

# Non-coherent Rayleigh fading MIMO channels: Capacity and optimal input

Rasika R. Perera, Tony S. Pollock\*, and Thushara D. Abhayapala\*

Department of Information Engineering  
Research School of Information Sciences and Engineering  
The Australian National University, ACT 0200, AUSTRALIA

**Abstract**—Information transfer over a discrete time uncorrelated Rayleigh fading multiple input multiple output (MIMO) channel is considered, where neither the transmitter nor the receiver has the knowledge of the channel state information (CSI) except the fading statistics. We derive a capacity supremum with the receive antenna number at any signal to noise ratio (SNR) using Lagrange optimisation. We show that the asymptotic capacity is double logarithmic when the input power is large. We prove that to achieve the capacity, the amplitude of the multiple input needs to have a discrete distribution with a finite number of mass points, one of them necessarily located at the origin. We show how to compute the capacity numerically in multi-antenna configuration at any SNR with the discrete input using the Kuhn-Tucker condition for optimality. Furthermore, we show that the capacity with two mass points is optimal at low SNR signifying on-off keying. As the number of receive antennas increases, the maximum SNR at which two mass points are optimal decreases.

**Index Terms**—Channel capacity, mutual information, Rayleigh fading, upper bound, SISO, MIMO, Lagrange optimisation, non-coherent, channel state information.

## I. INTRODUCTION

In wireless systems, the knowledge of CSI is considered to be vital part in information transfer across the channel. In slowly changing channels, the CSI can be accurately obtained using pilot symbols. It is also beneficial if the statistics of the channels are known. However, the statistics of wireless channels are highly variant and finding a general model which holds in all scenarios seems to be very difficult if not an impossible task. Therefore, there are scenarios or applications where coherent detection is not plausible.

The capacity achieving input distribution of non-coherent Rayleigh fading MIMO channels has been an open problem for some time. Early work of [1], [2] using a block fading model gave some insights into the characteristics of the optimal input, with explicit calculations for a single input single output (SISO) system at high SNR. In [1] it is shown that no capacity gain is achieved by increasing the number of transmitter antennas beyond the channel coherence time. The general structure of the input signal matrix that achieves the capacity was given, along with the capacity asymptotically in channel coherence time for a SISO system and the signal density that achieves it. The authors in [2] computed the asymptotic capacity at high SNR in terms of the channel coherence time, and the number of transmit and receive antennas. It

was shown that the non-coherent and coherent capacities are asymptotically equal at high SNR. The non-coherent channel capacity is compared to the promised capacity increase using MIMO in coherent Rayleigh fading channels [3], [4].

In [5], the maximum capacity loss due to lack of receiver CSI for a wideband MIMO channel in Rayleigh fading is considered. The maximum penalty to be paid in terms of capacity not having the CSI at the receiver is shown. Furthermore, it is conjectured that on-off signaling is optimal, however no proof was given. Upper and lower bounds of time selective MIMO systems at high SNR is reported in [6]. The SISO non-coherent Rayleigh fading channel is extensively studied in [7]. It was shown that the optimal input is discrete with a finite number of mass points. Capacity is computed numerically choosing the optimal number of mass points, their probabilities and locations. In addition, Taricco [8] studied the capacity supremum and confirmed that the attainable input distribution is discrete in agreement with Abou-Faycal's results in [7].

The contributions of this paper as follows: 1) We optimise the mutual information using Lagrange optimisation method and show the capacity supremum for a given number of receivers at any SNR. 2) We prove that the asymptotic capacity supremum is double logarithmic at high SNR, similar to the results shown in [9]. 3) We prove for the first time, that the capacity achieving amplitude input distribution is discrete with a finite number of mass points, one necessarily located at the origin. 4) The numerical simulation work in optimising the channel capacity is described in brief, extending the work presented in [7] for a single antenna to multiple antennas. 5) Finally, we show that at low SNR, on-off keying is optimal.

## II. CHANNEL MODEL

The input output relationship of a MIMO channel can be written as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{N}, \quad (1)$$

where the output  $\mathbf{Y}$  is  $n_r \times 1$ , the channel gain matrix  $\mathbf{H}$  is  $n_r \times n_t$ . The input  $\mathbf{X}$  is  $n_t \times 1$  and the noise  $\mathbf{N}$  which is assumed to be zero mean complex Gaussian is  $n_r \times 1$ . Each element of  $\mathbf{H}$ ,  $h_{ij}$ ,  $i = 1, \dots, n_r$ ,  $j = 1, \dots, n_t$  and  $\mathbf{N}$  is assumed to be zero mean circular complex Gaussian random variables with a unit variance in each dimension.

**Notation:**  $X = |\mathbf{X}|$  and  $Y = |\mathbf{Y}|$  denote the random scalar variables where  $|\cdot|$  is the Euclidean norm.  $x$  and  $y$  represent each realisation of  $X$  and  $Y$  (i.e.  $x \in X$  and  $y \in Y$ ). The input is power limited with an average

T. S. Pollock and T. D. Abhayapala have appointments with National ICT Australia (NICTA).

power constraint  $\int x^2 p_X(x) dx \leq P$ .  $n_t$  and  $n_r$  denote the number of transmit and receive antennas respectively and  $\gamma = -\int_0^\infty e^{-y} \log y dy \approx 0.5772...$ , is Euler's constant.  $\Gamma(\cdot)$  and  $\Psi(\cdot)$  are the Gamma and Psi functions respectively [10]. It is assumed that the channel coherence time is one.

### III. CAPACITY UPPER BOUND

#### A. Mutual Information

The conditional output probability density function (pdf) of the non-coherent Rayleigh fading MIMO channel with  $n_r$  uncorrelated receivers is given by

$$p_{Y|X}(y|x) = \frac{y^{2n_r-1} \exp\left[-\frac{y^2}{2(1+x^2)}\right]}{2^{n_r-1} \Gamma(n_r) (1+x^2)^{n_r}}, \quad (2)$$

and represents the distribution of the magnitudes [11]. The mutual information between the channel input and output of the model (1)

$$I(\mathbf{X}; \mathbf{Y}) = -\int_0^\infty p_Y(y) \log p_Y(y) dy - \frac{1}{2} \mathbb{E}_x \left\{ \log(1+x^2) \right\} - \log \left[ \frac{\Gamma(n_r)}{\sqrt{2}} \right] + \left( n_r - \frac{1}{2} \right) \Psi(n_r) - n_r, \quad (3)$$

can be used to compute the transmission rate for a given input distribution  $p_X(x)$  [11]. For the channel capacity, (3) needs to be maximised over all input distributions subject to given constraints.

#### B. Capacity supremum

The maximisation of (3) subject to the constraints  $\int_0^\infty p_X(x) dx = 1$  and  $\int_0^\infty x^2 p_X(x) dx = P$  does not provide a valid input pdf for either SISO [8] or MIMO systems. To overcome this difficulty, we adopt a similar optimisation technique used in [8] for SISO non-coherent Rayleigh fading channels to optimise the mutual information in (3). We ascertain the following constraints,  $\int_0^\infty p_Y(y) dy = 1$ ,  $\int_0^\infty y^2 p_Y(y) dy = 2n_r(1+P)$ , and

$$\int_0^\infty p_Y(y) \log y dy = \frac{1}{2}(\beta + \Psi(n_r) + \log 2) \quad (4)$$

in order to optimise (3) where  $\beta = \mathbb{E}_x \left\{ \log(1+x^2) \right\}$ . Note that the second constraint is the average mean squared power of  $y \in Y$ , which is considered as the induced power at the output by the input, channel gain and noise. Using the Lagrange variable  $L$  and the multipliers  $\lambda_1$ ,  $\lambda_2$  and  $\lambda_3$ , we define

$$L = I(\mathbf{X}; \mathbf{Y}) + \lambda_1 \left[ \int_0^\infty p_Y(y) dy - 1 \right] + \lambda_2 \left[ \int_0^\infty y^2 p_Y(y) dy - 2n_r(1+P) \right] + \lambda_3 \left[ \int_0^\infty p_Y(y) \log y dy - \frac{1}{2}(\beta + \Psi(n_r) + \log 2) \right]. \quad (5)$$

Solving (5) for  $p_Y(y)$ , we can obtain the optimum output pdf

$$p_Y(y) = \exp \left[ (\lambda_1 - 1) + \lambda_2 y^2 + \lambda_3 \log y \right] \quad (6)$$

for the mutual information in (3) in terms of Lagrange variables. Using the constraints defined on the output, we obtain the output pdf which optimises (3)

$$p_Y(y) = \frac{\zeta^\zeta y^{2\zeta-1}}{n_r(1+P)^\zeta \Gamma(\zeta)} \exp \left[ -\frac{\zeta y^2}{2n_r(1+P)} \right], \quad \zeta > 0, \quad (7)$$

where  $\zeta = -\lambda_2 n_r(1+P)$  and  $\lambda_2 < 0$ . An expression for  $\beta$ , defined above is given by [11]

$$\beta = \log \left\{ \frac{n_r(1+P)}{\zeta} \right\} + \Psi(\zeta) - \Psi(n_r). \quad (8)$$

Substituting the optimum  $p_Y(y)$  from (7) and  $\beta$  from (8) into (3), we obtain the non-coherent channel capacity

$$C(\zeta) = G(\zeta) - G(n_r), \quad (9)$$

where  $G(\tau) = \log \Gamma(\tau) + \tau(1 - \Psi(\tau))$ . Note that  $0 \leq \beta \leq \log(1+P)$ , where the upper bound is apparent through Jensen's inequality [12]. The capacity  $C(\zeta)$  is a monotonically decreasing function of  $\zeta$  since  $C'(\zeta) = -\zeta \Psi_1(\zeta)$ , where  $\Psi_n(\cdot)$  is the  $n^{\text{th}}$  derivative of  $\Psi(\cdot)$  [10, page 253-255]. Furthermore,  $\beta$  in (8) is a monotonically increasing function of  $\zeta$  where  $\beta'(\zeta) = -1/\zeta + \Psi_1(\zeta)$ . Therefore, the supremum of (9)

$$C_{\text{sup}} = C(\zeta_s) = G(\zeta_s) - G(n_r) \quad (10)$$

is obtained with  $\beta = 0$  where the corresponding  $\zeta_s$  is given by

$$\Psi(\zeta_s) - \log(\zeta_s) = \Psi(n_r) - \log[n_r(1+P)]. \quad (11)$$

The input distribution which provides the capacity supremum (9) is given by

$$p_Y(y) = \int_0^\infty p_X(x) p_{Y|X}(y|x) dx = \frac{\zeta^\zeta y^{2\zeta-1}}{n_r(1+P)^\zeta \Gamma(\zeta)} \exp \left[ -\frac{\zeta y^2}{2n_r(1+P)} \right]. \quad (12)$$

The integral in (12) takes the form  $\int_a^b K(s, t) f(t) dt = g(s)$ , a well known Fredholm equation of the first kind [13] where  $K(s, t)$  is the *kernel*. The *kernel* in (12) is analytic in  $y$  over the whole plane for any  $n_r$ . However, the right hand side of (12) and its derivative with respect to  $y$  is infinite when  $y \rightarrow 0$  for any  $n_r$  and  $\zeta \in (0, 1)$ . Therefore, (12) does not provide a continuous solution for  $p_X(x)$  in which the  $C_{\text{sup}}$  in (10) is attained. This leads us to find a discrete input distribution in the form of  $p_X(x) = \sum_{i=1}^N p_i \delta(x - x_i)$  where  $p_i$  and  $x_i$  are to be obtained solving  $g(s) = \sum_{i=1}^N p_i K(s, t_i)$ . If the solution exists, it will provide a good lower bound to  $C_{\text{sup}}$ . The capacity supremum in (10) can be treated as an upper bound for the capacity of non-coherent Rayleigh fading MIMO channels.

#### C. Asymptotic Analysis

We consider the asymptotic analysis of the capacity supremum (10) when  $P \rightarrow \infty$ . Note that the  $P$  in (11) approaches

$\infty$  when  $\zeta_s \rightarrow 0$  since  $\lim_{\zeta_s \rightarrow 0} \zeta_s / e^{\Psi(\zeta_s)} \rightarrow \infty$  [10]. Also when  $\zeta_s = 1$ ,

$$P = \frac{1}{2n_r} \exp[\Psi(n_r) + \log 2 + \gamma] - 1, \quad (13)$$

where the input power is zero for  $n_r = 1$  and a non-zero quantity for  $n_r > 1$ . Therefore,  $\zeta_s = 1$  is not valid in this case and the  $\zeta_s$  which produces  $P = 0$  can be found by solving  $\log \zeta_s - \Psi(\zeta_s) = \log(n_r) - \Psi(n_r)$  for each  $n_r$ . We find an expression for  $\zeta_s$ , when  $P$  approaches infinity, as a result of  $\zeta_s$  approaching zero. From (11) we get

$$\log \zeta_s = \Psi(\zeta_s) + \log n_r(1 + P) - \Psi(n_r). \quad (14)$$

Multiplying both sides of (14) by  $\zeta_s$  and using  $\lim_{\zeta_s \rightarrow 0} \zeta_s \log \zeta_s = 0$  and  $\lim_{\zeta_s \rightarrow 0} \zeta_s \Psi(\zeta_s) = -1$ , [10], we get  $\zeta_s \approx 1/\log n_r(1 + P)$ . Substituting this  $\zeta_s$  in (10) we get the asymptotic capacity

$$C_{\text{sup}} \approx \log[\log n_r(1 + P)]. \quad (15)$$

This double logarithmic behavior is also reported in [6], [8], [9], [14], for both the MIMO and SISO schemes.

#### IV. CAPACITY AND OPTIMAL INPUT

##### A. Mutual Information

The mutual information of the channel (1) is given by

$$I(\mathbf{X}; \mathbf{Y}) = - \int_0^\infty p_Y(y; G) \log p_Y(y; G) dy - \frac{1}{2} \int_0^\infty \log(1 + x^2) dG_X(x) + V(n_r), \quad (16)$$

where  $V(n_r) = -\log[\Gamma(n_r)/\sqrt{2}] + (n_r - 1/2)\Psi(n_r) - n_r$ , and  $G_X(x) \triangleq \int p_X(x) dx$  is the cumulative input distribution function [11]. The channel capacity

$$C = \sup_{G_X(\cdot) \in \mathcal{E}\{x^2\} \leq P} I(G_X) \quad (17)$$

is the supremum of (16) over the set of all input distributions satisfying the input power constraint  $\int x^2 p_X(x) dx \leq P$  where  $I(G_X) \triangleq I(\mathbf{X}; \mathbf{Y})$  and  $p_Y(y; G) = \int_0^\infty p_{Y|X}(y|x) dG_X(x)$  is the marginal probability density induced by the input distribution  $G_X$ . The existence of an optimal amplitude distribution achieving the supremum in (17) can be shown proving i) the mutual information is continuous and concave in the input distribution function, and ii) the set of input distribution functions that meet the constraint is compact [7]. The following lemma gives a necessary and sufficient condition for an amplitude distribution  $G_0 \in \mathcal{G}_X$  to be optimal.

**Lemma 1:** For the uncorrelated Rayleigh fading MIMO channel with the input average power constraint  $P$ ,  $G_0$  is the capacity achieving input amplitude distribution if and only if there exist  $\lambda$  such that the following is satisfied  $\forall x \geq 0$

$$\int_0^\infty p_{Y|X}(y|x) \log p_Y(y; G_0) dy + \frac{1}{2} \log(1 + x^2) + C - V(n_r) + \lambda(x^2 - P) \geq 0 \quad (18)$$

with equality if  $x \in E_0$  where  $E_0$  is the set of points of increase of  $G_0$ .

The condition (18) is known as the Kuhn-Tucker condition for the optimal input distribution and can be used to characterise its behavior.

*Proof:* See appendix VII-A. ■

##### B. Input Distribution

The optimal input  $X^*$  should possess one of the following properties: i) the support set contains an interval, ii) it is discrete, with an infinite number of mass points on some bounded interval, iii) it is discrete and infinite, but with only a finite number of mass points on any bounded interval or, iv) it is discrete with a finite number of mass points. However, the proof is not a straightforward extension from single antenna [7] to multi-antenna systems.

Let's assume i) or ii) holds, and define  $u = 1/(1 + x^2)$ ,  $z = y^2/2$ . The support set  $E_0$  has infinite number of distinct points and the Kuhn Tucker condition holds with equality for all real  $u \in E_0$  [7]. Therefore, we get

$$\int_0^\infty e^{-uz} \left\{ z^{n_r-1} \log[\sqrt{2z} p_Z(z)] \right\} dz = \frac{\Gamma(n_r)}{u^{n_r}} \times \left[ -\lambda \left( \frac{1}{u} - 1 - P \right) - C + \frac{1}{2} \log u + V(n_r) \right] \quad (19)$$

where the left hand side (LHS) of (19) is the Laplace transform of the function  $z^{n_r-1} \log[\sqrt{2z} p_Z(z)]$ . Taking the inverse transform, we get the output pdf

$$p_Y(y) = \frac{\sqrt{2}}{y} M \exp\left(\frac{-\lambda y^2}{2n_r}\right) \quad (20)$$

where  $M = \exp\{\lambda(1 + P) - C + V(n_r) + \Psi(n_r)/2\}$ . However, for any  $\lambda/2n_r$ , the integral over  $(0, \infty)$  is infinite and hence the  $p_Y(y)$  in (20) cannot be a valid pdf, negating our original assumptions i) and ii).

Assume case iii) holds for  $X^*$ , then the support set  $U^*$  can be written as a sequence  $\{u_i\}$  converging to 0. With  $Pr[U = u_i] = p_i \neq 0$ ,  $i = 1, 2, \dots$ , we get

$$p_Y(y) = \sum_{i=0}^\infty p_i \left\{ \frac{y^{2n_r-1} u_i^{n_r} \exp\left(-\frac{u_i y^2}{2}\right)}{2^{n_r-1} \Gamma(n_r)} \right\} > p_i \left\{ \frac{y^{2n_r-1} u_i^{n_r} \exp\left(-\frac{u_i y^2}{2}\right)}{2^{n_r-1} \Gamma(n_r)} \right\}, \forall y \geq 0, \quad (21)$$

and the LHS of (18) can be lower bounded as

$$\text{LHS} \geq \frac{(\lambda - n_r u_i)}{u} - \lambda(1 + P) + C - V(n_r) + \left( n_r - \frac{1}{2} \right) \Psi(n_r) - n_r + \log \left[ \frac{\sqrt{2} p_i}{\Gamma(n_r)} \left( \frac{u_i}{u} \right)^{n_r} \right] \quad (22a)$$

$$= \frac{(\lambda - n_r u_i)}{u} + O\left(\frac{1}{u}\right). \quad (22b)$$

This lower bound diverges to  $\infty$  for  $\lambda > n_r u_i$  when  $u \rightarrow 0$ , but the LHS of (18) is zero on the support set  $U^*$ , hence by contradiction  $\lambda \leq n_r u_i$ , where  $\lambda \leq 0$  when  $u \rightarrow 0$ . Therefore, the only possibility is that  $\lambda = 0$  if the input  $X^*$  is discrete with infinite mass points. However, the Kuhn Tucker theorem [15] for convex functions (the mutual information and hence the channel capacity is concave [7]) states that the Lagrangian multiplier  $\lambda \geq 0$  on the support set which optimises the objective function, negating the original assumption confirming the optimal input is discrete with a *finite* set of mass points.

**Lemma 2:** The optimal input distribution  $X^*$  of a non-coherent uncorrelated Rayleigh fading MIMO channel contains necessarily a mass point located at the origin.

*Proof:* Since the optimal input  $X^*$  is discrete with a finite number of mass points, using the distribution function  $G_X^*(x) = \sum_{i=1}^N p_i \delta(x - x_i)$ , where  $0 \leq x_0 < x_1 < \dots < x_N$ , the mutual information can be written as

$$I(X; Y) = \sum_{i=0}^N p_i \int_0^\infty p_{Y|X}(y|x_i) \log \left[ \frac{p_{Y|X}(x|x_i)}{\sum_j p_j p_{Y|X}(x|x_j)} \right] dz. \quad (23)$$

Let's define

$$J(z) = \log \left[ \frac{p_{Z|X}(z|x_0)}{\sum_j p_j p_{Y|X}(x|x_j)} \right], \quad (24)$$

differentiating (23) with respect to  $x_0 \geq 0$ , we get

$$\frac{\partial I(X; Z)}{\partial x_0} = \frac{2x_0 p_0}{(1 + x_0^2)} \int_0^\infty [z - n_r(1 + x_0^2)] \times p_{Z|X}(z|x_0) J(z) dz \quad (25)$$

where  $n_r(1 + x_0^2)$  is the mean value of  $p_{Z|X}(z|x)$ . It can be shown that the function  $J(z)$  is decreasing. Using Lemma 1 in [7], we conclude that the derivative is negative with respect to  $x_0$  for  $0 \leq x_0 < x_1 < \dots < x_N$ . Therefore, the  $X^*$  with  $x_0 \geq 0$  can not produce a local maximum, hence the input distribution  $G_X^*(x)$  necessarily has a mass point located at the origin.

## V. NUMERICAL RESULTS

Fig. 1 shows the capacity supremum in (10) against the input power for different  $n_r$ , equating  $\zeta_s$  in (11). It is clear that the channel capacity is not promising with SNR even for a large number of receivers as in coherent scheme [3], [4]. At high SNR, the capacity asymptotically converges to (15) for any  $n_r$  showing the double logarithmic behaviour.

The optimal input amplitude, discrete with a finite number mass points can be used to compute the MIMO channel capacity numerically. The capacity is achievable once the optimal number of mass points, their probabilities and locations are found satisfying the Kuhn-Tucker condition stated in Lemma 1. Fig. 2 depicts the channel capacity as a function of input power for  $n_r = \{1, 2, 3, 5\}$ . The capacity results obtained for both two and three mass points are shown. It is clear

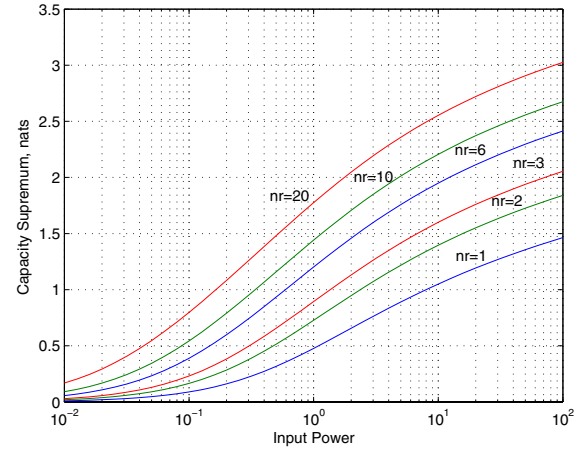


Fig. 1. Capacity supremum of the non-coherent Rayleigh fading MIMO channel with the input power for different number of receive antennas  $n_r$ .

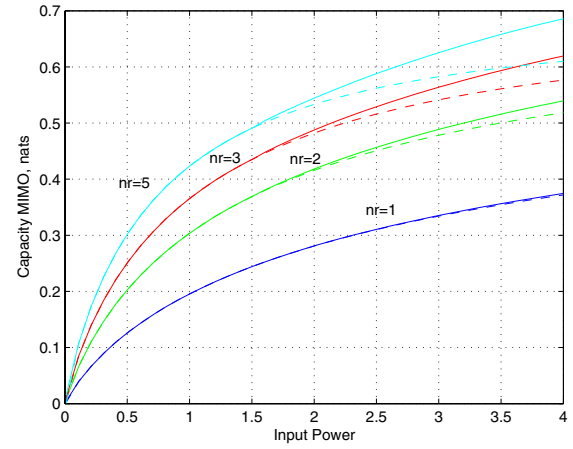


Fig. 2. Capacity of non-coherent Rayleigh fading MIMO channel vs input power using two and three mass points for different number of receiver antennas  $n_r = \{1, 2, 3, 5\}$ . The dashed lines show the capacity having two mass points for each receiver configuration.

that at low input power, there is no difference in capacity in either case. Also, it is evident that as number of receivers increases, the maximum input power at which two mass points are inadequate decreases. In this analysis we conclude that at very low SNR, the optimal input distribution has two mass points, one located at the origin. Hence the on-off keying is optimal similar to SISO systems [7].

Fig. 3 depicts the Kuhn-Tucker condition (18) for  $P = 1.4$  with a single receiver. As claimed in Lemma 1, it is above zero except at the optimal mass point locations where the Kuhn-Tucker condition equals to zero.

## VI. CONCLUSIONS

A capacity supremum is derived for non-coherent (no CSI) Rayleigh fading MIMO channels in terms of the number of receivers and the SNR. It was shown that at high SNR, the asymptotic behavior of the capacity supremum is double

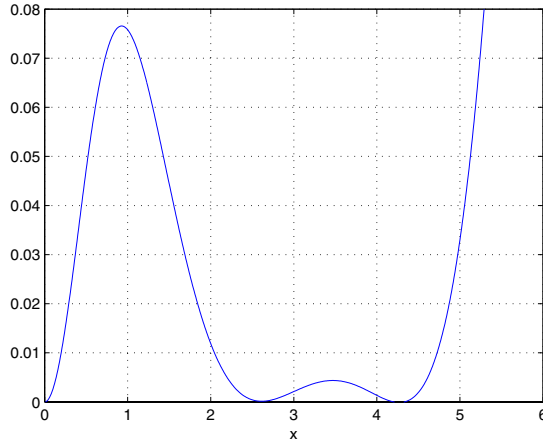


Fig. 3. The Kuhn-Tucker condition (18) for  $P = 1.4$ ,  $\lambda = 0.08819532394409$  and  $p_X(x) = 0.79549244782344\delta(x) + 0.20444300113415\delta(x - 2.6160) + 0.0000645510564\delta(x - 4.1451)$ . The channel capacity  $C=0.23549289482526$ .

logarithmic. Also, the number of transmitters has no bearing on capacity supremum.

It is shown that the capacity achieving input distribution is discrete with a finite number of mass points, one necessarily located at the origin. The channel capacity is computed numerically finding the optimal number of mass points, their probabilities and locations. The optimality is guaranteed when the optimal input distribution satisfies the Kuhn-Tucker condition. The conjectured capacities in [7] are actually achieved. Furthermore, the simulation is extended to MIMO systems and the capacities are presented against the input power for multiple receivers.

## VII. APPENDIX

### A. Proof of lemma 1

The following definition is given in [16] for the weak differentiability on a convex space.

**Definition 1:** Let  $S$  be a convex space,  $L$  a functional from  $S$  into the real line  $\mathbb{R}$ ,  $x_0$  a fixed element of  $S$ , and  $\theta$  a number in  $[0, 1]$ . Suppose there exists a map  $L'_{x_0} : \rightarrow \mathbb{R}$  such that

$$L'_{x_0}(x) \triangleq \lim_{\theta \rightarrow 0} \frac{L[(1-\theta)x_0 + \theta x] - L(x_0)}{\theta} \quad (26)$$

for all  $x$  in  $S$ . Then  $L$  is said to be weakly differentiable in  $S$  at  $x_0$  and  $L'_{x_0}$  is the weak derivative in  $S$  at  $x_0$ . If  $L$  is weakly differentiable in  $S$  at  $x_0$  for all  $x_0$  in  $S$ ,  $L$  is said to be weakly differentiable in  $S$  or simply weakly differentiable.

The following theorem [15, Page 139] shows the necessary and sufficient condition for a weakly differentiable convex function to have an optimum.

**Theorem 1:** Suppose  $Q$  is weakly differentiable, so that for all  $x_0, y$  in its domain  $S$ ,

$$Q(y) \geq Q(x_0) + Q'(x_0)(y - x_0). \quad (27)$$

Let  $X$  denote the feasible set, i.e.

$$X = \{x | Q_i(x) \leq 0, i = 1, \dots, m, h_i(x) = 0, i = 1, \dots, p\}, \quad (28)$$

then  $x_0$  is optimal if and only if  $x_0 \in X$  and

$$Q'(x_0)(y - x_0) \geq 0 \text{ for all } y \in X. \quad (29)$$

Geometrically, if  $Q'(x_0) \neq 0$  it means that  $-Q'(x_0)$  defines a supporting hyperplane to the feasible set at  $x_0$ .

Using Theorem 1 on weakly differentiable concave functional, we get the following.

**Corollary 1:** Assume  $L$  is weakly differentiable, concave functional on a convex set  $S$ . If  $L$  achieves its maximum on  $S$  at  $x_0$ , then a necessary and sufficient condition for  $L(x_0) = \max_{x \in S} L(x)$  is that  $L'_{x_0} \leq 0$  for all  $x$  in  $S$ .

The following shows the Lagrangian theorem [15, Page 215-218] commonly being used to find optimal solutions in both convex and non convex functions. Lagrangian theorem is valuable since it always provides a lower bound, and most cases the optimal solution in the absence of a duality gap.

**Theorem 2:** Let  $X$  be a linear vector space,  $Z$  a normed space,  $\Omega$  a convex subset of  $X$ , and  $P$  the positive cone in  $Z$ . Assume that  $P$  contains an interior point. Let  $f$  be a real valued concave functional on  $\Omega$  and  $G$  a convex mapping from  $\Omega$  into  $Z$ . Assume the existence of a point  $x_1 \in \Omega$  for which  $G(x_1) < 0$ . Let

$$\mu_0 = \sup_{\substack{x \in \Omega \\ G(x) < 0}} f(x) \quad (30)$$

and assume  $\mu_0$  is finite. Then there is an element  $Z_0^* > 0$  in  $Z$  (the dual space of  $Z$ ) such that

$$\mu_0 = \sup_{x \in \Omega} \{f(x) - \langle G(x), Z_0^* \rangle\}. \quad (31)$$

Furthermore, if the supremum is achieved in (30) by an  $x_0 \in \Omega$ ,  $G(x_0) \leq 0$ , it is achieved by  $x_0$  in (31) and  $\langle G(x_0), Z_0^* \rangle = 0$ .

Using the theorem 2, we can pose the optimisation problem for channel capacity with  $\lambda \geq 0$

$$C = \sup_{\substack{G_X \in \mathcal{G} \\ E\{|x|^2\} \leq P}} I(G_X) = \sup_{G_X \in \mathcal{G}} I(G_X) - \lambda \phi(G_X) \quad (32)$$

$$= \sup_{G_X \in \mathcal{G}} I(G_X) - \lambda \left( \int_0^\infty x^2 dG_X(x) - P \right). \quad (33)$$

Note in here that all the conditions of the Lagrangian theorem are satisfied. The set of input distributions of nonnegative random variables forms a convex set, the mutual information is a concave function of the input distribution [7, Appendix I-B], and input power constraint is convex since it is a linear functional of the input distribution. Next we will show that both mutual information  $I(\cdot)$  and the input constraint  $\phi(\cdot)$  are weakly differentiable functions.

**Lemma 3:** The mutual information  $I(\cdot)$  defined in (16) and  $\phi(\cdot)$  defined in (33) are weakly differentiable functionals on  $\mathcal{G}$

with weak derivatives

$$I'_{G_0}(G_X) = - \int_0^\infty p_Y(y; G_X) \log p_Y(y; G_0) dy - \frac{1}{2} \int_0^\infty \log(1+x^2) dG_X(x) - I(G_0) + V(n_r) \quad (34)$$

and

$$\phi'_{G_0}(G_X) = \phi(G_X) - \phi(G_0). \quad (35)$$

*Proof:* We define

$$G_\theta = (1-\theta)G_0 + \theta G, \quad \theta \in (0,1) \quad (36)$$

where  $G_0$  is the optimal input distribution. The difference in mutual information obtained with two distributions  $G_\theta, G_0$  is given by

$$I(G_\theta) - I(G_0) = \int_0^\infty p_Y(y; G_0) \log p_Y(y; G_0) dy - \int_0^\infty p_Y(y; G_\theta) \log p_Y(y; G_\theta) dy + \frac{1}{2} \left\{ \int_0^\infty \log(1+x^2) dG_0(x) - \int_0^\infty \log(1+x^2) dG_\theta(x) \right\}.$$

Using  $p_Y(y; G_\theta) = (1-\theta)p_Y(y; G_0) + \theta p_Y(y; G)$  and  $dG_\theta = (1-\theta)dG_0 + \theta dG$ , we get

$$\lim_{\theta \rightarrow 0} \left[ \frac{I(G_\theta) - I(G_0)}{\theta} \right] = \int_0^\infty p_Y(y; G_0) \log p_Y(y; G_0) dy - \int_0^\infty p_Y(y; G) \log p_Y(y; G_0) dy + \frac{1}{2} \left\{ \int_0^\infty \log(1+x^2) dG_0(x) - \int_0^\infty \log(1+x^2) dG(x) \right\}. \quad (37)$$

Also note that

$$I(G_0) = - \int_0^\infty p_Y(y; G_0) \log p_Y(y; G_0) dy - \frac{1}{2} \int_0^\infty \log(1+x^2) dG_0(x) + V(n_r). \quad (38)$$

Substituting  $I(G_0)$  from (38) into (37), we get (34). Similarly we can write the first derivative of  $\phi(\cdot)$ ,

$$\lim_{\theta \rightarrow 0} \left[ \frac{\phi(G_\theta) - \phi(G_0)}{\theta} \right] = \phi(G_X) - \phi(G_0), \quad (39)$$

proving (35). ■

Therefore  $I(G_X)$  and  $\phi(G)_X$  are weakly differentiable functions on  $\mathcal{G}$ . Using Corollary 1, and the weak differentiability of  $I(G_X)$  and  $\phi(G)_X$ , (33) achieves its maximum if and only if

$$I'_{G_0}(G_X) - \lambda \phi'_{G_0}(G_X) \leq 0. \quad (40)$$

Using the results obtained for  $I'_{G_0}(G_X)$  in (34) and  $\phi'_{G_0}(G_X)$  in (35) we get the following inequality

$$\int_0^\infty \left\{ \int_0^\infty p_{Y|X}(y|x) \log p_Y(y; G_0) dy \right\} dG_X(x) + \frac{1}{2} \int_0^\infty \log(1+x^2) dG_X(x) + C - V(n_r) + \lambda \int_0^\infty (x^2 - P) dG_X(x) \geq 0 \quad \forall G \in \mathcal{G} \quad (41)$$

in order to have a optimal point. The following theorem is given in [7].

**Theorem 3:** Let  $E_0$  be the points of increase of a distribution function  $F_0$ . Then

$$\int [I(x; F_0) - \lambda x^2] dF_X(x) \leq C - \lambda P \quad (42)$$

for all  $F \in \mathcal{F}$  if and only if

$$I(x; F_0) \leq C + \lambda(x^2 - P), \quad \forall x, \quad (43)$$

and

$$I(x; F_0) = C + \lambda(x^2 - P), \quad \forall x \in E_0. \quad (44)$$

Using Theorem 3 in (41), we obtain the necessary and sufficient condition (18) for the function (33) to have a maximum. This is known as the Kuhn-Tucker condition.

## VIII. ACKNOWLEDGEMENTS

National ICT Australia (NICTA) is funded through the Australian Government's *Backing Australia's Ability Initiative*, in part through the Australian Research Council.

## REFERENCES

- [1] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. on Info. Theory*, vol. 45, no. 01, pp. 139–157, Jan. 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 Info. Theory*, vol. 48, no. 02, pp. 359–383, Feb. 2002.
- [3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications*, vol. 10, no. 06, pp. 585–595, Nov. 1999.
- [4] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 06, pp. 311–335, Mar. 1998.
- [5] S. Ray, M. Medard, L. Zheng, and J. Abounadi, "On the sublinear behavior of MIMO channel capacity at low SNR," *IEEE Int. Symp. on Info. Theory*, pp. 1031–1034, Oct. 2004.
- [6] Y. Liang and V. V. Veeravalli, "Capacity of noncoherent time selective Rayleigh fading channels," *IEEE Trans. on Info. Theory*, vol. 50, no. 12, pp. 3095–3110, Dec. 2004.
- [7] I. C. Abou-Faycal, M. D. Trott, and S. Shamai, "The capacity of discrete time memoryless Rayleigh fading channels," *IEEE Trans. on Info. Theory*, vol. 47, no. 04, pp. 1290–1301, May 2001.
- [8] G. Taricco and M. Elia, "Capacity of fading channel with no side information," *IEE Electronics Letters*, vol. 33, no. 16, pp. 1368–1370, July 1997.
- [9] A. Lapidoth and S. Moser, "On the fading number of multi-antenna systems over flat fading channels with memory and incomplete side information," *IEEE Int. Symp. on Info. Theory*, vol. Lausanne, Switzerland, pp. 478–478, June-July 2002.
- [10] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, Inc, New York, 1965.
- [11] R. R. Perera, T. S. Pollock, and T. D. Abhayapala, "Upper bound on non-coherent MIMO channel capacity in Rayleigh fading," in *Proc. 11th Asia-Pacific Conference on Communications, APCC 2005*, pp. 72–76, Perth, Australia, Oct. 3-5, 2005.
- [12] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, John Wiley, New York, 1991.
- [13] L. M. Delves and J. L. Mohamed, *Computational Methods for Integral Equations*, Cambridge University Press, Great Britain, 1985.
- [14] A. M. Sengupta and P. P. Mitra, "Capacity of multivariate channels with multiplicative noise," *Tech. Repo., AT&T Bell Labs*, 2004.
- [15] S. Boyd and L. Vandenberghe, *Convex Optimisation*, Cambridge University Press, United Kingdom, 2004.
- [16] J. G. Smith, "The information capacity of peak and average power constrained Gaussian channels," *Inform. Contr.*, vol. 18, no. 107, pp. 203–219, 1971.