ON THE CAPACITY OF LINEAR TIME-VARYING CHANNELS

Sergio Barbarossa, Anna Scaglione

Infocom Dept., Univ. of Rome "La Sapienza", Roma, ITALY
{sergio, annas}@infocom.ing.uniroma1.it

ABSTRACT

Linear time-varying (LTV) channels are often encountered in mobile communications but, as opposed to the linear time-invariant (LTI) channels case, there is no a well established theory for computing the channel capacity, or providing simple bounds to the maximum information rate based only on the channel impulse response, or predicting the structure of the channel eigenfunctions. In this paper, we provide: i) a method for computing the mutual information between blocks of transmitted and received sequences, for *any finite* block length; ii) the optimal precoding (decoding) strategy to achieve the maximum information rate; iii) an upper bound for the channel capacity based only on the channel time-varying transfer function; iv) a time-frequency representation of the channel eigenfunctions, revealing a rather intriguing, but nonetheless intuitively justifiable, *bubble* structure.

1. INTRODUCTION

The knowledge of the channel capacity has a clear importance in digital communications because it establishes the value of the maximum information rate that can be transmitted through the channel with an arbitrarily low bit error rate (BER), provided that sufficient redundancy is added to the transmitted sequence via precoding. For linear time-invariant (LTI) channels with additive Gaussian noise, it is well known how to compute the channel capacity, how to obtain an upper bound based only on the channel transfer function and which is the form of the channel eigenfunctions (complex exponentials with linear phase). Unfortunately, due to inherent mathematical tractability problems, the theoretical framework for LTV channels, in spite of their increasing importance in mobile communications, is not as well established as in the case of LTI channels. Indeed, in [4] the general case of linear channels is considered in the continuous-time domain but, due to evident difficulties to solve the integrals, no closed form solutions are provided neither for the capacity, nor for an easy form for its bound, nor for the channel eigenfunctions. There are more recent works addressing the computation of the capacity of LTV channels, where the channel variability is modeled in a statistical sense and the capacity is evaluated using ensemble averages. However, in some cases, e.g. [5] and [6], the LTV channels are memoryless, whereas in [7], to simplify the mathematical tractability, only the two-ray propagation model is considered. In this work, we assume that the channel is modeled, as usually in mobile communications, as the superposition of a discrete number of paths, each one characterized by a complex amplitude, a time delay and a Doppler frequency shift and we provide: i) an explicit formula for the mutual information, for *any* block length P and its asymptotic behavior for P going to infinity; ii) the optimal precoding (decoding) strategy maximizing the information rate; iii) an upper bound for the channel capacity, based only on the knowledge of the channel time-varying transfer function; iv) the properties of the channel eigenfunctions. To evaluate the impact of the channel variability on the capacity, we compare the performance of *equivalent* LTI and LTV channels (where the LTI channel is derived from the LTV channel simply setting the Doppler frequencies to zero).

This work was supported by ARL contract no. DAAL01-98-Q-1053.

2. CHANNEL MODEL AND OPTIMAL CODING STRATEGY

The most general form for a continuous-time (CT) LTV channel is given by the following input/output relationship

$$y(t) = \int_{-\infty}^{\infty} h(t;\tau)x(t-\tau)d\tau + v(t) \tag{1}$$

where $v(t)$ is additive noise and the channel is completely characterized by the time-varying impulse response $h(t;\tau)$ or its counterpart, the time-varying transfer function $H(t;f)$, defined as the Fourier Transform (FT) of $h(t;\tau)$ with respect to τ. In mobile communications, where the channel is often characterized by multiple propagations, with the generic kth path described by a complex amplitude h_k, a delay τ_k and a Doppler frequency shift f_k, the time-varying impulse response and transfer function are:

$$h(t;\tau) = \sum_{k=0}^{K-1} h_k \delta(\tau-\tau_k) e^{j2\pi f_k t} \tag{2}$$

and

$$H(t;f) = \sum_{k=0}^{K-1} h_k e^{j2\pi(f_k t - f\tau_k)}, \tag{3}$$

respectively. The equivalent discrete-time model of a causal time-varying channel is $y(n) = \sum_{k=0}^{\infty} h(n,k)x(n-k)+v(n)$. In this paper, we assume that the channel memory is finite, that is $h(n,k)=0$ for $k > L$. The most convenient form for expressing the I/O relationship in the DT case is the matrix form. Specifically, considering a block of M received symbols $\boldsymbol{y}$, the output is related to the input as $\boldsymbol{y} = \boldsymbol{T}\boldsymbol{x}+\boldsymbol{v}$, where $\boldsymbol{T}$ is the $M \times P$ channel matrix, $\boldsymbol{x}$ is the $P \times 1$ input vector, $\boldsymbol{v}$ is the $M \times 1$ noise vector and $P = M + L$.

The average mutual information between the transmitted and received sequences $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$I = \lim_{P\to\infty} \frac{1}{P} I(\boldsymbol{x},\boldsymbol{y}), \tag{4}$$

where $I(\boldsymbol{x},\boldsymbol{y})$ is the mutual information between the random vectors $\boldsymbol{x}$ and $\boldsymbol{y}$. The channel capacity is the maximum value assumed by the average mutual information over all possible probability distribution of the transmitted symbols, consistent with some input constraints (in general the average transmitted power). Borrowing a result from [2], derived for LTI channels and white Gaussian noise, with slight modifications to incorporate the cases of time-varying channels and colored noise, we can prove that, given an additive Gaussian noise $\boldsymbol{v}$, the mutual information between $\boldsymbol{x}$ and $\boldsymbol{y}$ is maximum when the input vector $\boldsymbol{x}$ is also Gaussian and is equal to [2]:

$$I(\boldsymbol{x};\boldsymbol{y}) = \frac{1}{P}\log_2 |(\boldsymbol{R}_{xx}^{\dagger} + \boldsymbol{T}^H \boldsymbol{R}_{vv}^{-1}\boldsymbol{T})\boldsymbol{R}_{xx}|, \tag{5}$$

where $\boldsymbol{R}_{xx}$ is the covariance matrix of $\boldsymbol{x}$ and $\boldsymbol{R}_{vv}$ is the noise covariance matrix. The matrix $(\boldsymbol{R}_{xx}^{\dagger}+\boldsymbol{T}^H\boldsymbol{R}_{vv}^{-1}\boldsymbol{T})\boldsymbol{R}_{xx}$ may be rank deficient and, in such a case, as in the evaluation of the entropy of Gaussian random vectors having a singular covariance matrix, the determinant has to be substituted by the product of the nonzero singular values of $(\boldsymbol{R}_{xx}^{\dagger} + \boldsymbol{T}^H\boldsymbol{R}_{vv}^{-1}\boldsymbol{T})\boldsymbol{R}_{xx}$ [2, Appendix II]. The channel capacity can be evaluated by maximizing the average mutual information with respect to all possible covariance matrices $\boldsymbol{R}_{xx}$ consistent with the constraint on the average transmitted power:

$$C = \lim_{P\to\infty} \max_{\boldsymbol{R}_{xx}} I(\boldsymbol{x};\boldsymbol{y}), \tag{6}$$

with $I(\boldsymbol{x};\boldsymbol{y})$ given by (5). For an arbitrarily LTV channel, the limit in (6) need not exist and capacity is only defined if the limit does exist [4] [1]. However, for the model generally adopted in wireless communications, e.g. (2), we have always observed the convergence of the limit.

In [8] we derived the optimal $\boldsymbol{R}_{xx}$ maximizing the mutual information rate, in the LTI channels case, providing also the optimal coding strategy to achieve the maximum mutual information. The same approach can be extended to the LTV case because it does not depend on the structure of the matrix $\boldsymbol{T}$. In [8] we assumed a block transmission strategy and introduced the minimal amount of redundancy, in the form of guard intervals, to avoid inter-block interference (IBI). More specifically, the generic nth block of information symbols $\boldsymbol{s}(n)$, of dimension $M\times 1$, is initially multiplied by a $P\times M$ precoding matrix $\boldsymbol{F}$, with $P = M + L$, where L is the channel memory, and the received block is decoded by multiplication for an $M\times P$ decoding matrix $\boldsymbol{G}$. Denoting by $\boldsymbol{H}$ the $P\times P$ channel matrix, and by $\hat{\boldsymbol{s}}(n)$ the $M\times 1$ decoded sequence, the input/output relationship is [8]

$$\hat{\boldsymbol{s}}(n) = \boldsymbol{GHFs}(n) + \boldsymbol{Gv}(n)\ . \tag{7}$$

Input $\boldsymbol{s}(n)$ and additive Gaussian noise (AGN) $\boldsymbol{v}(n)$ are generally complex, mutually uncorrelated, stationary with full rank covariance matrices $\boldsymbol{R}_{ss}$ and $\boldsymbol{R}_{vv}$, respectively ($\sigma_{ss}^2\boldsymbol{I}$ and $\sigma_{vv}^2\boldsymbol{I}$ when white). Setting $\boldsymbol{T} := \boldsymbol{GH}$, $\boldsymbol{x}(n) = \boldsymbol{Fs}(n)$ and $\boldsymbol{w}(n) := \boldsymbol{Gv}(n)$, Eqn.(7) can be written equivalently as $\hat{\boldsymbol{s}}(n) = \boldsymbol{Tx}(n) + \boldsymbol{w}(n)$.

In [8] we derived the optimal pair $(\boldsymbol{F},\boldsymbol{G})$ that, for given $\boldsymbol{H}$, $\boldsymbol{R}_{ss}$, and $\boldsymbol{R}_{vv}$, maximizes the information rate, subject to a fixed average transmitted power constraint and assuming the transmission through an LTI channel. Generalizing that result to the LTV channel, we have the following (the proof follows the same guidelines as [8])

Theorem 1: Assuming that the channel matrix $\boldsymbol{H}$, the input symbol covariance matrix $\boldsymbol{R}_{ss}$ and the noise covariance matrix $\boldsymbol{R}_{vv}$, be given and imposing a constraint on the average transmit power $\mathcal{P}_0 := tr(\boldsymbol{F}\boldsymbol{R}_{ss}\boldsymbol{F}^H)$, denoting by $\boldsymbol{U}$, $\boldsymbol{V}$ the unitary matrices, and by $\boldsymbol{\Delta}$, $\boldsymbol{\Lambda}$ the diagonal matrices resulting from the eigen- decompositions:

$$\boldsymbol{R}_{ss} \;=\; \boldsymbol{U\Delta U}^H, \qquad \boldsymbol{H}^H\boldsymbol{R}_{vv}^{-1}\boldsymbol{H} \;=\; \boldsymbol{V\Lambda V}^H\ , \tag{8}$$

the optimum $(\boldsymbol{F},\boldsymbol{G})$ filterbank pair maximizing (5) is:

$$\boldsymbol{F}_{opt} = \boldsymbol{V\Phi U}^H, \qquad \boldsymbol{G}_{opt} = \boldsymbol{U\Gamma\Lambda}^{-1}\boldsymbol{V}^H\boldsymbol{H}^H\boldsymbol{R}_{vv}^{-1}\ , \tag{9}$$

where $\boldsymbol{\Gamma}$ denotes an arbitrary invertible matrix and $\boldsymbol{\Phi}$ is a diagonal matrix with entries:

$$\phi_{ii} \;=\; \sqrt{\max\left(\frac{\mathcal{P}_0 + tr(\boldsymbol{\Lambda}^{-1})}{M\delta_{ii}} - \frac{1}{\lambda_{ii}\delta_{ii}}\ ,\ 0\right)}\ , \tag{10}$$

and $\lambda_{ii}(\delta_{ii})$ is the ith diagonal entry of $\boldsymbol{\Lambda}$ ($\boldsymbol{\Delta}$).

The optimum pair $(\boldsymbol{F}_{opt},\boldsymbol{G}_{opt})$ is non-unique, and matrix $\boldsymbol{\Gamma}$ in (9) offers extra degrees of freedom which can be exploited to satisfy added requirements such as the zero-forcing (ZF), or, the Minimum Mean-Square Error (MMSE) conditions [8]. In [8] we also proved that with $\boldsymbol{\Gamma}$ diagonal, the optimal coding pair $(\boldsymbol{F}_{opt},\boldsymbol{G}_{opt})$ renders the block transmission ISI channel model equivalent to M independent parallel ISI-free scalar subchannels, each with flat fading gains $\phi_{ii}\gamma_{ii}$ and uncorrelated AGN samples $\beta_i(n)$ with variance $1/\lambda_{ii}$; i.e.,

$$\hat{s}_i(n) \;=\; \phi_{ii}\,\gamma_{ii}\,s_i(n) \;+\; \gamma_{ii}\,\beta_i(n)\ . \tag{11}$$

Because $s_i(n)$ has variance δ_{ii}, the SNR_i at the output of the ith subchannel is:

$$SNR_i = \frac{\delta_{ii}|\phi_{ii}|^2|\gamma_{ii}|^2}{\lambda_{ii}^{-1}|\gamma_{ii}|^2} = \delta_{ii}|\phi_{ii}|^2\lambda_{ii}\ . \tag{12}$$

The independence of the parallel subchannels implies a corresponding decomposition of the maximum mutual information as (see [8]):

$$I(\boldsymbol{x};\hat{\boldsymbol{s}}) = \frac{1}{P}\sum_{i=1}^{M}\log_2(1+SNR_i)\ , \tag{13}$$

where SNR_i and ϕ_{ii} are given by (12) and (10), respectively. From (10) it turns out that, especially at low SNR, some values of Φ_{ii} may be equal to zero. This means that the constraint on the average transmitted power prevents the transmission over the most faded sub-channels. As a consequence, the available power must be properly re-allocated over the other channels, in order to use all the available average transmitted power. Proper *loading* algorithms are then used to distribute the power across the channels (see, e.g. [8], for example).

As an example of application, we have considered a numerical transmission with symbol rate $1/T$ over a three-rays

[1] It is important to remind that, for LTV channels, the coding theorem need not apply, so that one could construct, in principle, examples of channels with the capacity given by (6) such that, at rates below the capacity, the bit error rate cannot be made arbitrarily small. However, the converse of the coding theorem applies, so that if the source entropy is greater than the capacity, arbitrarily small error probabilities cannot be achieved [4].

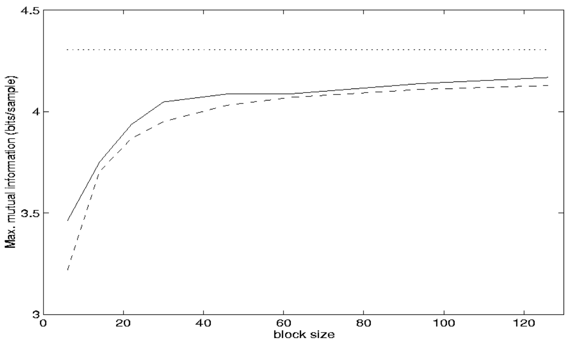

Figure 1: Maximum mutual information vs. block size.

channel, as in (2), with amplitudes $h_0 = 1$, $h_1 = .9e^{j\pi/4}$, $h_2 = .9e^{-j\pi/4}$, Doppler frequencies $f_0 = 0$, $f_1 = 2.3/(64T)$, $f_2 = 1.4/(64T)$ and delays $\tau_0 = 0$, $\tau_1 = T$, $\tau_2 = 2T$. The SNR is 10 dB. The corresponding maximum mutual information, after power loading, is reported in Fig. 1 as a function of the block size. In particular, solid line refers to the LTV channel described above whereas the dashed line refers to the *equivalent* LTI channel, defined as the channel having the same parameters as the LTV channel, except for the Doppler frequencies which are set equal to zero (the dotted line is an upper bound which will be introduced in the ensuing section). With reference to Fig. 1, it is important to make the following remarks, whose validity is not restricted to the specific case analyzed in this example because it has occurred in all the cases where the channel was modeled as in (2): i) the maximum average mutual information tends to a horizontal asymptote, which can then be taken as the channel capacity; ii) the channel capacity is approximately the same for the two equivalent LTI and LTV channels. This last comment means that the channel variability does not imply any significant impairment on the maximum information rate, provided that proper coding is applied to the transmitted sequence.

3. CHANNEL EIGENFUNCTIONS AND CAPACITY BOUNDS

It is well known that the eigenfunctions of LTI channels are complex exponentials with constant amplitude and linear phase. The knowledge of the eigenfunctions leads also to a very simple bound for the channel capacity. In fact, in the ideal case of infinite duration blocks and considering the transmission of i.i.d. symbols, with variance $P_x = E\{|x(n)|^2\}$, through an LTI channel with additive white Gaussian noise, (13) generalizes into [4], [9]:

$$I_{LTI} = \int_0^{1/2} \log_2 \left(1 + \frac{P_x}{\sigma_n^2}|H(f)|^2\right) df \qquad (14)$$

where σ_v^2 is the noise variance and $H(f)$ is the channel transfer function (f is the frequency normalized with respect to the sampling rate). In this case, in fact, the eigenvalues λ_{ii} tend simply to $|H(f)|^2$. To generalize such a result to the LTV case, we need to know the properties of the channel eigenfunctions first. It turns out that, while the FT plays a prominent role with LTI channels, with LTV channels it is more meaningful to work with time-frequency distributions. Specifically, we may prove the following properties [3]: **P1**) the channel eigenvalues are comprised between the minimum and maximum value of $|H(t;f)|^2$; **P2**) the Wigner-Ville Distribution (WVD) of the channel eigenfunction corresponding to the generic eigenvalue λ is strongly concentrated along the curve in the time-frequency plane satisfying the following implicit equation: $|H(t;f)|^2 = \lambda$. Indeed, the eigenfunctions of some simple LTV channels can be found rather easily and their WVD satisfies the previous statement, as confirmed by the following examples.

a) Two-ray model: Let us assume that the LTV channel is given by (2), with $K = 2$, $\tau_0 = 0$, $f_0 = 0$ and generic values of the other parameters. It is easy to prove that the eigenfunctions of such a channel are *chirp signals* whose sweep rate is $\mu = f_1/\tau_1$. In fact, the eigenfunction must satisfy, by definition, the following identity:

$$\lambda x_\lambda(t) = h_0 x(t) + h_1 x(t - \tau_1) e^{j2\pi f_1 t}. \qquad (15)$$

If we set $x_\lambda(t) = e^{j\pi(2\nu_\lambda t + f_1/\tau_1 t^2)}$, we may verify by substitution, that equation (15) is satisfied with

$$\lambda = h_0 + h_1 e^{-j2\pi\tau_1(\nu_\lambda - f_1/2)}. \qquad (16)$$

The WVD of the eigenfunctions $x_\lambda(t)$ is concentrated along lines of equation $f = \nu_\lambda + f_1 t/\tau_1$, which coincide with the lines where $|H(t;f)|^2$, as given by (3) specialized to the present case, is constant.

b) Multiplicative channel: If we set in (2) all the lags τ_k equal to zero, the channel is a purely multiplicative channel, i.e. $y(t) = x(t)z(t)$, with $z(t) = \sum_{k=0}^{K-1} h_k e^{j2\pi f_k t}$. In such a case, the eigenfunctions are Dirac functions, i.e. $x_\lambda(t) = \delta(t - t_\lambda)$, with $\lambda = \sum_{k=0}^{K-1} h_k e^{j2\pi f_k t_\lambda}$. In this case the WVD of the eigenfunctions is maximally concentrated along lines of equation $t = t_\lambda$, that is the lines where $|H(t;f)|^2 \equiv |H(t)|^2$ is constant.

c) LTI channel: The latter case is the dual case of LTI channels, whose eigenfunctions are sinusoids and then have a WVD maximally concentrated along lines of equation $f = f_\lambda$, where $|H(t;f)|^2 \equiv |H(f)|^2$ is also constant.

To assess the validity of the previous properties in the most general case, we provide now some numerical result. To get rid of artifacts in the WVD due to the finite length observation, we used the Reassigned Smoothed Pseudo Wigner-Ville Distribution (RSPWVD), for its good localization properties and low inner interferences [1]. As an example, we considered the transmission of independent symbols through an AWGN channel with 3 rays, with the following parameters: $h_0 = 1$, $h_1 = .9\exp(j\pi/4)$, $h_2 = .9\exp(-j\pi/4)$, $\tau_o = 0$, $\tau_1 = T$, $\tau_2 = 2T$, $f_0 = 0$, $f_1 = 2.3/64T$, and $f_2 = 1.4/64T$ (T is the symbol duration). We considered blocks of $M = 62$ information symbols and computed the eigen-decomposition of the channel matrix $\boldsymbol{H}$, as in (8). In such a case, the matrix $\boldsymbol{U}$ in (8) is the identity matrix and the columns $\boldsymbol{f}_i$ of the coding matrix $\boldsymbol{F}$ are equal to the channel eigenvectors $\boldsymbol{v}_i$, given by the columns of $\boldsymbol{V}$, multiplied by Φ_{ii}. In Fig.2 we report the contour lines of $|H(t;f)|^2$ (left column) relative to two generic different levels (eigenvalues) λ_{kk} (top) and λ_{mm} (bottom) and the contour plots (right column) of the RSPWVD of the eigenvectors $\boldsymbol{v}_k$ and $\boldsymbol{v}_m$ associated to λ_{kk} and λ_{mm}, respectively. We can observe a good agreement between the two

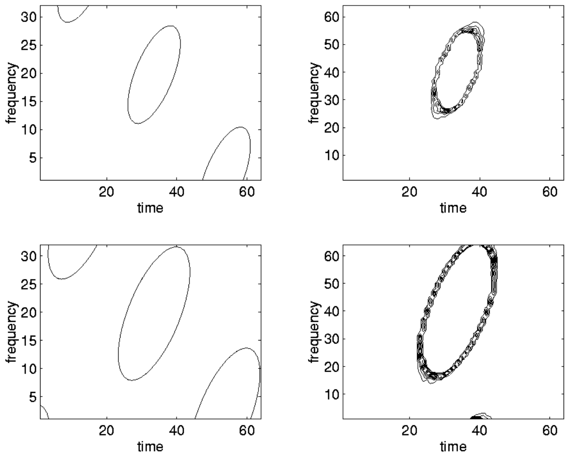

Figure 2: Contour lines of $|H(t;f)|^2$ (left column) relative to two different levels (eigenvalues) λ_{kk} and λ_{ll} and contour plots of the time-frequency distribution (RSPWVD) (right column) of the corresponding eigenvectors $\boldsymbol{v}_k$ and $\boldsymbol{v}_l$.

behaviors. It is also interesting to observe the *bubble* structure of the energy distribution of the channel eigenvectors in the time-frequency plane. Indeed this behavior generalizes what is known for LTI (multiplicative) channels whose eigenfunctions have non-overlapping support in the time-frequency domain, but they are completely overlapped in the time (frequency) domain, where their support is unlimited. In the generic LTV case, eigenfunctions corresponding to different eigenvalues still have non-overlapping support in the time-frequency domain (because of **P2**), but they may also be limited, depending on the form of $H(t;f)$, in *both time and frequency* domains. Indeed the bubble structure deriving from **P2** is a perfectly reasonable way to construct orthogonal functions [2]. This interpretation is also the key point for extending the *water-filling* principle [4] to LTV channels and generalizing the upper bound (14). In fact, starting again from (13) and (12), whose validity is completely general, and exploiting **P2**, we arrive at the following upper bound for the channel capacity [3]:

$$I_{LTV} = \lim_{T\to\infty} \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} \int_{0}^{\frac{1}{2}} \log_2 \left(1 + \frac{P_x}{\sigma_n^2} |H(t;f)|^2\right) dt df. \tag{17}$$

In Fig. 1 we had already reported this bound (dotted line), together with the maximum information rate, for a given channel configuration. To check the validity of this bound in a less specific case, we considered a multipath channel with 4 rays (K=4) and coefficients h_k modeled as independent complex Gaussian random variables (normalized to have unit norm: $\sum_k |h_k|^2 = 1$). The sets of delays and Doppler frequencies are $(0, T, 2T, 3T)$ and $(0, -1.3/64, 3.5/64, .54/64)/T$, respectively. In Fig. 3 we show the average values of the capacity bound (dotted line) and of the maximal information rates for LTV (solid line) and equivalent LTI (dashed line) channels, as a function of

[2] Using Moyal's property, the square modulus of the scalar product between two signals is equal to the scalar product, in the time-frequency domain, between their WVD's. Hence signals with non-overlapping WVD's are orthogonal.

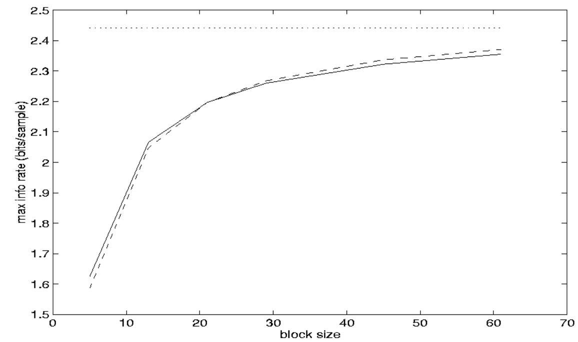

Figure 3: Average value of maximum mutual information vs. block size for LTV (solid) and LTI (dashed) channels - average capacity bound (dotted line).

the block size, evaluated according to (13) and (12), and averaged over 100 independent channel realizations (SNR is 4 dB). We can observe a clear asymptotic behavior and the closeness of the capacity bound with the asymptote of the maximal information rates.

4. REFERENCES

[1] F. Auger and P. Flandrin, "Improving the Readability of Time-Frequency and Time-Scale Representations by the Reassignment Method", *IEEE Trans. on Signal Proc.*, Vol.43, No.5, May 1995, pp.1068-1089.

[2] N. Al-Dhahir, J.M. Cioffi, "Block Transmission over Dispersive Channels: Transmit Filter Optimization and Realization, and MMSE-DFE Receiver Performance," *IEEE Trans. Inform. Theory*, pp. 137-160, Jan. 1996.

[3] S. Barbarossa and A. Scaglione, "On the capacity of linear time-varying channels," *IEEE Trans. on Info. Theory*, 1999 (submitted).

[4] R. Gallager, *Elements of Information Theory*, Section 8, New York: Wiley, 1968.

[5] B. Hughes, P. Narayan, "The Capacity of a Vector Gaussian Arbitrarily Varying Channel," *IEEE Trans. on Info. Theory,* pp. 995-1003, Sept. 1988.

[6] T. Huschka, M. Reinhardt, J. Lindner, "Channel Capacity of Fading Radio Channels", *Proc. of Intl. Conf. on Personal, Indoor and Mobile Radio Comm., PIMRC '96*, pp. 467-471, 1996.

[7] L. H. Ozarow, S. Shamai, A. D. Wyner, "Information Theoretic Considerations for Cellular Mobile Radio", *IEEE Trans. on Vehicular Technology*, pp.359-378, May 1994.

[8] A. Scaglione, S. Barbarossa, and G. B. Giannakis, "Filterbank Transceivers Optimizing Information Rate in Block Transmissions over Dispersive Channels," *IEEE Trans. on Info. Theory*, 1999 (to appear).

[9] S. Shamai (Shitz), L. H. Ozarow, A. D. Wyner, "Information Rates for a Discrete-Time Gaussian Channel with Intersymbol Interference and Stationary Inputs", *IEEE Trans. on Inform. Theory*, pp.1527-1539, Nov. 1991.