On achievable performance of spatial diversity fading channels *

Suhas N. Diggavi
180, Park Avenue, AT&T Shannon Laboratories,
Bldg 103, Florham Park, New Jersey, NJ 07932, USA.
Email:suhas@research.att.com - Tel.: (973)-360-8492 Fax: (973)-360-8178

Abstract

Channel time-variation and frequency selectivity (causing ISI) are two major impairments in transmission for a wireless communication environment. Spatial diversity on the transmitter or the receiver side has been traditionally used to combat multipath fading. Recent results indicate significant gains in using multiple transmitter and *receiver antenna diversity. By deriving the mutual information and cut-off rate we characterize the gains on these channels. We show that gains linear in the number of antennas can be achieved either when the SNR becomes very large or when the number of antennas becomes large. We show that some of these gains can be achieved by lower complexity linear receiver structures. By evaluating the cut-off rate for PSK constellations we further quantify the gains of using spatial diversity at both the transmitter and the receiver. Next we examine the expected mutual information for slowly fading ISI channels where the channel is assumed to be block time-invariant. We then examine the impact of fast channel time-variation (time-variation within a transmission block) on multicarrier transmission schemes. We derive the average mutual information for Orthogonal Frequency Division Multiplexing (OFDM) in time-varying ISI environments. Using this we examine the impact of transmitter and receiver diversity on OFDM transmission over time-varying ISI channels. We also study the effect of time-variation on OFDM packet-size design.*

Editorial area: Communications

*This research was presented in part at the International Symposium on Information Theory (ISIT) 1998 [1], and in the International Conference on Communications (ICC) 1997 [2].

I Introduction

The growing demand of wireless communications has created the need for reliable transmission over time-varying environments. In the wireless medium channel time-variation due to multipath fading severely impacts reliable transmission. In addition, delay spread which is due to multiple scatterers in the radio-wave propagation environment, translates to inter-symbol interference (ISI) in a digital communication scheme. To handle these two transmission impairments, diversity and equalization techniques have been developed over the past few decades [3].

Achievable performance over fading channels has been a subject of interest for several decades [3, 4, 5, 6, 7] and references therein. Spatial receive diversity is most commonly analyzed. More recently, performance with spatial transmit diversity has been examined [8, 9] where only the transmitter has spatial diversity. Recent results in [10, 11, 12, 13, 14, 15] suggest significant advantages in using both transmitter and receiver spatial diversities. In [10], the information-theoretic performance of both transmitter and receiver diversities was considered. We compare our research with this work in the subsequent paragraphs. In [11], practical coding schemes that take advantage of the potential gains have been developed. In [13], the capacity of multiple antenna fading channels was first established. In this report, it was shown that the capacity could be numerically computed using Laguerre polynomials. In [14], the capacity of *time-invariant* multiple-input-multiple output ISI channels was considered. In [15], neither the transmitter nor the receiver has access to the channel-state information, but the underlying probability distribution of the fading channel was assumed to be known. In this paper, we assume that we have access to the channel-state information at the receiver (but not at the transmitter).

There has been a large body of work devoted to data transmission over frequency selective channels [3]. Transmitter and receiver diversity in time-invariant ISI channels have been examined in [16, 17, 14] and references therein. Recently, reliable transmission over time-varying ISI channels have been studied in [18, 19, 4]. In [18, 19] it is assumed that both the transmitter and the receiver instantaneously know the channel realization. In [4] only the receiver is assumed to know the channel state. However, a quasi-static assumption where the channel is assumed to be time-invariant over the transmission block is used, making it suitable for slowly time-varying channels.

In this paper we first study the mutual information and cut-off rate for transmitter and receiver diversity fading channels. Recently it has been reported [10] that the mutual information grows linearly with the number of spatial diversity elements, asymptotically as the number of antennas become very large. Using an asymptotic decoupling argument we give an alternate approach for obtaining this result. In this we show that using a "linear detector", *i.e.* a matched filter followed by decoupled detection, we still get mutual information that grows linearly with the number of spatial diversity elements. However, the linear growth assumes that the channel gain becomes unbounded resulting in unbounded achievable rates. Consequently we examine the channel where the average gain is unity and find that the mutual information grows linearly with signal-to-noise ratio (SNR) as the number of diversity elements becomes large. Additionally we show that the linear gain in the number of spatial diversity elements (on both the transmitter and the receiver) also occurs when the SNR becomes very large. This has also been observed in the context of time-invariant channels [14]. The cut-off rate is considered an important parameter in practical code design [3]. We derive the cut-off rate of the diversity channel when the transmitter knows only the spatial correlation behavior of the fading channel. Using the cut-off rate we can evaluate the gains in using coding over multiple antennas, *e.g.* space-time coding [11].

Next, we study the mutual information for time-varying ISI channels. We derive the achievable rate for multiple transmitter and receiver diversity in slowly fading channels. Recently the use of multicarrier transmission over diversity channels has been proposed [14, 20]. We establish the achievable rate for OFDM in time-varying channels. Using this result we examine the impact of transmitter diversity and receiver diversity on OFDM transmission in time-varying channels. The main advantage of OFDM transmission in time-invariant channels is in the simplicity of the decoder, where the equalizer is just a single tap filter in the frequency domain. This occurs due to the fact that the Fourier basis is the eigenbasis for time-invariant channels. However, for time-varying channels the Fourier basis need not be the eigenbasis and hence we would have loss of orthogonality (of the carriers) at the receiver, and hence have inter-carrier interference (ICI). We examine the performance of the OFDM scheme both when we *equalize* the channel and when we ignore the ICI. As expected we show that the performance, when we ignore the ICI, depends directly on the amount of inter-carrier interference (ICI). If the channel is almost time-invariant in the transmission block, the loss in performance in neglecting the ICI is small. In OFDM there is a cyclic prefix (or a guard interval) equal to the length of the channel with every transmitted block. This results in an overhead that becomes a smaller fraction of the total rate when the block size increases. For fast time-varying channels, if we use longer packets the loss would be greater and hence there is an inherent trade-off of performance with transmission overhead. Moreover, if we decode the signals jointly (*i.e. equalize* the channel) then we get performance enhancement at the cost of higher complexity. Therefore, these trade-offs are important in packet size design as well as transceiver design. These results give insights into the role of equalization in time-varying ISI channels.

A brief outline of this paper is as follows. In Section II we describe the spatial diversity channel model suitable for time-varying channels. In Section III we develop the results on flat fading diversity channels. Section IV focuses on multicarrier transmission over fading ISI channels. In Section V we provide numerical examples for the results developed. We conclude with a brief discussion in Section VI. Some of the detailed proofs are given in the appendices.

II Channel model

The input data $\{x(k)\}$ is passed through the filter $g(t)$ to produce the transmitted signal $s(t)$. The received signal can be written as,

$$y_c(t) = \int h_c(t;\tau)s(t-\tau)d\tau + z(t) \tag{1}$$

where $h_c(t;\tau)$ is the impulse response of the time-varying channel and $z(t)$ is the additive Gaussian noise. We collect sufficient statistics through Nyquist sampling. The basic idea is that if we sample at a rate larger than $2(W_I + W_s)$, where W_I is the input bandwidth and W_s is the bandwidth of the channel time-variation [21], then we get Nyquist sampling. A careful argument about the sampling rate required for time-varying channels can be found in [19]. In this paper we assume that this criterion is met and therefore we have the following discrete-time model:

$$y(k) = y_c(kT_s) = \sum_{l=0}^{\nu-1} h(k;l)x(k-l) + z(k) \tag{2}$$

where $h(k;l)$ represents the sampled time-varying channel impulse response (which combines the transmit filter $g(t)$ with the physical channel $h_c(t;\tau)$). The approximation of having a finite impulse response in (2) can be made as good as we need by choosing ν [19]. In this paper we focus on the discrete-time model given in (2). For the case of multiple transmitter and receiver diversity channel with M_r receive and M_t transmit antennas, we can easily generalize this model to,

$$\mathbf{y}(k) = \sum_{l=0}^{\nu-1} \mathbf{H}(k;l)\mathbf{x}(k-l) + \mathbf{z}(k) \tag{3}$$

where $\mathbf{H}(k;l) \in \mathbb{C}^{M_r \times M_t}$ is the l^{th} tap of the matrix response with $\mathbf{x}(k) \in \mathbb{C}^{M_t}$ as the input, $\mathbf{y}(k) \in \mathbb{C}^{M_r}$ is the output and $\mathbf{z}(k) \in \mathbb{C}^{M_r}$ is the complex additive temporally white Gaussian noise with $\mathbf{z}(k) \sim \mathbb{C}\mathcal{N}(0, \mathbf{R}_z)$, *i.e.* a complex Gaussian vector with mean $\mathbf{0}$ and covariance $\mathbf{R}_z$. Throughout this paper we impose an average power constraint on the input, i.e. $\mathbb{E}[||\mathbf{x}(k)||^2] \leq P$. The specific structure of $\{\mathbf{H}(k;l)\}$

could be constructed by assigning a special structure to $\mathbf{H}_c(t;\tau)$ (for example a discrete multipath channel). We assume a statistical description of the channel in an attempt to balance practical utility and analytical tractability. Also, throughout this paper we assume that we have natural logarithms when we use the notation $\log(\cdot)$.

A Flat Fading channel

In this case we have $\nu = 1$ and hence the single tap $\mathbf{H}(k;0)$ is denoted for brevity by $\mathbf{H}(k)$. Hence we can rewrite (3) as

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{z}(k) \tag{4}$$

We assume that the elements of $\mathbf{H}(k)$ are i.i.d. complex Gaussian, *i.e.* $\mathbf{H}(k) = [\mathbf{h}_1(k), \ldots, \mathbf{h}_{M_t}(k)]$ has $\mathbf{H}_{i,j}(k) \sim \mathbb{C}\mathcal{N}(0,1)$ i.i.d. elements. This could be justified when the antennas are separated far enough apart so that the fading on each of the links are independent. Note that the there is a linear "array" gain associated with this model in that the average channel gain grows linearly with the number of receive antennas ($\mathbb{E}||\mathbf{h}_i(k)||^2 = M_r$). This gain captures the effect of gathering more energy when we add more receive antennas. We also consider a model when we have captured all the energy transmitted and we do not have an average gain from the channel. In this case we consider a model where the $\mathcal{L}_2$ norm of each column of $\mathbf{H}(k)$ is unity, *i.e.* $\mathbf{H}_{i,j}(k) \sim \mathbb{C}\mathcal{N}(0, 1/M_r)$, which reflects a passive channel having no average gain. We also assume for the flat fading channel considered in Section III that we have ideal interleaving, *i.e.* $\mathbf{H}(k)$ is temporally i.i.d.

B The ISI Channel

In the problem considered in Section IV, we have a time-varying ISI channel described in (3). For convenience we use the Wide-Sense Stationary Uncorrelated Scattering (WSSUS) model commonly used in describing scalar fading channels [22]. In this model, the channel $\{\mathbf{H}(k;l)\}$ is modeled as a Gaussian stochastic process with the property that $\mathbb{E}[\mathbf{H}(k;n)\mathbf{H}^H(k;m)] = \mathbb{E}[\mathbf{H}(k;n)\mathbf{H}^H(k;m)]\delta[n-m]$, *i.e.* each tap fades independently. In this case we cannot assume symbol-by-symbol ideal interleaving because of ISI. However, when we transmit using OFDM packets we can assume that we have ideal packet interleaving. We invoke this model in Section IV to gain insight into transmission over fading channels. In both the flat fading and the ISI cases, the assumption of ideal interleaving is not critical, though it simplifies achievable rate arguments. The mutual information expressions hold when the channel is an ergodic process [4, 6].

III Achievable Performance in flat fading channels

In this section we examine the advantages of using spatial diversity in flat-fading environments. In subsection A we review the capacity of this channel. In subsection B we show that the linear asymptotic growth of the mutual information with number of antennas occurs when we use a decoupled detection scheme. A passive channel where the average channel gain is normalized, is examined in Section C. In this case the mutual information grows linearly with SNR asymptotically with the number of antennas. In subsection D we show that the linear gain also occurs when the SNR becomes very large. Finally in subsection E we examine the cut-off rate and examine the coding gains for some finite PSK constellations.

A Capacity

In this section we assume the flat fading channel model described in Section IIA. The receiver is assumed to have perfect channel state information (CSI) and the transmitter only knows the statistics of the channel. We also assume that we have ideal interleaving so that the fading process is memoryless. Using these assumptions the mutual information for a block of n time samples can be written as

$$\begin{aligned}\frac{1}{n}I(\mathbf{X}^{(n)};\mathbf{Y}^{(n)},\mathbf{H}^{(n)}) &= \frac{1}{n}\left[I(\mathbf{X}^{(n)};\mathbf{H}^{(n)}) + I(\mathbf{X}^{(n)};\mathbf{Y}^{(n)}|\mathbf{H}^{(n)})\right] \\ &\overset{(a)}{=} \frac{1}{n}\mathbb{E}_{\mathcal{H}}[I(\mathbf{X}^{(n)};\mathbf{Y}^{(n)}|\mathcal{H}^{(n)} = \{\mathbf{H}^{(n)}\})] \\ &\overset{(b)}{=} \mathbb{E}_{\mathcal{H}}[\log(\frac{|\mathbf{R}_z + \mathbf{H}\mathbf{R}_x\mathbf{H}^H|}{|\mathbf{R}_z|})],\end{aligned} \tag{5}$$

where (a) follows from the fact that the input $\{\mathbf{x}(k)\}$ is independent of the fading process (as the transmitter does not have CSI) and (b) follows from the memoryless property of the vector Gaussian channel obtained by conditioning on $\mathbf{H}(k)$. We also use i.i.d. Gaussian input $\{\mathbf{x}(k)\}$ with $\mathbf{R}_x = \mathbb{E}[\mathbf{x}(k)\mathbf{x}(k)^H]$, as this maximizes the mutual information conditioned on $\mathbf{H}$. In general it is difficult to evaluate (5) except for special cases. If we assume $\mathbf{R}_z = \sigma^2\mathbf{I}$, and if we have independent diversity, *i.e.* if $\mathbf{H}(k)$ consists of iid Gaussian elements it can be shown [13] that

$$C = \mathbb{E}_{\mathcal{H}}[\log(|I + \frac{P}{M_t\sigma^2}\mathbf{H}\mathbf{H}^H|)] \tag{6}$$

is the capacity of the fading matrix channel, where $\mathbf{R}_x = \frac{P}{M_t}\mathbf{I}$. This expression can be evaluated using properties of Wishart matrices and represented in a numerically computable form [13, 23, 24, 12].

B Decoupled detection

To achieve the capacity given in (6) we require joint optimal (maximum-likelihood) decoding of all the receiver elements. In this section we explore the performance of a sub-optimal "linear decoding" scheme which is similar in flavor to the matched filter receiver studied in multi-user detection [25]. We show that due to the decoupling properties of the channel, this detector still retains the linear growth rate of the optimal decoding scheme. However, we do pay a price in terms of rate growth with SNR.

In the following we will assume that $\mathbf{H}(k)$ has iid elements, ($\mathbf{H}_{i,j}(k) \sim \mathbb{C}\mathcal{N}(0,1)$) and $\mathbf{R}_z = \sigma^2\mathbf{I}$. In some cases we have $M_t = M_r$ and we will state when this is true. Note that from the data model given in (4), the maximum likelihood decoding rule for $\mathbf{R}_z = \sigma^2\mathbf{I}$ is $\min_{\{\hat{\mathbf{x}}(k)\}} \sum_k ||\mathbf{y}(k) - \mathbf{H}(k)\hat{\mathbf{x}}(k)||^2$. Hence a sufficient statistic would be $\tilde{\mathbf{y}}(k) = \mathbf{H}^H(k)\mathbf{y}(k)$, and thus,

$$\tilde{\mathbf{y}}(k) = \mathbf{H}^H(k)\mathbf{H}(k)\mathbf{x}(k) + \mathbf{H}^H(k)\mathbf{z}(k) \tag{7}$$

In [10] it has been stated that the mutual information grows linearly with M_r (with $M_r = M_t$) as $M_r \to \infty$, *i.e.* $\lim_{M_r\to\infty} \frac{I(\mathbf{x};\mathbf{y},\mathbf{H})}{M_r}$ = constant. A proof outline is provided in [10] for this using a layered architecture wherein the problem is treated akin to a multiple access channel and decoding is done using an onion peeling scheme with orthogonal projection. In the following we provide an alternative approach which demonstrates the linear growth and also shows the importance of "interference suppression". We observe from the strong law of large numbers (SLLN) that $\lim_{M_r\to\infty} \mathbf{h}_i^H(k)\mathbf{h}_j(k)/M_r = \delta_{i-j}$ *a.s.*, where δ_k is the Kronecker delta function and we have $\mathbf{H}_{i,j}(k) \sim \mathbb{C}\mathcal{N}(0,1)$. As the channels asymptotically decouple we investigate the rate achievable if we ignore the cross-coupling of the channels when decoding. If we denote $\tilde{\mathbf{z}}(k) = \mathbf{H}^H(k)\mathbf{z}(k)$, we can write the i^{th} component of $\tilde{\mathbf{y}}(k)$ in (7) as

$$\tilde{\mathbf{y}}_i(k) = ||\mathbf{h}_i(k)||^2\mathbf{x}_i(k) + \sum_{\substack{j=1\\ j\neq i}}^{M_t} \mathbf{h}_i^H(k)\mathbf{h}_j(k)\mathbf{x}_j(k) + \tilde{\mathbf{z}}_i(k), \quad i = 1,\dots,M_t. \tag{8}$$

By ignoring the cross-coupling between the channels we decode $\hat{\mathbf{x}}_i$ as $\min_{\hat{\mathbf{x}}_i} \sum_k |\tilde{\mathbf{y}}_i(k) - ||\mathbf{h}_i(k)||^2\hat{\mathbf{x}}_i(k)|^2$ and hence we include the "interference" from $\{\mathbf{x}_j\}_{j\neq i}$ as part of the noise. This is identical to the matched filter receiver studied in multiuser detection schemes [25]. Using almost identical arguments as in (5) we can show that the mutual information $I(\mathbf{x}_i;\tilde{\mathbf{y}}_i,\mathbf{H})$ can be written as

$$I(\mathbf{x}_i;\tilde{\mathbf{y}}_i,\mathbf{H}) = \mathbb{E}\left[\log\left(1 + \frac{(||\mathbf{h}_i||^2)^2 P/M_t}{||\mathbf{h}_i||^2\sigma^2 + (\sum_{\substack{j=1\\ j\neq i}}^{M_t} |\mathbf{h}_i^H\mathbf{h}_j|^2)P/M_t}\right)\right]. \tag{9}$$

Hence the total rate achievable (R_I) when we ignore the cross-coupling is given by

$$
\begin{aligned}
R_I &= \mathbb{E}\left[\sum_{i=1}^{M_t}\log\left(1+\frac{(||\mathbf{h}_i||^2)^2P/M_t}{||\mathbf{h}_i||^2\sigma^2+(\sum_{\substack{j=1\\j\neq i}}^{M_t}|\mathbf{h}_i^H\mathbf{h}_j|^2)P/M_t}\right)\right], \qquad (10)\\
&\stackrel{(a)}{=} M_t\mathbb{E}\left[\log\left(1+\frac{(||\mathbf{h}_i||^2)^2P/M_t}{||\mathbf{h}_i||^2\sigma^2+(\sum_{\substack{j=1\\j\neq i}}^{M_t}|\mathbf{h}_i^H\mathbf{h}_j|^2)P/M_t}\right)\right]
\end{aligned}
$$

where (a) follows due to the i.i.d. assumption on the fading channels. Next we study the asymptotics on the transmit and receive antennas with the restriction that $M_t = \lfloor \alpha M_r \rfloor$, $0 < \alpha < \infty$, with the understanding that $M_t = 1$ if $\lfloor \alpha M_r \rfloor = 0$. We still denote this with $\lfloor \alpha M_r \rfloor$ to avoid more cumbersome notation. We believe that these results are also true when $\frac{M_t}{M_r} \to \alpha$ without the specific relationship. Now $\lim_{\substack{M_t\to\infty\\M_t=\lfloor\alpha M_r\rfloor}} ||\mathbf{h}_i(k)||^2/M_t = \frac{1}{\alpha}$ a.s., by the SLLN and we show in Appendix A that $\lim_{\substack{M_t\to\infty\\M_t=\lfloor\alpha M_r\rfloor}} \sum_{\substack{i=1\\i\neq j}}^{M_t} |\mathbf{h}_i^H(k)\mathbf{h}_j(k)/M_r|^2 = \alpha$ a.s. Hence using these two facts we have

$$
\lim_{\substack{M_t\to\infty\\M_t=\lfloor\alpha M_r\rfloor}} \log\left(1+\frac{(||\mathbf{h}_i||^2)^2P/M_t}{||\mathbf{h}_i||^2\sigma^2+(\sum_{\substack{j=1\\j\neq i}}^{M_t}|\mathbf{h}_i^H\mathbf{h}_j|^2)P/M_t}\right) = \log\left(1+\frac{\frac{P}{\sigma^2\alpha}}{1+\frac{P}{\sigma^2}}\right) \text{ a.s.} \qquad (11)
$$

as $\log(\cdot)$ is a continuous function. It is shown in Appendix A that we can exchange limits and expectations to get,

$$
\begin{aligned}
\lim_{\substack{M_t\to\infty\\M_t=\lfloor\alpha M_r\rfloor}} R_I/M_t &= \lim_{\substack{M_t\to\infty\\M_t=\lfloor\alpha M_r\rfloor}} \mathbb{E}\left[\log\left(1+\frac{(||\mathbf{h}_i||^2)^2P/M_t}{||\mathbf{h}_i||^2\sigma^2+(\sum_{\substack{j=1\\j\neq i}}^{M_t}|\mathbf{h}_i^H\mathbf{h}_j|^2)P/M_t}\right)\right] \qquad (12)\\
&= \mathbb{E}\left[\lim_{\substack{M\to\infty\\M_t=\lfloor\alpha M_r\rfloor}} \log\left(1+\frac{(||\mathbf{h}_i||^2)^2P/M_t}{||\mathbf{h}_i||^2\sigma^2+(\sum_{\substack{j=1\\j\neq i}}^{M_t}|\mathbf{h}_i^H\mathbf{h}_j|^2)P/M_t}\right)\right].
\end{aligned}
$$

Hence we have proved the following result.

Proposition III.1 *If* $\mathbf{H}_{i,j}(k) \sim \mathbb{C}\mathcal{N}(0,1)$, *then*

$$
\lim_{\substack{M_t\to\infty\\M_t=\lfloor\alpha M_r\rfloor}} \frac{1}{M_t} I(\mathbf{y},\mathbf{H};\mathbf{x}) \geq \lim_{\substack{M_t\to\infty\\M_t=\lfloor\alpha M_r\rfloor}} R_I/M_t = \log\left(1+\frac{\frac{P}{\sigma^2\alpha}}{1+\frac{P}{\sigma^2}}\right).
$$

■

Hence, using a decoupled detection scheme which has significantly lower complexity than the optimal scheme, we still obtain asymptotically linear growth rate in mutual information with number of antenna elements. The analysis also demonstrates the importance of joint decoding. We see that if the other transmit channels are regarded as noise, even though the channels asymptotically decouple, the contribution from the "interference" limits performance. Hence the improvement in rate with SNR requires joint detection of all the transmit codebooks. Also note that with decoupled detection, α plays an important role in the growth rate. Also note that if we use more powerful detectors such as MMSE linear detectors [25], we would expect better performance due to interference suppression. In this subsection, we considered the model where the channel elements are $\mathbf{H}_{i,j}(k) \sim \mathbb{C}\mathcal{N}(0,1)$ and with this we get the capacity growth behaviour seen in Proposition III.1. In the next subsection (Section C), we consider the model which has an average channel gain of unity and investigate the behaviour of capacity under such a model. We expect that the capacity would not grow as fast under such a model, and this intuition is quantified in the next subsection.

C Passive Channel

In the above analysis channel gain goes to infinity and hence so does the rate which grows asymptotically linearly in M_t (and M_r). Next, we investigate the behavior in the case where the channel gain is unity, *i.e.* the $\mathcal{L}_2$ norm of each of the columns of $\mathbf{H}(k)$ is unity ($\mathbf{H}_{i,j}(k) \sim \mathbb{C}\mathcal{N}(0, 1/M_r)$). A general upper bound on the mutual information in this case can be obtained through the Jensen's inequality.

$$\mathbb{E}[\log|\mathbf{I} + \frac{P}{M_t\sigma^2}\mathbf{H}\mathbf{H}^H|] \leq \log|\mathbf{I}(1 + \frac{P}{M_t\sigma^2})| \tag{13}$$

where we have used Jensen's inequality on the concave function $\log\det[\cdot]$ and the fact that $\mathbb{E}[\mathbf{H}\mathbf{H}^H] = \mathbf{I}$. Hence letting $M_t = M_r \to \infty$ in (13) we get the relationship,

$$\lim_{M_t=M_r\to\infty} \mathbb{E}[\log|\mathbf{I} + \frac{P}{M_t\sigma^2}\mathbf{H}\mathbf{H}^H|] \leq \frac{P}{\sigma^2} \tag{14}$$

In the sequel we explore the tightness of this upper bound. In (6) if we let $M_r \to \infty$, then we would get $\lim_{M_r\to\infty} I(\mathbf{x}; \mathbf{y}, \mathbf{H}) = M_t \log(1 + \frac{P}{M_t\sigma^2})$. Since $\lim_{M_r\to\infty} \mathbf{H}^H\mathbf{H} = \mathbf{I}_{M_t}$ a.s.. Now if we let $M_t \to \infty$ we get

$$\lim_{M_t\to\infty}\lim_{M_r\to\infty} I(\mathbf{x}; \mathbf{y}, \mathbf{H}) = \frac{P}{\sigma^2}. \tag{15}$$

This argument demonstrates that the channel behaves like M_t decoupled straight wire channels as $M_r \to \infty$ and hence resembles the infinite bandwidth Gaussian channel result [26]. This is also true if we let $M_t \to \infty$ first and then $M_r \to \infty$. However, when we have $M_r = M_t$ and then let them go to infinity the above argument is incorrect, but a similar result can be proved with a more technical argument shown below.

In [10] a lower bound on $|\mathbf{I}+\frac{P}{M_t\sigma^2}\mathbf{H}\mathbf{H}^H|$ is developed and an informal proof outline was provided. Here we provide a more complete proof of this inequality. In the following we assume that $M_r = M_t$ (similar arguments work for $M_r \geq M_t$ but not for $M_r < M_t$).

Proposition III.2 *If* $\mathbf{x} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \lambda\mathbf{I}_{M_r})$ *and* $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_K] \in \mathbb{C}^{M_r \times K}$ *is a random matrix independent of* $\mathbf{x}$ *and* $\mathbf{a}_i^H\mathbf{a}_j = \delta_{i-j}$*, then* $\mathbf{A}^H\mathbf{x} \sim \mathbb{C}\mathcal{N}(0, \lambda\mathbf{I}_K)$*.*

Proof: Let $\mathbf{z} = \mathbf{A}^H\mathbf{x}$, then we can write its conditional density function as,

$$f_{\mathbf{z}|\mathbf{A}}(\mathbf{z}|\mathbf{A}) = \frac{\exp(-\mathbf{z}^H\mathbf{z}/\lambda)}{(\pi)^K\lambda^K} \tag{16}$$

since $\mathbf{x}$ is independent of $\mathbf{A}$, conditioned on $\mathbf{A}$, $\mathbf{z}$ is Gaussian with mean zero and covariance $\mathbb{E}[\mathbf{A}^H\mathbf{x}\mathbf{x}^H\mathbf{A}|\mathbf{A}]$. As $\mathbf{A}^H\mathbf{A} = \mathbf{I}_K$ and $\mathbf{x} \sim \mathbb{C}\mathcal{N}(\mathbf{0}, \lambda\mathbf{I}_{M_r})$ we obtain the covariance of $\mathbf{z}$ conditioned on $\mathbf{A}$ to be $\lambda\mathbf{I}_K$, hence we get the conditional distribution given in (16). Averaging over the distribution of $\mathbf{A}$, as (16) does not depend on $\mathbf{A}$ we obtain the desired result. ■

This proposition can be used to formalize the lower bound on $|\mathbf{I}+\frac{P}{M_t\sigma^2}\mathbf{H}\mathbf{H}^H|$ (or equation (20) seen later) using the onion-peeling with projection idea in [10]. We can write (4) as,

$$\mathbf{y}^{(1)}(k) = \mathbf{y}(k) = \mathbf{h}_1(k)\mathbf{x}_1(k) + \ldots + \mathbf{h}_{M_t}(k)\mathbf{x}_{M_t}(k) + \mathbf{z}(k) \tag{17}$$

Thus as in [10] in the l^{th} stage we have already decoded $\mathbf{x}_1, \dots, \mathbf{x}_{l-1}$ and we have subtracted its contributions from $\mathbf{y}(k)$ to obtain $\mathbf{y}^{(l)} = \sum_{i=l}^{M_t}\mathbf{h}_i(k)\mathbf{x}_i(k) + \mathbf{z}(k)$. To decode $\mathbf{x}_l(k)$ we project $\mathbf{y}^{(l)}$ onto the space orthogonal to span$\{\mathbf{h}_{l+1}(k), \dots, \mathbf{h}_{M_t}(k)\}$. The orthonormal basis for this space is a *random* basis which is a deterministic function of the random vectors $\{\mathbf{h}_{l+1}(k), \dots, \mathbf{h}_{M_t}(k)\}$ and hence is independent of $\mathbf{h}_l(k)$. Hence if we denote the projection matrix by $\mathbf{A}_l(k)$ and we denote $\tilde{\mathbf{y}}^{(l)}(k) = \mathbf{A}_l^H(k)\mathbf{y}^{(l)}(k)$ as the projected vector we obtain,

$$\tilde{\mathbf{y}}^{(l)}(k) = \mathbf{A}_l^H(k)\mathbf{h}_l(k)\mathbf{x}_l(k) + \mathbf{A}_l^H(k)\mathbf{z}(k) \tag{18}$$

as $\mathbf{A}_l^H(k)\mathbf{h}_i(k) = 0, i = l+1, \dots, M_t$. Now using Proposition III.2 we have that $\mathbf{A}_l^H(k)\mathbf{h}_l(k) \sim \mathbb{C}\mathcal{N}(0, \mathbf{I}_l/M_r)$ and $\mathbf{A}_l^H(k)\mathbf{z}(k) \sim \mathbb{C}\mathcal{N}(0, \sigma^2\mathbf{I}_l)$. Thus following arguments almost identical to (5) we have,

$$R^{(l)} = I(\tilde{\mathbf{y}}^{(l)}, \mathbf{H}; \mathbf{x}_l) = \mathbb{E}[\log(1 + \frac{P}{\sigma^2 M_t M_r}\chi^2_{2l})] \tag{19}$$

where χ^2_{2l} is a chi-squared random variable with $2l$ degrees of freedom and $\mathbb{E}[\chi^2_{2l}] = l$ [1]. Thus we obtain the overall rate as $R = \sum_{l=1}^{M_t} R^{(l)}$ which is achievable and a lower bound to $I(\mathbf{x}; \mathbf{y}, \mathbf{H})$ and hence we obtain

[1] Note that this is not standard notation

the following,

$$I(\mathbf{x};\mathbf{y},\mathbf{H}) \geq \mathbb{E}[\sum_{i=1}^{M_t} \log(1+\frac{P}{\sigma^2 M_t M_r}\chi^2_{2i})] \tag{20}$$

where χ^2_{2i} is a chi-squared random variable with $2i$ degrees of freedom and $\mathbb{E}[\chi^2_{2i}] = i$. If we denote $\chi^2_{2i} = \sum_{j=1}^{i} |U_j|^2$ where $U_j \sim \mathbb{C}\mathcal{N}(0,1)$ i.i.d. then it is shown in Appendix C that

$$\lim_{M_t=M_r\to\infty} \sum_{i=1}^{M_t} \log(1+\frac{P}{\sigma^2 M_t M_r}\chi^2_{2i}) = \lim_{M_r=M_t\to\infty} \sum_{i=1}^{M_t} \frac{P}{\sigma^2 M_t M_r} \sum_{j=1}^{i} |U_j|^2 = \frac{P}{2\sigma^2} \quad \text{a.s.} \tag{21}$$

Using this and by exchanging limits and expectation in (20) (as explained in Appendix C) we have the following result.

Proposition III.3 *If* $\mathbf{H}_{i,j}(k) \sim \mathbb{C}\mathcal{N}(0,1/M_r)$, *then*

$$\frac{P}{\sigma^2} \geq \lim_{M_t=M_r\to\infty} I(\mathbf{y},\mathbf{H};\mathbf{x}) \geq \frac{P}{2\sigma^2}.$$

■

Here we get a factor of half over the result in (15) because of the inequality in (20). In the above results we have let the number of diversity elements become very large.

D Finite Diversity

In this section we investigate the case where the SNR becomes very large, but the number of diversity elements is finite. Let R_{M_r,M_t} be the rate achievable for M_r receiver and M_t transmitter antennas. In the following proposition we consider the case where, $\mathbf{H}_{i,j}(k) \sim \mathbb{C}\mathcal{N}(0,1)$, and its extension to the case where $\mathbf{H}_{i,j}(k) \sim \mathbb{C}\mathcal{N}(0,1/M_r)$ is straightforward.

Proposition III.4 $\liminf_{P/\sigma^2\to\infty} \frac{R_{M,M}}{R_{M,1}} \geq M$

Proof:

$$R_{M,1} = \mathbb{E}[\log(1+||\mathbf{h}||^2 P/\sigma^2)] \stackrel{(a)}{\leq} \log(1+\mathbb{E}[||\mathbf{h}||^2]P/\sigma^2) = \log(1+MP/\sigma^2) \tag{22}$$

where (a) is due to Jensen's inequality. Now we lower bound $R_{M,M}$ as

$$\begin{aligned} R_{M,M} &= \mathbb{E}[\log(|\mathbf{I}+\mathbf{H}\mathbf{H}^H\frac{P}{M\sigma^2}|)] \\ &= \mathbb{E}[\sum_{i=1}^{M}\log(1+\lambda_i(\mathbf{H}\mathbf{H}^H)\frac{P}{M\sigma^2})] \\ &\stackrel{(a)}{\geq} \mathbb{E}[M\log(1+\lambda_{min}(\mathbf{H}\mathbf{H}^H)\frac{P}{M\sigma^2})] \\ &\stackrel{(b)}{\geq} M\log(1+\epsilon\frac{P}{M\sigma^2})\Pr(\lambda_{min}(\mathbf{H}\mathbf{H}^H)>\epsilon), \end{aligned} \tag{23}$$

where $\lambda_i(\cdot)$ is the eigenvalue and (b) is true for all $\epsilon > 0$. Now to investigate the $P/\sigma^2 \to \infty$ we let $P/\sigma^2 = \exp(1/\epsilon)$. When we let $\epsilon \to 0$, $P/\sigma^2 \to \infty$. Now, by combining (22), (23) and using $P/\sigma^2 = \exp(1/\epsilon)$ we obtain

$$\frac{R_{M,M}}{R_{M,1}} \geq M\frac{\log(1+\epsilon\exp(1/\epsilon)/M)}{\log(1+M\exp(1/\epsilon))}\Pr(\lambda_{min}(\mathbf{H}\mathbf{H}^H)>\epsilon) \tag{24}$$

By letting $\epsilon \to 0$, we let $P/\sigma^2 \to \infty$. Now $\lim_{\epsilon\to 0}\Pr(\lambda_{min}(\mathbf{H}\mathbf{H}^H) > \epsilon) = 1$ and we can show by using L'Hospital's rule that $\lim_{\epsilon\to 0}\frac{\log(1+\epsilon\exp(1/\epsilon)/M)}{\log(1+M\exp(1/\epsilon))} = 1$. By substituting these results into (24) we get the desired result. ■

In the above result, if $M_t \neq M_r$, by a simple extension of the above argument it can be shown that $\liminf \frac{R_{M_r,M_t}}{R_{M_r,1}} \geq \min(M_r, M_t)$. Thus the gain is dictated by the number of parallel channels created by the matrix channel. The basic intuition behind this result is that we have $\min(M_r, M_t)$ cross-coupled channels when there are M_t transmitters and M_r receivers. Hence when $M_t = 1$ (or $M_r = 1$) we have effectively only one channel and the gain in using multiple antennas at both the transmitter and the receiver end is in creating more cross-coupled channels. These results indicate the advantages of using multiple spatial diversity elements at both the transmitter and the receiver.

E Cut-off rate

Cut-off rate is considered an important measure of system performance [27]. We therefore compute the cut-off rate for the diversity fading channel. This is also motivated by two other reasons, first the cut-off rate allows us to compare the performance of inputs modulated by finite constellations. Secondly we can observe the effects of correlation in the channel and noise.

The pairwise error probability between two sequences $\{\mathbf{x}(k)\}$ and $\{\hat{\mathbf{x}}(k)\}$ can be upper bounded by the

the Chernoff bound as

$$\Pr(\{\mathbf{x}(k)\} \to \{\hat{\mathbf{x}}(k)\}|\mathbf{H}) \leq \prod_k \exp(-\gamma(\gamma-1)\mathbf{h}^H(k)\mathbf{E}^H(k)\mathbf{R}_z^{-1}\mathbf{E}(k)\mathbf{h}(k)) \tag{25}$$

where γ is the Chernoff parameter, $\mathbf{E}(k) = \mathbf{e}^H(k) \otimes \mathbf{I}_{M_r}$, $\mathbf{h}(k) = \text{vec}(\mathbf{H}(k))$, $\mathbf{e}(k) = \mathbf{x}(k) - \hat{\mathbf{x}}(k)$, $\text{vec}(\cdot)$ stacks the matrix into a vector column-by-column [28] and $\otimes$ indicates the Kronecker product. Hence using the fact that $\mathbf{h}(k) \sim \mathbb{C}\mathcal{N}(0, \mathbf{R}_h)$ iid, and optimizing over γ we obtain,

$$\Pr(\{\mathbf{x}(k)\} \to \{\hat{\mathbf{x}}(k)\}) \leq \prod_k \frac{1}{|\mathbf{I} + \frac{1}{4}(\mathbf{e}(k)\mathbf{e}^H(k) \otimes \mathbf{R}_z^{-1})\mathbf{R}_h|} \tag{26}$$

Using the definition of cut-off rate $R_0 = -\frac{1}{n}\log \mathbb{E}[\Pr(\{\mathbf{x}(k)\} \to \{\hat{\mathbf{x}}(k)\})]$ we obtain,

$$R_0 = -\log \mathbb{E}_{\mathbf{e}}[\frac{1}{|\mathbf{I} + \frac{1}{4}(\mathbf{e}\mathbf{e}^H \otimes \mathbf{R}_z^{-1})\mathbf{R}_h|}] \tag{27}$$

This result is used in the numerical examples presented in Section V. It is used both in comparison between achievable rates for Gaussian codebooks and for finite constellation modulation schemes (*e.g.* PSK constellations).

IV Frequency Selective fading

Reliable transmission in frequency selective fading channels has been studied extensively in literature, [3, 4] and references therein. The most common assumption in studying these schemes is that of slow time-variation (*i.e.* Bandwidth $\gg$ Doppler spread). The rate of reliable information for the scalar channel has been derived in [4] in terms of the expected mutual information. We begin this section with a simple extension of this result to the case of transmitter and receiver diversity in Section IVA. In Section IVB we focus on the impact of time-variation within a transmission block and mainly analyze performance of multicarrier transmission schemes in such a scenario. We specialize the results of Section IVB to the WSSUS model (see Section IIB) in Section IVC.

A Slowly Fading Channels

Consider the model specified in Section II (3). If we use the average power constraint given by $\mathbb{E}[||\mathbf{x}(k)||^2] \leq P$, we can define for a block of size n,

$$R_n = \frac{1}{n} I(\mathbf{x}^{(n)}; \mathbf{y}^{(n)}, \mathbf{H}^{(n)}) \tag{28}$$

where $\mathbf{x}^{(n)} = [\mathbf{x}(0)^T, \dots, \mathbf{x}(n-1)]^T$, $\mathbf{y}^{(n)} = [\mathbf{y}(0)^T, \dots, \mathbf{y}(n-1)]^T$ and $\mathbf{H}^{(n)} = \{\mathbf{H}(k;l)\}_{k=0}^{n-1}, l \in [0, \dots, \nu-1]$. In the ISI channel we assume that a guard interval of ν samples exists between the transmission blocks. Note that this fixed interval does not lower the transmission rate asymptotically in the block size n. In the above we have assumed that the guard interval $\mathbf{x}(-\nu), \dots \mathbf{x}(-1)$ is a deterministic function of $\mathbf{x}^{(n)}$. By using arguments identical to (5) we can show that if the transmitter has no knowledge of the channel realization then,

$$R_n = \frac{1}{n}\mathbb{E}[I(\mathbf{x}^{(n)}; \mathbf{y}^{(n)} | \mathcal{H} = \mathbf{H}^{(n)})]. \tag{29}$$

Conditioned on $\mathbf{H}^{(n)}$ the channel (3) becomes an additive Gaussian channel. The slowly time-varying channel assumption is that the channel is time-invariant over a transmission block, *i.e.* $\mathbf{H}(k;l) = \mathbf{H}(r;l), \forall r, k \in [0, \dots, n-1], \forall l$. Now, to evaluate (29) we can use the DFT-based approach for the achievable rate of discrete time Gaussian channels developed in [29]. Let us first assume that we have appended a circular prefix in the guard interval, *i.e.* $\mathbf{x}(-k) = \mathbf{x}(n-k), k = 1, \dots, \nu$, which is like OFDM [30]. This is called as the N-circular Gaussian channel in [29] and using similar arguments we can show that the achievable rate ($\tilde{R}_n$) for this channel is,

$$\begin{aligned} \tilde{R}_n &= \frac{1}{n}\mathbb{E}[I(\mathbf{x}^{(n)}; \tilde{\mathbf{y}}^{(n)} | \mathcal{H} = \mathbf{H}^{(n)})] \\ &\overset{(a)}{=} \frac{1}{n}\mathbb{E}[\sum_{p=0}^{n-1} \log(|\tilde{\mathbf{H}}(p)\mathbf{S}(p)\tilde{\mathbf{H}}^H(p)/\sigma^2 + \mathbf{I}|) \end{aligned} \tag{30}$$

where we have defined the output of the circular channel as $\tilde{\mathbf{y}}^{(n)}$, (a) is obtained by using Gaussian inputs $\mathbf{x}^{(n)}$ and defining $\tilde{\mathbf{H}}(p) = \sum_{l=0}^{n-1} \mathbf{H}(0;l)\Omega^{-pl}$ where $\Omega = \exp(-j2\pi/n), j = \sqrt{-1}$. As in (5) the Gaussian input maximizes the mutual information and using the block time-invariance assumption we use a stationary Gaussian codebook with correlation function $\mathbf{R}_x(l) = \mathbb{E}[\mathbf{x}(k)\mathbf{x}^H(k+l)]$. We have also defined the input power spectral density as $\mathbf{S}(p) = \sum_{l=0}^{n-1} \mathbf{R}_x(l)\Omega^{-pl}$. We can easily show using properties of Riemann integrals (see [29] for details of this argument) that,

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n}[I(\mathbf{x}^{(n)}; \mathbf{y}^{(n)} | \mathcal{H} = \mathbf{H}^{(n)})] &= \lim_{n\to\infty} \frac{1}{n} I(\mathbf{x}^{(n)}; \tilde{\mathbf{y}}^{(n)} | \mathcal{H} = \mathbf{H}^{(n)}) \\ &= (2\pi)^{-1} \int_0^{2\pi} \log(|\mathbf{H}(f)\mathbf{S}(f)\mathbf{H}^H(f)/\sigma^2 + \mathbf{I}|)df \end{aligned} \tag{31}$$

where $\mathbf{H}(f)$ and $\mathbf{S}(f)$ are the Fourier transforms of $\{\mathbf{H}(0;l)\}$ and $\{\mathbf{R}_x(l)\}$ respectively. Note that the discrete frequency versions $\{\tilde{\mathbf{H}}(p)\}_{p=0}^{n-1}\{\mathbf{S}(p)\}_{p=0}^{n-1}$ converge to $\mathbf{H}(f), \mathbf{S}(f)$ as $n \to \infty$. Hence we can show that,

$$\lim_{n\to\infty} R_n = \lim_{n\to\infty} \tilde{R}_n = (2\pi)^{-1}\mathbb{E}[\int_0^{2\pi} log(|\mathbf{H}(f)\mathbf{S}(f)\mathbf{H}^H(f)/\sigma^2 + \mathbf{I}|)df] \tag{32}$$

using arguments similar to those in Appendix C to exchange limits and expectations. Thus for slowly time-varying channels the achievable rate is well approximated by (32). In general maximizing (32) with respect to $\mathbf{S}(f)$ is a hard problem and further simplifying assumptions need to be made for solving this problem. However, schemes using a flat input spectrum in both time and "space" may be practical [11, 20] and therefore it is worthwhile to study them. In time-varying channels there is an inherent conflict between increasing the transmission block length (for coding arguments) and the block time-invariance assumption. This is a topic we will explore in the next section.

B Impact of fast time-variation

The relation (29) holds even when we do not invoke the block time-invariance assumption. To gain an understanding of this problem, let us consider the scalar channel with $M_r = M_t = 1$. Let us assume that the transmitter chooses an orthonormal basis for transmission. For time-invariant channels it is known that asymptotically the Fourier basis is optimal as this is the eigenbasis of linear time-invariant channels [31]. For time-varying channels when the channel is unknown at the transmitter, this remains an open problem. However, due to the low complexity of using the Fourier basis and the prevalence of practical schemes using it, this is an interesting case to focus on. In time-invariant channels, the Fourier basis allows us to form parallel ISI-free channels and this scheme is illustrated in Figure 1. However, in *time-varying* channels, the Fourier basis is not in general an eigenbasis. This loss of orthogonality causes Inter-Carrier Interference (ICI). In this section we derive the information rates achievable in the presence of ICI.

We can write the output of the DFT at the receiver for a time-block $[-(\nu-1), \dots, N-1]$ (where N is the number of carriers) as,

$$Y(p) = G(p,p)X(p) + \sum_{\substack{q=0 \\ p\neq q}}^{N-1} G(p,q)X(q) + Z(p) \tag{33}$$

for $p = 0, \dots, N-1$, where $Y(p), X(p)$ and $Z(p)$ are the DFTs of $\{y(k)\}, \{x(k)\}$ and $\{z(k)\}$ respectively. We have also defined $G(p,q)$ as the $(p,q)^{th}$ element of $\mathbf{G} = \mathbf{Q}\bar{\mathbf{H}}^{(N)}\mathbf{Q}^H/N$. Here $\mathbf{Q}$ is the DFT matrix defined as $\mathbf{Q}^H = [\mathbf{q}_0, \dots \mathbf{q}_{N-1}]$ and $\mathbf{q}_s = [1, \dots, \exp(j2\pi s(N-1)/N)]^T$, and $\bar{\mathbf{H}}^{(N)}$ is the equivalent channel matrix including the effects of the cyclic prefix (used in OFDM) defined as

$$\bar{\mathbf{H}}^{(N)} = \begin{bmatrix} h(0;0), & 0, & \dots, & 0, & h(0;\nu-1), & \dots, & \dots, & h(0;1) \\ h(1;1), & h(1;0), & 0, & \dots, & 0, & h(1;\nu-1), & \dots, & h(1;2) \\ & & & \vdots & & & & \\ h(\nu-1;\nu-1), & \dots, & h(\nu-1;0), & 0, & \dots, & 0, & \dots, & 0 \\ & & & \vdots & & & & \\ 0, & \dots, & 0, & \dots, & 0, & h(N-1;\nu-1), & \dots, & h(N-1;0) \end{bmatrix}. \quad (34)$$

We can easily evaluate $G(m,s)$ as,

$$G(m,s) = \frac{1}{N} \sum_{r=0}^{N-1} \sum_{l=0}^{\nu-1} h(r;l) e^{j2\pi r(s-m)/N} e^{-j2\pi sl/N} \quad (35)$$

Note that the form of the model in (33) is applicable to more general cases than using OFDM. We can replace $\mathbf{Q}$ by any arbitrary matrix $\mathbf{B}$ and for a given structure of the guard interval (prefix), we can find the specific structure of $\mathbf{G}$ (as in (35) for OFDM) for that case. We can rewrite (33) as,

$$\mathbf{Y} = \mathbf{G}\mathbf{X} + \mathbf{Z} \quad (36)$$

where $\mathbf{Y} = [Y(0), \dots, Y(N-1)]^T$, $\mathbf{X} = [X(0), \dots, X(N-1)]^T$ and $\mathbf{Z} = [z(0), \dots, z(N-1)]^T$. Thus we can write R_N given in (29) as,

$$R_N = \mathbb{E}[\log |\mathbf{I} + \mathbf{G}\mathbf{G}^H P/\sigma^2|] \quad (37)$$

where we have assumed independent Gaussian codebooks on each of the frequency bins. Note that the rate given in (37) is achievable if we assume ideal packet interleaving (*i.e.* the matrix $\bar{\mathbf{H}}$ is identically distributed from packet to packet and is independent between packets). Here all the subcarriers need to be decoded jointly, which would imply that equalization (such as MLSE [32]) in the frequency domain needs to be employed. Therefore, a natural question that arises is the rate loss that occurs if we ignored the ICI while decoding, as is typically done in OFDM. The question is similar to that posed in Proposition III.1 for diversity channels and hence we expect a similar answer. Using an argument almost identical to that used in obtaining (9), we can show that for an OFDM of packet size N, the rate achievable per transmitted sample is,

$$R_{OFDM,N} = \frac{1}{N} \mathbb{E}_{\mathbf{G}}\Big[\sum_{p=0}^{N-1} \log\Big(1 + \frac{|G(p,p)|^2 P}{\sigma^2 + \sum_{\substack{q=0 \\ q \neq p}}^{N-1} |G(p,q)|^2 P}\Big)\Big] \quad (38)$$

Thus we see that the rate loss directly depends upon the amount of ICI as we would intuitively expect. In the above we have assumed that we use a Gaussian input codebook and that the codebooks for each subcarrier is independent. The easiest coding theorem proof for the above mutual information rate assumes independent fading on successive transmission blocks. Though this could be justified by an ideal packet interleaving assumption, a more general proof can be based on ergodicity assumptions on the channel impulse response. Note that we have not made any assumptions on the independence between $G(p,p)$ and $\{G(p,q)\}$ in this relationship. Also we need only the *instantaneous* SNR at the receiver to achieve this rate. We know from the results in flat fading channels that spatial diversity reduces variability of the channel. The next question is whether having transmit diversity helps us in this problem. For transmit diversity the expression in (38) can be easily modified. We examine the case when the number of transmit antennas M_t is very large and we have independent fading channels. This model is similar to the one used in IIA. In this case when $M_t \to \infty$ we have,

$$R^{(M_t)}_{OFDM,N} = \frac{1}{N}\sum_{p=0}^{N-1} \mathbb{E}_{\mathbf{G}}[\log(1+\frac{P\sum_{d=0}^{M_t-1}|G^{(d)}(p,p)|^2/M_t}{\sigma^2+P\sum_{\substack{q=0\\q\neq p}}^{N-1}\sum_{d=0}^{M_t-1}|G^{(d)}(p,q)|^2/M_t})] \quad (39)$$

$$\overset{M_t\to\infty}{\to} \frac{1}{N}\sum_{p=0}^{N-1} log(1+\frac{P\mathbb{E}[|G(p,p)|^2]}{\sigma^2+P\sum_{\substack{q=0\\q\neq p}}^{N-1}\mathbb{E}[|G(p,q)|^2]})$$

where $\mathbf{G}^{(d)} = \mathbf{Q}\bar{\mathbf{H}}^{(d)}\mathbf{Q}^H$ and $\bar{\mathbf{H}}^{(d)}$ is the channel given in (34) for the d^{th} diversity element. We have assumed that independent Gaussian codebooks of power P/M_t are used at each of the transmit antennas and OFDM subcarriers. We get the above by using strong law of large numbers (SLLN) and exchanging limits and expectations (easily justified in a manner similar to Appendix A). An interesting phenomenon occurs here due to the averaging effects of transmit diversity. If there were no ICI, such averaging would always increase the information rate by Jensen's inequality. However, as both the ICI and the signal are averaged it is not necessary that the rate increases. This property is illustrated in the numerical example given in Section V. A similar result has been observed in fading multiple access channels in the context of broadband vs narrowband transmission [6]. Thus transmit diversity would not always help if the ICI is ignored. However, we do expect it to help if the amount of ICI is small (*i.e.* slowly varying channels).

In the case when we have transmit and receive diversities we can modify (33) as,

$$\mathbf{Y}(p) = \mathbf{G}(p,p)\mathbf{X}(p) + \sum_{\substack{q=0\\q\neq p}}^{N-1}\mathbf{G}(p,q)\mathbf{X}(q) + \mathbf{Z}(p) \quad (40)$$

where $[\mathbf{G}(p,q)]_{i,j} = [\mathbf{Q}\mathbf{H}^{(i,j)}\mathbf{Q}^H/N]_{p,q}$ and $\mathbf{H}^{(i,j)}$ is the equivalent channel matrix (as in (34)) for the time-varying ISI channel from the i^{th} transmitter to the j^{th} receiver, $\{h^{(i,j)}(k;l)\}$. Also the received vector for the

p^{th} frequency bin is $\mathbf{Y}(p) = [Y^{(1)}(p), \ldots, Y^{(M_r)}(p)]^T$, $\mathbf{X}(p) = [X^{(1)}(p), \ldots, X^{(M_t)}(p)]^T$ is the transmitted vector for the p^{th} frequency bin and $\mathbf{Z}(p) = [Z^{(1)}(p), \ldots, Z^{(M_r)}(p)]^T$ is the noise. Hence we can easily write the achievable rate (if we use independent Gaussian codebooks on the different frequencies and the diversity elements) as,

$$\begin{aligned} R_N^{(M_r \times M_t)} &= \frac{1}{N}\mathbb{E}\left[\log|\mathbf{I} + \tilde{\mathbf{G}}\tilde{\mathbf{G}}^H \frac{P}{M_t\sigma^2}|\right] \\ &\stackrel{(a)}{=} \frac{1}{N}\mathbb{E}\left[\log|\mathbf{I} + \tilde{\mathcal{H}}\tilde{\mathcal{H}}^H \frac{P}{M_t\sigma^2}|\right] \end{aligned} \tag{41}$$

where $\tilde{\mathbf{G}}$ is a block matrix consisting of $M_r \times M_t$ blocks of $N \times N$ matrices with the $(i,j)^{th}$ block given by $\mathbf{Q}\mathbf{H}^{(i,j)}\mathbf{Q}^H/N$. Thus we can write $\tilde{\mathbf{G}} = \tilde{\mathbf{Q}}_1\tilde{\mathcal{H}}\tilde{\mathbf{Q}}_2/N$, with $\tilde{\mathcal{H}}$ having $M_r \times M_t$ blocks of the type $\mathbf{H}^{(i,j)}$, and $\tilde{\mathbf{Q}}_1$ and $\tilde{\mathbf{Q}}_2$ are block diagonal matrices having M_r and M_t blocks respectively with $\mathbf{Q}$ as the block elements. Using the fact that $\tilde{\mathbf{Q}}_1$ and $\tilde{\mathbf{Q}}_2$ are unitary matrices we get (a) in (41). Though $\tilde{\mathcal{H}}\tilde{\mathcal{H}}^H$ is Wishart distributed [23], there does not seem to be a simple closed form solution to (41). To achieve the rate in (41), we need to do joint decoding of all the codebooks. Due to the huge complexity, if there is significant ICI we can decode using the "onion-peeling" principle by treating (40) as a matrix multiple access channel. The other option is to ignore the ICI and as in (38) we can write the rate achievable (if the ICI is considered part of the noise) as

$$R_{OFDM,N}^{(M_r \times M_t)} = \frac{1}{N}\sum_{p=0}^{N-1}\mathbb{E}\log\left[\frac{|\mathbf{I} + \sum_q \mathbf{G}(p,q)\mathbf{G}^H(p,q)P/M_t\sigma^2|}{|\mathbf{I} + \sum_{\substack{q=0 \\ q\neq p}}^{N-1}\mathbf{G}(p,q)\mathbf{G}^H(p,q)P/M_t\sigma^2|}\right]. \tag{42}$$

Here we can envisage several techniques which are popular in spatio-temporal processing to get better performance for reasonable complexity. We can suppress the ICI using the receive sensors and interference suppression techniques [33, 34]. Therefore, we expect that receive diversity will always help performance even when we ignore the ICI.

C The WSSUS Channel

We specialize the results of Section B to the WSSUS model described in Section II and focus on the scalar channel. This allows us to gain insight into the impact of fast time-variation on OFDM transceivers. We notice from (35) and the WSSUS model that $\{G(m,n)\}$ are jointly Gaussian. We can write the information rate described in (38) as,

$$R = \mathbb{E}_{\mathbf{G}}[\log(\sigma^2 + P\sum_q |G(p,q)|^2)] - \mathbb{E}_{\mathbf{G}}[\log(\sigma^2 + P\sum_{p\neq q}|G(p,q)|^2)] \tag{43}$$

We denote $\sum_q |G(p,q)|^2 = \mathbf{g}_p^H \mathbf{g}_p$ and $\sum_{q \neq p} |G(p,q)|^2 = \bar{\mathbf{g}}_p^H \bar{\mathbf{g}}_p$. We notice that $\mathbf{g}_p = [G(p,0), \ldots, G(p,N-1)]^T$ is Gaussian, and so is $\bar{\mathbf{g}}_p$ which is of dimension $N-1$ (constructed by deleting the element $G(p,p)$ from $\mathbf{g}_p$). Hence using the fact that $\mathbf{g}_p$ and $\bar{\mathbf{g}}_p$ are Gaussian vectors, we can easily evaluate (43). Using (35) and the WSSUS channel (see Appendix D for details) we can write,

$$\mathbb{E}[G(m,s)G^*(m,q)] = \frac{1}{N^2} \sum_{r_1=0}^{N-1} \sum_{r_2=0}^{N-1} r_h(r_1 - r_2) e^{j2\pi r_1 (s-m)/N} e^{-j2\pi r_2 (q-m)/N} \sum_{l=0}^{\nu-1} e^{-j2\pi l (s-q)/N} \tag{44}$$

where $r_h(r_1 - r_2) = \mathbb{E}[h(r_1; l) h^*(r_2; l)]$. Let $\mathbf{R}_1 = \mathbb{E}[\mathbf{g}_m \mathbf{g}_m^H]$ and $\mathbf{R}_2 = \mathbb{E}[\bar{\mathbf{g}}_m \bar{\mathbf{g}}_m^H]$ and for simplicity assume that $\mathbf{R}_1$ and $\mathbf{R}_2$ have no repeated eigenvalues. Then (see Appendix D) we can write (43) as,

$$R = -\sum_{q=0}^{N-1} \delta_q^{(1)} exp(\sigma^2 / P\lambda_q^{(1)}) E_i(-\sigma^2 / P\lambda_q^{(1)}) + \sum_{q=1}^{N-1} \delta_q^{(2)} exp(\sigma^2 / P\lambda_q^{(2)}) E_i(-\sigma^2 / P\lambda_q^{(2)}) \tag{45}$$

where $E_i(x) = \int_{-\infty}^{x} e^t/t \quad dt$ is the exponential integral function [4]. Here $\{\delta_q^{(1)}\}$ are the residues of the characteristic function of $\mathbf{g}_m^H \mathbf{g}_m$ at $\{\lambda_q^{(1)}\}$ which are the eigenvalues of $\mathbf{R}_1$. This is written for the case where the eigenvalues $\{\lambda_q^{(1)}\}$ of $\mathbf{R}_1$ are distinct. Similarly $\{\delta_q^{(2)}\}$ are the residues of the characteristic function of $\bar{\mathbf{g}}_m^H \bar{\mathbf{g}}_m$ at $\{\lambda_q^{(2)}\}$ which are the eigenvalues of $\mathbf{R}_2$. The expression in (45) can be easily modified for the general repeated eigenvalue case, though the expression is more complicated and does not provide much further insight. The details of the above expression are given in Appendix D.

V Numerical Results

In this section we provide numerical examples which evaluate the expressions derived in in Sections III and IV. In the results for Section III we use the channel model with $\mathbf{H}_{i,j}(k) \sim \mathbb{C}\mathcal{N}(0, 1/M_r)$ iid and $\mathbf{R}_z = \sigma^2 \mathbf{I}$. For the results in Section IV we use the WSSUS model described in Section II.

We first compare the cut-off rate for a $M_r = 2 = M_t$ case with the mutual information for the $M_r = 2, M_t = 1$ and $M_r = 1, M_t = 2$ cases. These numerical results reinforce the advantages of using multiple spatial diversity elements at both the transmitter and the receiver. We consider the case when $\mathbf{R}_z = \sigma^2 \mathbf{I}$, $\mathbf{R}_h = \mathbf{I}/M_r$ and thus we can rewrite the cut-off rate given in (27) as,

$$R_0 = -\log \mathbb{E}_{\mathbf{e}} \frac{1}{(1 + ||\mathbf{e}||^2 / (4 M_r \sigma^2))^{M_r}} \tag{46}$$

This can be evaluated for finite constellations as well as for Gaussian input symbols. In the latter case the result is in terms of hypergeometric functions [35]. In the case with Gaussian symbols, $\mathbf{e} \sim \mathbb{C}\mathcal{N}(0, \frac{2P}{M_t}\mathbf{I})$. The distribution of $||\mathbf{e}||^2$ can be written by using the characteristic function of $||\mathbf{e}||^2 = \sum_{i=1}^{M_t} |e_i|^2$. Then, the expectation needed in (46) can be taken with respect to this distribution. Doing this we find that the integral can be expressed in terms of a hypergeometric function, as $R_0 = -\log(C_{exp})$ where $C_{exp} = \eta^{-M_t}\Psi(M_t, M_t+1-M_r, \frac{1}{\eta})$, $\eta = \frac{P}{2M_tM_r\sigma^2}$, and $\Psi(\cdot,\cdot,\cdot)$ is the confluent hypergeometric function (also called the Kummer function) defined as $\Psi(a,b,z) = \frac{1}{(a-1)!}\int_0^\infty \frac{e^{-zt}t^{a-1}}{(1+t)^{1+a-b}}dt$ [35].

We denote the case of M_r receive and M_t transmit antennas as the $M_r \times M_t$ case. Figure 2 shows the cut-off rate for 2×2,2×1 and the 1×2 cases for Gaussian symbols. We have also plotted the AWGN rate and the fading capacity for the 2×1, 1×2 cases for reference. Here we see that there is a distinct advantage of using spatial diversity at both the transmitter and the receiver, and the advantage grows with SNR as predicted in Proposition III.4. In Figure 3 we plot the cut-off rate for various PSK constellations. The curves show that for a rate 2/3 coded PSK we can achieve gains up to 10dB using random codes. This is a comparison of the gains when we use a 2/3 coded PSK on each of the transmit antennas with $M_r = M_t = 4$ and the case where we have the same transmit scheme and $M_r = 1, M_t = 4$. It has been shown in [36] that gains of at most 3dB can be achieved with random codes when only receive diversity is used. In Figure 4 we plot the cut-off rate for PSK symbols against the number of transmitter and receiver sensors. This shows a linear growth in cut-off rate for PSK symbols. These results demonstrate the advantages in using spatial diversity at both the transmitter and the receiver. Space-time code designs in [11] have demonstrated the advantages using practical trellis codes.

Next we turn our attention to the fading ISI channel and evaluate the expression derived in Section IV. This allows us to plot the information rate as a function of Doppler shift and block size. Using these plots we illustrate the trade-off between receiver complexity and overhead.

We use a WSSUS channel with three taps, each of which has energy of 1. We assume a signal bandwidth of 30kHz. The signal-to-noise ratio (P/σ^2) is fixed at 20dB and the transmit OFDM spectrum is flat. We have chosen a scenario where there is significant time-variation within a block of transmission. The parameters chosen are similar to that seen in second generation TDMA scenarios (IS-54 for example). The channel time-statistics are assumed to be represented by $r_h(k) = J_0(\omega_d k T_s)$ (Jakes' model [22]). Here $J_0(\cdot)$ is the zeroth order Bessel function of the first kind, $\omega_d = 2\pi v/\lambda$ is the Doppler spread, and v is the mobile velocity. A carrier wavelength of $\lambda = 0.3$ (*i.e.*, 1 GHz carrier frequency) was assumed. In Figure 5 we plot the information rate per transmitted sample, $RN/(N+\nu)$, as a function of the packet size N and various Doppler spreads. For very low velocity the time-invariant assumption is quite valid and there is little loss due to ICI. However, for larger time-variation the loss due to ICI is quite significant. This indicates that block sizes need to be quite small in fast time-varying channels and therefore the overhead for OFDM could be quite large. For reference, the information rate for an AWGN channel with the same channel gain and SNR is 8.23 bits/transmitted sample. The corresponding rate for the scalar slowly fading channel (Section IVA) is 7.42 bits/transmitted sample. In Figure 6 we have plotted the information rates for infinite transmit diversity. Comparing this to Figure 5 we see that at high velocities there is not much gain due to diversity. For example at 80mph and packet size of 256 for single transmit antenna (Figure 5), the rate achievable is 0.43 bits/transmitted sample and with infinite diversity it (Figure 6) is 0.36 bits/transmitted sample. This shows that the averaging effect of diversity on the ICI offsets the gain of the averaging effect on the signal. For improved performance we would need a multi-tap frequency domain equalizer, which increases the receiver complexity. This demonstrates a trade-off between transmission overhead and receiver complexity.

VI Discussion

In this paper we studied fading diversity channels using an information theoretic approach. The utility of both transmitter and receiver spatial diversity was illustrated through the rate advantages it provides. In particular, we examined a low complexity decoding scheme which is similar in flavor to the linear detectors used in multiuser detection. This showed that a linear gain in the number of antennas could be obtained using simpler detection schemes which might be attractive in practice. This result was asymptotic in the number of transmitting (and receiving) antennas. We also showed that when the SNR becomes very large, we obtain a linear gain in the number of transmitting (and receiving) antennas. In ISI channels, we studied multicarrier transmission in fast time-varying channels. This allowed us to examine the trade-off between equalization (complexity) and overhead (packet size). By doing this in terms of achievable rate, we can use this analysis for packet size and receiver structure design.

There are several important open questions which remain unanswered. The first is the achievable rate and code-design criterion for spatially correlated fading channels. Here one would intuitively expect that one would design codes that transmit along the "preferred" spatial directions rather than omnidirectional (spatially white) codes. Next, an explicit form for achievable rate for time-varying ISI channels is desirable. Upto now expressions for quasi-static (slow time-varying) or finite packet sizes have been developed. Finally the question of the best basis for transmission on time-varying channels is an open one. We have seen that there is rate loss associated with ignoring the ICI. One approach would be to choose a basis which has lower ICI and thus a lower rate loss with complexity similar to OFDM.

ACKNOWLEDGEMENT

We gratefully acknowledge stimulating discussions with Prof. Tom Cover. We greatly appreciate careful reading and insightful comments on the paper by Boon Chong Ng, Assaf Zeevi and the anonymous reviewers.

Appendix A Details of Proposition III.1

Proposition A.1 *If* $\mathbf{H}(k) = [\mathbf{h}_1(k), \ldots, \mathbf{h}_{M_t}(k)] \in \mathbb{C}^{M_r \times M_t}$ *and* $\mathbf{h}_l(k) \sim \mathbb{C}\mathcal{N}(0, \mathbf{I}_{M_r}), l = 1, \ldots, M_t$*, are i.i.d. then*

$$\lim_{\substack{M_r \to \infty \\ M_t = \lfloor \alpha M_r \rfloor}} \sum_{\substack{j=1 \\ j \neq i}}^{M_t} |\frac{\mathbf{h}_i^H(k)\mathbf{h}_j(k)}{M_r}|^2 = \alpha \quad \text{almost surely.}$$

Proof: For brevity we will suppress the time index k and denote $V_j = \frac{|\mathbf{h}_i^H \mathbf{h}_j|^2}{M_r} - 1$. Let us examine,

$$\begin{aligned} \mathbb{E}[\frac{1}{M_r}\sum_{\substack{j=1 \\ j\neq i}}^{M_t} V_j]^4 &= \frac{1}{M_r^4}\mathbb{E}[\sum_{k_1}\sum_{k_2}\sum_{k_3}\sum_{k_4} V_{k_1}V_{k_2}V_{k_3}V_{k_4}] \quad \text{(A.1)} \\ &\overset{(a)}{=} \frac{1}{M_r^4}\{(\bar{n}^4 - 6\bar{n}^3 + 11\bar{n}^2 - 6\bar{n})\mathbb{E}[V_{k_1}V_{k_2}V_{k_3}V_{k_4}] + (6\bar{n}^3 - 18\bar{n}^2 + 12\bar{n})\mathbb{E}[V_{k_1}^2 V_{k_3}V_{k_4}] \\ &+ \bar{n}\mathbb{E}[V_{k_1}^4] + (3\bar{n}^2 - 3\bar{n})\mathbb{E}[V_{k_1}^2 V_{k_2}^2] + (4\bar{n}^2 - 4\bar{n})\mathbb{E}[V_{k_1}^3 V_{k_2}]\} \end{aligned}$$

where $\bar{n} = M_t - 1$, (a) follows by the expansion of the product and using the i.i.d. property, and the notation $V_{k_1}V_{k_2}V_{k_3}V_{k_4}$ is shorthand for the case $k_1 \neq k_2 \neq k_3 \neq k_4$. Note that $\mathbb{E}|V|^2 = 1$, $\mathbb{E}|V|^4 = 2$ for $V \sim \mathbb{C}\mathcal{N}(0,1)$ having i.i.d. real and imaginary parts. Using this and the i.i.d. property of $\mathbf{h}_j \sim \mathbb{C}\mathcal{N}(0, \mathbf{I}_{M_r})$ we can show after some algebra (see Appendix B) that,

$$\begin{aligned} \mathbb{E}[V_{k_1}V_{k_2}V_{k_3}V_{k_4}] &= \mathbb{E}[(\mathbf{h}_i^H\mathbf{h}_i/M_r - 1)^4] = O(\frac{1}{M_r^2}) \quad \text{(A.2)} \\ \mathbb{E}[V_{k_1}^2 V_{k_3}V_{k_4}] &= \mathbb{E}[(\mathbf{h}_i^H\mathbf{h}_i/M_r)^2(\mathbf{h}_i^H\mathbf{h}_i/M_r - 1)^2] = O(\frac{1}{M_r}) \\ \mathbb{E}[V_{k_1}^4] &= O(1) \\ \mathbb{E}[V_{k_1}^2 V_{k_2}^2] &= O(1) \\ \mathbb{E}[V_{k_1}^3 V_{k_2}] &= O(1) \end{aligned}$$

where $h(n) = O(g(n))$ means $\limsup_{n\to\infty} |\frac{h(n)}{g(n)}| < \infty$ [37]. Now we choose $M_t = \lfloor \alpha M_r \rfloor$, then $M_t \leq \alpha M_r + 1$, and $\bar{n} = M_t - 1 \leq \alpha M_r$. Using the above and (A.2) in (A.1) we obtain,

$$\mathbb{E}\left[\frac{1}{M_r}\sum_{\substack{j=1\\ j\neq i}}^{M_t} V_j\right]^4 = O(\frac{1}{M_r^2}) \tag{A.3}$$

which implies that $\sum_{M_r} \mathbb{E}[\frac{1}{M_r}\sum_{\substack{j=1\\ j\neq i}}^{M_t} V_j]^4 < \infty$. This means that $\sum_{\substack{j=1\\ j\neq i}}^{M_t} V_j/M_r \to 0$, a.s. using Theorem 6.5 in [37]. Thus we have,

$$\lim_{\substack{M_r\to\infty\\ M_t=\lfloor\alpha M_r\rfloor}} \sum_{\substack{j=1\\ j\neq i}} \frac{1}{M_r^2}|\mathbf{h}_i^H\mathbf{h}_j|^2 = \lim_{\substack{M_r\to\infty\\ M_t=\lfloor\alpha M_r\rfloor}} (M_t-1)/M_r = \alpha, \text{a.s.} \tag{A.4}$$

■

To prove (12) let us define $W_{M_r,M_t} = \log(1 + \frac{(||\mathbf{h}_i||^2)^2 P/M_t}{||\mathbf{h}_i||^2\sigma^2 + (\sum_{\substack{j=1\\ j\neq i}}^{M_t} |\mathbf{h}_i^H\mathbf{h}_j|^2)P/M_t})$. We have shown in (11) that W_{M_r,M_t} converges almost surely as $M_r \to \infty$, $\frac{M_r}{M_t} \to \alpha$. Notice that,

$$W_{M_r,M_t} \leq \log(1 + ||\mathbf{h}_i||^2 P/M_t\sigma^2) \stackrel{(a)}{\leq} ||\mathbf{h}_i||^2 P/M_t\sigma^2 \tag{A.5}$$

where (a) follows because $\log(1+x) \leq x$ for $x \geq 0$. This can be seen by defining $g(x) = x - \log(1+x)$ and noticing that $g(0) = 0, dg(x)/dx = x/(1+x) \geq 0$ for $x \geq 0$. Hence the inequality holds as $g(x)$ is an increasing function. Using (B.5) we have,

$$\mathbb{E}W_{M_r,M_t}^2 \leq (P/\sigma^2)^2\mathbb{E}(||\mathbf{h}_i||^2/M_t)^2 = (1+1/M_r)(P/\sigma^2)^2(\frac{M_r}{M_t})^2 \leq 2(P/\sigma^2)^2(\frac{2}{\alpha})^2, \forall M_t = \lfloor\alpha M_r\rfloor, \tag{A.6}$$

where the last inequality is because $M_t, M_r \geq 1$ and $\alpha M_r - 1 \leq M_t$. Hence using Theorems 13.3(a) and 13.7 in [37] $\{W_{M_r,M_t}\}_{M_t=\lfloor\alpha M_r\rfloor}$ is uniformly integrable and hence

$$\lim_{\substack{M_r\to\infty\\ M_t=\lfloor\alpha M_r\rfloor}} \mathbb{E}W_{M_r,M_t} = \mathbb{E}\lim_{\substack{M_r\to\infty\\ M_t=\lfloor\alpha M_r\rfloor}} W_{M_r,M_t}$$

giving us (12).

Appendix B

In this section we outline the algebra to show (A.2). Using $V_j = \frac{1}{M_r}|\mathbf{h}_i^H\mathbf{h}_j|^2 - 1$ we have,

$$\begin{aligned}\mathbb{E}[V_{k_1}V_{k_2}V_{k_3}V_{k_4}] &= \mathbb{E}[(|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2/M_r - 1)(|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2/M_r - 1)(|\mathbf{h}_i^H\mathbf{h}_{k_3}|^2/M_r - 1)(|\mathbf{h}_i^H\mathbf{h}_{k_4}|^2/M_r - 1)] \quad \text{(B.1)}\\ &\stackrel{(a)}{=} \mathbb{E}\{|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2|\mathbf{h}_i^H\mathbf{h}_{k_3}|^2|\mathbf{h}_i^H\mathbf{h}_{k_4}|^2/M_r^4 - 4|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2|\mathbf{h}_i^H\mathbf{h}_{k_3}|^2/M_r^3\\ &+ 6|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2/M_r^2 - 4|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2/M_r + 1\}\end{aligned}$$

where (a) follows by using the fact that $\{\mathbf{h}_j\}$ are i.i.d. and $k_1 \neq k_2 \neq k_3 \neq k_4$. Now we have, $\mathbb{E}|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2|\mathbf{h}_i^H\mathbf{h}_{k_3}|^2|\mathbf{h}_i^H\mathbf{h}_{k_4}|^2 = \sum_{l_1}\sum_{l_2}\sum_{l_3}\sum_{l_4}\mathbb{E}|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_2}|^2|\mathbf{h}_{il_3}|^2|\mathbf{h}_{il_4}|^2$ by using the fact that $\mathbb{E}[\mathbf{h}_{jl}] = 0$, $\mathbb{E}[|\mathbf{h}_{jl}|^2] = 1$ and $\mathbf{h}_{k_1}, \mathbf{h}_{k_2}, \mathbf{h}_{k_3}, \mathbf{h}_{k_4}$ are independent. In a similar manner we can show that, $\mathbb{E}|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2|\mathbf{h}_i^H\mathbf{h}_{k_3}|^2 = \sum_{l_1}\sum_{l_2}\sum_{l_3}\mathbb{E}|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_2}|^2|\mathbf{h}_{il_3}|^2$, $\mathbb{E}|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2 = \sum_{l_1}\sum_{l_2}\mathbb{E}|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_2}|^2$ and $\mathbb{E}|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2| = \sum_{l_1}\mathbb{E}|\mathbf{h}_{il_1}|^2$. Thus using (B.1) we get,

$$\mathbb{E}[V_{k_1}V_{k_2}V_{k_3}V_{k_4}] = \mathbb{E}[\sum_l |\mathbf{h}_{il}|^2/M_r - 1]^4 \tag{B.2}$$

Now, from a combinatorial count and using $\mathbb{E}[|\mathbf{h}_{il}|^2] = 1$ and $\mathbb{E}[|\mathbf{h}_{il}|^4] = 2$ we have,

$$\begin{aligned}
\sum_{l_1,\dots,l_4}\frac{1}{M_r^4}\mathbb{E}[|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_2}|^2|\mathbf{h}_{il_3}|^2|\mathbf{h}_{il_4}|^2] &= \frac{M_r^4 - 6M_r^3 + 11M_r^2 - 6M_r}{M_r^4}\mathbb{E}|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_2}|^2|\mathbf{h}_{il_3}|^2|\mathbf{h}_{il_4}|^2 \\
&+ \frac{6M_r^3 - 18M_r^2 + 12M_r}{M_r^4}\mathbb{E}|\mathbf{h}_{il_1}|^4|\mathbf{h}_{il_2}|^2|\mathbf{h}_{il_3}|^2 + O(1/M_r^2) \\
&= 1 + 6/M_r + O(1/M_r^2)
\end{aligned} \tag{B.3}$$

Similarly we have,

$$\begin{aligned}
\sum_{l_1}\sum_{l_2}\sum_{l_3}\mathbb{E}[|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_2}|^2|\mathbf{h}_{il_3}|^2]/M_r^3 &= (M_r^3 - 3M_r^2 + 2M_r)/M_r^3\mathbb{E}|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_2}|^2|\mathbf{h}_{il_3}|^2 \\
&+ (3M_r^2 - 3M_r)/M_r^3\mathbb{E}|\mathbf{h}_{il_1}|^4|\mathbf{h}_{il_2}|^2 + O(1/M_r^2) \\
&= 1 + 3/M_r + O(1/M_r^2)
\end{aligned} \tag{B.4}$$

$$\begin{aligned}
\sum_{l_1}\sum_{l_2}\mathbb{E}|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_2}|^2/M_r^2 &= (M_r(M_r-1))/M_r^2\mathbb{E}|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_2}|^2 + (M_r)/M_r^2\mathbb{E}|\mathbf{h}_{il_1}|^4 \\
&= 1 + 1/M_r
\end{aligned} \tag{B.5}$$

$$\sum_{l_1}\mathbb{E}|\mathbf{h}_{il_1}|^2|/M_r = 1 \tag{B.6}$$

Thus using (B.3–B.6) in (B.1) we get, $\mathbb{E}[V_{k_1}V_{k_2}V_{k_3}V_{k_4}] = O(1/M_r^2)$.

To show that $\mathbb{E}[V_{k_1}^2V_{k_2}V_{k_3}] = O(1/M_r)$, let us consider,

$$\begin{aligned}
\mathbb{E}[V_{k_1}^2V_{k_2}V_{k_3}] &= \mathbb{E}[(|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2/M_r - 1)^2(|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2/M_r - 1)(|\mathbf{h}_i^H\mathbf{h}_{k_3}|^2/M_r - 1)] \\
&\overset{(a)}{=} \mathbb{E}\left\{|\mathbf{h}_i^H\mathbf{h}_{k_1}|^4|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2|\mathbf{h}_i^H\mathbf{h}_{k_3}|^2/M_r^4 - 2|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2|\mathbf{h}_i^H\mathbf{h}_{k_3}|^2/M_r^3\right. \\
&- \left. 2|\mathbf{h}_i^H\mathbf{h}_{k_1}|^4|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2/M_r^3 + |\mathbf{h}_i^H\mathbf{h}_{k_1}|^4/M_r^2 + 5|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2/M_r^2 - 4|\mathbf{h}_i^H\mathbf{h}_{k_1}|^2/M_r + 1\right\}
\end{aligned} \tag{B.7}$$

where (a) follows by using the fact that $\{\mathbf{h}_j\}$ are i.i.d. and $k_1 \neq k_2 \neq k_3$. We have,

$$\begin{aligned}
\mathbb{E}[|\mathbf{h}_i^H\mathbf{h}_{k_1}|^4|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2|\mathbf{h}_i^H\mathbf{h}_{k_3}|^2]/M_r^4 &= \frac{1}{M_r^4}\sum_{l_1,\ldots,l_6}\mathbb{E}\mathbf{h}_{il_1}^*\mathbf{h}_{il_2}\mathbf{h}_{il_3}^*\mathbf{h}_{il_4}|\mathbf{h}_{il_5}|^2|\mathbf{h}_{il_6}|^2\mathbb{E}\mathbf{h}_{k_1l_1}\mathbf{h}_{k_1l_2}^*\mathbf{h}_{k_1l_3}\mathbf{h}_{k_1l_4}^* \quad \text{(B.8)}\\
&\stackrel{(a)}{=} \frac{1}{M_r^4}\sum_{l_5,l_6}\Bigg\{\sum_{l_1}\mathbb{E}[|\mathbf{h}_{il_1}|^4|\mathbf{h}_{il_5}|^2|\mathbf{h}_{il_6}|^2]\mathbb{E}[|\mathbf{h}_{k_1l_1}|^4]\\
&+ 2\sum_{l_1,l_3\neq l_1}\mathbb{E}[|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_3}|^2|\mathbf{h}_{il_5}|^2|\mathbf{h}_{il_6}|^2]\mathbb{E}[|\mathbf{h}_{k_1l_1}|^2|\mathbf{h}_{k_1l_3}|^2]\Bigg\}\\
&\stackrel{(b)}{=} 2\sum_{l_1,\ldots,l_4}\mathbb{E}|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_2}|^2|\mathbf{h}_{il_3}|^2|\mathbf{h}_{il_4}|^2 \stackrel{(c)}{=} 2(1+6/M_r)+O(1/M_r^2)
\end{aligned}$$

where (a) follows from the fact that h_{kl} are i.i.d., (b) by using $\mathbb{E}|h_{kl}|^2 = 1, \mathbb{E}|h_{kl}|^4 = 2$ and (c) from (B.3). Similarly we have,

$$\begin{aligned}
\mathbb{E}[|\mathbf{h}_i^H\mathbf{h}_{k_1}|^4|\mathbf{h}_i^H\mathbf{h}_{k_2}|^2/M_r^3] &= \sum_{l_1,\ldots,l_5}\mathbb{E}[\mathbf{h}_{il_1}^*\mathbf{h}_{il_2}\mathbf{h}_{il_3}^*\mathbf{h}_{il_4}|\mathbf{h}_{il_5}|^2]\mathbb{E}[\mathbf{h}_{k_1l_1}\mathbf{h}_{k_1l_2}^*\mathbf{h}_{k_1l_3}\mathbf{h}_{k_1l_4}^*]\mathbb{E}[|\mathbf{h}_{k_2l_5}|^2]/M_r^3 \quad \text{(B.9)}\\
&= \frac{1}{M_r^3}\sum_{l_5}\Bigg\{\sum_{l_1}\mathbb{E}[|\mathbf{h}_{il_1}|^4|\mathbf{h}_{il_5}|^2]\mathbb{E}[|\mathbf{h}_{k_1l_1}|^4]\\
&+ 2\sum_{l_1,l_3\neq l_1}\mathbb{E}[|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_3}|^2|\mathbf{h}_{il_5}|^2]\mathbb{E}[|\mathbf{h}_{k_1l_1}|^2|\mathbf{h}_{k_1l_3}|^2]\Bigg\}\\
&= 2\sum_{l_1,\ldots,l_3}\mathbb{E}|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_2}|^2|\mathbf{h}_{il_3}|^2 = 2(1+3/M_r)+O(1/M_r^2)
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}[|\mathbf{h}_i^H\mathbf{h}_{k_1}|^4/M_r^2] &= \sum_{l_1,\ldots,l_4}\mathbb{E}[\mathbf{h}_{il_1}^*\mathbf{h}_{il_2}\mathbf{h}_{il_3}^*\mathbf{h}_{il_4}]\mathbb{E}[\mathbf{h}_{k_1l_1}\mathbf{h}_{k_1l_2}^*\mathbf{h}_{k_1l_3}\mathbf{h}_{k_1l_4}^*]/M_r^2 \quad \text{(B.10)}\\
&= \frac{1}{M_r^2}\Bigg\{\sum_{l_1}\mathbb{E}[|\mathbf{h}_{il_1}|^4]\mathbb{E}[|\mathbf{h}_{k_1l_1}|^4] + 2\sum_{l_1,l_3\neq l_1}\mathbb{E}[|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_3}|^2]\mathbb{E}[|\mathbf{h}_{k_1l_1}|^2|\mathbf{h}_{k_1l_3}|^2]\Bigg\}\\
&= 2\sum_{l_1,l_2}\mathbb{E}|\mathbf{h}_{il_1}|^2|\mathbf{h}_{il_2}|^2 = 2(1+1/M_r)
\end{aligned}$$

Using (B.3–B.6) and (B.8–B.10) in (B.7) we get $\mathbb{E}[V_{k_1}^2V_{k_2}V_{k_3}] = O(1/M_r)$. By similar algebra we can show the other relations given in (A.2). In these expressions we use similar combinatorial count arguments and the fact that for $\mathbf{h}_{il} \sim \mathbb{C}\mathcal{N}(0,1)$, $\mathbb{E}|\mathbf{h}_{il}|^{2p}$ is a known finite constant for $p = 1,\ldots,4$.

Appendix C Details of Proposition III.3

Proposition C.1 *If* $\lim_{N\to\infty}\sum_{i=1}^{N}\sum_{j=1}^{i}|U_j|^2(C/N^2) = \lambda$ *and* $\lim_{N\to\infty}\max_{1\le i\le N}\sum_{j=1}^{i}|U_j|^2(C/N^2) = 0$ *then* $\lim_{N\to\infty}\sum_{i=1}^{N}\log(1+\sum_{j=1}^{i}|U_j|^2(C/N^2)) = \lambda$, *where* C *is any constant.*

Proof: Let $b_{i,N} = \sum_{j=1}^{i}|U_j|^2(C/N^2)$ hence $\lim_{N\to\infty}\max_{1\le i\le N}|b_{i,N}| = 0$. Thus $\forall\epsilon > 0$ there is N_ϵ s.t., $|b_{i,N}| < \epsilon, \forall N \ge N_\epsilon$. Now, using the fact that $|\log(1+z)-z| \le |z|^2$ for $|z| \le 1/2$ (see for *e.g.* [37], 18.3(c)). Using $\epsilon < 1/2$, *s.t.*, $|b_{i,N}| < \epsilon < 1/2 \quad \forall N \ge N_\epsilon$ we have $|\log(1+b_{i,N}) - b_{i,N}| \le \epsilon^2$, and hence by letting $\epsilon = 1/N$ we obtain

$$\sum_{i=1}^{N} b_{i,N} - (1/N) \le \sum_{i=1}^{N}\log(1+b_{i,N}) \le \sum_{i=1}^{N} b_{i,N} + (1/N) \tag{C.1}$$

Hence by a sandwich argument we have the desired result. ∎

Proposition C.2 *If* $U_i \sim \mathbb{C}\mathcal{N}(0,1)$ *is i.i.d. then* $\lim_{N\to\infty}\sum_{i=1}^{N}\sum_{j=1}^{i}|U_j|^2(C/N^2) = C/2$, a.s.

Proof: By exchanging the two finite sums over i and j we have,

$$\begin{aligned}\sum_{i=1}^{N}\sum_{j=1}^{i}|U_j|^2(C/N^2) &= \sum_{j=1}^{N}\sum_{i=j}^{N}|U_j|^2(C/N^2) = \sum_{j=1}^{N}(N-j+1)|U_j|^2 C/N^2 \\ &= C\left[\frac{1}{N}\sum_{j=1}^{N}|U_j|^2 + \frac{1}{N^2}\sum_{j=1}^{N}|U_j|^2 - \frac{1}{N}\sum_{j=1}^{N}|U_j|^2(j/N)\right]\end{aligned} \tag{C.2}$$

Now consider,

$$\begin{aligned}\mathbb{E}[\frac{1}{N}\sum_{j=1}^{N}(|U_j|^2(j/N)-1/2)]^4 &\overset{(a)}{=} \frac{1}{N^4}\sum_{j=1}^{N}\mathbb{E}[|U_j|^2(j/N)-1/2]^4 \\ &+ \frac{1}{N^4}\sum_{j_1}\sum_{j_2\ne j_1}\mathbb{E}[|U_{j_1}|^2(j_1/N)-1/2]^2\mathbb{E}[|U_{j_2}|^2(j_2/N)-1/2]^2 \\ &\overset{(b)}{=} O(1/N^2)\end{aligned} \tag{C.3}$$

where (a) follows by the i.i.d. nature of $\{U_j\}$ and (b) follows because $\mathbb{E}[|U_j|^2]^4 < \infty$. Hence as in the proof of Proposition A.1, by using the Borel-Cantelli lemma we get,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{j=1}^{N}(|U_j|^2(j/N)-1/2) = 0, \text{a.s.} \tag{C.4}$$

Thus $\lim_{N\to\infty} \frac{1}{N} \sum_{j=1}^{N} |U_j|^2 (j/N) = 1/2$ a.s. Using this and $\lim_{N\to\infty} \sum_j |U_j|^2/N = 1$ a.s. (SLLN), in (C.2) we have the result. ∎

By using Propositions C.1 and C.2 we obtain,

$$\lim_{N\to\infty} \sum_i \log(1 + \frac{P}{\sigma^2 N^2} \sum_{j=1}^{i} |U_j|^2) = P/2\sigma^2, \text{a.s.} \tag{C.5}$$

It is easy to show that $\{\sum_i \log(1 + \frac{P}{\sigma^2 N^2} \sum_{j=1}^{i} |U_j|^2)\}$ is uniformly integrable (using arguments similar to that used in Appendix A) and thus we can exchange limits and expectation and obtain the result of Proposition III.3.

Appendix D

In this section we outline the algebra leading to (44) and (45). If $\mathbf{q}_s = [1, \ldots, exp(j2\pi s(N-1)/N)]^T$ we can show using (34) that

$$(\mathbf{H}\mathbf{q}_s)_r = \sum_{l=0}^{\nu-1} h(r;l)\exp(j2\pi s((r-l))_N/N) \tag{D.1}$$

where $((\cdot))_N$ denotes modulo-N addition. Hence as $G(m,s) = \mathbf{q}_m^H \mathbf{H}\mathbf{q}_s/N$ and $\exp(j2\pi s((r-l))_N/N) = \exp(j2\pi s(r-l)/N)$ we can write,

$$G(m,s) = \frac{1}{N} \sum_{r=0}^{N-1} \sum_{l=0}^{\nu-1} h(r;l)\exp(j2\pi r(s-m)/N)\exp(-j2\pi sl/N) \tag{D.2}$$

Now, using the WSSUS assumption we have $\mathbb{E}[h(r_1;l_1)h(r_2;l_2)] = r_h(r_1 - r_2)\delta(l_1 - l_2)$ and using this and (D.2) we get (44).

Next, we outline the steps leading to (45). From (D.2) it is clear that $\mathbf{g}_m = [G(m,0), \ldots, G(m,N-1)]^T \sim \mathbb{C}\mathcal{N}(0, \mathbf{R}_1)$ where $\mathbf{R}_1$ is determined from (44). Similarly as $\bar{\mathbf{g}}_m \sim \mathbb{C}\mathcal{N}(0, \mathbf{R}_2)$, our problem in evaluating (43) reduces to finding $\mathbb{E}\log(\sigma^2 + P||\mathbf{w}||^2)$ for $\mathbf{w} \sim \mathbb{C}\mathcal{N}(0, \mathbf{R})$. To this end we can write $\mathbf{w}^H\mathbf{w} = \psi^H \mathbf{R}\psi = \tilde{\psi}\Lambda\tilde{\psi}$, where $\psi \sim \mathbb{C}\mathcal{N}(0, \mathbf{I})$, $\tilde{\psi} \sim \mathbb{C}\mathcal{N}(0, \mathbf{I})$ and $\mathbf{R} = \mathbf{U}\Lambda\mathbf{U}^H$ is its eigendecomposition. Thus $\mathbf{w}^H\mathbf{w} = \sum_q \lambda_q |\psi_q|^2$ where $\{\lambda\}$ are the eigenvalues of $\mathbf{R}$. As $\{\tilde{\psi}\}$ are independent Gaussians we can write the characteristic function of $\mathbf{w}^H\mathbf{w}$ as

$$\Phi(\omega) = \prod_q \frac{1}{1 - j\omega\lambda_q} \stackrel{(a)}{=} \sum_q \frac{\delta_q}{1 - j\omega\lambda_q} \tag{D.3}$$

where we have assumed distinct eigenvalues to get the partial fraction expansion (a). Thus the probability density function of $\mathbf{w}^H\mathbf{w}$ is,

$$f_{\mathbf{w}^H\mathbf{w}}(\beta) = \sum_q (\delta_q/\lambda_q)\exp(-\beta/\lambda_q), \beta \geq 0. \tag{D.4}$$

Hence,

$$\mathbb{E}[\log(\sigma^2 + P||\mathbf{w}||^2) = \sum_q \delta_q \int_0^\infty \frac{1}{\lambda_q} \log(\sigma^2 + P\beta^2) \exp(-\beta/\lambda_q) d\beta \tag{D.5}$$

Using the fact that $\int_0^\infty \frac{1}{\lambda_q} \log(\sigma^2 + P\beta^2) \exp(-\beta/\lambda_q) d\beta = -\exp(\sigma^2/P\lambda_q) E_i(-\sigma^2/P\lambda_q)$ [35] where $E_i(\cdot)$ is the exponential integral function we obtain (45).

References

[1] S. N. Diggavi, "On achievable performance over spatial diversity fading channels," in *Proc. International Symposium on Information Theory*, p. 396, August 1998.

[2] S. N. Diggavi, "Analysis of multicarrier transmission in time-varying channels," in *Proc. ICC*, pp. 1191–1195, June 1997.

[3] J. G. Proakis, *Digital Communications.* New York: McGraw Hill, 3 ed., 1995.

[4] L. H. Ozarow, S. Shamai, and A. D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Transactions on Vehicular Technology*, vol. 43, pp. 359–378, May 1994.

[5] G. Kaplan and S. Shamai, "Achievable performance over the correlated Rician channel," *IEEE Transactions on Communications*, vol. 42, pp. 2967–2978, Nov 1994.

[6] S. Shamai and A. D. Wyner, "Information-Theoretic Considerations for symmetric, Cellular, Multiple Access Fading Channels, Parts I,II," *IEEE Transactions on Information Theory*, vol. 43, pp. 1877–1911, November 1997.

[7] G. Caire, G. Taricco, J. Ventura-Traveset, and E. Biglieri, "A multiuser approach to narrowband cellular communications," *IEEE Transactions on Information Theory*, vol. 43, pp. 1503–1517, September 1997.

[8] G. Wornell and M. Trott, "Efficient Signal Processing techniques for exploiting transmit antenna diversity on fading channels," *IEEE Transactions on Signal Processing*, vol. 45, pp. 191–205, Jan 1997.

[9] A. Narula, M. D. Trott, and G. W. Wornell, "Information-theoretic analysis of multiple-antenna transmission diversity for fading channels," in *Proceedings International Symposium on Information Theory and its Applications(ISITA'96)*, 1996.

[10] G. Foschini, "Layered space-time architecture for wireless communication in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, pp. 41–59, September 1996.

[11] V. Tarokh, N. Seshadri, and A. Calderbank, "Space-time codes for high data rate wireless communications: Performance criterion and code construction," *IEEE Transactions on Information Theory*, vol. 44, pp. 744–765, March 1998.

[12] D. Tse Private communications, October 1997.

[13] I. E. Telatar, "Capacity of multiple antenna gaussian channels." AT&T Technical Memorandum, 1995.

[14] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication.," *IEEE Transactions on Communications*, vol. 46, pp. 357–366, March 1998.

[15] T. Marzetta and B. M. Hochwald, "Fundamental limitations on multiple-antenna wireless links in Rayleigh fading," in *Proc. IEEE International Symposium on Information Theory (ISIT), Cambridge, Massachussetts*, p. 310, 1998.

[16] J. Yang and S. Roy, "On joint transmitter and receiver optimization for multiple-input-multiple-output (mimo) transmission systems," *IEEE Transactions on Communications*, vol. 42, pp. 3221–3231, December 1994.

[17] P. Balaban and J. Salz, "Optimum diversity combining and equalization in digital data transmission with applications to cellular mobile radio," *IEEE Transactions on Communications*, vol. 40, pp. 885–894, May 1992.

[18] A. Goldsmith, *Design and performance of high-speed communication systems over time-varying radio channels.* PhD thesis, University of California, Berkeley, CA., 1994.

[19] M. Medard, *Capacity of time-varying multiple user channels in wireless communications.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1995.

[20] L. J. Cimini, B. Daneshrad, and N. R. Sollenberger, "Clustered OFDM with transmitter diversity and coding," in *GLOBECOM*, pp. 703–707, 1996.

[21] T. Kailath, "Channel characterization: time-variant dispersive channels," in *Lectures on Communication System Theory* (E. J. Baghdady, ed.), pp. 95–123, New York: McGraw-Hill Book Co., 1961.

[22] W. C. Jakes, *Microwave Mobile Communications.* New York: John Wiley and Sons, 1974.

[23] R. J. Muirhead, *Aspects of multivariate statistical theory.* New York: John Wiley and Sons, 1982.

[24] A. Edelman, *Eigenvalues and Condition numbers of random matrices.* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, 1989.

[25] S. Verdu, *Multiuser Detection.* United Kingdom: Cambridge University Press, 1998.

[26] T. M. Cover and J. A. Thomas, *Elements of Information Theory.* New York: John Wiley and Sons, Inc., 1991.

[27] J. L. Massey, "Coding and modulation in digital communications," in *Proc.International Zurich seminar on digital communications*, pp. E2(1)–E2(4), 1974.

[28] J. W. Brewer, "Kronecker products and Matrix Calculus in System Theory," *IEEE Transactions on Circuits and systems*, vol. 25, pp. 772–781, September 1978.

[29] W. Hirt and J. L. Massey, "Capacity of the discrete-time gaussian channel with inter-symbol interference," *IEEE Transactions on Information Theory*, vol. 34, pp. 380–388, May 1988.

[30] J. M. Cioffi, "Digital data communications." In preparation, book to be published, notes for class EE 379, Stanford University, 1994.

[31] R. G. Gallager, *Information theory and reliable communications*. New York: John Wiley and Sons, 1968.

[32] G. D. Forney, "Maximum-Likelihood Sequence Estimation of Digital Sequences in the Presence of Intersymbol Interference," *IEEE Transactions on Information Theory*, vol. 18, pp. 363–378, May 1972.

[33] J. H. Winters, J. Salz, and R. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Transactions on Communications*, vol. 42, pp. 1740–1751, Feb/Mar/Apr 1994.

[34] S. N. Diggavi, B. C. Ng, and A. Paulraj, "Joint channel-data estimation and interference suppression," in *International Conference on Communications ICC*, pp. 465–469, June 1998.

[35] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series and Products*. San Diego: Academic Press, 1994.

[36] J. Ventura-Traveset, G. Caire, E. Biglieri, and G. Taricco, "Impact of diversity reception in fading channels with coded modulation – Part I: Coherent detection," *IEEE Transactions on Communications*, vol. 45, pp. 563–572, May 1997.

[37] D. Williams, *Probability with Martingales*. New York: Cambridge University Press, 1991.

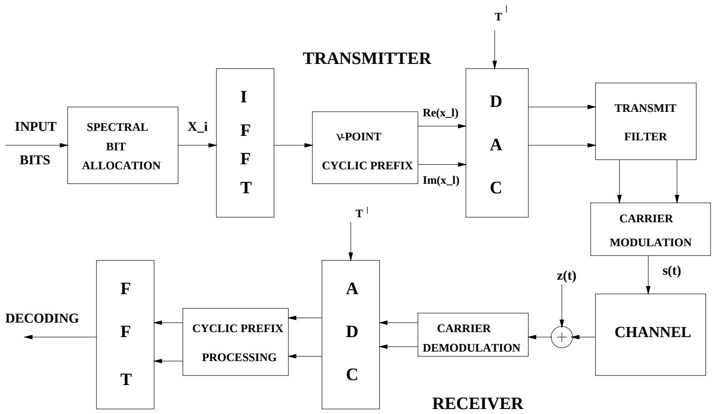

Figure 1: An OFDM based transmission scheme

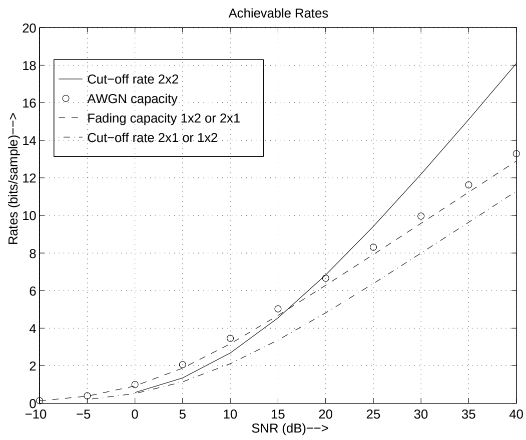

Figure 2: Mutual information and cut-off rates for fading diversity channels.

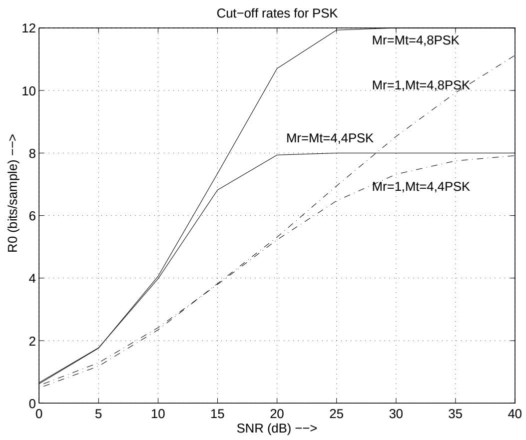

Figure 3: Cut-off rate for 4PSK and 8PSK modulations.

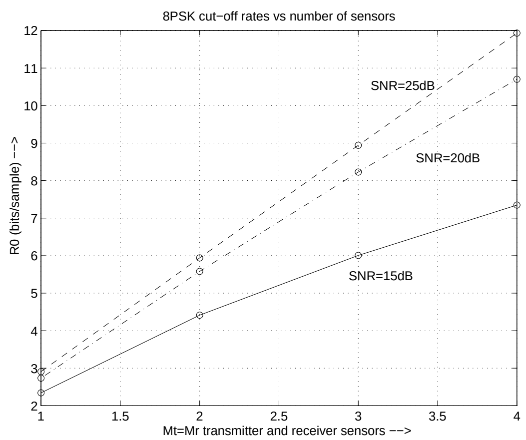

Figure 4: Cut-off rate vs number of sensors.

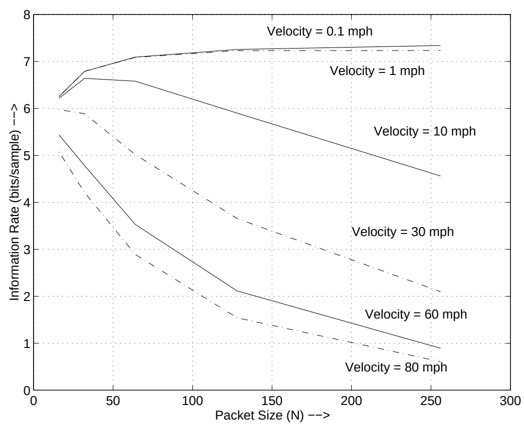

Figure 5: Information rates for various block sizes and Doppler shifts.

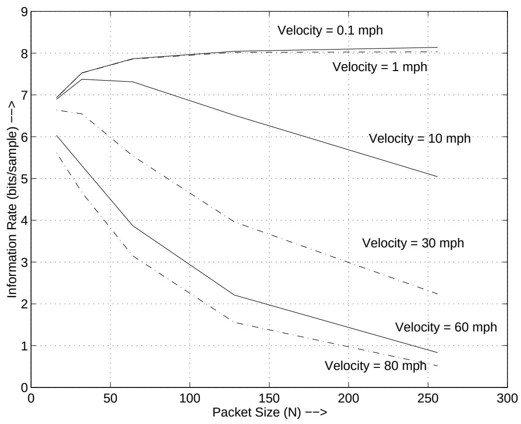

Figure 6: Information rates with large diversity for various block sizes and Doppler shifts.