Asymptotic Capacity of Underspread and Overspread Stationary Time- and Frequency-Selective Channels

Umer Salim and Dirk Slock
Mobile Communications Department
Eurecom
BP 193, 06904 Sophia Antipolis Cedex, France
{umer.salim,dirk.slock}@eurecom.fr

Abstract— **In this paper, we consider stationary time- and frequency-selective channels. No channel knowledge neither at the transmitter nor at the receiver is assumed to be available. We investigate the capacity behavior of these doubly selective channels as a function of the channel parameters delay spread, Doppler bandwidth and channel spread factor (the product of the delay spread and the Doppler bandwidth). We shed light on different capacity regimes at high values of signal to noise ratio (SNR) in which the dominant capacity term is either of order log(SNR) or log(log(SNR)), depending on the channel conditions (delay spread, Doppler Bandwidth and channel spread factor). For critically spread channels (channel spread factor of 1), it is widely believed that the dominant term of the high-SNR expansion of the capacity is of order log(log(SNR)) or in other words, that the pre-log (the coefficient of log(SNR)) is zero. We provide a very simple scheme that shows that even for critically spread channels a non-zero pre-log might exist under certain conditions. We also specify these conditions in terms of Doppler bandwidth and delay spread. We also show that a non-zero pre-log might exist even for over-spread channels (channel spread factor greater than 1). We specify the channel conditions which govern the range of existence of the log(SNR) regime. At higher channel spread factor, the log(SNR) term vanishes and a log(log(SNR)) term becomes the dominant capacity term. We specify the range of this log(log(SNR)) regime and also provide bounds for the coefficient of this log(log(SNR)) term (the pre-loglog).**

I. Introduction

Information theoretic capacity analysis for different types of channel models started with the somewhat unusual assumption that the channel is perfectly known at the receiver (channel state information at the receiver (CSIR)) or sometimes even assuming that the channel is known at the transmitter (channel state information at the transmitter (CSIT)). But inherently all channels are non-coherent in nature and they need some kind of estimation to get CSIR and then some kind of feedback and/or estimation to have CSIT. The area of capacity analysis for non-coherent (no CSIR nor CSIT) fading channels has received considerable attention in recent years.

Usually block fading models are assumed for obtaining capacity bounds in the no CSIR case. In the standard version of this model [1], the fading remains constant over blocks consisting of T symbol periods, and changes independently from block to block. Capacity bounds are obtained by introducing training segments in an ad hoc fashion. For the standard block fading model, the capacity is shown [1], [2] to grow logarithmically with SNR at high values of SNR, thus log(SNR) was shown to be the dominant term of capacity. Later Liang and Veeravalli [3] allowed the fading to vary inside the block with a certain correlation matrix characterized by its rank Q and showed for SISO channels that the capacity pre-log is $(1-Q/T)$. For block constant frequency selective channels with L taps, the pre-log was shown to be $(1-L/T)$ in [4].

Non-coherent capacity has also been analyzed with the channel fading process taken to be (symbol-by-symbol) stationary. In this model, fading is not independent but time selective without any block structure. Surprisingly, this model leads to very different capacity results: contrary to log(SNR) capacity growth in block fading channels, here the capacity grows only double logarithmically with SNR at high values of SNR [5], [6], [7] when the fading process is non-bandlimited (the Doppler spectrum spans the full transmission bandwidth; in this case the channel prediction error is non-zero even if the infinite channel past is known.

For symbol-by-symbol stationary Gaussian fading channels, if the Doppler spectrum is bandlimited (of limited support), then the fading process is called non-regular and the prediction error given the infinite past goes to zero. Lapidoth [8] studied the SISO case for this kind of fading processes showing that the capacity grows logarithmically with SNR and the capacity pre-log is the Lebesgue measure of the frequencies where the spectral density of the fading process (Doppler spectrum) has nulls.

Etkin and Tse [9] study the same channel model of bandlimited fading for MIMO systems; they show that the pre-log exists even for MIMO systems with no CSIR but they only give a lower bound of the capacity pre-log.

All of the above mentioned studies except [4] deal with flat fading, so the (symbol rate) discrete-time channel response filter has a single non-zero tap, varying at each time instant with the Doppler spread of the channel. We are interested in studying non-coherent doubly selective channels where the channel has multiple taps varying in time as stationary processes characterized by a (Doppler) spectrum. Their coherent counterparts have a pre-log of one. For such non-coherent channels under the strictly underspread assumption, we show that the loss in pre-log is equal to the spread factor of the channel (the product of the delay spread and the Doppler bandwidth). This result is not counter-intuitive as channel spread in time or frequency introduces more channel parameters that need to be estimated for (coherent) detection of the data. This result shows that the pre-log should be zero when channel spread factor becomes 1 but we present a simple scheme which shows the existence of log(SNR) for overspread channels.

We should emphasize that the fading processes considered in this paper are stationary and doubly selective. The rest of the paper is organized as follows. In sections II and III, we provide the system model and its representation using a basis expansion model (BEM) for the channel. Section IV presents the capacity analysis for underspread channels. In section V, a simple transmission scheme is introduced showing the existence of the pre-log for overspread channels, with the associated conditions for the existence of this pre-log. In section VI, we discuss the optimality of our transmission scheme. Then in section VII, we specify the boundaries of the high SNR capacity regimes of log(SNR) and log(log(SNR)) and give simple bounds for the pre-loglog factor. The section VIII provides an analogy between transmission for frequency-selective SISO and for frequency-flat MIMO chanels. The paper ends with some concluding remarks in section IX.

Notation: $\mathbb{E}$ denotes statistical expectation. Lowercase letters represent scalars, boldface lowercase letters represent vectors, and boldface uppercase letters denote matrices. $\mathbf{A}^\dagger$ ($\mathbf{A}^\mathbf{T}$) denotes the Hermitian (transpose) of matrix $\mathbf{A}$. The determinant of $\mathbf{A}$ is denoted as $|\mathbf{A}|$.

II. System Model

We consider a discrete-time single input single output (SISO) fading channel at symbol rate, having L taps whose time-k output $y[k] \in \mathbb{C}$ is given by

$$y[k] = \sqrt{\mathrm{SNR}} \sum_{l=0}^{L-1} h[k,l]x[k-l] + z[k] \quad (1)$$

where $x[k] \in \mathbb{C}$ denotes the time-k channel input, the complex scalar $h[k,l] \in \mathbb{C}$ represents the l-th coefficient of the FIR (finite impulse response) channel filter at time k consisting of circularly symmetric complex Gaussian components of zero mean and unit variance, and $z[k] \in \mathbb{C}$ denotes the additive white Gaussian noise. Here $\mathbb{C}$ denotes the complex field.

The channel fading process $\{h[k,l]\}$ for each tap l is assumed to be stationary, ergodic and bandlimited. They are independent and identically distributed (i.i.d.) across different taps l. The hypthesis of the bandlimitedness of the fading process is motivated by the physical limitations on the mobile speed. For a mobile speed v, the maximum Doppler frequency magnitude f_{max} for each path is $f_{max} = v/\lambda_c$ where λ_c is the carrier wavelength. The bandwidth of each fading process will be upper bounded by the two-sided Doppler bandwidth $2f_{max}$. We define the normalized Doppler bandwidth as $B_d = 2f_{max}T_s$ where T_s represents the symbol period, assuming the Doppler spectrum has support between the two extreme Doppler shifts. In general, B_d will denote the support of the Doppler spectrum. The hypothesis of bandlimited Doppler spectrum is an approximation because the Doppler shifts do not remain constant. Similarly, the hypothesis of limited delay spread is an approximation. Limited values for Doppler and delay spreads can be justified at a given working SNR.

The system is normalized so that the channel input has an average power constraint of $\mathbb{E}[|x[k]|^2] \leq 1$.

The capacity pre-log is normally defined as

$$PreLog = \lim_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log(\mathrm{SNR})} \quad (2)$$

whevever $C(\mathrm{SNR})$ is of order $\log(\mathrm{SNR})$, and the capacity pre-loglog is given by

$$PreLogLog = \lim_{\mathrm{SNR}\to\infty} \frac{C(\mathrm{SNR})}{\log(\log(\mathrm{SNR}))} \quad (3)$$

whevever $C(\mathrm{SNR})$ is of order $\log(\log(\mathrm{SNR}))$.

III. Representation Using Basis Expansion Model

We shall assume here w.l.o.g. that the Doppler spectrum is contiguous and that the demodulation is synchronized to the lower edge of the Doppler Spectrum. To get a proper model for the doubly selective channel, we start by considering block transmission with block length N. Continuous transmission results will then be obtained by letting the block size N grow to infinity. Observing a signal over a block can always be thought of as if the block considered is one period of a periodic process, in which case the signal has a Fourier series expansion. This leads to a Basis Expansion Model (BEM) for the time-varying channel coefficients in which the basis functions are complex exponentials with frequencies at the multiples of $1/N$ [10]. As the Doppler spectrum is bandlimited, we shall take the BEM to be correspondingly bandlimited. We should note here that we do not necessarily demand of the BEM to provide an exact description of the channel statistics over the block of length N, as long as the description becomes exact as the block length tends to infinity. The BEM leads to the following representation for the channel coefficients over a block that start at time zero w.l.o.g.,

$$h[k,l] = \sum_{n=0}^{N_d-1} g[n,l]e^{j2\pi kn/N} \quad , \quad k = 0, 1, \cdots, N-1 \quad (4)$$

where $N_d = \lceil N B_d \rceil$. In the above equation, $g[n,l]$ are independent, uncorrelated, zero mean proper complex Gaussian random variables whose variances are the values of the spectrum of the corresponding fading process at the respective frequencies n/N. If the block transmission is alternatively thought of as an isolated block (instead of a period of a periodic process), then the windowing in time domain with a rectangular block of size N leads to an interpolation in frequency domain between the frequencies n/N with $\frac{\sin \pi N f}{N \sin \pi f}$

$$\mathbf{H} = \begin{bmatrix} h[0,0] & & h[0,L{-}1] & \cdots & h[0,1] \\ \vdots & h[1,0] & & \ddots & \vdots \\ \vdots & & & & h[L{-}2,L{-}1] \\ h[L{-}1,L{-}1] & \vdots & & & \\ & h[L,L{-}1] & \ddots & & \\ & & \ddots & & \\ & & & & h[N{-}1,0] \end{bmatrix}$$

which leads to something non-bandlimited, as indeed a signal cannot be both time- and bandlimited. However, the process becomes bandlimited as the block size N tends to infinity (see also [3]). To avoid inter-block interference and facilitate the description in the frequency-domain, we add a cyclic prefix of length $L-1$ making the total block length $N+L-1$. At the receiver the first $L-1$ received samples corresponding to the prefix get neglected and the remaining N outputs, the inputs and the noise get collected in vector form as $\mathbf{y} = [y[0]\, y[1]\, \cdots\, y[N-1]]^T$, $\mathbf{x} = [x[0]\, x[1]\, \cdots\, x[N-1]]^T$, $\mathbf{z} = [z[0]\, z[1]\, \cdots\, z[N-1]]^T$, leading to the system equation

$$\mathbf{y} = \sqrt{\text{SNR}}\,\mathbf{H}\mathbf{x} + \mathbf{z} \tag{5}$$

where $\mathbf{H} \in \mathbb{C}^{N\times(N+L-1)}$ is the channel matrix for this block and has the circulant structure shown at the top of the page, resulting from the equality $[x[-(L{-}1)]\, \cdots\, x[-1]] = [x[N{-}(L{-}1)]\, \cdots\, x[N{-}1]]$.

We also need a system representation in which the roles of channel and input are reversed. For this, we define a diagonal matrix $\mathbf{X_i} = diag(x[i], x[i+1], \cdots, x[i+N-1])$ and $\mathbf{X} = [\mathbf{X_0}\ \mathbf{X_{-1}}\ \cdots\ \mathbf{X_{-(L-1)}}]$. Hence $\mathbf{X} \in \mathbb{C}^{N\times(NL)}$ is the system input for one block of length N. If $\mathbf{h_l} = [h[0,l]\, h[1,l] \cdots h[N{-}1,l]]^T$ and $\mathbf{h} = [\mathbf{h_0}^T\, \mathbf{h_1}^T \cdots \mathbf{h_{L-1}}^T]^T$ then (5) can be written as

$$\mathbf{y} = \sqrt{\text{SNR}}\,\mathbf{X}\mathbf{h} + \mathbf{z} \tag{6}$$

Similarly by putting uncorrelated coefficients of BEM in vectors $\mathbf{g_l} = [g[0,l]\ \cdots\ g[N_d,l]]^T$, (4) takes the form of $\mathbf{h_l} = \mathbf{F}\mathbf{g_l}$ where $\mathbf{F} \in \mathbb{C}^{N\times N_d}$ is the (partial) IDFT matrix. By regrouping BEM coefficients of all channel taps in a vector $\mathbf{g} = [\mathbf{g_0^T}\ \mathbf{g_1^T}\ \cdots\ \mathbf{g_{L-1}^T}]^T$, we can write

$$\mathbf{h} = \mathbf{F_c}\mathbf{g} \tag{7}$$

where $\mathbf{F_c} = \mathbf{I_L} \otimes \mathbf{F}$ and $\otimes$ represents the Kronecker product. With this (6) can be written as

$$\mathbf{y} = \sqrt{\text{SNR}}\,\mathbf{X}\mathbf{F_c}\mathbf{g} + \mathbf{z} \tag{8}$$

IV. Underspread Channels

Typically wireless channels are underspread in nature [11], so first of all we study the capacity pre-log for doubly selective channels when they are underspread (the product of the delay spread and the normalized Doppler bandwidth is strictly less than one). We derive lower and upper bounds for the mutual information of non-coherent doubly selective channels and specify the corresponding pre-log.

A. Lower Bound of Mutual Information

Using the BEM developed in section III, we give a lower bound for the mutual information in Appendix A. System input is selected as Gaussian i.i.d. satisfying the average power constraint imposed and the result is

$$\lim_{\text{SNR}\to\infty} \frac{1}{N} I(\mathbf{x};\mathbf{y}) \geq \left(1 - \frac{LN_d}{N}\right)\log(\text{SNR}) + O(1) \tag{9}$$

where $O(1)$ represents a term which does not grow with SNR.

B. Upper Bound of Mutual Information

The upper bound of the mutual information for our time- and frequency selective channel model is given in Appendix B. The main point in the derivation is the intelligent splitting of the mutual information in two parts, where one term gives the growth with $\log(\text{SNR})$ as given by a coherent channel and the other term is shown to have no growth with $\log(\text{SNR})$.

$$\lim_{\text{SNR}\to\infty} \frac{1}{N} I(\mathbf{x};\mathbf{y}) \leq \left(1 - \frac{LN_d}{N}\right)\log(\text{SNR}) + O(1) \tag{10}$$

where $O(1)$ indicates that there might be lower order terms which depend upon SNR but they are negligible as compared to log(SNR) at very high values of SNR.

C. Pre-Log and its Large Block Length Asymptotic

Based upon the above two bounds on the mutual information of strictly underspread channels, one can conclude that the pre-log is given by

$$PreLog = \left(1 - \frac{LN_d}{N}\right) \tag{11}$$

Now we can also let the block length N go to infinity. The factor N_d which is the total number of Fourier coefficients required to describe a single channel tap over block length N has its dependence upon N and the limiting value of N_d/N with large block length turns out to be $2f_{max}T_s$, a quantity we described as the normalized Doppler bandwidth in section II. So the capacity pre-log for underspread channels becomes

$$PreLog = 1 - LB_d \tag{12}$$

It shows that the loss in pre-log for a non-coherent SISO channel is equal to the channel spread factor which is the average number of channel parameters per symbol time that can parameterize the channel.

V. Overspread Channels

In this section we treat the case of a channel which is overspread. Hence the channel spread factor (the product of the delay spread of the channel and the normalized Doppler bandwidth) is greater than one which would imply that the pre-log obtained in the previous section for such doubly selective channels $(1-LB_d)$ becomes zero. In fact according to the pre-log expression of $(1-LB_d)$, the pre-log will become zero as soon as the channel is critically spread ($LB_d = 1$). Below we give a very simple scheme which shows that the log(SNR) term exists for overspread channels under certain conditions.

A. Transmission Scheme

Our transmission scheme to realize log(SNR) growth for overspread channels is based upon zero padding. The zero padding is done in such a manner that at the receiver side, each transmitted symbol appears without inter-symbol interference (ISI) for at least one symbol time. So to achieve this one output sample free of ISI, we transmit an input symbol and then do zero padding of $\lfloor L/2 \rfloor$ symbols. That means each information symbol is followed by $\lfloor L/2 \rfloor$ deterministic zeros. If we analyze carefully, after transmission of one particular symbol at the transmitter it appears with no ISI at $(\lfloor L/2 \rfloor + 1)$-th symbol instant. Now one may focus attention on the input information symbols transmitted at the transmitter and the ISI free received symbols at the receiver delayed by $(\lfloor L/2 \rfloor + 1)$ symbol intervals. For this scheme $\lfloor L/2 \rfloor$ input symbols are wasted (zero-padded) corresponding to each single information symbol transmitted but the good thing is that the effective channel is frequency flat and each ISI free symbol at the receiver comes multiplied with the same channel tap, the $(\lfloor L/2 \rfloor + 1)$-th tap. This scheme is explained in Figure 1.

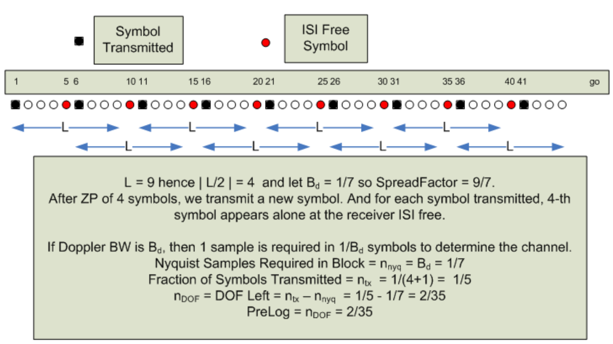

Fig. 1. Transmission Scheme Example for Overspread Channels

From figure, it is clear that each time at the selected output sample this is the same channel tap which appears multiplied with the input symbol. Hence effectively we are using a frequency flat channel with no ISI but increased Doppler bandwidth due to zero padding of symbols at the transmitter side. Now we need to see what fraction of symbols we are able to transmit in this zero-padded scheme where $\lfloor L/2 \rfloor$ symbols get wasted for each single information symbol. So the fraction of the information symbols is

$$n_{tx} = \frac{1}{\lfloor L/2 \rfloor + 1} \tag{13}$$

Now keeping in mind that here we are interested in only a single channel tap (which appears with ISI free output symbol) requiring $N_d = 2M+1$ BEM coefficients to be estimated to be fully known over a block length N as we argued in section III. And to estimate a single channel tap, per symbol coefficients required N_d/N was shown to be equal to the normalized Doppler bandwidth B_d in section IV-C. We denote this fraction by n_{nyq}, the minimum number of samples required to estimate the channel

$$n_{nyq} = \lim_{N\to\infty} \frac{N_d}{N} = B_d \tag{14}$$

If we want to estimate the channel by sending pilot symbols, we need to transmit B_d fraction of pilots among the non-zero transmit symbols and then this particular channel tap can be estimated by estimating its BEM coefficients. But in this scheme, the total number of information symbols transmitted is the fraction $1/(\lfloor L/2 \rfloor + 1)$ per unit length. Now there is the possibility that some degrees of freedom (DOF) are left even after estimating this particular channel tap but it will be depending upon the relative values of the channel delay spread L and the normalized Doppler bandwidth B_d.

$$n_{DOF} = n_{tx} - n_{nyq} = \frac{1}{\lfloor \frac{L}{2} \rfloor + 1} - B_d \tag{15}$$

So we can have coherent transmission albeit with imperfect channel estimate over this fraction n_{DOF} (if this number is non-zero, of course) and so it corresponds to a coherent channel where pre-log exists. Hence pre-log per symbol time is given by

$$PreLog = n_{DOF} = \frac{1}{\lfloor \frac{L}{2} \rfloor + 1}\left(1 - B_d(\lfloor \frac{L}{2} \rfloor + 1)\right) \tag{16}$$

Formal information theoretic proof for the achievability of the above pre-log for overspread channels has been given in Appendix C.

B. Conditions for the Existence of the PreLog

First of all, the Doppler spectrum should not be of full support i.e. the normalized Doppler bandwidth should not be 1. So the condition is to have $B_d \leq 1$. If normalized Doppler bandwidth is one, even for frequency flat channels, channel estimation becomes impossible hence coherent regime can never come into play and the log(SNR) term does not exist [8].

But our scheme gives more strict restriction on the normalized Doppler bandwidth. In our zero-padded transmission scheme, we transmit a fraction $1/(\lfloor L/2 \rfloor + 1)$ number of symbols and the fractional number of Nyquist samples required for minimal channel representation is B_d. Hence the number of transmitted symbols over any block length should be greater than Nyquist symbols required to have some positive DOF where coherent operation can be carried out to obtain capacity growth with log(SNR). So this gives us the condition

$$B_d \leq \frac{1}{\lfloor \frac{L}{2} \rfloor + 1} \tag{17}$$

We can also find out the channel parameter values where the pre-log given by the zero-padded transmission scheme surpasses the pre-log $(1 - LB_d)$ derived in section IV-C. This gives us a lower bound on the normalized Doppler bandwidth. Combining this lower bound with the upper bound given above, we get

$$\frac{\lfloor \frac{L}{2} \rfloor}{(\lfloor \frac{L}{2} \rfloor + 1)(L-1)} \leq B_d \leq \frac{1}{\lfloor \frac{L}{2} \rfloor + 1} \tag{18}$$

The left inequality shows the condition for an underspread channel where the pre-log of this zero-padding scheme takes over the classical pre-log of $(1-LB_d)$ and the right inequality shows the condition under which an overspread channel shows positive pre-log with this scheme. The multiplication of the above inequality with L gives us the corresponding bounds on the channel spread factor.

VI. Optimality of Zero Padded Transmission Scheme

In our transmission scheme with zero padding, we transmit one information symbol in each block of $(\lfloor L/2 \rfloor + 1)$ symbols. One can argue if more than 1 symbol is transmitted and zero padding of the same size is done, there might be the possibility of having more DOF and resultantly a higher pre-log factor. In Figure 2, we explain this modified transmission scheme and develop generalized pre-log expressions when more symbols are transmitted and from this analysis we show the optimality of the scheme proposed in V-A.

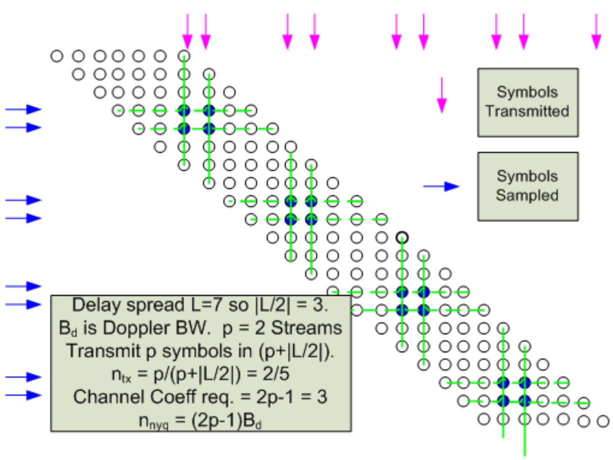

Fig. 2. Channel Matrix with Modified Transmission Scheme

As we transmit p symbols and do zero padding of $\lfloor L/2 \rfloor$ symbols so the fraction of the symbols transmitted is

$$n_{tx} = \frac{p}{\lfloor L/2 \rfloor + p} \tag{19}$$

From the figure, we see that if p symbols are transmitted, to detect these p symbols at the receiver involves the estimation of at least $2p-1$ channel taps. For one channel tap, the BEM coefficients required per symbol interval is B_d. Hence for our case where we have $2p-1$ channel taps involved, the number of observations required is

$$n_{nyq} = (2p-1)B_d \tag{20}$$

So DOF or the pre-log (the number of symbols available for coherent detection after estimating the minimum required Nyquist samples) is given by

$$PreLog = n_{tx} - n_{nyq} = \frac{p}{\lfloor L/2 \rfloor + p} - (2p-1)B_d \tag{21}$$

A. Optimality for Critically Spread Channels

The above expression of pre-log can be specialized to critically spread channels (the spread factor of 1) which gives $B_d = 1/L$. Hence in that case, the pre-log is given by

$$PreLog = \begin{cases} \frac{L-(2p-1)^2}{2L(\lfloor L/2 \rfloor + p)} & \text{for L odd integer} \\ \frac{L-2p(2p-1)}{2L(\lfloor L/2 \rfloor + p)} & \text{for L even integer} \end{cases}$$

This expression of pre-log gets maximized for $p = 1$ which gives the transmission scheme given in section V-A hence proving the optimality of our scheme at spread factor of 1.

B. General Condition

If the scheme with p ($p \geq 2$) streams is better than zero-padding scheme described in the previous section, then its pre-log should be higher than the pre-log of that scheme (16) which gives us the following condition after some manipulations

$$B_d \leq \frac{1}{2(\lfloor \frac{L}{2} \rfloor + 1)} \frac{\lfloor \frac{L}{2} \rfloor}{\lfloor \frac{L}{2} \rfloor + p} \tag{22}$$

On the other hand, the pre-log of the scheme with p streams should also be higher than the conventional pre-log of $1-LB_d$ which gives another condition on B_d

$$B_d \geq \frac{1}{L+1-2p} \frac{\lfloor \frac{L}{2} \rfloor}{\lfloor \frac{L}{2} \rfloor + p} \tag{23}$$

Combining the above two inequalities with some algebra, we get the following condition

$$\frac{1}{L+1-2p} \leq B_d \left(\frac{\lfloor \frac{L}{2} \rfloor + p}{\lfloor \frac{L}{2} \rfloor} \right) \leq \frac{1}{2(\lfloor \frac{L}{2} \rfloor + 1)} \tag{24}$$

If we pick the terms on the extreme left and the extreme right of the above inequalities, after some manipulation we get

$$p \leq \frac{L}{2} - \lfloor \frac{L}{2} \rfloor - 1 \tag{25}$$

which is always false for any positive value of p. So we prove that the pre-log for any scheme with p information symbols and zero padding of $\lfloor L/2 \rfloor$ symbols never beats the conventional pre-log of $1 - LB_d$ and the pre-log of zero-padding scheme given in (16) at the same time.

Thus among all such schemes which may employ zero padding at higher spread factors to reduce the number of active channel taps, the scheme described in V-A is the optimal one.

C. The PreLog for The Transient Regime

For the channels with very high spread factors (nearly critically spread channels to overspread channels where range was specified in section V-B), the achievability of the pre-log was proved in Appendix C. We also showed some optimality conditions of this zero padding scheme in previous subsections. Although we don't have a proof for the upper bound of the pre-log for this transient regime but we conjecture that the pre-log given by this zero-padding scheme is also the upper bound of the pre-log in this regime. Hence for this transient regime, the actual pre-log is

$$PreLog = \frac{1}{\lfloor \frac{L}{2} \rfloor + 1} \left(1 - B_d(\lfloor \frac{L}{2} \rfloor + 1) \right) \tag{26}$$

VII. Different Regimes of Capacity with SNR

In this section we characterize the boundaries of different regimes like log(SNR) and log(log(SNR)). We showed the optimality of our zero padding transmission scheme among other schemes which may employ zero padding in the previous section. This scheme is an extreme case of zero padding and corresponds to the worst case scenario with higher spread factors so the boundaries of different capacity regimes like log(SNR) and log(log(SNR)) will be the same as given by this scheme.

A. Boundaries of log(SNR)

For underspread channels, the dominant term of capacity is log(SNR). From our scheme which we explained in the previous sections, we conclude that the log(SNR) regime will be there as long as the following condition is satisfied.

$$B_d \leq \frac{1}{\lfloor \frac{L}{2} \rfloor + 1} \tag{27}$$

It is important to mention that even overspread channels might satisfy this condition and in that case show capacity growth with log(SNR).

B. Boundaries of log(log(SNR))

The regime where the dominant term of the capacity is log(log(SNR)) starts when the logarithmic regime log(SNR) ends. Now we want to know when this double logarithmic regime also ceases to exist. According to our transmission scheme, no matter how large is the delay spread, we just do zero padding in a manner such that at the receiver side, we get at least one ISI free sample and the information symbols can be thought of passing through a frequency flat channel. Now as in zero padding, there are unused symbols so channel estimation might become impossible due to the increased Doppler spread. But even non-coherent detection can give us log(log(SNR)) growth of capacity. So according to our scheme, this log(log(SNR)) regime will only stop when channel faces an infinitely long delay spread. And when this is the case, high SNR capacity will be bounded giving no increase with increasing SNR.

C. Bounds for pre-loglog

For overspread channels in the regime where log(log(SNR)) becomes the dominant term, a trivial upper bound of the pre-loglog is 1 which is the pre-loglog factor of the critically spread non-coherent channels with $B_d = 1$ and $L = 1$.

A very trivial lower bound can be given by the scheme where only one symbol is transmitted in each block length of L, completely removing the ISI. In this scheme, each input symbol will appear at the output with no ISI and multiplied with one channel tap. Although for the pre-loglog point of view, it is sufficient to focus on a single tap. Due to zero padding with $L-1$ zeros, the pre-loglog will be $1/L$.

A less trivial lower bound on the pre-loglog is given by our zero padding transmission scheme explained in section V-A where one symbol is transmitted in each block of ($\lfloor L/2 \rfloor + 1$) symbols, hence giving

$$PreLogLog = \frac{1}{\lfloor L/2 \rfloor + 1} \tag{28}$$

VIII. Analogy with MIMO Systems

Suppose we are working with a MIMO system having n_t transmit and n_r receive antennas. And each channel coefficient of $n_r \times n_t$ MIMO matrix is frequency flat and has normalized Doppler bandwidth of B_d.

When the channel is largely underspread ($B_d < 1$), from [12] the pre-log of this MIMO system can be represented as

$$PreLog = n_t'(1 - n_t' B_d) \tag{29}$$

where n_t' is given by the following expression

$$n_t' = \min\left(n_t, n_r, \frac{1}{2B_d} \right) \tag{30}$$

n_t' is in fact the optimal number of transmit antennas which need to be activated to get this pre-log corresponding to the capacity of this channel. Now this expression for n_t' shows that the active number of transmit antennas depends upon the channel spread factor (which is equal to the normalized Doppler bandwidth for frequency flat channels) in a fashion that as channel spread factor increases, one needs to activate lesser and lesser number of transmit antennas.

The intuitive reasoning for this is that the existence of the log(SNR) dominant regime requires coherent detection of the data, hence one needs to minimize the number of channel coefficients to be estimated to get into this regime. So practically the active number of transmit antennas should be reduced with the increase in Doppler bandwidth. Although it reduces the spatial signal dimensions but also the average number of channel parameters which need to be estimated. The same reasoning makes our scheme work where we compromise over temporal signal dimensions to reduce the number of active channel parameters, getting the benefit of coherent detection and resultantly achieve pre-log.

IX. Concluding Remarks

In this contribution we derived the pre-log expression for underspread time- and frequency selective channels. We proved the existence of log(SNR) regime of capacity growth for overspread channels under certain conditions of the delay spread and the Doppler bandwidth of the channel with the help of a very simple transmission scheme utilizing zero padding. The optimality of the scheme was shown over other schemes employing zero padding for highly spread channels. We also gave the boundary where log(SNR) regime converts to log(log(SNR)) regime and further when there is no growth of capacity with SNR.

Appendix A
Achievability for Underspread Channels

To show achievability, we select Gaussian i.i.d. inputs, denoted as $\mathbf{x^G}$. The mutual information between the input and the output of the doubly selective channel (5) over the block length N is given by

$$\begin{aligned} I(\mathbf{x^G};\mathbf{y}) &= I(\mathbf{x^G},\mathbf{H};\mathbf{y}) - I(\mathbf{H};\mathbf{y}|\mathbf{x^G}) \\ &= I(\mathbf{x^G};\mathbf{y}|\mathbf{H}) + I(\mathbf{H};\mathbf{y}) - I(\mathbf{H};\mathbf{y}|\mathbf{x^G}) \\ &\geq I(\mathbf{x^G};\mathbf{y}|\mathbf{H}) - I(\mathbf{H};\mathbf{y}|\mathbf{x^G}) \end{aligned} \quad (31)$$

Equalities here follow from the introduction of the channel matrix $\mathbf{H}$ and using the chain rule of mutual information multiple times and the inequality follows from the positivity of the mutual information.

First term in the above inequality is the mutual information when the channel is known and can be evaluated readily

$$I(\mathbf{x^G};\mathbf{y}|\mathbf{H}) = h(\mathbf{y}|\mathbf{H}) - h(\mathbf{y}|\mathbf{H},\mathbf{x^G}) = h(\mathbf{y}|\mathbf{H}) - h(\mathbf{z})$$

As the input $\mathbf{x^G}$ has been selected as i.i.d. Gaussian so $(\mathbf{y}|\mathbf{H})$ is also Gaussian distributed with zero mean and its covariance is $\mathbb{E}[\mathbf{y}\mathbf{y}^\dagger|\mathbf{H}] = \mathrm{SNR}\mathbf{H}\mathbf{H}^\dagger + \mathbf{I_N}$, so

$$I(\mathbf{x^G};\mathbf{y}|\mathbf{H}) = \mathbb{E}\log|\mathrm{SNR}\mathbf{H}\mathbf{H}^\dagger + \mathbf{I_N}| \quad (32)$$

$\mathbf{H}\mathbf{H}^\dagger$ will be a full rank matrix due to its block diagonal structure and Gaussian entries hence at high SNR, the above mutual information can be approximated as

$$\lim_{\mathrm{SNR}\to\infty} I(\mathbf{x^G};\mathbf{y}|\mathbf{H}) = N\log(\mathrm{SNR}) + O(1) \quad (33)$$

Now we bound the second mutual information term in (31) using the model in (6).

$$I(\mathbf{H};\mathbf{y}|\mathbf{x^G}) = I(\mathbf{h};\mathbf{y}|\mathbf{X^G}) = h(\mathbf{y}|\mathbf{X^G}) - h(\mathbf{y}|\mathbf{h},\mathbf{X^G})$$

where the entropy of $(\mathbf{y}|\mathbf{h},\mathbf{X^G})$ is equal to the entropy of the i.i.d. Gaussian noise vector $\mathbf{z}$ because of the invariability of the entropy due to deterministic translations [13] and $(\mathbf{y}|\mathbf{X^G})$ is zero mean Gaussian distributed with covariance $\mathbb{E}[\mathbf{y}\mathbf{y}^\dagger|\mathbf{X^G}] = \mathrm{SNR}\mathbf{X}\mathbf{K_h}\mathbf{X}^\dagger + \mathbf{I_N}$ where $\mathbf{K_h}$ denotes the covariance matrix of the NL length channel vector $\mathbf{h}$.

$$\begin{aligned} I(\mathbf{H};\mathbf{y}|\mathbf{x^G}) &\overset{a}{=} \mathbb{E}\log|\mathrm{SNR}\mathbf{X}\mathbf{K_h}\mathbf{X}^\dagger + \mathbf{I_N}| \\ &\overset{b}{=} \mathbb{E}\log|\mathrm{SNR}\mathbf{K_h}\mathbf{X}^\dagger\mathbf{X} + \mathbf{I_{NL}}| \\ &\overset{c}{\leq} \log|\mathrm{SNR}\mathbf{K_h}\mathbb{E}(\mathbf{X}^\dagger\mathbf{X}) + \mathbf{I_{NL}}| \\ &\overset{d}{=} \log|\mathrm{SNR}\mathbf{K_h} + \mathbf{I_{NL}}| \end{aligned} \quad (34)$$

Equality (b) follows from the determinant identity $|\mathbf{I}+\mathbf{AB}| = |\mathbf{I}+\mathbf{BA}|$, in (c) we use the Jensen's inequality and (d) follows as $\mathbb{E}(\mathbf{X}^\dagger\mathbf{X}) = \mathbf{I_{NL}}$. As $\mathbf{h} = \mathbf{F_c}\mathbf{g}$ so $\mathbf{K_h} = \mathbf{F_c}\mathbf{K_g}\mathbf{F_c^\dagger}$ where $\mathbf{K_g}$ is the diagonal covariance matrix of LN_d length BEM coefficient vector $\mathbf{g}$ due to its uncorrelated entries and the above equation becomes

$$\begin{aligned} I(\mathbf{H};\mathbf{y}|\mathbf{x^G}) &\overset{a}{\leq} \log|\mathrm{SNR}\mathbf{F_c}\mathbf{K_g}\mathbf{F_c^\dagger} + \mathbf{I_{NL}}| \\ &\overset{b}{=} \log|\mathrm{SNR}\mathbf{K_g}\mathbf{F_c^\dagger}\mathbf{F_c} + \mathbf{I_{LN_d}}| \\ &\overset{c}{=} \log|\mathrm{SNR}\mathbf{K_g} + \mathbf{I_{LN_d}}| \\ &\overset{d}{=} \sum_{i=1}^{LN_d} \log[\mathrm{SNR}\mathbb{E}(g_i g_i^\dagger) + 1] \end{aligned} \quad (35)$$

In (b), we again use the determinant identity $|\mathbf{I}+\mathbf{AB}| = |\mathbf{I}+\mathbf{BA}|$ and (c) follows as $\mathbf{F_c^\dagger}\mathbf{F_c} = \mathbf{I_{LN_d}}$. At high SNR, this term can be approximated as

$$\lim_{\mathrm{SNR}\to\infty} I(\mathbf{H};\mathbf{y}|\mathbf{x^G}) \leq LN_d\log(\mathrm{SNR}) + O(1) \quad (36)$$

Combining equations (31),(33) and (36), we get the following lower bound of the mutual information

$$\lim_{\mathrm{SNR}\to\infty} I(\mathbf{x^G};\mathbf{y}) \geq (N - LN_d)\log(\mathrm{SNR}) + O(1) \quad (37)$$

Appendix B
Upper Bound of MI for Underspread Channels

To derive the upper bound on the mutual information between the input and the output of the channel (5) over the block length N, we split the output vector $\mathbf{y} \in \mathbb{C}^N$ in two vectors, one consisting of first LN_d entries and the other having the rest of $N - LN_d$ entries, respectively denoted as $\mathbf{y_1}$ and $\mathbf{y_2}$. The noise vector $\mathbf{z} \in \mathbb{C}^N$ is divided in $\mathbf{z_1}$ and $\mathbf{z_2}$ in the same manner. The input vector $\mathbf{x} \in \mathbb{C}^{N+L-1}$ is split in two vectors, $\mathbf{x_1} = [x[-(L-1)] \cdots x[LN_d-1]]^T$ and $\mathbf{x_2} = [x[LN_d] \cdots x[N-1]]^T$.

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) &= I(\mathbf{x};\mathbf{y_1},\mathbf{y_2}) \\ &= I(\mathbf{x};\mathbf{y_1}) + I(\mathbf{x};\mathbf{y_2}|\mathbf{y_1}) \end{aligned} \quad (38)$$

Now we try to upper bound both of the terms in the above equation separately. We treat the second term as

$$\begin{aligned} I(\mathbf{x};\mathbf{y_2}|\mathbf{y_1}) &\overset{a}{=} h(\mathbf{y_2}|\mathbf{y_1}) - h(\mathbf{y_2}|\mathbf{y_1},\mathbf{x}) \\ &\overset{b}{\leq} h(\mathbf{y_2}) - h(\mathbf{y_2}|\mathbf{y_1},\mathbf{x},\mathbf{H}) \\ &\overset{c}{\leq} (N - LN_d)h(y[n]) - h(\mathbf{z_2}) \end{aligned} \quad (39)$$

(a) is the definition of MI in terms of differential entropy, (b) follows because conditioning reduces the entropy and (c) uses the independence bound [13] and because with $\mathbf{x}$ and $\mathbf{H}$

known, the randomness in $\mathbf{y}$ is only due to the noise. $y[n]$ is zero mean and its variance is $\mathbb{E}[y[n]y[n]^{\dagger}] = \text{SNR}L+1$ using the independence of different channel taps and that they are zero mean Gaussian distributed with unit variance. Hence

$$I(\mathbf{x};\mathbf{y_2}|\mathbf{y_1}) \leq (N-LN_d)\{\log[\pi e(\text{SNR}L+1)] - \log(\pi e)\}$$

At high SNR, this gives

$$\lim_{\text{SNR}\to\infty} I(\mathbf{x};\mathbf{y_2}|\mathbf{y_1}) \leq (N-LN_d)\log(\text{SNR}) + O(1) \quad (40)$$

For the first term in (38), we decompose it again using the chain rule of MI

$$\begin{aligned} I(\mathbf{x};\mathbf{y_1}) &\stackrel{a}{=} I(\mathbf{x_1};\mathbf{y_1}) + I(\mathbf{x_2};\mathbf{y_1}|\mathbf{x_1}) \\ &\stackrel{b}{=} I(\mathbf{x_1};\mathbf{y_1}) \end{aligned} \quad (41)$$

(a) follows from the chain rule and (b) follows because given $\mathbf{x_1}$, the only randomness in $\mathbf{y_1}$ is due to the corresponding channel coefficients and noise both of which are independent of $\mathbf{x_2}$ and hence $I(\mathbf{x_2};\mathbf{y_1}|\mathbf{x_1}) = 0$. The mutual information term $I(\mathbf{x_1};\mathbf{y_1})$ represents an overspread channel as number of observations available are LN_d and same is the number of minimal independent BEM coefficients which need to be estimated. So this term gives no growth with $\log(\text{SNR})$ and hence at high SNR, the upper bound can be approximated as

$$\lim_{\text{SNR}\to\infty} I(\mathbf{x};\mathbf{y}) \leq (N-LN_d)\log(\text{SNR}) + O(1) \quad (42)$$

APPENDIX C
ACHIEVABILITY FOR OVERSPREAD CHANNELS

Here we derive a lower bound on the achievable data rate of overspread channels when we use the zero-padded transmission scheme described in section V-A. In our scheme, we transmit one symbol and do zero padding of $\lfloor L/2 \rfloor$ symbols and so on. Thus the input vector $\mathbf{x}$ can be split in two vectors, one vector $\mathbf{x_a}$ containing all the non-zero input samples and the other $\mathbf{x_b}$ containing all the zero-padded input samples. So $\mathbf{x_a}$ has samples of $\mathbf{x}$ from indices $\{i(\lfloor L/2\rfloor+1), i = 0, 1, \cdots N/(\lfloor L/2\rfloor+1)\}$. Similarly we split the output samples in two vectors, ones which appear with no ISI and the other samples where we get multiple channel coefficients with inputs. We denote $\mathbf{y_a}$ as the vector of output samples which appear without ISI and hence they contain sample values of $\mathbf{y}$ from indices $\{j(\lfloor L/2\rfloor+1)+\lfloor L/2\rfloor, j = 0, 1, \cdots N/(\lfloor L/2\rfloor+1)\}$. $\mathbf{y_b}$ is the vector of output samples which appear with ISI and which we neglect. So the achievable data rate is

$$\begin{aligned} R_N &\stackrel{a}{=} I(\mathbf{x};\mathbf{y}) = I(\mathbf{x_b};\mathbf{y}) + I(\mathbf{x_a};\mathbf{y}|\mathbf{x_b}) \\ &\stackrel{b}{=} I(\mathbf{x_a};\mathbf{y}|\mathbf{x_b}) \\ &\stackrel{c}{=} I(\mathbf{x_a};\mathbf{y_a}|\mathbf{x_b}) + I(\mathbf{x_a};\mathbf{y_b}|\mathbf{x_b},\mathbf{y_a}) \\ &\stackrel{d}{\geq} I(\mathbf{x_a};\mathbf{y_a}|\mathbf{x_b}) \end{aligned} \quad (43)$$

(b) follows as $\mathbf{x_b}$ is deterministically zero, giving $I(\mathbf{x_b};\mathbf{y}) = 0$ and (d) follows from the non-negativity of the mutual information.

All the elements in $\mathbf{x_a}$ and $\mathbf{y_a}$ have a one-to-one relationship of the form

$$y_a[j] = \sqrt{\text{SNR}}x_a[i]h[j, \lfloor L/2\rfloor] + z[j] \quad (44)$$

where $j = i + \lfloor L/2 \rfloor$. This equation represents the input-output relationship for a frequency flat time varying channel for which high SNR capacity results are already known in the non-coherent case [3]. In the mutual information term $I(\mathbf{x_a};\mathbf{y_a}|\mathbf{x_b})$, both the input and the output have length $N/(\lfloor L/2\rfloor+1)$ which plays the role of the block length in this case. Now there is only a single channel tap which needs to be estimated for the coherent detection of the data and requires the estimation of $N_d = 2M+1$ BEM coefficients for this block and is the rank of the channel covariance matrix for this particular tap. Hence in a straightforward manner, using the result from [3], we can write

$$\lim_{\text{SNR}\to\infty} R_N \geq \left(\frac{N}{\lfloor\frac{L}{2}\rfloor+1} - N_d\right)\log(\text{SNR}) + O(1) \quad (45)$$

This is the rate over the block length of N symbol intervals, so the pre-log per symbol time is

$$PreLog \geq \frac{1}{\lfloor\frac{L}{2}\rfloor+1}\left(1 - (\lfloor\frac{L}{2}\rfloor+1)\frac{N_d}{N}\right) \quad (46)$$

While deriving large block length asymptotics for underspread channels, we showed that for very large values of N, the factor N_d/N is equal to B_d, the normalized Doppler bandwidth. Hence the pre-log for our zero-padded transmission scheme becomes

$$PreLog \geq \frac{1}{\lfloor\frac{L}{2}\rfloor+1}\left(1 - (\lfloor\frac{L}{2}\rfloor+1)B_d\right) \quad (47)$$

ACKNOWLEDGMENTS

Institut Eurcom's research is partially supported by its industrial members: BMW Group Research & Technology BMW Group Company, Bouygues Telecom, Cisco Systems, France Telecom , Hitachi, SFR, Sharp, STMicroelectronics, Swisscom, Thales. The research work leading to this paper has also been partially supported by the European Commission under the ICT research network of excellence NewCom++ of the 7th Framework programme and by the French ANR project APOGEE.

REFERENCES

[1] T. Marzetta and B. Hochwald, "Capacity of a mobile multiple-antenna communications link in rayleigh flat fading," *IEEE Trans. on Information Theory*, vol. 45, pp. 139–157, January 1999.

[2] L. Zheng and D. N. C. Tse, "Communication on the grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. on Information Theory*, vol. 48, pp. 359–383, February 2002.

[3] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time-selective rayleigh-fading channels,," *IEEE Trans. on Information Theory*, vol. 50, pp. 3095–3110, December 2004.

[4] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Transactions on Information Theory*, vol. 52, pp. 2572–2583, September 2004.

[5] A. Lapidoth and S. Moser, "Capacity bounds via duality with applications to multi-antenna systems on flat fading channels," *IEEE Trans. on Information Theory*, vol. 49, pp. 2426–2467, October 2003.
[6] T. Koch and A. Lapidoth, "The fading number and degrees of freedom in non-coherent mimo fading channels: a peace pipe," *IEEE International Symposium on Information Theory*, vol. 52, pp. 437–453, September 2005.
[7] A. Lapidoth and S. Moser, "The fading number of single-input multiple-output fading channels with memory," *IEEE Trans. on Information Theory*, vol. 52, pp. 437–453, February 2006.
[8] A. Lapidoth, "On the asymptotic capacity of stationary gaussian fading channels," *IEEE Trans. on Information Theory*, vol. 51, pp. 437–446, February 2005.
[9] R. Etkin and D. N. C. Tse, "Degrees of freedom in some underspread mimo fading channels," *IEEE Trans. on Information Theory*, vol. 52, pp. 1576–1608, April 2006.
[10] M. Guillaud and D.T.M. Slock, "Channel Modeling and Associated Inter-Carrier Interference Equalization for OFDM Systems with High Doppler Spread," in *Proc. Int'l Conf. Acoustics, Speech and Sig. Proc. (ICASSP)*, Hong Kong, Apr. 6-10 2003.
[11] D. Tse and P. Viswanath, "Fundamentals of wireless communications," *Cambridge, U.K. Cambridge Univ. Press*, 2005.
[12] U. Salim and D. Slock, "Capacity pre-log achieving schemes for stationary frequency-flat multi-antenna channels," *Allerton Conf. on Communication, Control, and Computing*, 2007.
[13] T.M. Cover and J.A. Thomas, "Elements of information theory," *New York: John Wiley and Sons*, 1991.