Survey on limits and results for the capacity for MIMO systems

Dimitrios Katselis
EE8510
St.Id. 3256501

Abstract: This survey is motivated by the need to gather and understand all the fundamental results concerning the capacity limits and the asymptotic performance of MIMO systems. MIMO systems based on the multi-antenna array (MEA) technology, offer significant increase in the spectral efficiency of wireless systems by exploitation and processing of the spatial dimension, except from the temporal. We state known results under different correlation models and channel state information assumptions, based on the existing literature. Furthermore, we present results concerning the transmitter optimization problem (i.e. optimization of the transmit covariance) and we state capacity results for the cases that these results are known. Then, we examine closely the Ricean channel model and we try to analyze the impact of the Ricean K factor on the capacity limits.

INTRODUCTION

Nowadays, there is a great expansion of technology all over the world. The planet "earth" becomes a small village through the great evolution of transportation and communication systems. The economy follows the path, which leads to a global picture of the society structures, among different nations. As economy expands, it is inevitable that technology will track and follow the expansion. Wireless systems obtain a central role to this "homogeneity" transformation. This "compact" process of evolution combined with the annual increase of the world population, yields an increase of the number of users that they enter the pool of wireless technology use. Consequently, in order to accommodate the new users, especially in crowded cities and the new applications of wireless technologies, there is an increasing need in ever higher data rates. This goal has to be achieved under certain challenging assumptions imposed by the existing structures on wireless LANs i.e. power and bandwidth constraints. Also, since in mobile systems the users are moving people

using small devices, the next challenging assumptions that is imposed is that of complexity.

MIMO techniques offer great increase in spectral efficiency. Originally, Foschini, [1], and Telatar, [2], predicted this huge increase in spectral efficiency, when the wireless system employs multiple receive and/or transmit antennas. The basic assumption that characterizes an environment such that the gain in spectral efficiency is significant is an environment of rich scattering. Another great issue, in combination with the above, is how the coherence time of the channel is compared with the burst time or in other words how fast the channel changes. One technique that can be used in MIMO systems is that of "training of the channel". Through this technique and by the use of pilot signaling from the transmitter, the receiver can perform channel estimation and it may use a fast feedback link to send this valuable information to the transmitter. The implicit assumption almost always is that the feedback link is noise free. Hence, by this two-way communication scheme the transmitter obtains channel state information (CSI) (global or partial) and it can use this information to achieve higher capacity. In mobile systems, this can be the case to either uplink or downlink scenarios.

The general issue that comes into picture, in wireless communication systems is "multipath". Gesbert in [3] tries to answer the question: "Multipath: Curse or blessing?". So the natural question in our case is: "How does multipath connect to the capacity achieved by MIMO systems? ". Basically, multipath is a spatial concept, connected to the general issue of scattering. A propagation environment, which is characterized by rich scattering offers the valuable operating condition of independent transmission paths. This independence condition in combination with some "homogeneity" of the propagation environment leads to transmission paths of independent and statistically identical behavior. Consequently, we may describe the channel with such a model that the complex channel gains are i.i.d rvs, which in turn may lead to a simplification and mathematical tractability of the system's analysis.

So far, there are several contributions available for MIMO systems. Practical channel models that they have been analyzed are the Rayleigh and the Ricean fading models. The general channel model is: $\underline{y} = H\underline{x} + \underline{z}$ (1) **, where** $\underline{x} \in C^M$ **,** $\underline{y} \in C^N$**,** $\underline{z} \in C^N$ **and** $H \in C^{NxM}$**. Obviously this is a stochastic model, representing the output of the channel (received vector) as a function of the input (transmit vector) in the baseband. This model fits to the representation of a MIMO system with M transmit antennas and N receive antennas, under the implicit assumption of a single user matrix channel. The channel matrix** H **denotes the fading effect of the channel on the channel input, while the vector** z **denotes a additive noise vector, at the front end of the receiver. The noise vector is assumed to contain independent, circularly symmetric, complex gaussian entries with zero mean and variance** σ^2**. In the Rayleigh fading model,** H **is assumed to contain independent, circularly symmetric, complex gaussian entries with zero mean and**

variance α ,while in the Ricean case the independent, circularly symmetric, complex gaussian entries are assumed to have nonzero mean and so the channel matrix H can be written in the following form: $H = \sqrt{\frac{K}{K+1}} H_{LOS} + \sqrt{\frac{1}{K+1}} H_{NLOS} = D + \tilde{H}$, where $H_{LOS}(D)$ is a deterministic NxM matrix, denoting the line_of_sight (LoS) component and $H_{NLOS}(\tilde{H})$ is a random matrix with the same description as that in the Rayleigh fading model. K is the Ricean K factor. The channel model given by (1), contains another constraint, mentioned in the beginning of this introduction. The input vector is subject to an average power constraint i.e $E[xx^{+}] \leq P$, where $+$ denotes the hermitian transposition.

The channel model, given in (1), is the general channel model for MIMO systems in the baseband, as we mentioned earlier. This matrix model occurs naturally from the original motivation of MIMO systems i.e. the diversity gain, which is obtained by introducing in the system several transmit and/or receive antennas, increasing in this manner artificially the spatial dimensions of the entire communication scheme. Antenna diversity leads to efficient utilization of the available bandwidth, while it can also lead to limitation of the transmit power and reduction of the co-channel interference. Diversity, at last, ends to increased capacity i.e. increased throughput of the hole system in its "global" picture. Basically, if one wants to use antenna diversity at the receiver, the exploitation of this set up is straightforward: the receiver can always perform channel estimation from the received data. The use of antenna diversity at the transmitter is difficult, since the transmitter cannot in general guess the channel state in a magic way (the transmitter does not have any data that have absorbed the channel effect). In this case, the problem can be approached by various techniques such as the Time Division Duplexing (TDD) or Frequency Division Duplexing (FDD), so the transmitter can figure out the channel trough feedback information from the receiver, unless the designer of the system wants to avoid "closed loop" schemes of communications, trough the use of space-time codes (STC).

After, the above general introduction, we state in the following sections the main results concerning the capacity analysis of MIMO systems and we focus on the Ricean case. The results, presented below, are generally known in the literature, while in each case we give the reference and we state various types of channel knowledge at the transmitter and/or the receiver.

Capacity of MIMO systems

In the previous section, we stated the channel model for single user MIMO systems. After having the appropriate framework, we may first proceed in describing some general concepts and giving some definitions in the MIMO case.

First of all, we consider z to be white (independence assumption based on the gaussianity and the existence of no correlation among the additive noise vector components). Using the observation in [4], we may also result in the same model in the case of nonwhite noise covariance. If the noise covariance is denoted by K_Z, this can be achieved by multiplying $\underline{x}$ with $K_Z^{-1/2}$. This transformation yields to an effective channel matrix $K_Z^{-1/2}H$ and a white noise vector.

Also, another interesting observation in [4] justifies the aforementioned statistical model, especially in the case where the channel matrix H is assumed to be zero mean. This observation is based on a very useful insight of how a model that one selects to describe a real situation might highlight the special nature of the time-space evolution of a particular phenomenon. The authors ,in [4], mention that the statistical model depends on the time scale of interest. The explanation is that the structure of the channel matrix H, i.e. the values of the entries, reflects the geometry of a particular propagation environment. While in the short term, the channel coefficients might be nonzero mean and correlated random variables, in the long term, they can be described as zero mean uncorrelated random variables due to averaging over several propagation environments. This capitalizes on the fact that the user in a wireless system generally moves. The motion generates different geometries while all his/her possible movements generate an infinite number of channel realizations. Note that the possible movements of a particular user depend on his/her will to follow a particular direction. Even as person, a user has infinite many choices which are countable according to his/her understanding of the world as a discrete reality, its own inherent nature as generating motive of the statistical phenomenon of the channel variations, has the "power of continuous". Even this rather philosophical approach seems to be contradictive, one can verify that it's not after some thinking. The averaging procedure is a "smoothing", which has the effect of gradually erasing the special form of any instantaneous channel realization. This explanation/justification of the selected model is somehow intuitive and it may based on some experimental results i.e. reduces to the well-known strategy used in science that justifies the use of certain model, based on how well the model "fits" the data. But it is generally easily understood that a model has an axiomatic foundation, while from its very own nature, it has the ability of bringing in the surface certain aspects/details that the engineer/scientist may want to highlight. The above aspect is rather philosophical and capitalizes on the general principle that nature can never be described exactly, only approximately. For example, even if we describe certain phenomena statistically, the nature of any phenomenon is deterministic depending

on the infinite many variables/parameters which act deterministically but cannot be tracked so in the large scale they are described under the umbrella of stochastic theory.

Suppose now that the channel distribution is generally known as a functional relationship but depends on a generally unknown parameter θ, that characterizes the transmission environment. By [4], we have the following definition of the "ergodic capacity" : "The ergodic capacity C is the capacity $C(\theta)$ averaged over different realizations of θ: $C = E[C(\theta)]$ (3)**.**
By [4], we also have that the computation of $C(\theta)$ under a general $p_\theta()$ is generally a hard problem. Hence, the research interest has been focused on the following three models:

(a) **Zero mean spatial white (ZMSW):** $E[H] = 0, H = H^W$**;** H^W **is a white channel matrix.**
(b) **Channel mean information (CMI):** $E[H] = \overline{H}, H = \overline{H} + \sqrt{a}H^W$ **; this model can naturally fit in the Ricean channel case, where** a **is the estimation error.**
(c) **Channel covariance information (CCI):** $E[H] = 0, H = R_r^{1/2} H^W R_t^{1/2}$**, where** R_r, R_t **are the receive and transmit fade covariance matrices, which are assumed to be known.**

In the following, we state the capacity results for the three above cases. These results are summarized based on [4]. Before getting involved with the above argument, just in favor of intuition, we give the "constant MIMO channel capacity" ([4]):
"When the channel is constant and known perfectly at the transmitter and the receiver, the capacity is:

$$C = \max_{Q:tr(Q)=P} \log | I_N + HQH^+ | \qquad (4)$$

where Q **is the input covariance matrix .**

As mentioned in [4], it was earlier proved by Telatar in [2], that the above channel, under the previous mentioned assumptions can be transformed into a bank of parallel SISO(single input-single output) branches/channels, achieving interference cancellation through an SVD of the channel matrix at the transmitter. This singular value decomposition yields to $\min\{M, N\}$ **parallel channels. If we denote the singular values of** H **by** σ_i $i = 1,..., \min\{M, N\}$**, the waterpouring solution determines the optimal power allocation([4]):**

$$P_i = (\mu - \frac{1}{\sigma_i^{\,2}})^+, i = 1,..., \min\{M, N\} \qquad (5)$$

where μ **is the waterfill level,** P_i **the power of the** i^{th} **eigenmode of the channel and** $(x)^+$ **denotes the positive part of x i.e.** $(x)^+ = \max\{x, 0\}$**. Then capacity is ([4]):**

$$C = \sum_{i=1}^{\min\{M,N\}} (\log(\mu\sigma_i^{\ 2}))^+ \qquad (6)$$

We can observe in the above case that the capacity expression doesn't depend on the eigenmodes of the channel but only on its singular values. Basically, the interesting to note is how a certain definition of the channel matrix H ,as a matrix $\in C^{NxM}$ which is just a certain placing of complex numbers in a rectangular form, yields to the so significant from practical perspective channel diagonalization. Basically, the eigenmodes are characterized by orthogonality and reflect the spatial structure of the transmit linear array of antennas (interelement spacing), the spatial structure of the receive linear array, their distance from each other, the scattering peculiarities of the specific propagation environment and how these parameters in combination with the transmitted power, yield such a wave superposition at the receiver, that the transmitted eigenmode maintains its directionality when passing through the propagation environment. Channel diagonalization has the desirable effect of interference cancellation. We can see by (5) that as the singular value $\sigma_i^{\ 2}$ increases, P_i increases. This shows that the singular values of the channel represents the inverse of the level of power absorbency in the originally transmitted power, that is imposed by the channel to the i^{th} eigenmode. We can observe that this absorbency is equal to all the transmitted vector components, which means that the original channel decomposition yields a coupled and "compact" view of the channel at the receiver. Keep in mind this explanation, because it might be useful later on.

The above results, in the constant MIMO case, set the foundation for the study of fading MIMO channels. Obviously, in the constant MIMO case, with transmitter channel state knowledge, the channel capacity case doesn't contain any ensemble averaging. Note that the capacity is a concept that is closely connected only with the general view of the transmitter about the communication protocol. All the capacity expressions that will be stated later on, are dictated from the degree of knowledge that the transmitter has about the channel state or the channel covariance/uncertainty. And this is a consequence of the fact that the data rate is a quantity determined by the sender.

For the fading case, we state the results for the capacity expressions based on the different assumptions for the channel state information, given in [4]:
(a) *Capacity with perfect channel state information at the transmitter (CSIT) and at the receiver (CSIR):* This model implies that the channel varies slowly. Slow variation of the channel might imply slow or no motion at the receiver in a downlink scenario or the transmitter, in an uplink scenario. This operating condition leads to the ability of the receiver to perform channel estimation (assume no error) and to feed the channel state information back to the transmitter perfectly, through a noiseless feedback link. Obviously, even if the channel condition is known at the transmitter, the channel fade process is a reality. Hence, we may obtain an ergodic capacity expression for this scenario ([4]):

$$C = E_H[\max_{Q:tr(Q)=P} \log|I_N + HQH^+]| \qquad (7)$$

Note that the expectation operator $E_H[.]$ precedes the expectation operator over the transmit covariance, which implies that the channel variation is so slow that the transmitter can use efficiently the CSI that it has in the next transmission. In real operating conditions, the last expression given in (7), is accurate only in the case that the channel remains the same during the channel estimation at the receiver, the feedback to the transmitter and during the next transmission.

(b)*Capacity with perfect channel state information at the receiver (CSIR) and channel distribution knowledge at the transmitter(CDIT):* This is the ZMSW case. Here, the capacity expression is:

$$C = \max_{Q:tr(Q)=P} C(Q) \qquad (8)$$

where $C = E_H[\log | I_N + HQH^+ |] \qquad (9)$. The matrix Q is the transmit covariance. As mentioned in [4], for a given covariance matrix, the mutual information $C(Q)$ is achieved by transmitting independent, circularly symmetric, complex gaussian symbols along the eigenvectors of Q. The powers allocated to each eigenvector are given by the eigenvalues of Q. In this case, we can see that the expectation operator $E_H[.]$ comes after the maximization. This reflects the fact that there is no feedback from the receiver. Hence, as mentioned in [4] and was proved in [2], the optimum input covariance has the form $Q = \frac{P}{M} I$ and the capacity expression becomes:

$$C = E_H[\log | I_N + \frac{P}{M} HH^+ |] \qquad (10).$$

The effect of selecting as an optimum input covariance a scaled identity matrix, is to split the power equally among all the transmit antenna elements. Since the transmitter doesn't have any knowledge of the channel matrix, the best it can do, is to use the multipath. So what it does, is that it considers the channel to be equally good in all directions and puts equal amount of powers to propagate through different spatial paths. This is intuitively correct, since under these circumstances, the argument is that this selection yields a maximization of the power that each receive antenna collects. In this case, Telatar proved in [2] that capacity grows linearly in min{M,N} as M,N increase simultaneously.

The operating conditions, reflected by this selection of the covariance, is that the channel may varies fast enough so that the receiver cannot feedback this information to the transmitter on time or the feedback link is too noisy.

(c) *Capacity with perfect CSIR and CDIT: CMI and CCI models:*

The assumptions here are exactly the same, as in the previous case, so again we have:

$$C = \max_{Q:tr(Q)=P} C(Q)$$

where $C = E_H[\log | I_N + HQH^+ |]$.

By [4], the optimum input covariance might be full (high) rank. In the case that its rank becomes unity, capacity can be achieved via beamforming.

There are some significant advantages associated with the beamforming. The general results that occur are that beamforming leads to a very simple implementation of the transmitter so one can reduce complexity. Also, beamforming

may lead to scalar coding of each bit stream in each transmit antenna element. In the scalar case, there are coding schemes that achieve performance near to capacity. These are the turbo codes. Based on these advantages, Jafar in [5] explores the aforiomentioned cases and proves necessary and sufficient conditions for which beamforming is optimal in a MIMO system, in the sense that beamforming achieves capacity. We present here theorems 2 and 4 from this paper:

Theorem 2: Under the CCI model, the input covariance that maximizes (8) has unit rank if and only if:

$$\frac{\sigma^2}{P\lambda_2^{\Sigma}} \geq \frac{N}{1-(\frac{\sigma^2}{P\lambda_1^{\Sigma}})^N e^{\sigma^2 / P\lambda_1^{\Sigma}} \Gamma(1-N, \frac{\sigma^2}{P\lambda_1^{\Sigma}})} - 1$$

where $\Gamma(\alpha,\beta) = \int_{\beta}^{\infty} x^{\alpha-1} e^{-x} dx$ is the incomplete Gamma function. The channel matrix is of the form $H = H_W \Sigma^{1/2}$ where H_W is a white matrix and Σ is the transmit fade covariance. λ_1^{Σ} and λ_2^{Σ} are the two largest eigenvalues of Σ, while if the noise vector z given by (1) is assumed to be white, then σ^2 is the variance of each noise vector component.

Theorem 4: Under the CMI model i.e. the model for which $H = H_\mu + H_w$, the input covariance matrix that maximizes (8) has unit rank if and only if:

$$E[\frac{1}{1+\frac{Ph_\mu}{\sigma^2} w}] \leq \frac{1+\frac{P}{\sigma^2}(1-N)}{1+\frac{P}{\sigma^2}}$$

where w is a noncentral chi-squared random variable with 2N degrees of freedom and noncentrality parameter $\delta = h_\mu^{\ 2}$ where h_μ is the nonzero eigenvalue of H_μ.

After we have stated the general results for MIMO channels, we are now ready to proceed furthermore to the Ricean channel model and state the available results for this case. The Ricean distribution dictates the structure or the kind of the random variables consisting the channel matrix. In this case, the complex channel gains are gaussian with nonzero mean. Showing explicitly the dependence of the channel matrix on the Ricean K factor, we may write the channel matrix in the following form:

$$H = \sqrt{\frac{K}{K+1}} H_{LOS} + \sqrt{\frac{1}{K+1}} H_{NLOS} \qquad (11)$$

The component matrices H_{LOS}, H_{NLOS} have been explained previously, while the K factor represents the ratio of the power of the LoS component over the NloS component. We can observe in (11) that as $K \to \infty$, $\frac{K}{K+1} \to 1$ and $\frac{1}{K+1} \to 0$. So as $K \to \infty$, we approach the deterministic case, while as $K \to 0$, the channel model reduces to Rayleigh fading.

At this point, it is crucial to make the following extension: when the channel model describes with enough accuracy the reality, it is significant to analyze it. The natural process of human perception, concerning the models, works in two directions: A model serves as a representation and a parametrization of the reality, while the same model, when it fits the reality, it helps in gaining greater insight in how the inner mechanism that drives the reality works. This is something that a naturalist philosopher, like Andres Breton or an ancient Greek philosopher, like Aristotelis, would call the “closed loop of perception”. Or even in later years, we may recall Haxley and the “Doors of perception”. It is the engineer or the scientist, who in the process of understanding and control nature, tries to “break the old, stacked doors of perception”. The understanding of the model leads to better exploitation of the natural resources. One may claim that for the specific problem of channel capacity, it is the model perception that will lead us to greater data rates.

In the Ricean case, we may state the known results based on [6]. In order to be consistent with the notation of this source, we rewrite the channel model as follows:

$$\underline{y} = \sqrt{\frac{\rho}{n_t}} H \underline{x} + \underline{n} \qquad (12)$$

where we have n_t co-located transmit antennas and n_r co-located receive antennas. The channel matrix H is again of the form given by (11) and we may also assume that $n_r, n_t \to \infty$ and $\beta = \frac{n_r}{n_t}$ constant.

Furthermore, another assumption that is stated in [6], is that the sequence of the empirical eigenvalue distribution of the matrix $\frac{H_{LOS} H_{LOS}^{+}}{n_t}$ converges to a deterministic limit function $F_{\frac{H_{LOS}}{\sqrt{n_t}}}$. Finally, the mutual information per receiving antenna is given by [6]:

$$J(H,Q) = \frac{1}{n_r} \log\det(I_{n_r} + \frac{\rho}{n_t} HQH^{+}) \qquad (13)$$

We state now the different asymptotic capacity results for the Ricean channel based on [6].

Perfect channel knowledge at the transmitter:

[6]:Theorem 1: Given a channel H, the channel capacity per receiving antenna converges almost surely as $n_r = \beta n_t \to \infty$ to

$$C1 = \frac{1}{\ln 2} \int_{\frac{1}{\mu^*}}^{+\infty} \ln \lambda \mu^* dF_{\frac{H}{\sqrt{n_t}}}(\lambda) = \rho \qquad (15)$$

and $F_{\frac{H}{\sqrt{n_t}}}(\lambda)$ ***is the limit distribution function of the eigenvalues of*** $\frac{HH^{+}}{n_t}$***, whose Stieltjes transform*** $m_{\frac{H}{\sqrt{n_t}}}(z)$ ***is the unique solution of the fixed point equation:***

$$m\frac{H}{\sqrt{n_t}}(z) = \int \frac{dF_{\frac{HLOS}{\sqrt{n_t}}}(\lambda)}{\frac{K\lambda}{\beta m_{\frac{H}{\sqrt{n_t}}}(z)+1+K} - z\left(\frac{\beta m_{\frac{H}{\sqrt{n_t}}}(z)}{K+1}+1\right) + \frac{1-\beta}{K+1}} \qquad (16)$$

such that $\mathrm{Im}(m_{\frac{H}{\sqrt{n_t}}}(z)) > 0$ ***for*** $\mathrm{Im}(z) > 0$**.**

Also by [6], if the distribution function $F(\lambda)$ ***has a continuous derivative, it is related to its Stieltjes transform*** $m(z)) = \int \frac{dF(\lambda)}{\lambda - z}$ ***by***

$$\frac{dF}{d\lambda} = \frac{1}{\pi}\lim \mathrm{Im}(m(\lambda + iy)) \qquad (17)$$

Perfect knowledge of NloS component:

By [6],the channel capacity for a Ricean channel when the Tx has knowledge of the channel mean is achieved by gaussian vectors having the same eigenvectors with $\frac{H_{LOS}H_{LOS}^{+}}{n_t}$**. This is a general result and it can be stated as the general problem of the eigenspace matching. Basically, since the Tx has knowledge of the channel mean, its tries to exploit this information since the channel mean denotes the average channel behavior. Intuitively, it is correct that the eigenvectors of the optimal transmitted covariance are the same with those of the channel mean. This process leads to a sort of coupling the transmission with the peculiarities of the channel average behavior. What is the gain from finding the eigenvectors of a channel matrix? What is the advantage? Our intuition suggests that the propagation environment, if assumed to be fixed, determines the wireless channel. A fixed environment leads to a fixed channel matrix that reflects the whole set up of this environment. We should scan the whole 3-D space in order to find such directions of forcing the power to move so as the superposition at the Rx will yield a vector which is just a scaled version of what we introduced to the input. It is amazing to understand that the power moves in a certain propagation environment and it is absorbed in certain amounts or with certain rates from the existing materials, it is reflected to move to other directions and finally reaches the receiver. This whole microscopic behavior can be predicted and used by an engineer in the macroscopic fashion by a so simple mathematical construction/concept as the channel matrix.**

By [6], we state the second theorem for this case:

Theorem 2: Let H_{LoS} ***be the line of sight component of H, with SVD*** $H_{LOS} = V\hat{Q}U^{+}$ ***and*** $\hat{Q} \in R^{n_r x n_t}$ ***be a diagonal matrix. As*** $n_r, n_t \to \infty$ ***with*** $\frac{n_r}{n_t} \to \beta$***, the mutual information of the channel per receiving antenna with gaussian inputs having covariance matrix*** $U\hat{Q}U^{+}$ ***converges almost surely to a deterministic value, given by:***

$$J(H, U\hat{Q}U^{+}) = \frac{1}{\ln 2}\int_0^{\rho}\frac{1}{x}(1-\frac{1}{x}m_2(-\frac{1}{x}))dx \qquad (18)$$

where $m_2(z)$ **is the unique function solution to:**

$$\begin{cases} m_1(z) = \dfrac{\beta(m_1(z)-z(K+1))q(K+1)dF_{\frac{H}{\sqrt{n_t}}}(\lambda)}{(K+1+\beta m_2(z)q)(m_1(z)-z(K+1))+(K+1)K\lambda q} + \dfrac{(1-\beta)q_0(K+1)}{(K+1+\beta m_2(z)q_0)} \end{cases}$$

$$\begin{cases} m_2(z) = \displaystyle\int \dfrac{(K+1+q\beta m_2(z))(K+1)dF_{\frac{H}{\sqrt{n_t}}}(\lambda)}{(K+1+\beta m_2(z)q)(m_1(z)-z(K+1))+\lambda q(K+1)K} \end{cases}$$

$q(\lambda, F(\lambda), K, \beta)$ **denotes the diagonal entries of** $\hat{Q}$ **and** $q_o = q(0, F(\lambda), K, \beta)$

In order to get the asymptotic capacity in this case, we just need to maximize over $\hat{Q}$**, the expression given above for the mutual information.**

In [6], what is also stated is the case the Tx has knowledge of the limiting eigenvalue distribution of $\frac{H_{LOS}H_{LOS}^{+}}{n_t}$**. What is important to note is that the significant knowledge is the eigenstructure of the channel matrix. The eigenvectors of the channel matrix give the proper transmitting directions. A knowledge, such as the limiting eigenvalue distribution of the mean is not important. So, as intuition suggests, since the Tx doesn't have any information about the channel matrix, the best thing to do is to split all the transmit power equally among all the Tx antenna elements. Normalizing the transmitting power, we get an optimal covariance** $Q = I_{n_t}$ **and the asymptotic mutual information per receiving antenna, as** $n_r, n_t \to \infty$ ***and*** $\frac{n_r}{n_t} \to \beta$ ***is given by*** $I_3 = \frac{1}{\ln 2}\int_0^{\rho}\frac{1}{x}(1-\frac{1}{x}m_{\frac{H}{\sqrt{n_t}}}(-\frac{1}{x}))dx \qquad (19)$ **with** $m_{\frac{H}{\sqrt{n_t}}}(z)$ **be the unique solution of equation (16) such that** $\mathrm{Im}(m_{\frac{H}{\sqrt{n_t}}}(z)) > 0$**.**

Based now on further observations made in [6], we have that the Rice MIMO channel with rank –1 line of sight behaves as a Rayleigh MIMO channel in the asymptotic regime, i.e as $n_r, n_t \to \infty$. The result is intuitively correct since as $n_r, n_t \to \infty$ we have more and more transmission paths averaged, so we get the zero mean effect leading to Rayleigh behavior.

Monotonicity results concerning the capacity of Ricean MIMO channels

After having stated all the previous results available in the literature for the asymptotic Ricean case analysis, we may also mention some significant results about the monotonicity behavior of the channel capacity in the Ricean case. Before stepping into this argument, we may clarify that for the Ricean case, the asymptotic analysis is suitable since it makes the approach more tractable, in mathematical sense. At this point, one may wishes to touch upon the previous claim. Basically, if we observe the results for the CMI model and we compare it with the CCI model or the Rayleigh fading model, then we may understand that the zero mean makes things easier. The asymptotic analysis offers the "almost zero mean behavior", since, as we mentioned earlier, in the asymptotic case, one has more and more spatial transmission paths to average over the channel effect. Furthermore, its important to understand that in many cases where the analysis is performed in the asymptotic regime, for example in estimation problems when one tries to find expressions for asymptotic variance of a scalar estimator or the asymptotic covariance of a vector estimator, this asymptotic regime might be reasonably reachable from a practical perspective. In our case, we wish to underline that the asymptotic regime can be considered a large but reasonable number of antenna elements in either the transmitter and/or the receiver.

Now, about the singular values of the mean in the Ricean case we have some monotonicity results. Why it is important to study the effect of the singular values of the mean and not the behavior of the entire channel matrix? The channel matrix is characterized by two components: one deterministic and one random. If we consider the random component matrix fixed, then a study of the singular values of the whole matrix is important. The singular values of the mean determine the volume/shape of the hyperellipsoid that is the basic subset of the min{M,N} space that the channel mean transforms unit vectors. This hyperellipsoid characterizes the mean channel behavior. Basically, if we could assume a fixed random component, with the same left and right singular vectors as the channel mean, the this random component component would determine a random process according which we would have hyperellipsoid perturbations. Actually, the $\log\det(.)$ **expressions that we have in the previous capacity formulas is nothing else but the volume of a hyperellipsoid. What we try to do is to maximize the volume of a hyperellipsoid. This fact, in combination**

with the known result that the singular values represent the actual lengths of the semiaxes, may lead to the conclusion that the channel capacity depends on the singular values of the mean in a nondecreasing fashion. This conclusion is verified by the results, we state below, based on [7],[8].

Lapidoth in [8], gives the following corollary:

Corollary 1.3: Let $D, \tilde{D} \in C^{mxn}$ ***bet two LoS component matrices with decreasingly ordered singular values***

$$\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_{\min\{m,n\}}$$

and $$\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq ... \geq \tilde{\sigma}_{\min\{m,n\}}$$

respectively. Then

$$\left(\tilde{\sigma}_i \geq \sigma_i, \forall 1 \leq i \leq \min\{m,n\}\right) \Rightarrow C(\mathrm{E}, \tilde{D}) \geq C(\mathrm{E}, D)$$

where $C(\mathrm{E}, D) = \sup_K I(K, D)$***,*** K ***the input covariance subject to a power constraint*** $tr(K) \leq E$ ***and*** $I(K, D)$ ***the mutual information which is given by the following expression:*** $I(K, D) = E[\log\det(I_m + \frac{1}{\sigma^2}(D + H_{Rayl})K(D + H_{Rayl}))]$***.In this expression*** σ^2 ***is the additive noise variance per component and the channel matrix in this case is written in the form*** $H = D + H_{Rayl}$ ***where the Ricean*** K ***factor has been absorbed by the matrix components.***

It is also worthy to record the following theorem from this paper:

Theorem 1.4: The set of eigenvectors of the capacity achieving covariance matrix K, for the coherent MIMO Ricean channel must coincide with those of $D^+ D$***.***

Basically, the above two results verify our earlier conclusion. The important concept of "ergodic capacity" in the Ricean MIMO case increases with larger singular values of the mean, while theorem 1.4 states that the optimal input covariance, when the set up is such that allows coherent reception (i.e. the receiver does perform channel estimation) must be simultaneously diagonalizable with the channel mean. This is something reasonable, because from our so far experience on MIMO systems, we know that the major motivation is to align ourselves with the eigenstructure of the channel mean matrix or the channel matrix in general, achieving in such a way orthogonal signaling or "multibranch channel decoupling".

In [8], the authors state also monotonicity results when our operating definition of the channel capacity is the outage capacity. We didn't define earlier the outage capacity and so we prefer not to record these results. If the reader of this report is interested in learning more on this topic, he/she may read the reference paper.

From the above results, the conclusion is that the singular values of the channel mean provides us with certain quantitative results on the channel "quality" or the increase in capacity that can be achieved when the transmitter has knowledge of the channel mean and it can use this knowledge to find the optimal input covariance. This last conclusion yields the solution of the transmitter optimization problem., in this case.

At this point, we would like to share with the reader of this survey some thoughts. It is obvious from the above analysis that information theory in these areas is not yet well established. The general ideas that drive this research are based on general motivations that exist in a wide area of analysis on wireless systems. The idea of MIMO systems goes a few years back and so the information theoretic perspectives of MIMO systems are now proceed into more depth. When one wants to built up a theory so as to explain certain behaviors either of physical phenomena or even social phenomena, the first thing he/she has to do is to set up the axiomatic foundation and then proceed to basic definitions. Note here that this set up doesn't apply in the case that someone wants to explain phenomena of individual behavior, because as Paul Dirac claimed sometime in his lifetime, "human beings do not act according to probabilities. They act according to possibilities." And just to make a perenthesis, this is the major reason that there is no global economical theory to describe and predict the financial behavior of large crowds, while the existing ones has huge prediction gaps. From Shannon's years, information theory has obtained the form of a whole science. It has all the formulation needed to claim that it is actually a theory. Many brilliant people worked throughout forty years so this theory to expand and cover all the areas of communication theory. From the above results, concerning the monotonicity results on the capacity of MIMO Ricean channels, we just have a first step in understanding the major parameters that have a significant impact on quantities, such as the channel capacity, in the MIMO case.

Approach of the monotonicity behavior of the MIMO Ricean channel capacity, with respect to the Ricean K factor: A new result

As stated in the proposal of this survey, an interesting thing to examine, is the monotonicity of the channel capacity with respect to the Ricean K factor. Professor Nihar Jindal made the following conjecture: "If the receiver has CSI, while the transmitter knows H_{LOS} and if the number of receive and transmit antenna elements is equal, the channel capacity should be nondecreasing in K".

This proposition/conjecture is based on an intuitive observation: From the channel matrix formula, we see that as $K \to \infty$, the only component that survives is the deterministic, while as $K \to 0$, the only term that survives is the random

component. We may now observe carefully the equations (4),(7),(8) and (9). Obviously, $\log\det(.)$ **is a concave function, while the expectation operator is a generalized convex combination of all possible realizations of H. So, the deterministic capacity given by (4), is an upper bound to the other cases, based on Jensen's inequality. So, the deterministic case, with transmitter CSI, outperforms the other cases, in the channel capacity sense. Consequently, the claim of Professor Nihar Jindal is intuitively correct.**

Furthermore, yesterday, during the last presentation, a classmate gave numerical examples where the actual channel capacity in the Ricean case, increases with K but at a specific point it saturates. From a practical perspective, the channel capacity indeed saturates with increasing K. Theoretically, it never stops to increase, as we approach the deterministic case. What is happening, is that the rate of increase of the channel capacity reduces gradually to zero, yielding this "saturation behavior". Actually, we are never going to reach the deterministic case in reality. This is like the ancient paradox of the race between the turtle and the rabbit! So, what we will try to prove is that the channel capacity, in the Ricean case, is nondecreasing in K.

We give again the channel matrix formula:

$$H = \sqrt{\frac{K}{K+1}}H_{LOS} + \sqrt{\frac{1}{K+1}}H_{NLOS} \qquad (20)$$

First of all, $\lim_{K\to\infty}\frac{K}{K+1} = \lim_{K\to\infty}(\frac{K+1-1}{K+1}) = \lim_{K\to\infty}(1-\frac{1}{K+1}) = 1$**. From this relationship, we can see that** $f(K) = \frac{K}{K+1}$ **does a lower approximation to** 1**. This leads to the conclusion that** $f(K)$ **is a nondecreasing function of** K**. To prove this is straight forward since** $f(K)$ **is a continuous and differentiable function of** K**, in** $[0.\infty)$**. By taking the first derivative of** $f(K)$ **with respect to** K**, we have:**

$$\frac{df(K)}{K} = \frac{K+1-K}{(K+1)^2} = \frac{1}{(K+1)^2} \succ 0, \quad \forall K \in [0,\infty) \qquad (21)$$

So, since the derivative is strictly positive in the interval $[0,\infty)$**,** $f(K)$ **is an increasing function of** K**.**

Furthermore, we will not bound ourselves in the case of square H_{LOS} **with equal eigenvalues. Suppose that** H_{LOS} **is a rectangular matrix. Also, suppose that we examine the channel capacity in the asymptotic regime i.e. we let the number of receive and transmit antennas to increase to infinity, by maintaining a constant ratio, say** β**. The singular value decomposition of the LOS matrix is given by:**

$$H_{LOS} = V\Sigma U^{+} \qquad (22)$$

where $V \in C^{n_r x n_r}, U \in C^{n_t x n_t}$ **are two unitary matrices and** $\Sigma \in R^{n_t x n_r}$ **is a diagonal matrix, with ordered entries** $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_{\min\{m,n\}}$**. We now prove the following lemma:**

Lemma 1: If the LoS component matrix has an SVD given by (22), then the scaled LoS matrix ξH_{LOS}***, where*** ξ ***is a positive scalar, has an SVD given by the same formula, where the unitary matrices*** U, V ***are the same but the diagonal matrix*** $\tilde{\Sigma}$ **contains in this case the entries:** $\tilde{\sigma}_i = \xi\sigma_i, \ \ i = 1,...,\min\{n_t, n_r\}$**.** ***Since the scalar*** ξ ***is positive, the ordering of the singular values is maintained:*** $\tilde{\sigma}_1 \geq \tilde{\sigma}_2 \geq ... \geq \tilde{\sigma}_{\min\{m,n\}}$**.**

Proof

By (22), the LOS matrix can be written as follows:

$H_{LOS} = \sum_{i=1}^{\min\{n_t,n_r\}} \sigma_i v_i u_i^{+}$ *, where + denotes the hermitian transposition. Using the "orthonormality" property of the vectors* v_i, u_i *we may have:*

$v_i^{+} H_{LOS} u_i = \sigma_i, \ \ i = 1,...,\min\{n_t, n_r\}$.

Now,

$\xi H_{LOS} = \sum_{i=1}^{\min\{n_t,n_r\}} \xi\sigma_i v_i u_i^{+}$ *, so* $v_i^{+}\xi H_{LOS} u_i = \xi\tilde{\sigma}_i, \ \ i = 1,...,\min\{n_t, n_r\}$*. Since* ξ *is positive, the ordering of the singular values is maintained.* **Q.E.D.**

By lemma 1 and by (20), assuming that $\xi = \sqrt{\frac{K}{K+1}} \succ 0$**, we have that the singular values of** H_{LOS} **maintain their ordering as** $K \to \infty$**. Since** $\xi \uparrow$ **with** K**, we have that if** $K_1 \succ K_2$**,** $\xi_1 = \sqrt{\frac{K_1}{K_1+1}}, \xi_2 = \sqrt{\frac{K_2}{K_2+1}}$ **and** $\tilde{H}_{1,LOS} = \xi_1 H_{LOS}$**,** $\tilde{H}_{21,LOS} = \xi_2 H_{LOS}$**, then** $\tilde{\sigma}_{1,i} \succ \tilde{\sigma}_{2,i}, \ \ i = 1,...,\min\{n_t, n_r\}$**, where** $\tilde{\sigma}_{1,i}, \tilde{\sigma}_{2,i}, \ \ i = 1,...,\min\{n_t, n_r\}$ **are the singular values of** $\tilde{H}_{1,LOS}, \tilde{H}_{2,LOS}$ **respectively. By employing corollary 1.3 given previously, (Lapidoth [8]), we get the result that the channel capacity in the MIMO Ricean case is nondecreasing in K.**

Conclusion

After we have seen all the above results, our experience is sufficient to get into final concluding remarks and extensions to general issues. Every conclusion is like the end of a trip either in reality or in a depth of time, through life. Without conclusion, we don't have a complete understanding about the meaning of performing research or performing life. When a transmitter emits a signal into the environment, its wish is the receiver to get the signal in the best possible condition. The receiver has actually the same wish. Between these two wishes, there is the actual signal, which experiences the "trip" i.e. the multiplicative and additive effects of the channel. What is unfair into this "picture", is that the signal cannot perform a "conclusion" of itself. The conclusion, about the condition of the signal, is going to be given from the receiver. Any receiver cannot give the same estimate of what effects the signal experienced, when passing trough the channel. So, the quality of the estimation of the signal condition depends on how "smart" or how "experienced" the receiver is. This point-to-point communication scheme can be generalized into MIMO systems. What is happening here is that we have a bank of receivers (actually one receiver in the single user case, the smartest we can think) which have the same quality in performing conclusions i.e. they are equally smart. Then, we reveal the signal to all of them, with the difference that each one receives the same signal in different condition. Then, each one gives its conclusion. Hence, from this combined procedure of equally smart conclusions for the same signal, we hope that we can improve the final conclusion, in comparison with the single transmitter –single receiver case. The projection to real life is the following: If we say the same story to a number of equally smart people, with small differences, the combined conclusion, based on their opinions is going to be better, than speak only to one person. In MIMO systems, multiple transmit antennas might transmit parts of a transmission vector with no overlapping in the information that they transmit. At the receiver's side, multiple antennas collect the same signal i.e. all the components transmitted from each single transmit antenna element. But due to different propagation paths through the wireless medium, each receive antenna collects the same signal in different condition. Then, the processing of the multiple received replicas is performed in a centralized fashion, except from some preprocessing that might take place in each receive antenna.

The question is now, what do we win from a capacity perspective by this process? By the above "preprocessing", i.e. the introductive previous paragraph ,it is obvious that by this introduction in the system of multiple antennas, we increase the hole throughput of the system. How do we achieve this argument? We achieve this argument by having multiple receive antennas to collect power from many different directions. Or hence we maximize the amount of information that passes through the channel by maintaining its "structure". We can make the extension that if we have many people hearing the same story, we maximize our opportunity not to have "details" to be forgotten. I.e. we maximize the probability that someone will remember a certain detail, which might important for us in the future. We can also

extend this "picture" to physical constructions that engineers try to achieve. For example, well-known pulses are the solitons. Solitons, as mechanical or optical waves, have the ability to travel trough huge distances by maintaining their "shape". Hence they maximize our ability to keep the transmitted information "compact", while they minimize the communication cost, since they reduce the need in regenerative repeaters. Consequently, the MIMO systems lead to gain in capacity, when compared with single antenna systems. The above justifies the original prediction of Telatar and Foschini, concerning huge gains in channel capacity via MIMO techniques.

As far as the technical part is concerned, we stated most of the known results concerning capacity limits and bounds for MIMO systems. We saw that these results apply in a straightforward manner to Rayleigh fading channels. We stated results concerning the transmitter optimization problem and we recorded the conditions under which, beamforming is optimal transmission strategy, in the case of CCI model, as well as in the CMI case. Then, we gave asymptotic results concerning the capacity of single user MIMO systems under different assumptions for the channel knowledge of the transmitter, as well as the receiver. Finally, we stated some monotonicity results concerning the channel capacity in the Ricean case. We saw that the capacity is nondecreasing in the singular values of the mean, while the optimal transmit covariance must be aligned with the eigenspace of $D^{+}D$**, where** D **is the channel. Finally our own contribution is the proof of the fact that the channel capacity in the Ricean case is nondecreasing in** K**, where** K **is the Ricean K factor.**

Future work might be to explore the multi-user case, for the Ricean channel model and examine the impact of parameters such as the multiple channel means and their singular values, the Ricean K factor and the received SNR in the capacity region. We may also try to analyze the channel capacity region, in the asymptotic regime, with equal but large number of receive antennas in each receiver and large number of transmit antennas. Till then, just remember: "Karpe diem!!!"(Or in plain English "Seize the day")

References

1)G.J.Foschini : “Layered space-time architecture for wireless communication in fading environments, when using multi-antenna elements” Bell Labs Tech J., pp 41-59, 1996
2)E.Telatar: “Capacity of multi-antenna Gaussian channels” Eur. Trans. Telecomm. ETT, vol. 10, no 6, pp 585-596, Nov 1999

3)Gesbert: “Multipath: Curse or blessing? A system performance analysis of MIMO systems” in proceedings of International Zurich seminar on Communications (IZS), Switzerland, vol.2 , Feb.2001

4) “Capacity limits of MIMO channels”, Andrea Goldsmith, Syed Ali Jafar, Nihar Jindal and Siram Vishwanath, IEEE Journal on Selected areas in communications, vol. 21, No.5, June 2003

5) “Transmitter optimization and optimality of beamforming for Multiple Antenna Systems”, Syed Ali Jafar, Andrea Goldsmith, IEEE transactions on wireless communications, vol.3, No.4, July 2004.

6) “On the capacity of MIMO rice channels”, Laura Cottatelucci, Merouane Debbah.

7) “The capacity of Ricean MIMO channel is monotonic is monotonic in the singular values of the mean”, Daniel Hösli, Amos Lapidoth.

8)”Monotonicity results for coherent MIMO Ricean channels”, Daniel Hösli, Young-Hen Kim, Amos Lapidoth.