}\left[\frac{\mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_j^{*\mathrm{T}}\right]\vec{X}_j^*\right|^2\right]}{\sigma_Z^4}\right]\right. \\ & \left. + \frac{\mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_i^{*\mathrm{T}}\right]\vec{X}_i^*\right|^2\right]}{2\sigma_Z^4}\right\} \end{aligned} \tag{39}$$

Finally, denote $\mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_j^{*\mathrm{T}}\right]\vec{X}_j^*\right|^2\right]$ by $g(i,j)$ and note that $g(i,j)=g(j,i)$. Thus, we have

$$\begin{aligned} \sum_{i=1}^{n}\sum_{j=1}^{i-1} g(i,j) &= \sum_{j=1}^{n}\sum_{i=j+1}^{n} g(i,j) \\ &= \sum_{i=1}^{n}\sum_{j=i+1}^{n} g(i,j) \end{aligned} \tag{40}$$

The first two expressions are equal since both sum over i and j such that $i<j$. The final two are equal by interchanging i and j and using $g(i,j)=g(j,i)$. By replacing the first sum in (39) with half that sum, plus half the sum over $i>j$, we get (29). To establish (30), note that for each sample value of the input, $|\vec{x}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_j^{*\mathrm{T}}\right]\vec{x}_j^*|^2 \leq \mathsf{E}\left[|\vec{x}_i^{\mathrm{T}}\vec{F}_i\vec{F}_j^{*\mathrm{T}}\vec{x}_j^*|^2\right]$. Thus,

$$\begin{aligned} & \mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_j^{*\mathrm{T}}\right]\vec{X}_j^*\right|^2\right] \leq \\ & \mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\vec{F}_i\vec{F}_j^{*\mathrm{T}}\vec{X}_j^*\right|^2\right] = \mathsf{E}\left[\left|\vec{U}_i^{\mathrm{T}}\vec{U}_j^*\right|^2\right] \end{aligned}$$

Proof of theorem 1: To complete the proof of theorem 1, we expand each of the fourth moment terms in (29) as follows:

$$\begin{aligned} & \mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_j^{*\mathrm{T}}\right]\vec{X}_j^*\right|^2\right] = \\ & \mathsf{E}\left[\left|\sum_k\sum_m X_{i-k}\ \mathsf{E}\left[F_{i,k}F_{j,m}^*\right]X_{j-m}^*\right|^2\right] \\ \leq \ & \sum_k\sum_m\sum_{k'}\sum_{m'}\left|\ \mathsf{E}\left[F_{i,k}F_{j,m}^*\right]\right|\ \left|\ \mathsf{E}\left[F_{i,k'}F_{j,m'}^*\right]\right| \\ & \mathsf{E}\left[\left|X_{i-k}X_{j-m}^*\right|\left|X_{i-k'}X_{j-m'}^*\right|\right] \end{aligned} \tag{41}$$

The expected values here can now be upper bounded by repeated applications of the Cauchy inequality,

$$\begin{aligned} & \mathsf{E}\left[\left|X_{i-k}X_{j-m}^*\right|\left|X_{i-k'}X_{j-m'}^*\right|\right] \\ \leq \ & \left\{\ \mathsf{E}\left[\left|X_{i-k}X_{j-m}^*\right|^2\right]\ \mathsf{E}\left[\left|X_{i-k'}X_{j-m'}^*\right|^2\right]\right\}^{1/2} \\ \leq \ & \left\{\ \mathsf{E}\left[|X_{i-k}|^4\right]\ \mathsf{E}\left[|X_{j-m}|^4\right]\right. \\ & \left.\ \mathsf{E}\left[|X_{i-k'}|^4\right]\ \mathsf{E}\left[|X_{j-m'}|^4\right]\right\}^{1/4} \\ \leq \ & \gamma \end{aligned} \tag{42}$$

where we used (4) for the final bound on the expected value of the fourth power of each input variable. Substituting (42) into (41),

$$\begin{aligned} & \mathsf{E}\left[\left|\vec{X}_i^{\mathrm{T}}\ \mathsf{E}\left[\vec{F}_i\vec{F}_j^{*\mathrm{T}}\right]\vec{X}_j^*\right|^2\right] \\ \leq \ & \gamma\sum_k\sum_m\sum_{k'}\sum_{m'}\left|\ \mathsf{E}\left[F_{i,k}F_{j,m}^*\right]\right| \\ & \left|\ \mathsf{E}\left[F_{i,k'}F_{j,m'}^*\right]\right| \\ = \ & \gamma\left[\sum_k\sum_m\left|\ \mathsf{E}\left[F_{i,k}F_{j,m}^*\right]\right|\right]^2 \end{aligned} \tag{43}$$

Substituting this into (29),

$$I(\vec{X}^n;\vec{Y}^n) \leq \gamma\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\left[\sum_k\sum_m\left|\ \mathsf{E}\left[F_{i,k}F_{j,m}^*\right]\right|\right]^2}{2\sigma_Z^4}$$

This can be further upper bounded by summing j from $-\infty$ to $+\infty$. Because of the stationarity of F, each term in the sum over i is then the same, and equal to β as given in (3). Thus, $I(\vec{X}^n;\vec{Y}^n) \leq n\gamma\beta/(2\sigma_Z^4)$, completing the proof.

IV. Interpretation of theorem 1

To get the simplest interpretation of our result, assume that for each j, $\mathsf{E}\left[F_{0,k}F_{j,m}^*\right]=0$ for $k\neq m$. This is reasonable since $F_{j,k}$ is the response at time j to the set of paths whose delay is approximately k/W. We expect these path responses to be uncorrelated with those paths at some other delay m/W. With this assumption, β in (3) simplifies to

$$\beta = \sum_{j=-\infty}^{\infty} \left[\sum_{k=0}^{L-1} \mid \mathsf{E}\left[F_{0,k}F_{j,k}^*\right] \mid \right]^2 \qquad (44)$$

One of the standard ways [3] of representing fading multipath channels is by the *two frequency correlation function*, $\mathcal{R}(\{,\sqcup)$, defined as the correlation between the response at time 0 to a sinusoid of unit power at the carrier frequency c and the response at time t to a sinusoid of unit power at $c+f$. This correlation function at $f=0$ is simply the correlation function of the response to a unit sinusoid at the carrier frequency. In our baseband representation, this unit sinusoid is simply $X_i = 1/\sqrt{W}$ for all i. Thus,

$$\mathcal{R}(\prime,|/\mathcal{W}) = \mathcal{W}\ \mathsf{E}\left[\mathcal{U}_\prime\mathcal{U}_|^*\right] = \ \mathsf{E}\left[\sum_{\|=\prime}^{\mathcal{L}-\infty} \mathcal{F}_{\prime,\|}\mathcal{F}_{|,\|}^*\right] \qquad (45)$$

If we assume for the moment either that $\mathsf{E}\left[F_{0,k}F_{j,k}^*\right]$ is real for all j,k or that $L=1$, then we can substitute (45) into (44), getting

$$\beta = \sum_{j=-\infty}^{\infty} |\mathcal{R}(\prime,|/\mathcal{W})|^2 \qquad (46)$$

As j increases, $\mathcal{R}(\prime,|/\mathcal{W})$ decreases because of the randomly changing phases and delays on the multiple paths contributing to each of the channel taps. The *channel coherence time*, t_c, is the time until this correlation becomes small, i.e., the value of j/W at which $\mathcal{R}(\prime,|/\mathcal{W})$ gets small relative to $\mathcal{R}(\prime,\prime)$. This correlation usually drops off gradually with increasing j, and thus t_c is simply a single number characterization of the extent of the correlation function. For our purposes here, it is convenient to define this number[1] as

$$t_c = \frac{\int_0^\infty |\mathcal{R}(\prime,\sqcup)|^2\, dt}{\mathcal{R}^{\in}(\prime,\prime)} \qquad (47)$$

Assuming that t_c is much larger than $1/W$, the integral in (47) can be approximated by a sum, and comparing (46) with (47), we have

$$\beta = 2Wt_c\mathcal{R}^{\in}(\prime,\prime) \qquad (48)$$

This equation is based on the assumption that $\mathsf{E}\left[F_{0,k}F_{j,k}^*\right]$ is real. However, if W is chosen so that $1/W$ is greater than the multipath spread (assuming the typical case in which the multipath spread is much less than t_c), then the only significant term in the inner sum of (44) is that for $k=0$. Thus, moving the absolute value sign in (44) outside of the inner sum is a good approximation in this case, thus justifying (46) and (48). Substituting (48) into (5),

$$\frac{1}{n} I(\vec{X}^{\,n};\vec{Y}^{\,n}) \le \frac{\gamma W t_c \mathcal{R}^{\in}(\prime,\prime)}{\sigma_Z^4} \qquad (49)$$

Let γ satisfy (6) and $\sigma_Z^2 = N_0$. Substituting this into (49), the mutual information per unit time (rather than per sample) becomes,

$$\frac{W}{n} I(\vec{X}^{\,n};\vec{Y}^{\,n}) \le \frac{\alpha t_c S_W^2 \mathcal{R}^{\in}(\prime,\prime)}{N_0^2} \qquad (50)$$

The term $\mathcal{R}(\prime,\prime)$ above is simply the power gain (or attenuation) from transmitter to receiver. One usually normalizes the input and output levels to make this term equal to 1, but we leave it in here to avoid confusion. In the section to follow, we look at a broad bandwidth as a collection of many smaller bandwidth slices of fixed size. The input variables are then constrained (approximately) both to a fixed bandwidth W and time $1/W$. With the constraint (4), we will then see that the mutual information per unit time decreases with increasing overall bandwidth.

[1] t_c is also often defined as the time τ at which $|\ \mathcal{R}(\prime,\tau)\ |^{\in}$ drops to some fixed fraction κ of $R(0,\tau)\ |^2$. The fraction κ is assigned various values between 0.37 and 0.9 [4], [6], [12]. This does not suffice here since we need a measure involving how $|\ \mathcal{R}(\prime,\tau)\ |^{\in}$ goes to zero with τ.

V. Mutual information over multiple frequency slices

Consider an arbitrary number b of frequency slices. The continuous time input $\{\mathcal{X}(t); -\infty < t < \infty\}$, at passband, is then constrained to a bandwidth bW'. Let $\mathcal{X}^a(t)$ be the continuous time passband input in the a^{th} of the b slices. Thus $\mathcal{X}(t) = \sum_{a=1}^{b} \mathcal{X}^a(t)$. Aside from the constraints, which we discuss later, viewing the input as b slices, each of bandwidth W', is simply an analytical tool and has nothing to do with the actual choice of the input within the overall bandwidth constraint bW'.

Let $\mathcal{F}(t,\tau)$ be the impulse response of the fading channel at time t to an input at time $t-\tau$. Then the response of the channel to the input slice $\mathcal{X}^a(t)$ is $\int_\tau \mathcal{X}^a(t-\tau)\mathcal{F}(t,\tau)d\tau$. If the overall Doppler spread is ΔB, then this response is bandlimited to a bandwidth $W = W' + \Delta B$. Let

$$\mathcal{Y}^a(t) = \int_\tau \mathcal{X}^a(t-\tau)\mathcal{F}(t,\tau)d\tau + \mathcal{Z}^a(t)$$

where $\{\mathcal{Z}^a(t)\}$ is a stationary real Gaussian noise process whose spectral density is flat over the bandwidth W of interest and is zero elsewhere. The baseband version of $\mathcal{X}^a(t)$ and $\mathcal{Y}^{\dashv}(\sqcup)$, sampled at rate W, corresponds to the model of the previous section. For the b passband systems just defined, consider the somewhat artificial system $\breve{\mathcal{Y}}(t) = \sum_{a=1}^{b} \breve{\mathcal{Y}}^a(t)$, where

$$\breve{\mathcal{Y}}^a(t) = \int_\tau \mathcal{X}^a(t-\tau)\mathcal{F}(t,\tau)d\tau + \breve{\mathcal{Z}}^a(t)$$

and where the noise processes $\{\breve{\mathcal{Z}}^a(t)\}$, even though overlapping in frequency, are independent between different values of a. The frequency bands occupied by adjacent outputs, say $\breve{\mathcal{Y}}^a(t)$ and $\breve{\mathcal{Y}}^{a+1}(t)$ overlap by the Doppler spread, ΔB, and therefore the process $\{\breve{\mathcal{Y}}(t)\}$ does not necessarily specify the individual processes $\{\breve{\mathcal{Y}}^a(t)\}$. Because of the data processing theorem, the average mutual information per unit time between $\{\mathcal{X}(t)\}$ and $\{\breve{\mathcal{Y}}(t)\}$ is less than or equal to the average mutual information per unit time between the set of inputs $\{\mathcal{X}^1(t), \mathcal{X}^2(t), \ldots, \mathcal{X}^b(t)\}$ and $\{\breve{\mathcal{Y}}^1(t), \breve{\mathcal{Y}}^2(t), \ldots, \breve{\mathcal{Y}}^b(t)\}$.

Since $\mathcal{X}(t) = \sum_{a=1}^{b} \mathcal{X}^a(t)$, we can represent $\breve{\mathcal{Y}}(t)$ by

$$\breve{\mathcal{Y}}(t) = \int_\tau \mathcal{X}(t-\tau)\mathcal{F}(t,\tau)d\tau + \breve{\mathcal{Z}}(t)$$

where $\breve{\mathcal{Z}}(t) = \sum_{a=1}^{b} \breve{\mathcal{Z}}^a(t)$. We assume that $W' > \Delta B$, so that only adjacent bands overlap. This entails no essential loss of generality, since W' was arbitrary up until this point. This means that the spectral density of $\breve{\mathcal{Z}}(t)$ is twice as large in the overlap regions as in the non-overlapped regions of the band. The actual received waveform $\mathcal{Y}(t)$, on the other hand, is given by

$$\mathcal{Y}(t) = \int_\tau \mathcal{X}(t-\tau)\mathcal{F}(t,\tau)d\tau + \mathcal{Z}(t)$$

where $\mathcal{Z}(t)$ has spectral density $N_0/2$ over the received band.

Now suppose we define the spectral density of each noise process $\breve{\mathcal{Z}}^a(t)$ above to be $N_0/4$. In that case, $\breve{\mathcal{Z}}(t)$ has spectral density $N_0/2$ in the overlap regions and $N_0/4$ in the non-overlap regions. We can get the true output $\mathcal{Y}(t)$ from the artificial output $\breve{\mathcal{Y}}(t)$ by adding stationary Gaussian noise of spectral density $N_0/4$ in each of the non-overlap regions. By the data processing theorem, again, the mutual information per unit time between $\{\mathcal{X}(t)\}$ and $\{\mathcal{Y}(t)\}$ is then upper bounded by that between $\{\mathcal{X}(t)\}$ and $\{\breve{\mathcal{Y}}(t)\}$, and that is further upper bounded by that between $\{\mathcal{X}^1(t), \mathcal{X}^2(t), \ldots, \mathcal{X}^b(t)\}$ and $\{\breve{\mathcal{Y}}^1(t), \breve{\mathcal{Y}}^2(t), \ldots, \breve{\mathcal{Y}}^b(t)\}$.

Next we represent each of the b pairs $\mathcal{X}^a(t)$, $\breve{\mathcal{Y}}^a(t)$ above by a discrete time baseband channel. Let $X_{i,a}$, $Y_{i,a}$, and $Z_{i,a}$ be the i^{th} time sample in the complex baseband representation of $\mathcal{X}^a(t)$, $\breve{\mathcal{Y}}^a(t)$, and $\breve{\mathcal{Z}}^a(t)$. Let $\vec{X}_a^n = (X_{1,a}, \ldots, X_{n,a})^{\mathrm{T}}$ and $\vec{Y}_a^n = (Y_{1,a}, \ldots, Y_{n,a})^{\mathrm{T}}$ be the vector input and output over the a^{th} frequency band. Also, let $\vec{X}^{n,a} = (\vec{X}_1^n, \ldots, \vec{X}_a^n)^{\mathrm{T}}$ and $\vec{Y}^{n,a} = (\vec{Y}_1^n, \ldots, \vec{Y}_a^n)^{\mathrm{T}}$. We want to find an upper bound on $I(\vec{X}^{n,b}; \vec{Y}^{n,b})$, which, as we have seen, is an upper bound on the mutual information between $\mathcal{X}(t)$ and $\mathcal{Y}(t)$ over an interval of duration n/W. As before, we can expand this mutual information as

$$I\left(\vec{X}^{n,b}; \vec{Y}^{n,b}\right) = I\left(\vec{Y}^{n,b}; \vec{X}^{n,b}\right)$$

$$= \sum_{a=1}^{b}\sum_{i=1}^{n} I\left(Y_{i,a}; \vec{X}^{n,b} \mid \vec{Y}_a^{i-1}, \vec{Y}^{n,a-1}\right) \quad (51)$$

At this point, we have reduced the continuous time channel to a vector discrete time channel. We have been somewhat cavalier about truncating ideal bandlimited processes, but this is a familiar problem in capacity arguments, and the fading multipath does not change that problem in any critical way.

To complete the model, let $F_{i,a,m}$ be the m^{th} tap of the a^{th} baseband equivalent channel filter at time i, let $\vec{F}_{i,a} = (F_{i,a,0}, \ldots, F_{i,a,L-1})^{\mathrm{T}}$, and, as before, let $\vec{X}_{i,a}$ denote $(X_{i,a}, \ldots, X_{i-L+1,a})^{\mathrm{T}}$. Let $U_{i,a} = \vec{X}_{i,a}^{\mathrm{T}}\vec{F}_{i,a}$ be the faded signal at time i for frequency slice a. Then

$$Y_{i,a} = U_{i,a} + Z_{i,a} = \vec{X}_{i,a}^{\mathrm{T}}\vec{F}_{i,a} + Z_{i,a} \quad (52)$$

As before, we impose the constraint

$$\mathsf{E}\left[|X_{i,a}|^4\right] \leq \gamma$$

for all i, a for some fixed γ. We also assume (since $\breve{\mathcal{Z}}^a(t)$ has spectral density $N_0/4$) that $\mathsf{E}\left[|Z_{i,a}|^2\right] = \sigma_Z^2 = N_0/2$ for all i, a.

Lemma 1 generalizes with no change, except for the additional conditioning on $\vec{Y}^{n,a-1}$. The following lemma gives this generalization; the proof is omitted since it is the same as lemma 1.

Lemma 3: Let $\widehat{F}_{i,a}(\vec{x}^{\,n,b}, \vec{y}_a^{\,i-1}, \vec{y}^{\,n,a-1})$ be the conditional mean of $\vec{F}_{i,a}$, conditional on $\vec{Y}_a^{\,i-1} = \vec{y}_a^{\,i-1}$, $\vec{Y}^{\,n,a-1} = \vec{y}^{\,n,a-1}$, and $\vec{X}^{\,n,b} = \vec{x}^{\,n,b}$. Let $\widetilde{F}_{i,a} = \vec{F}_{i,a} - \widehat{F}_{i,a}(\vec{x}^{\,n,b}, \vec{y}_a^{\,i-1}, \vec{y}^{\,n,a-1})$ be the corresponding fluctuation. Let $\mathrm{B_{i,a}(\vec{x}^{n,b})}$ be the covariance matrix of this conditional fluctuation and abbreviate $\vec{Y}_a^{\,i-1}, \vec{Y}^{\,n,a-1})$ by $\vec{Y}$. Then

$$\begin{aligned}
& h(Y_{i,a} \mid \vec{x}^{\,n,b}, \vec{Y}) \\
= \;& \ln\left[\pi e\left(\sigma_Z^2 + \vec{x}_{i,a}^{\mathrm{T}}\mathrm{B_{i,a}(\vec{x}^{n,b})}\,\mathrm{\vec{x}_{i,a}^{*}}\right)\right] \\
& I(Y_{i,a}; \vec{X}^{\,n,b} \mid \vec{Y}) \\
\leq \;& \frac{\mathsf{E}\left[\left|\vec{X}_{i,a}^{\mathrm{T}}\widehat{F}_{i,a}(\vec{X}^{\,n,b}, \vec{Y})\right|^2\right]}{\sigma_Z^2} \\
+ \;& \frac{\mathsf{E}\left[\left|\vec{X}_{i,a}^{\mathrm{T}}\,\mathsf{E}\left[\widetilde{F}_{i,a}\widetilde{F}_{i,a}^{*\mathrm{T}}\right]\vec{X}_{i,a}^{*}\right|^2\right]}{2\sigma_Z^4} \qquad (53)
\end{aligned}$$

The following lemma is a slightly less straightforward generalization of lemma 2, and we give a proof for those details that are different.

Lemma 4:

$$\begin{aligned}
& I(\vec{X}^{\,n,b}; \vec{Y}^{\,n,b}) \\
\leq \;& \sum_{a=1}^{b}\sum_{a'=1}^{b}\sum_{i=1}^{n}\sum_{j=1}^{n} \\
& \frac{\mathsf{E}\left[\left|\vec{X}_{i,a}^{\mathrm{T}}\,\mathsf{E}\left[\widetilde{F}_{i,a}\widetilde{F}_{j,a'}^{*\mathrm{T}}\right]\vec{X}_{j,a'}^{*}\right|^2\right]}{2\sigma_Z^4} \qquad (54) \\
& I(\vec{X}^{\,n,b}; \vec{Y}^{\,n,b}) \\
\leq \;& \sum_{a=1}^{b}\sum_{a'=1}^{b}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\mathsf{E}\left[|U_{i,a}U_{j,a'}^{*}|^2\right]}{2\sigma_Z^4} \qquad (55)
\end{aligned}$$

Proof: Consider the numerator of the first term in (53) for a fixed input $\vec{X}^{\,n,b} = \vec{x}^{\,n,b}$. Abbreviate $\widehat{F}_{i,a}(\vec{x}^{\,n,b}, \vec{Y}_a^{\,i-1}, \vec{Y}^{\,n,a-1})$ by $\widehat{F}_{i,a}$. Thus, we want to upper bound $\mathsf{E}\left[\left|\vec{x}_{i,a}^{\mathrm{T}}\widehat{F}_{i,a}\right|^2\right] = \vec{x}_{i,a}^{\mathrm{T}}\,\mathsf{E}\left[\widehat{F}_{i,a}\widehat{F}_{i,a}^{*\mathrm{T}}\right]\vec{x}_{i,a}^{*}$, where the expectation is over $\vec{Y}$ for fixed $\vec{x}^{n,b}$. Using the same argument as in (34)

$$\mathsf{E}\left[\left|\vec{x}_{i,a}^{\mathrm{T}}\widehat{F}_{i,a}\right|^2\right] \leq \sigma_Z^{-2}\left\|\vec{x}_{i,a}^{\mathrm{T}}\,\mathsf{E}\left[\vec{F}_{i,a}\vec{Y}^{\,*\mathrm{T}} \mid \vec{x}^{\,n,b}\right]\right\|^2 \quad (56)$$

The vector $\vec{Y}$ here has $i-1+n(a-1)$ components, consisting both of the first $i-1$ received variables in frequency band a plus all n received variables in each of the bands 1 to $a-1$. Thus $\vec{x}_{i,a}^{\mathrm{T}}\,\mathsf{E}\left[\vec{F}_{i,a}\vec{Y}^{\,*\mathrm{T}} \mid \vec{x}^{\,n,b}\right]$ is an $i-1+n(a-1)$ dimensional row vector whose j^{th} component, $1 \leq j \leq i-1$, is given by

$$\begin{aligned}
& \vec{x}_{i,a}^{\mathrm{T}}\,\mathsf{E}\left[\vec{F}_{i,a}Y_{j,a}^{*} \mid \vec{x}^{n,b}\right] \\
= \;& \vec{x}_{i,a}^{\mathrm{T}}\,\mathsf{E}\left[\vec{F}_{i,a}(\vec{x}_{j,a}^{\mathrm{T}}\vec{F}_{j,a} + Z_{j,a})^{*}\right] \\
= \;& \vec{x}_{i,a}^{\mathrm{T}}\,\mathsf{E}\left[\vec{F}_{i,a}\vec{F}_{j,a}^{*\mathrm{T}}\right]\vec{x}_{j,a}^{*} \qquad (57)
\end{aligned}$$

The other components are indexed by j $(1 \leq j \leq n)$ and a', $(a' < a)$,

$$\vec{x}_{i,a}^{\mathrm{T}}\,\mathsf{E}\left[\vec{F}_{i,a}Y_{j,a'}^{*} \mid \vec{x}^{n,b}\right]$$

$$= \vec{x}_{i,a}^{\mathrm{T}} \; \mathsf{E}\left[\vec{F}_{i,a}(\vec{x}_{j,a'}^{\mathrm{T}}\vec{F}_{j,a'} + Z_{j,a'})^*\right]$$

$$= \vec{x}_{i,a}^{\mathrm{T}} \; \mathsf{E}\left[\vec{F}_{i,a}\vec{F}_{j,a'}^{*\mathrm{T}}\right] \vec{x}_{j,a'}^{*} \tag{58}$$

Substituting (58) and (57) into (56)

$$\mathsf{E}\left[\left|\vec{x}_{i,a}^{\mathrm{T}}\widehat{F}_{i,a}\right|^2\right]$$

$$\leq \sigma_Z^{-2}\sum_{j=1}^{i-1}\left|\vec{x}_{i,a}^{\mathrm{T}} \; \mathsf{E}\left[\vec{F}_{i,a}\vec{F}_{j,a}^{*\mathrm{T}}\right]\vec{x}_{j,a}^{*}\right|^2$$

$$+ \sigma_Z^{-2}\sum_{j=1}^{n}\sum_{a'=1}^{a-1}\left|\vec{x}_{i,a}^{\mathrm{T}} \; \mathsf{E}\left[\vec{F}_{i,a}\vec{F}_{j,a'}^{*\mathrm{T}}\right]\vec{x}_{j,a'}^{*}\right|^2 \tag{59}$$

The expected value of this, over $X_{i,a}$, is the numerator of the first term of (53). Thus, substituting this expectation into (53), we have

$$I(Y_{i,a};\vec{X}^{n,b} \mid \vec{Y}_a^{i-1},\vec{Y}^{n,a-1}) \leq \sum_{j=1}^{i-1}\frac{\mathsf{E}\left[\left|\vec{X}_{i,a}^{\mathrm{T}} \; \mathsf{E}\left[\vec{F}_{i,a}\vec{F}_{j,a}^{*\mathrm{T}}\right]\vec{X}_{j,a}^{*}\right|^2\right]}{\sigma_Z^4} + \sum_{j=1}^{n}\sum_{a'=1}^{a-1}\frac{\mathsf{E}\left[\left|\vec{X}_{i,a}^{\mathrm{T}} \; \mathsf{E}\left[\vec{F}_{i,a}\vec{F}_{j,a'}^{*\mathrm{T}}\right]\vec{X}_{j,a'}^{*}\right|^2\right]}{\sigma_Z^4} + \frac{\mathsf{E}\left[\left|\vec{X}_{i,a}^{\mathrm{T}} \; \mathsf{E}\left[\vec{F}_{i,a}\vec{F}_{i,a}^{*\mathrm{T}}\right]\vec{X}_{i,a}^{*}\right|^2\right]}{2\sigma_Z^4} \tag{60}$$

Summing this over i and using the trick in (40),

$$I(Y_a;\vec{X}^{n,b} \mid \vec{Y}^{n,a-1}) \leq \sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\mathsf{E}\left[\left|\vec{X}_{i,a}^{\mathrm{T}} \; \mathsf{E}\left[\vec{F}_{i,a}\vec{F}_{j,a}^{*\mathrm{T}}\right]\vec{X}_{j,a}^{*}\right|^2\right]}{2\sigma_Z^4} + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{a'=1}^{a-1}\frac{\mathsf{E}\left[\left|\vec{X}_{i,a}^{\mathrm{T}} \; \mathsf{E}\left[\vec{F}_{i,a}\vec{F}_{j,a'}^{*\mathrm{T}}\right]\vec{X}_{j,a'}^{*}\right|^2\right]}{\sigma_Z^4} \tag{61}$$

Finally, summing over a and applying the trick in (40) to a,

$$I(\vec{X}^{n,b};\vec{Y}^{n,b}) \leq \sum_{a=1}^{b}\sum_{a'=1}^{b}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\mathsf{E}\left[\left|\vec{X}_{i,a}^{\mathrm{T}} \; \mathsf{E}\left[\vec{F}_{i,a}\vec{F}_{j,a'}^{*\mathrm{T}}\right]\vec{X}_{j,a'}^{*}\right|^2\right]}{2\sigma_Z^4} \tag{62}$$

This is (54), proving the first part of lemma 4. The second part follows as in lemma 2, completing the proof.

Using the Cauchy inequality in the same way as (41 - 43),

$$\mathsf{E}\left[\left|\vec{X}_{i,a}^{\mathrm{T}} \; \mathsf{E}\left[\vec{F}_{i,a}\vec{F}_{j,a'}^{*\mathrm{T}}\right]\vec{X}_{j,a'}^{*}\right|^2\right] \leq \gamma\left[\sum_k\sum_m\left| \; \mathsf{E}\left[F_{i,a,k}F_{j,a',m}^{*}\right]\right|\right]^2 \tag{63}$$

Substituting this into (62),

$$I(\vec{X}^{n,b};\vec{Y}^{n,b}) \leq \gamma\sum_{a=1}^{b}\sum_{a'=1}^{b}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\left[\sum_k\sum_m\left| \; \mathsf{E}\left[F_{i,a,k}F_{j,a',m}^{*}\right]\right|\right]^2}{2\sigma_Z^4} \tag{64}$$

This upper bound on mutual information makes no assumptions about the stationarity of the fading process. We now assume that the fading process is wide sense stationary, both in time and in frequency. This is a reasonable assumption for overall bandwidths less than 10% or so of the carrier frequency. In particular, we assume that

$$\mathsf{E}\left[\vec{F}_{i,a}\vec{F}_{j,a'}^{*\mathrm{T}}\right] = \mathsf{E}\left[\vec{F}_{0,0}\vec{F}_{j-i,a'-a}^{*\mathrm{T}}\right] \; ; \tag{65}$$

for all i, j, a, a'.

Theorem 2: Assume that the fading process is wide sense stationary, in time and frequency, and that the input variables satisfy $\mathsf{E}\left[|X_{i,a}|^4\right] \leq \gamma$. Then

$$I(\vec{X}^{n,b};\vec{Y}^{n,b}) \leq \frac{\gamma n b\widetilde{\beta}}{2\sigma_Z^4} \tag{66}$$

where

$$\widetilde{\beta} = \sum_{a'=-\infty}^{\infty}\sum_{j=-\infty}^{\infty}\left[\sum_k\sum_m\left| \; \mathsf{E}\left[F_{0,0,k}F_{j,a',m}^{*}\right]\right|\right]^2$$

Proof: We upper bound (64) by extending the sum over a' and j from $-\infty$ to $+\infty$. Then, using (65), the sum for each i and each a is the same, completing the proof.

In order to interpret what this result means, akin to the interpretation in section IV, we assume again that $\mathsf{E}\left[F_{j,0,k}F^*_{0,a',m}\right] = 0$ whenever $k \neq m$. This simplifies $\widetilde{\beta}$ to

$$\widetilde{\beta} = \sum_{a'=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \left[\sum_k | \; \mathsf{E}\left[F_{0,0,k}F^*_{j,a',k}\right] | \right]^2 \quad (67)$$

Recall that the two frequency correlation function, $\mathcal{R}(f,t)$ is the correlation between the response at time 0 to a sinusoid of unit power at the carrier frequency c and the response at time t to a sinusoid of unit power at $c+f$. Then,

$$\mathcal{R}(a'W', j/W) = \mathsf{E}\left[\sum_k F_{0,0,k}F^*_{j,a',k}\right]$$

As before, we assume either that these terms are real or that W is small enough that the tapped delay line representing the multipath has only one significant tap. Then (67) simplifies further to

$$\widetilde{\beta} = \sum_{a'=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |\mathcal{R}(a'W', j/W)|^2 \quad (68)$$

As one final simplifying assumption, assume that the correlation over time is independent of that over frequency, i.e., that

$$|\mathcal{R}(a'W', j/W)| = \frac{|\mathcal{R}(a'W', 0)\mathcal{R}(0, j/W)|}{\mathcal{R}(0,0)}$$

One can easily find situations in which this is not a good assumption, but it makes sense as an average over which wireless systems must operate. Using the definition (47) of coherence time, (68) becomes

$$\widetilde{\beta} = 2Wt_c \sum_{a'=-\infty}^{\infty} |\mathcal{R}(a'W', 0)|^2 \quad (69)$$

The frequency correlation, $\mathcal{R}(f,0)$ gets small as f becomes large because of the random strength and phase of different paths. The *frequency coherence* f_c is the frequency f at which $\mathcal{R}(f,0)$ becomes small, and we define it, somewhat arbitrarily as

$$f_c = \frac{\int_0^\infty |\mathcal{R}(f,0)|^2 \, df}{\mathcal{R}^2(0,0)} \quad (70)$$

f_c is proportional to the reciprocal of the time spread on the channel. Substituting (70) into (69), we get

$$\widetilde{\beta} = 4(W/W')t_c f_c \mathcal{R}^2(0,0) \quad (71)$$

Recall that the spectral density for the noise process in each slice was chosen to be $N_0/4$ so that $\sigma_Z^2 = N_0/2$ Substituting this and (71) into (66),

$$\frac{1}{n} I(\vec{X}^{\,n,b}; \vec{Y}^{\,n,b}) \leq \frac{8\gamma b W t_c f_c \mathcal{R}^2(0,0)}{W' N_0^2} \quad (72)$$

Let S be an overall power constraint on the input, and let $S_W = S/b$ be the power constraint in each slice. Then, from (6), we have $\gamma = \alpha S^2/(bW)^2$. Substituting this into (72), and recalling that $\mathcal{W} = bW'$ is the overall input bandwidth,

$$\begin{aligned} & \frac{W}{n} I(\vec{X}^{\,n,b}; \vec{Y}^{\,n,b}) \\ \leq \; & \frac{8\alpha t_c f_c S^2 \mathcal{R}^2(0,0)}{bW' N_0^2} \\ = \; & \frac{8\alpha t_c f_c S^2 \mathcal{R}^2(0,0)}{\mathcal{W} N_0^2} \end{aligned} \quad (73)$$

This shows that the upper bound is decreasing inversely with $\mathcal{W}$. The question we now have to answer is whether this upper bound is meaningful in any region of interest. In the broadband region, the capacity of a WGN channel without fading is S/N_0, and the bound in (73) becomes equal to the WGN capacity when $W = 8\alpha t_c f_c S/N_0$, For conventional cellular mobile communication and PCS, $f_c t_c$ ranges from about 50 to 10^4 and thus the bound is only meaningful when $\mathcal{W} > 10^7$. What this means is that these channels have so many degrees of freedom, in time and frequency, over which the channel remains relatively constant, that the channel can in principle be measured adequately.

The bandwidth at which this bound becomes significant decreases with decreasing t_c, f_c, and S/N_0. t_c is inversely proportional to Doppler shift, which is proportional both to carrier frequency and velocity of transmitters, receivers, and reflectors. f_c is inversely proportional to time spread, which increases as multiple paths are spread over larger distances. Thus, the bound becomes more significant in the regime of high carrier frequencies, rapid velocities, scattering over widely dispersed paths, and low signal to noise ratio.

VI. Spreading using CDMA

In this section, we consider the special case of direct sequence code division multiple access systems. Along with the importance of this special case, we can acquire addional insight into the general bounds of Theorems 1 and 2 for this simple case. Instead of creating another upper bound on mutual information, we assume a simplified channel model and develop a crude approximation to the mutual information, assuming IID antipodal inputs.

We still consider the model of section V where the available input bandwidth $\mathcal{W}$ is separated into b slices, each of input bandwidth W'. We choose W', which is otherwise arbitrary, to be the channel coherence bandwidth, f_c. With this choice, it is reasonable to approximate the fading to be flat over each slice, *i.e.*, to assume that the number of time varying filter taps L required to model a slice is given by $L = 1$.

Denote the single filter tap (*i.e.*, the channel strength) in slice a at time i by $F_{i,a}$. Assume that the time sequence $F_{i,a}$, $-\infty < i < \infty$ in each slice a is statistically independent and identically distributed with that of all other slices. The assumptions of flat fading within each frequency slice and independence between slices is the frequency analog of block-fading in time (see [17], [16]), where the channel is assumed constant within each time block and independent between time blocks.

Under the assumption that the frequency coherence is much larger than the Doppler shifts, we also assume that the Doppler spreading between frequency slices is negligible, so the output bandwidth W for a frequency slice is taken to be $W = W' = f_c$. This avoids the need to use the artificial noise process $\breve{\mathcal{Z}}^a(t)$ defined in section V. Thus, in this section, the noise variance will be taken to be $\sigma_Z^2 = N_0$ rather than $N_0/2$.

In CDMA, the channel input, after coding and spreading, is typically antipodal. We take some slight liberty with these antipodal inputs here by assuming that the baseband inputs on each frequency slice are antipodal. This means that the magnitude squared of the input $X_{i,a}$ in each time i and slice a is simply a constant, say X^2 for each slice. We assume that the inputs are independent over time and slice, and are $\pm X$, each with probability 1/2. The frequency slices can now be analyzed independently, so we drop the subscript a and analyze an arbitrary single frequency slice.

For each such slice, we adopt a Gauss-Markov channel model ([20]). Here each time sample evolves as

$$F_i = \eta F_{i-1} + \Xi_i, \tag{74}$$

where Ξ_i is a random variable representing the innovations process. The random variables Ξ_i are IID Gaussian with zero-mean and variance σ_Ξ^2. From (74), we see that

$$\sigma_\Xi^2 = \left(1 - |\eta|^2\right)\sigma_F^2 \tag{75}$$

Since

$$\eta^j = \frac{\mathsf{E}\left[F_{i+j}F_i^*\right]}{\mathsf{E}\left[|\ F_i\ |^2\right]},$$

the constant η represents how fast the given canonic channel decorrelates, and in particular,

$$\frac{1}{1-|\eta|^2} = \sum_{i=0}^{\infty} \frac{|\ \mathsf{E}\left[F_i F_0^*\right]\ |^2}{\mathsf{E}\left[|\ F_i\ |^2\right]^2}$$

Using the definition of t_c in (47) and approximating the above sum by an integral,

$$t_c = \frac{1}{(1-|\eta|^2)W} \tag{76}$$

The output Y_i from a given frequency slice is

$$Y_i = X_i F_i + Z_i \tag{77}$$

The idealized estimation of F_i from (74) and (77) can now be represented by the single dimensional Kalman filtering equations. In particular, let $\widehat{F}_i'(\vec{x}^n, \vec{y}^i) = \mathsf{E}\left[F_i \mid \vec{x}^n, \vec{y}^i\right]$. Note that this estimate is based on the current output as well as previous outputs and differs from the idealized estimate $\widehat{F}_i$ of (11), which does not depend on the current output. The estimate depends also on the inputs $\vec{x}^{\,i}$; the future inputs are irrelevant to the estimate. Let $\sigma^2_{K(i)}$ be the conditional variance of F_i around $\widehat{F}'_i(\vec{x}^n, \vec{y}^i)$. This variance is independent of y_i and $\vec{x}^n$ and satisfies the well known Kalman recursion equation,

$$\frac{1}{\sigma^2_{K(i)}} = \frac{1}{|\eta|^2 \sigma^2_{K(i-1)} + (1-|\eta|^2)\sigma^2_F} + \frac{X^2}{\sigma^2_Z} \quad (78)$$

As in (10), $\sigma^2_{K(i)}$ and $\widehat{F}'_i(\vec{x}^n, \vec{y}^i)$ are related by

$$\mathsf{E}\left[\left|\widehat{F}'_i(\vec{X}^n, \vec{Y}^i)\right|^2\right] = \sigma^2_F - \sigma^2_{K(i)}$$

Finally, we see from (74) that

$$\begin{aligned} &\mathsf{E}\left[\left|\widehat{F}_i(\vec{X}^n, \vec{Y}^{i-1})\right|^2\right] \\ &= |\eta|^2 \,\mathsf{E}\left[\left|\widehat{F}'_{i-1}(\vec{X}^n, \vec{Y}^{i-1})\right|^2\right] \\ &= |\eta|^2[\sigma^2_F - \sigma^2_{K(i-1)}] \end{aligned} \quad (79)$$

The variance $\sigma^2_{K(i)}$, in (78) approaches the following steady state value σ^2_K as i increases

$$\frac{1}{\sigma^2_K} = \frac{1}{\eta^2 \sigma^2_K + (1-|\eta|^2)\sigma^2_F} + \frac{X^2}{\sigma^2_Z} \quad (80)$$

Let $A = \sigma^2_F - \sigma^2_K$. This is the steady state value of $\mathsf{E}\left[\left|\widehat{F}'_{i-1}(\vec{X}^n, \vec{Y}^{i-1})\right|^2\right]$. Substituting this into (80) yields

$$\frac{1}{\sigma^2_F - A} = \frac{1}{\sigma^2_F - |\eta|^2 A} + \frac{X^2}{\sigma^2_Z} \quad (81)$$

Multiplying both sides, first, by the denominator of the left side and, second, by the denominator of the first term on the right, and simplifying,

$$A = \frac{X^2}{(1-|\eta|^2)\sigma^2_Z}(\sigma^2_F - A)(\sigma^2_F - |\eta|^2 A) \quad (82)$$

Each of the final terms in (82) are positive, and thus can be upper bounded by σ^2_F, leading to

$$A \le \frac{X^2 \sigma^4_F}{(1-|\eta|^2)\sigma^2_Z} \quad (83)$$

It can be seen by comparing (82) and (83) that (83) becomes a good approximation to A for X^2 small, *i.e.*, for a large number of slices. Thus, using (79),

$$\mathsf{E}\left[\left|\widehat{F}_i(\vec{X}^n, \vec{Y}^{i-1})\right|^2\right] \approx \frac{|\eta|^2 X^2 \sigma^4_F}{(1-|\eta|^2)\sigma^2_Z} \quad (84)$$

for X^2 small. The variance, $\mathrm{B_i}$, of the fluctuation of F_i around this idealized estimate is then

$$\mathrm{B_i} \approx \sigma^2_{\mathrm{F}} - \frac{|\eta|^2 \mathrm{X}^2 \sigma^4_{\mathrm{F}}}{(1-|\eta|^2)\sigma^2_{\mathrm{Z}}} \quad (85)$$

Note that this (and in fact the exact value of $\mathrm{B_i}$) is not a function of the particular input or output. If we now look at Lemma 1 again, we recall from (13) that

$$h(Y_i \mid \vec{x}^n, \vec{Y}^{i-1}) = \ln\{\pi e[\sigma^2_Z + X^2 \mathrm{B_i}]\}$$

Since this does not depend[2] on $\vec{x}^n$, we have

$$h(Y_i \mid \vec{X}^n, \vec{Y}^{i-1}) = \ln\{\pi e[\sigma^2_Z + X^2 \mathrm{B_i}]\} \quad (86)$$

The entropy $h(Y_i \mid \vec{Y}^{i-1})$ can be upper bounded by (23) as

$$\begin{aligned} &h(Y_i \mid \vec{Y}^{i-1}) \\ &\le \ln\left\{\pi e\left[X^2\,\mathsf{E}\left[|\widehat{F}_i(\vec{X}^n \vec{Y}^{\,i-1})|^2\right] + X^2 \mathrm{B_i} + \sigma^2_{\mathrm{Z}}\right]\right\} \end{aligned} \quad (87)$$

Combining (86) and (87),

$$\begin{aligned} &I(Y_i; \vec{X}^n \mid \vec{Y}^{\,i-1}) \\ &\le \ln\left\{1 + \frac{X^2\,\mathsf{E}\left[|\widehat{F}_i(\vec{X}^n \vec{Y}^{\,i-1})|^2\right]}{X^2 \mathrm{B_i} + \sigma^2_{\mathrm{Z}}}\right\} \\ &\le \frac{X^2\,\mathsf{E}\left[|\widehat{F}_i(\vec{X}^n \vec{Y}^{\,i-1})|^2\right]}{\sigma^2_Z} \end{aligned} \quad (88)$$

We next show that this upper bound is also a good approximation when X^2 is small. Note that

[2] There is a familiar subtlety here: this is a conditional entropy, conditional on the input, but its *value* does not depend on the particular input. Similarly, the entropy of output conditional on input for an ordinary Gaussian channel is simply the noise entropy. Its *value* does not depend on the particular input, but the conditional entropy is certainly different from the unconditional entropy.

$h(Y_i \mid \vec{X}^n, \vec{Y}^{i-1}) = h(Y_i \mid \vec{X}^i, \vec{Y}^{i-1})$ and $h(Y_i \mid \vec{Y}^{i-1}) \geq h(Y_i \mid \vec{X}^{i-1}\vec{Y}^{i-1})$. Thus,

$$I(Y_i; \vec{X}^n \mid \vec{Y}^{i-1}) \geq I(Y_i; X_i \mid \vec{X}^{i-1}\vec{Y}^{i-1}) \quad (89)$$

For a given sample value $\vec{X}^{i-1}{=}\vec{x}^{i-1}, \vec{Y}^{i-1}{=}\vec{y}^{i-1}$, we can view the mutual information $I(Y_i; X_i \mid \vec{x}^{i-1}\vec{y}^{i-1})$ as the mutual information between the antipodal input random variable $\pm X\widehat{F}(\vec{x}^{i-1}, y^{i-1})$ and the output random variable which is the sum of that input and a Gaussian random variable of variance $\mathrm{B_i} + \sigma_{\mathrm{Z}}^2$. It is well known that one can approach capacity on a Gaussian noise channel, in the limit of large bandwidth (small X^2), by using antipodal signals. Thus, for small X^2,

$$I(Y_i; X_i \mid \vec{x}^{i-1}\vec{y}^{i-1}) \approx \frac{X^2|\widehat{F}_i(\vec{x}^n\vec{y}^{i-1})|^2}{X^2\mathrm{B_i} + \sigma_{\mathrm{Z}}^2} \quad (90)$$

Averaging over $\vec{X}^{i-1}$ and $\vec{Y}^{i-1}$), and then combining with (88) and (89), we see that

$$I(Y_i; \vec{X}^n \mid \vec{Y}^{i-1}) \approx \frac{X^2 \ \mathsf{E}\left[|\widehat{F}_i(\vec{X}^n\vec{Y}^{i-1})|^2\right]}{\sigma_Z^2} \quad (91)$$

Combining this with (84),

$$I(Y_i; \vec{X}^n \mid \vec{Y}^{i-1}) \approx \frac{|\eta|^2 X^4 \sigma_F^4}{(1-|\eta|^2)\sigma_Z^4} \quad (92)$$

We may now relate (92) to (73). Since the signal power is spread over b slices, $X^4 = \frac{S^2}{b^2W^2}$. From (76), $Wt_c = 1/(1-|\eta|^2)$. Also, $\sigma_F^4 = \mathcal{R}^2(0,0)$ and $\sigma_Z^2 = N_0$. Thus,

$$I(Y_i; \vec{X}^n \mid \vec{Y}^{i-1}) \approx \frac{t_c|\eta|^2 S^2 \mathcal{R}^2(0,0)}{Wb^2N_0^2} \quad (93)$$

Let us now use (93) to derive an expression similar to (73). Multiplying (93) by b and W, and multiplying numerator and denominator by f_c,

$$\begin{aligned} \frac{W}{n} I(\vec{X}^{n,b}; \vec{Y}^{n,b}) &\approx \frac{f_c t_c |\eta|^2 S^2 \mathcal{R}^2(0,0)}{b f_c N_0^2} \\ &\approx \frac{f_c t_c |\eta|^2 S^2 \mathcal{R}^2(0,0)}{\mathcal{W} N_0^2} \end{aligned} \quad (94)$$

where we have used the fact that $f_c = W'$ and $bW' = \mathcal{W}$. For a rapidly varying channel, η is between 0.9 and 1. Using the fact that $\alpha = 1$ for the antipodal signalling here, the approximation in (94) is 9 or 10 dB tighter than the general bound in (73).

VII. Conclusions

Our results point to the fact that uniform signaling over time and frequency (as formalized by a fourth moment constraint) for time-varying channels over very broad bands does not achieve good channel utilization. These results indicate that ultra wideband systems using such signaling over GHz of bandwidth should only be used to operate over quasi-static channels.

Several questions spring from this result, the most natural being what is a practical and efficient way of transmitting over very wide spectra. The channel model here almost certainly breaks down for the bandwidths required to approach capacity for the impulsive signaling schemes of [14] and [32]. While the infinite bandwidth capacity for an AWGN channel is approached reasonably rapidly as bandwidth increases, the results in [32], using [9], indicate that the infinite bandwidth capacity for fading channels is approached impractically slowly. Thus, there is a large operating regime where the constrained fourth moment signals of this paper are not desirable but the very broadband results of [14] and [32] are not applicable. Moreover, the extremely impulsive signals required to operate in the regimes considered by [32] have great practical drawbacks.

A practical scheme may consist of combining traditional CDMA with frequency hopping, spreading using CDMA to a moderate extent and then hopping across the spectrum. In order to evaluate the effectiveness of this technique, one must first determine the range of bandwidths for which the type of signaling addressed in this paper is advantageous. While [21] begins to address this issue for channels which are block fading in time and frequency, finding tight bounds for advantageous spreading regimes for more general channels is an open problem.

The model here assumes no feedback, and it would be interesting to see how feedback changes the picture. We conjecture that the results would be basically the same, since the fourth moment constraint prevents the receiver from estimating the channel, feedback or not.

References

[1] I.C. Abou-Faycal, M.D. Trott, S. Shamai (Shitz), "The Capacity of Discrete-Time Rayleigh Fading Channels," *International Symposium on Information Theory 1997*, pg. 473.

[2] B. D. O. Anderson, J. B. Moore, *Optimal Filtering*, Prentice Hall, Englewood Cliffs, New Jersey, 1979.

[3] P.A. Bello, "Correlation Functions in a Tapped Delay Line Model of the Orbital Dipole Channel," *IEEE Transactions on Information Theory*, IT-9, 1963.

[4] P.A. Bello, B.D. Nelin, "The Effect of Frequency Selective Fading on the Binary Error Probabilities of Incoherent and Differentially Coherent Matched Filter Receivers", *IEEE Transactions on Communications*, vol. COM-11, no. 2, June 1963.

[5] E. Biglieri, J. Proakis, S. Shamai, "Fading Channels: Information-Theoretic and Communication Aspects" *IEEE Transactions on Information Theory*, Vol 44, No. 6, October 98.

[6] D.C. Cox, R.P. Leck, "Correlation Bandwidth and Delay Spread Multipath Propagation Statistics for 910 MHz Urban Mobile Radio", *IEEE Transactions on Communications*, vol. COM-23, no. 11, November 1975.

[7] R.C. Dixon, *Spread Spectrum Systems*, Second Edition, John Wiley & Sons, 1984.

[8] R.G Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, 1968.

[9] R.G. Gallager, *Energy Limited Channels: Coding, Multiaccess and Spread Spectrum*, M.I.T. LIDS Report LIDS-P-1714, November 1987.

[10] R.G. Gallager, M. Médard, "Bandwidth Scaling for Fading Channels," *ISIT 1997*, pg. 471.

[11] K. Gilhousen, I.M. Jacobs, R. Padovani, A.J. Viterbi, L.A. Weaver, C.E. Wheatley, "On the Capacity of a Cellular CDMA System," *IEEE Transactions on Vehicular Technology*, vol. 40, no. 2, May 1991.

[12] W.C. Jakes, *Microwave Mobile Communications*, IEEE Press, Piscataway, NJ, 1993.

[13] T.T. Kadota, A.D. Wyner, "Coding Theorem for Stationary, Asymptotically Memoryless, Continuous-Time Channels", *The Annals of Mathematical Statistics*, vol. 43, no. 5, May 1972.

[14] R.S. Kennedy, *Fading Dispersive Communication Channels*, Wiley-Interscience, 1969.

[15] J. Madsen, "CDMA vs. GSM," presented to the Telecommunications Policy Research Conference.

[16] T. Marzetta, B. Hochwald, "Capapcity of a Mobile Multiple-Antenna Communication Link in a Raleigh Flat-fading Environment," *IEEE Transactions on Informaton Theory*, Vol 45,January 1999.

[17] R.J. McEliece and W. E. Stark, "Channels with block interference," *IEEE Transactions on Information Theory*, Vol IT-30, No. 1, January 1984.

[18] M. Médard, *The Capacity of Time Varying Multiple User Channels in Wireless Communications*, Sc.D. Thesis, MIT, September 1995.

[19] M. Médard, "Bound on Mutual Information for DS-CDMA Spreading over Independent Fading Channels," Proceedings of *Asilomar Conference on Signals, Systems and Computers*, November 1997.

[20] M. Médard, "The Effect upon Channel Capacity in Wireless Communications of Perfect and Imperfect Knowledge of the Channel", *IEEE Transactions on Information Theory*, vol. 46, no. 3, May 2000.

[21] M. Médard, D. N.C. Tse, "Spreading in Block-Fading Channels", Proceedings of *Asilomar Conference on Signals, Systems and Computers*, October 2000.

[22] D. Nicholson, *Spread Spectrum Design: LPI and AJ Systems*, Computer Science Press, Rockville, MD., 1988.

[23] E. Pfaffelhuber, "Channels with Asymptoticaly Decreasing Memory and Anticipation", *IEEE Transactions on Information Theory*, vol. IT-17, no. 4, July 1971.

[24] R.L. Pickholtz, D.L. Schilling, L.B. Milstein, "Theory of Spread-Spectrum Communications- A Tutorial," *IEEE Transactions on Communications*, vol. COM-30, no. 5, May 1982.

[25] M.S. Pinsker, V.V. Prelov, E.C. van der Meulen, "Information Transmission over Channels with Additive-multiplicative Noise," *International Symposium on Information Theory 1998*, pg. 239.

[26] J.S. Richters, *Communication over Fading Dispersive Channels*, M.I.T. RLE Report 464, November 1967.

[27] R. Schneiderman, *Wireless Personal Communications*, IEEE Press, New York, 1994.

[28] M.K. Simon, J.K. Omura, R.A. Scholtz, B.K. Levitt, *Spread Spectrum Communications*, vol. 1, Computer Science Press, Rockville, MD, 1985.

[29] V. Subramanian, B. Hajek, "Capacity and Reliability Function per Fourth Moment Cost for WSSUS Fading Channels," *Proceedings of the 1999 IEEE Information Theory and Communications Workshop*, June 1999.

[30] V. Subramanian, B. Hajek,"Broadband Fading Channels: Signal Burstiness and Capacity," submitted to *IEEE Transactions on Information Theory.*

[31] V. Subramanian, B. Hajek, "Capacity and Reliability Function for Small Peak Signal Constraints ," submitted to *IEEE Transactions on Information Theory*

[32] I.E. Telatar, *Coding and Multiaccess for the Energy Limited Rayleigh Fading Channel*, M.S. Thesis, M.I.T., 1988.

[33] E. Telatar, D. Tse, "Capacity and Mutual Information of Wideband Multipath Fading Channels,"*IEEE Transactions on Information Theory*, vol. 46, no. 4, July 2000.

[34] S. Verdú, "On Channel Capacity per Unit Cost," *IEEE Transactions on Information Theory*, vol. IT-34, September 1990.

[35] A. Viterbi, *CDMA: Principles of Spread Spectrum Communication*, Addison Wesley, 1995.

Muriel Médard Muriel Medard (M) is an Assistant Professor in the Electrical Engineering and Computer Science at MIT and a member of the Laboratory for Information and Decision Systems. She was previously an Assistant Professor at the Electrical and Computer Engineering Department and a member of the Coordinated Science Laboratory at the University of Illinois Urbana-Champaign. From 1995 to 1998 she was a Staff Member at MIT Lincoln Laboratory in the Optical Communications and the Advanced Networking Groups. She received B.S. degrees in EECS and in Mathematics in 1989, a B.S. degree in Humanities in 1990, a M.S. degree in EE 1991, and a Sc.D. degree in EE in 1995, all from the Massachusetts Institute of Technology (MIT),Cambridge. Her research interests are in the areas of reliable communications, particularly for optical and wireless networks

Robert G. Gallager Dr. Robert Gallager (F) has been a Professor of Electrical Engineering at MIT since 1960 and was Co-Director of the MIT Laboratory for Information and Decision Systems from 1986 to 1998. He is the author of the textbooks Information Theory and Reliable Communication (New York: Wiley, 1968), Data Networks (Prentice-Hall, Ed. 2, 1992, co-authored with Bertsekas)), and Discrete Stochastic Processes (Kluwer, 1995). He has published widely and holds a number of patents in the fields of information theory, wireless networks, high speed data networks, and communication technology.

Dr. Gallager was President of the IEEE Information Theory Society in 1971. He was the Chairman of the Advisory committee to the NSF Division on Networking and Communication research and Infrastructure from 1989 to 1992. He is a life Fellow of the IEEE (1966), member of the National Academy of Engineering (1979), member of the National Academy of Sciences (1992), and member of the American Academy of Arts and Sciences (1999).

His honors include the IEEE Baker Prize Paper Award (1966), Guggenheim Fellow (1978) IEEE IT Shannon Award (1983), IEEE Centennial Medal (1984), the IEEE Bennet Prize Paper Award (1993), and two Golden Jubilee Paper Awards from the IEEE IT Society in 1998. He was the recipient of the IEEE's highest honor in 1990, the Medal of Honor, awarded for fundamental contributions to communications coding techniques. In 1999, he received the Harvey Prize in Science and Technology from the Technion, Haifa.