

# Kurtosis Constraints In Communication Over Fading Channels

Sibi Raj Bhaskaran

School of Computer and Communication Sciences  
Ecole Polytechnique Federal de Lausanne  
Switzerland  
Email: bsraj@lth.epfl.ch

Emre Telatar

School of Computer and Communication Sciences  
Ecole Polytechnique Federal de Lausanne  
Switzerland  
Email: emre.telatar@epfl.ch

**Abstract**—The kurtosis of a signal is a quantitative measure of how ‘peaky’ it is. In this paper we consider two scenarios of communication over fading channels with kurtosis constraints: in the first, we analyze a non-coherent Rayleigh fading channel where the input signal is required to satisfy a kurtosis constraint in addition to a power constraint. In the second, we find the ‘worst’ fading process that satisfies a kurtosis constraint and has a given second moment, while the fading coefficients are assumed to be known at the receiver. In both cases the transmitter is assumed ignorant of the instantaneous fading realization. The technique that enables our analysis is based on bounding mutual information between random variables which satisfy kurtosis and second moment constraints; the bound is tight in the low second moment regime and can be extended to multi-antenna communications.

## I. INTRODUCTION

<sup>1</sup> We consider two scenarios of communicating over fading channels where a kurtosis constraint is relevant. A kurtosis constraint prevents the underlying random variable from being too peaky, in an appropriate sense which we will clarify later. In the cases we consider, the transmitter has no information of the instantaneous channel realization. Our first consideration is a Rayleigh fading channel, where the input signal has a kurtosis constraint in addition to an average power constraint. This is of interest in some situations, which will be mentioned in the sequel. We obtain an upperbound on capacity, which is tight in the low power regime. Extensions to multi-antenna channels can be done by an appropriate definition of kurtosis constraint on the input vectors. The second scenario is where the channel coefficients have an arbitrary law and are known at the receiver, but the transmitter is forced to use independent Gaussian inputs. By using the same bounding technique mentioned above, we obtain the worst channel which satisfies a kurtosis constraint and having a given gain. Here also, we explain some extensions to multi-antenna channels.

The organization of the rest of the paper is as follows. Section 2 explains the kurtosis constraint in detail. Bounds for capacity of Rayleigh fading channel and signaling schemes which come close at low powers are presented in the section

<sup>1</sup>The work presented in this paper was supported by the National Competence Center in Research on Mobile Information and Communication Systems (NCCR-MICS), a center managed by the Swiss National Science Foundation under grant number 5005-67322.

3 and 4. Finding the worst channel in the coherent case using Gaussian inputs is explained in the next section. We conclude in section 6.

## II. KURTOSIS CONSTRAINT

Consider the complex scalar channel in (1) where  $y$  is the channel output,  $x$  the input. The input is corrupted by a multiplicative fading  $h$  which is unknown to both the transmitter and receiver except for its statistics. The additive noise is considered to be circularly symmetric white Gaussian of variance one. The fading  $h$  is assumed to be independent of both  $z$  and  $x$ .

$$\mathbf{y} = \mathbf{h}x + \mathbf{z} \quad (1)$$

If the multiplicative fading  $\mathbf{h}$  is of unit variance, the capacity per unit energy  $\hat{C}$  of this channel is same as standard additive white Gaussian noise (AWGN) channel ( $\mathbf{y} = x + \mathbf{z}$ ), when the input is power constrained [1],[5]. As the capacity per unit energy is one of the relevant performance measures for infinite bandwidth channels, one is tempted to think that at large bandwidths the fading and non-fading channels are equally good. However, [3]-[6] clearly shows that this statement is misleading, if applied to channels with large but finite bandwidth. Moreover, the input distribution that achieves capacity in fading channels becomes more and more ‘peaky’ as the power per degree of freedom decreases, while Gaussian inputs are optimal at all operating regions in an AWGN channel. The statement that a distribution is peaky can be characterized more precisely by the kurtosis of the underlying random variable, which is defined for a complex random variable  $\mathbf{w}$  as,

$$\kappa(\mathbf{w}) = \frac{E|\mathbf{w}|^4}{(E|\mathbf{w}|^2)^2} \quad (2)$$

By Jensen’s inequality,  $\kappa(\cdot) \geq 1$ .

A constraint on the average or peak power may seem natural in many situations. A kurtosis constraint makes sense when the requirements of communication prevents long durations of idleness (no transmissions), thus ruling out highly peaky signaling schemes. These requirements may be demanded by the communication standards, operational requirements at the transmitter or receiver such as synchronization etc. Another

situation where such a constraint is natural is when the ratio of peak to average power needs to be limited.

The capacity under an average or peak constraint on the input power will be a concave function, as can be seen by a simple time-sharing argument. However, an intriguing fact about the kurtosis constraint is that the capacity may no longer be a concave function of the transmit power, since the distributions which satisfy a kurtosis is not a convex set:

*Lemma 1:* Every non-trivial convex combination of two distributions, each of which individually satisfies the same kurtosis constraint with equality will have a strictly higher kurtosis.

*Proof:* Let  $\mathbf{x}_1$  and  $\mathbf{x}_2$  be two random variables with  $E[|\mathbf{x}_1|^2] = P_1$  and  $E[|\mathbf{x}_2|^2] = P_2$  and kurtosis  $\kappa(\mathbf{x}_1) = \kappa(\mathbf{x}_2) = \kappa$ .

$$\frac{E|\mathbf{x}_1|^4}{P_1^2} = \frac{E|\mathbf{x}_2|^4}{P_2^2} = \kappa \quad (3)$$

The kurtosis of a convex combination of  $X_1$  and  $X_2$  is given by

$$\frac{\lambda E|\mathbf{x}_1^4| + (1-\lambda)E|\mathbf{x}_2^4|}{(\lambda P_1 + (1-\lambda)P_2)^2} = \kappa \frac{\lambda P_1^2 + (1-\lambda)P_2^2}{(\lambda P_1 + (1-\lambda)P_2)^2} \geq \kappa \quad (4)$$

The last step follows from Jensen's Inequality applied to the function  $x \rightarrow x^2$ . This in turn means that for a fixed kurtosis, the capacity may be a non-concave function of power and thus the capacity per unit energy may not be achieved at vanishing powers. As an example, for the Rayleigh flat fading channel, the mutual information under a power constraint  $P$  and kurtosis constraint  $\kappa$  is bounded by [6],

$$I(\mathbf{x}; \mathbf{y}) \leq \frac{1}{2} \kappa(\mathbf{x}) P^2 \quad (5)$$

A non-negative concave function bounded by  $\frac{1}{2} \kappa(\cdot) P^2$  will necessarily be zero near  $P = 0$ . Thus we see that capacity is a non-concave function in this case.

We denote by  $\Theta_\rho$ , the set of all input distributions which satisfy a given power constraint  $\rho$ . Let  $\Pi_{\rho, \kappa}$  be the subset of  $\Theta_\rho$  which satisfies the power constraint with equality and has kurtosis less than  $\kappa$ . We drop the suffix and use  $\Pi$  for notational convenience. Since the additive noise is normalized, the terms power and SNR are used interchangeably.

### III. UPPER BOUND FOR RAYLEIGH FADING CHANNELS

For the channel model in (1), the mutual information can be written as,

$$I(\mathbf{x}; \mathbf{y}) = I(\mathbf{h}\mathbf{x}; \mathbf{h}\mathbf{x} + \mathbf{z}) - I(\mathbf{h}; \mathbf{h}\mathbf{x} + \mathbf{z}|\mathbf{x}) \quad (6)$$

The first term can be upper-bounded by the capacity of a corresponding AWGN channel with input  $u = h\mathbf{x}$ . The second term corresponds to a hypothetical channel between the channel fading and received signal, given the input. In the estimation parlance, this term captures the effectiveness of estimating  $\mathbf{h}$  from the channel input and output. The above decomposition can be extended to a more general multi-antenna model with part of the filter coefficients known at

the receiver.

When the multiplicative fading component in (1) is Gaussian circularly symmetric, the channel becomes the so called Rayleigh fading channel. Thus, given the input  $x$ , the output  $\mathbf{y}$  is zero mean Gaussian with variance  $(1 + |x|^2 E|\mathbf{h}|^2)$ . The normalized case has  $E|\mathbf{h}|^2 = 1$ .

A kurtosis constraint in addition to the usual power constraint is useful to avoid peaky signaling, as mentioned earlier. Under these constraints, finding the capacity is a maximization problem. Any chosen input distribution gives a lower bound. In particular, one can choose discrete distributions with support on a few values on the complex plane, so that the computation of mutual information is easy. We will see that this gives close bounds in the low power regime. Finding upperbounds which are tight for moderate powers is less straightforward. Nevertheless, for the power constrained case, bounds which are first order tight in power can be obtained using an elegant formula given in [5]. We use a simple technique which incorporates all the input moment constraints to obtain tighter bounds.

It is easy to see that the input phase conveys no information in the Rayleigh channel. Let symbols  $u$  and  $v$  denote respectively the input and output squared magnitudes of the Rayleigh channel. With this the channel is described by,

$$p(v|u) = \frac{1}{1+u} \exp \frac{-v}{1+u} \quad (7)$$

By letting  $Q$  denote the distribution of  $u$ , since the entropy of  $\mathbf{v}$  is less than that of a Gaussian with the same variance, we can bound (6) as,

$$I_Q(\mathbf{x}; \mathbf{y}) = I_Q(\mathbf{u}; \mathbf{v}) \leq \log(1+P) + E_Q \left[ \log \frac{1}{1+\mathbf{u}} \right] \quad (8)$$

Defining  $g(u) = \log(\frac{1}{1+u})$ ,

$$I_Q(\mathbf{u}; \mathbf{v}) \leq E_Q[g(\mathbf{u})] + \log(1+P) \quad (9)$$

An upperbound for capacity is obtained by taking the supremum over all input distributions  $Q \in \Pi$ , where  $\Pi$  is the input set defined in Section 2.

$$C \leq \sup_{Q \in \Pi} E_Q[g(\mathbf{u})] + \log(1+P) \quad (10)$$

When is SNR is very low, the bound for mutual information in (8) can be expanded using Taylor series. The second order approximation of logarithm can be used to lower-bound the second term of (8) as,

$$E[\log(1+\mathbf{u})] \geq E(\mathbf{u}) - \frac{1}{2} E(\mathbf{u}^2) \quad (11)$$

Substituting (11) in (8), along with  $\log(1+P) \leq P$  gives a capacity upperbound:

$$C(P, \kappa) \leq \frac{1}{2} \kappa P^2 \quad (12)$$

where  $C(P, \kappa)$  represent the capacity with a power constraint  $P$  and kurtosis constraint  $\kappa$ . This bound cannot be good at any but very small values of  $P$  as we know that for the non-

coherent Rayleigh channel, the capacity quickly falls into a double logarithmic behavior even when there is no kurtosis constraint [7]. The weakness of the bound here lies in (11) and we need to improve this expression. We note that the quantity  $E \log(1 + a\mathbf{x})$ ,  $a \in \mathbb{R}^+$  arises in many other situations too (see Sections 4 and 5).

We tackle the problem above by constructing a new bound, which stays close even at moderate values of power. The following theorem captures the essential idea on which the proposed bound here is based on. Recall the definition of set  $\Pi$ , which is the set of input distributions with power  $P$  and also satisfies the kurtosis constraint.

*Theorem 1:* Let  $s(u) = \sum_{i=0}^2 \alpha_i u^i$  be a second degree polynomial which bounds the function  $g(u) = -\log(1 + u)$  :  $u \geq 0$  from above and has  $\alpha_2 > 0$ . Then

$$\sup_{Q \in \Pi} E_Q(g(\mathbf{u})) \leq \alpha_0 + \alpha_1 E[\mathbf{u}] + \alpha_2 \kappa (E[\mathbf{u}])^2 \quad (13)$$

*Proof:* Since  $g(\cdot) \leq s(\cdot)$ ,

$$\begin{aligned} E[g(\mathbf{u})] &\leq E[s(\mathbf{u})] \\ &= E\left[\sum_{i=0}^2 \alpha_i \mathbf{u}^i\right] \end{aligned} \quad (14)$$

$$\leq \alpha_0 + \alpha_1 P + \alpha_2 \kappa P^2 \quad (15)$$

where the last inequality follows from the fact that  $\alpha_2 > 0$  and  $E[\mathbf{u}] = P$  in the set  $\Pi$ . This is true for any input distribution, in particular for the one which maximizes the left hand side. Thus Equation (8) along with Theorem 1 gives an upper bound for the capacity under our constraints of interest. We now proceed to construct the polynomial  $s(\cdot)$ . See [8] for a stronger version of Theorem 1.

### A. Bounding Polynomial

The supremum of  $E \log(1 + \mathbf{u})$  over all distributions in  $\Pi$  can be bounded as in the following theorem.

*Theorem 2:*

$$\sup_{Q \in \Pi} E_Q [\log(1 + \mathbf{u})] \geq \frac{1}{\kappa} \log(1 + \kappa P) \quad (16)$$

The proof of the above theorem is given in the appendix. Equation (16) together with (8) yields a new upperbound for the capacity of a Rayleigh fading channel.

*Corollary 1:* The information rates obtained using input signals with power and kurtosis constraints through a Rayleigh fading channel without channel side information is upper-bounded by

$$\sup_{Q \in \Pi} I[\mathbf{u}; \mathbf{v}] \leq \log(1 + P) - \frac{1}{\kappa} \log(1 + \kappa P) \quad (17)$$

Some observations follow from this. Consider an input signaling scheme with the input squared magnitudes ranging only two symbols,  $\{0, \kappa P\}$  and having average power  $P$ . The mutual information obtained by this is a lower bound to capacity. For a range of power levels in the low SNR regime, this lower bound is close with the upperbound in (17), see

figure below. From an operational standpoint, it means that limiting the kurtosis is equivalent to constraining the peak value of input squared magnitude by  $\kappa P$ .

Since scaling the input does not affect its kurtosis, one can immediately obtain the following corollary to bound the capacity of an unnormalized Rayleigh channel with  $\gamma = E[|\mathbf{h}|^2]$ .

*Corollary 2:*

$$\sup_{Q \in \Pi} I[\mathbf{x}; \mathbf{y}] \leq \log(1 + \gamma P) - \frac{1}{\kappa} \log(1 + \gamma \kappa P) \quad (18)$$

Some comparisons are in order here. The construction is different from the bound for the power constrained case as in [7], which is tight at high SNR. The bound here is tight only when power per degree of freedom is not high, so that the fourth moment constraint is active for input distributions that achieve good data rates. In addition to including the fourth moment constraints, the bound here also is in a compact form devoid of any further optimization over parameters. However, the methods in [7] can be combined with the technique used here to construct tight bounds for all power levels, thereby extending the results of [7]. The technique involves constructing polynomials which bound more complex functions than  $\log(1 + x)$ . However, this will not make a big difference in the very low power regime.

Capacity upperbounds in (12) and (17) along with above-mentioned lowerbound is shown in figure below, for a kurtosis constraint  $\kappa \leq 25$  and linear SNR scale. For comparison, the unconstrained capacity is also shown. In models where the

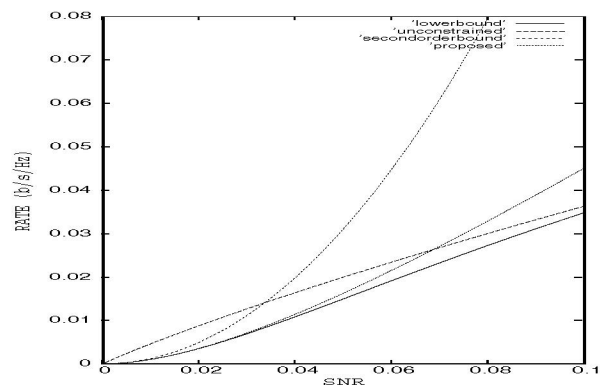


Fig. 1. Upper and Lower bounds of capacity with kurtosis  $\leq 25$

channel coefficients have two components,  $\mathbf{h} = \tilde{\mathbf{h}} + \bar{\mathbf{h}}$  where  $\bar{\mathbf{h}}$  is assumed to be known at the receiver and  $\tilde{\mathbf{h}}$  Gaussian of variance  $\gamma$ , we can extend the proposed bound. Since input phase is useful in this channel, we no longer have the freedom to work only with input squared magnitude. Using (6), the mutual information for the above model becomes,

$$I(\mathbf{x}; \mathbf{y}) \leq E \log(1 + (|\bar{\mathbf{h}}|^2 + \gamma)P) - I(\mathbf{h}; (\tilde{\mathbf{h}} + \bar{\mathbf{h}})\mathbf{x} + \mathbf{z} | \mathbf{x}, \bar{\mathbf{h}}) \quad (19)$$

$$= E \log(1 + (|\bar{\mathbf{h}}|^2 + \gamma)P) - I(\tilde{\mathbf{h}}; \tilde{\mathbf{h}}\mathbf{x} + \mathbf{z} | \mathbf{x}) \quad (20)$$

The last equality follows from the fact that  $\bar{\mathbf{h}}\mathbf{x}$  can be subtracted from  $\mathbf{y}$  and the remaining term at receiver is independent of  $\bar{\mathbf{h}}$ . We can conclude using Corollary 2 that,

$$\sup_{Q \in \Pi} I[\mathbf{x}; \mathbf{y}] \leq E \log(1 + |\bar{\mathbf{h}}|^2 P + \gamma P) - \frac{1}{\kappa} \log(1 + \gamma \kappa P) \quad (21)$$

#### IV. VECTOR MODEL

The results of the previous section can be generalized to the vector models, or the so called multi-antenna Gaussian fading channels.

$$\mathbf{y} = \mathbf{H} \begin{matrix} x \\ r \times t \end{matrix} + \mathbf{z} \begin{matrix} t \times 1 \\ r \times 1 \end{matrix} \quad (22)$$

Here  $\mathbf{H} = \bar{\mathbf{H}} + \tilde{\mathbf{H}}$ ,  $\tilde{\mathbf{H}}$  has circularly symmetric Gaussian entries independent of  $\bar{\mathbf{H}}$ . Further  $\mathbf{H}$  is also assumed *iid* across channel uses. Transmitter knows the channel statistics but input signals are independent of the channel realization. Receiver knows  $\bar{\mathbf{H}}$ , but not  $\tilde{\mathbf{H}}$ . In a similar fashion to equation (19), the mutual information can be written as the sum of two terms.

$$I(\mathbf{x}; \mathbf{y}|\bar{\mathbf{H}}) = I(\mathbf{H}\mathbf{x}; \mathbf{H}\mathbf{x} + \mathbf{z}|\bar{\mathbf{H}}) - I(\tilde{\mathbf{H}}; \tilde{\mathbf{H}}\mathbf{x} + \mathbf{z}|\mathbf{x}) \quad (23)$$

We also assume that the columns of  $\tilde{\mathbf{H}}$  are pairwise independent and  $E\tilde{\mathbf{h}}\tilde{\mathbf{h}}^\dagger = \Sigma$  for each column. Using these in (23), an input distribution of covariance matrix  $K$  will have mutual information,

$$\begin{aligned} I_K(\mathbf{x}; \mathbf{y}|\bar{\mathbf{H}}) \\ \leq E_{\bar{\mathbf{H}}} \log \det(I + \bar{\mathbf{H}}K\bar{\mathbf{H}}^\dagger + P\Sigma) - E_x \log \det(I + \Sigma\|\mathbf{x}\|^2) \end{aligned} \quad (24)$$

Imposing a kurtosis constraint  $\kappa$  on the norm of input vector  $\mathbf{x}$ , the bound can be simplified using Corollary 2 as,

$$\begin{aligned} I_K(\mathbf{x}; \mathbf{y}|\bar{\mathbf{H}}) \\ \leq E_{\bar{\mathbf{H}}} \log \det(I + \bar{\mathbf{H}}K\bar{\mathbf{H}}^\dagger + P\Sigma) - \frac{1}{\kappa} \log \det(I + \kappa P\Sigma) \\ \leq E_{\bar{\mathbf{H}}} \log \det(I + \bar{\mathbf{H}}K\bar{\mathbf{H}}^\dagger) + \log \det(I + P\Sigma) \\ - \frac{1}{\kappa} \log \det(I + \kappa P\Sigma) \end{aligned} \quad (25)$$

In the last expression, the inequality  $\log(I + A + B) \leq \log(I + A) + \log(I + B)$ , for  $A, B, C$  positive semidefinite is used. Denoting  $E[\bar{\mathbf{H}}^\dagger \bar{\mathbf{H}}] = \Delta$ , (25) implies,

$$\begin{aligned} I_K(\mathbf{x}; \mathbf{y}|\bar{\mathbf{H}}) \leq \text{trace}(\Delta K) + \log \det(I + P\Sigma) \\ - \frac{1}{\kappa} \log \det(I + \kappa P\Sigma) \end{aligned} \quad (26)$$

Notice that Jensen's inequality can be directly applied to the first term of the bound in (25), to move the expectation inside the  $\log \det(\cdot)$ , yielding another bound.

#### V. WORST CHANNEL IDENTIFICATION

In this section, we construct a lowerbound of the mutual information for the multiplicative fading model in (1), where the channel realization is known at the receiver, i.e, the coherent case. Some clarifications are needed before we proceed. The usage, 'a lowerbound to mutual information' is a cause

of confusion, as mutual information represents an achievable data rate, which can be made arbitrarily small by using bad codewords chosen from a distribution. Thus a lowerbound for achievable data rate is rather trivial. On the other hand, mutual information is defined as an abstract quantity determined by the input distribution and channel statistics. Thus a lowerbound to mutual information for a given channel means that there exists at least one coding scheme from the set of input distributions satisfying the constraints of interest, which achieves a data rate bigger than the lowerbound.

We derive a lowerbound to mutual information for the case where the input has iid Gaussian distribution. In the scalar case, this indeed is the capacity achieving distribution. In a number of multi-antenna situations too, independent Gaussian inputs are capacity achieving, thus our bounds give lowerbounds on capacity. For the multi-antenna models, we use a kurtosis constraint on the singular value distribution of  $H$ .

The fading coefficients here are arbitrarily distributed except for the fact that they obey a variance and kurtosis constraint. Let us denote  $E|\mathbf{h}|^2 = \gamma$  and thus  $E|\mathbf{h}|^4 \leq \kappa\gamma^2$ ,  $\kappa$  being the kurtosis constraint.

The mutual information in this case is given by

$$I_G(\mathbf{x}; \mathbf{y}) = E_h \log(1 + |\mathbf{h}|^2 P) \quad (27)$$

By using the bounding method of last section,

$$I_G(\mathbf{x}; \mathbf{y}) \geq \frac{1}{\kappa} \log(1 + \kappa\gamma P) \quad (28)$$

Furthermore, the lower bound is achieved when  $|\mathbf{h}|^2$  takes values in the set  $\{0, \kappa\gamma\}$  with appropriate probabilities. Thus the worst channels are the most peaky within the specified limits.

For the multi-antenna case, the model becomes

$$\mathbf{y} = \mathbf{H} \begin{matrix} \mathbf{x} \\ r \times t \end{matrix} + \mathbf{z} \begin{matrix} t \times 1 \\ r \times 1 \end{matrix} \quad (29)$$

The mutual information using independent Gaussian inputs are given by [4],

$$I_G(\mathbf{x}; \mathbf{y}) = E_H \log(I + \frac{P}{t} \mathbf{H}\mathbf{H}^\dagger) \quad (30)$$

If  $\lambda_1, \dots, \lambda_k$  are the eigenvalues of  $\mathbf{H}\mathbf{H}^\dagger$ , the mutual information becomes

$$I_G(\mathbf{x}; \mathbf{y}) = \sum_i E_{\lambda_i} \log(1 + \frac{P}{t} \lambda_i) \quad (31)$$

Thus the relevant term becomes  $E_{\lambda} \log(1 + \lambda)$ , where  $\lambda = \lambda_i$  with probability  $\frac{1}{k}$ . We impose a constraint on  $H$  by restricting the variance and kurtosis of  $\lambda$  as  $E|\lambda|^2 = \gamma$  and  $E|\lambda|^4 \leq \kappa\gamma^2$ . The preceding discussion then gives a lowerbound for this case. Thus the channel filter  $H$  influences the bound in varying the kurtosis of  $\lambda$  for a given variance.

#### VI. CONCLUSION

We have proposed upper bounds for possible data rates over some fading channels, which are popularly used in modeling.

While the capacity of some of the models can be numerically computed, the bounds give quick insights and are tight at low powers. Furthermore, the technique generalizes to many interesting scenarios.

#### REFERENCES

- [1] R. G. Gallager, 'Energy Limited Channels: Coding, Multiaccess and Spread Spectrum', Tech. Report p-1417, LIDS, M.I.T, 1987.
- [2] R. G. Gallager, *Information Theory and Reliable Communication*. New York, Wiley, 1968.
- [3] E. Telatar, 'Coding and Multiaccess for the Energy Limited Rayleigh Fading Channel', Master's thesis, EECS Dept, MIT, May 1988.
- [4] E. Telatar, 'Capacity of Multi-Antenna Gaussian channels', *Europ. Trans on Telecommunications*, Vol.10, No. 6, pp. 585-595, Nov/Dec 1999.
- [5] S. Verdú, 'On Channel Capacity per Unit Cost', *IEEE Transactions on Information Theory*, vol.36, pp. 1019-1030, September 1990.
- [6] S. Verdú, 'Spectral Efficiency in the Wideband Regime', *IEEE Transactions on Information Theory*, vol.48, pp. 1319-1343, June 2002.
- [7] A. Lapidoth and S. Moser, 'Capacity Bounds via Duality.' *IEEE Transactions on Information Theory*, vol.49, pp. 2426-2467, October 2003
- [8] S. Karlin and W. Studden, 'Tchebycheff Systems and Applications' Newyork, Interscience publishers, 1966.

#### APPENDIX

Let  $g(x) = -\log(1+x)$ ,  $x \geq 0$ . Observe that  $g(x)$  is a convex function of  $x$ . For a given  $x_0 > 0$ , define the affine function of  $x$ ,

$$h(x) = g(x_0) + (x - x_0)g'(x_0), x \geq 0, \quad (32)$$

and let

$$\nu(x, x_0) = \frac{g(x) - h(x)}{(x - x_0)^2} \quad (33)$$

Since  $g$  is convex and since  $h$  is tangent to it at  $x_0$ , it follows that  $\nu(x, x_0) \geq 0$ . Further observe that

$$g(x) = h(x) + (x - x_0)^2 \nu(x, x_0) \leq h(x) + (x - x_0)^2 \sup_x \nu(x, x_0). \quad (34)$$

With  $g(x) = -\log(1+x)$ ,  $\nu(x, x_0)$  can be written as  $(1+x_0)^{-2}(u - \log(1+u))/u^2$  with  $u = (x - x_0)/(1+x_0)$ , and thus maximizing  $\nu(x, x_0)$  over  $x \geq 0$  is equivalent to maximizing  $(u - \log(1+u))/u^2$  over  $u \geq -x_0/(1+x_0)$ . It is easy to see by differentiation that  $(u - \log(1+u))/u^2$  is a decreasing function for  $u > -1$ . We thus conclude that

*Lemma 2:* For any  $x_0 \geq 0$ ,

$$-\log(1+x) \leq -\log(1+x_0) + \frac{x-x_0}{1+x_0} + (x-x_0)^2 \left( \frac{\log(1+x_0) - \frac{x_0}{1+x_0}}{x_0^2} \right) \quad (35)$$

This bound is valid for any  $x_0$  in the interval  $(0, \infty)$ , the optimal one being that which yields the minimum of the right hand side of (15). This minimization is performed now.

*Lemma 3:* Let  $f(x_0) = \alpha_0(x_0) + \alpha_1(x_0)P + \alpha_2(x_0)\kappa P^2$ , where  $\alpha_i(x_0)$  is the coefficient of  $x^i$  in the right hand side of (35) for  $i = 0, 1, 2$ . Then

$$f(x_0) \geq f(\kappa P), \forall x_0 \in \mathbb{R}^+ \quad (36)$$

*Proof:* Substituting the value of polynomial coefficients,

$$f(x_0) = \frac{-P}{1+x_0} - 2 \left( \frac{\log(1+x_0) - \frac{x_0}{1+x_0}}{x_0} \right) P + \left( \frac{\log(1+x_0) - \frac{x_0}{1+x_0}}{x_0^2} \right) \kappa P^2 \quad (37)$$

Taking the derivative with respect to  $x_0$  and examining coefficients of the logarithmic term,

$$\frac{d}{dx_0} (f(x_0)) = 0 \Rightarrow \log(1+x_0) \left[ \frac{2P}{x_0^2} - \frac{2\kappa P^2}{x_0^3} \right] = 0 \quad (38)$$

Thus  $x_0 = \kappa P$  is an optimal point. The statement that this is a minimum can be verified by noting that the second derivative is positive at  $\kappa P$ . ■

Note that finding a minimum is not required to obtain a bound. By substituting  $x_0 = \kappa P$  itself yields the desired bound. In that sense, once we know that  $x_0 = \kappa P$  is a reasonable choice, Lemma 3 is redundant.