A Closed Form Expression for the Ergodic Capacity of MIMO Systems

E.Khan and C.Heneghan

Department of Electronic and Electrical Engineering,
University College Dublin, Belfield, Dublin 4, Ireland
Email: ejaz.khan@ee.ucd.ie, conor.heneghan@ucd.ie
Received February 1, 2005

Abstract—Multi-input Multi-output (MIMO) system are today regarded as one of the most promising research areas in wireless communications. This is due to the fact that MIMO channel can offer a significant capacity gain over traditional Single-input single-output (SISO) channel. This paper presents the capacity of an uncorrelated and correlated MIMO channel. The expression for the large system MIMO capacity is calculated in closed form using replica analysis and Grassmann variables. It is assumed that the receiver has perfect knowledge of the channel but no such information is available at the transmitter. It is shown that for large (but finite) number of antennas the capacity scales linearly with the number of antennas. Our expression gives very good approximation of capacity even for 2/3 antenna case. Besides this, using the same analysis we can find the bounds on ergodic capacity when we have noisy channel estimate at the receiver or even when there is co-channel interference (in addition to noise).

1. INTRODUCTION

Recent research in wireless communications has focused on potential performance gains achievable by using multiple element arrays at both the transmitter and receiver. For example, the use of multiple antennas is currently under consideration in the 802.11n standardization efforts. The reason for this interest is that in theory, much larger spectral efficiency can be achieved by utilizing spatial diversity at both the transmitter and the receiver as compared to the single-input-single-output (SISO) case. Specifically, Foschini and Gans showed that the channel capacity for a MIMO system could be bounded by quantities that scaled linearly with \( N \) (where \( N \) is taken to equal the number of transmit and receive antennas) \([5]\). These bounds indicate significant potential gains for MIMO processing over currently existing SISO systems. For example, in additive white Gaussian noise (AWGN) channel, it is well known that at high signal-to-noise ratio (SNR), 1/bit/s/Hz (bps/Hz) capacity gain can be achieved with every 3-dB increase in SNR. In contrast, for a multiple antenna system with \( M \) transmit and \( N \) receive antennas, with i.i.d. Rayleigh fading between all pairs, the capacity gain is potentially \( \min\{M, N\} \) bps/Hz for every 3dB SNR increase \([2]\). However, in the original presentation of Foschini and Gans \([5]\), their analysis led to useful lower bound on the ergodic channel capacity, to verify its scaling properties. In \([1]\), the author derived a closed form formula for the capacity calculation for unequal number of transmit and receive antenna with number of antennas on one side of the transmission system going to infinity. In \([10]\) the authors derived exact expression for the capacity of MIMO systems but their final equation involves integral which cannot be solved in closed form/exactly. In \([3]\) the authors also consider capacity when there are correlations. In this paper, we present a direct approximation of the ergodic channel capacity using two commonly used mathematical techniques in statistical physics: Grassmann variables, and replica analysis. Although our analysis is valid for large (but finite) number of antennas, it gives excellent approximation to even 2/3 number of antennas system. The results of this approximation are compared with channel capacities generated using Monte-Carlo numerical simulations.
2. SYSTEM MODEL

We will consider a discrete-time flat-fading channel model with $N$ receive and $N$ transmit antennas. The motivation for such a model is a point-to-point wireless communication link. We assume that communication occurs on a burst basis, and that the channel can be modeled as linear and time-invariant during a burst. Hence, a simplified discrete-time baseband model of the system can be written as:

$$y_j = \sum_{i=1}^{N} h_{ij} x_i + v_j,$$

where $y_j$ is the received signal at receiver antenna $j$, reflecting contributions from all transmit antenna signals $x_i$, and $v_j$ represents the additive white Gaussian noise seen at each receiver antenna. We will assume (without loss of generality) that the noise has unit generalized variance (i.e., $E[vv^T] = I_N$). The signal vector received at the output can be written in matrix form as

$$\begin{pmatrix}
  y_1 \\
  \vdots \\
  y_N
\end{pmatrix} =
\begin{pmatrix}
  h_{11} & \cdots & h_{N1} \\
  \vdots & \ddots & \vdots \\
  h_{1N} & \cdots & h_{NN}
\end{pmatrix}
\begin{pmatrix}
  x_1 \\
  \vdots \\
  x_N
\end{pmatrix} +
\begin{pmatrix}
  v_1 \\
  \vdots \\
  v_N
\end{pmatrix},$$

or in vector notation we have,

$$y = Hx + v,$$  \hspace{1cm} (2)

It is assumed that the channel is stationary, ergodic and independent of the channel inputs $x$ and the noise $v$.

The following notations are used in the paper: $I_N$ represents the $N \times N$ identity matrix, $\text{Tr}(X)$ is the trace of $X$, $\det(Y)$ is the determinant of $Y$, and $\otimes$ denotes the Kronecker product.

We assume that the overall input power at the transmitter is constrained to $\rho$,

$$\text{Tr}(E[xx^H]) \leq \rho.$$  \hspace{1cm} (3)

We shall now use this linear model to derive the capacity of a MIMO communication system, under certain further assumptions on the system parameters.

3. ERGODIC CAPACITY

We first assume that $H$ is a Gaussian random matrix whose realization is known at the receiver, or equivalently, that the channel output consists of the pair $y, H$. The input power is distributed equally over all transmitting antennas. Assuming a block fading model, then it is known that the ergodic capacity of a random MIMO channel is given by [1, 5]:

$$C = E_H \left\{ \log \det(I_N + \frac{\rho}{N} H^H H) \right\},$$

where $E_H$ denotes that the expectation is taken with respect to the ensemble statistics of $H$, which are Gaussian distributed in our case, i.e., $H \sim N(0, I_N \otimes I_N)$, and hence the envelopes are Rayleigh distributed. This is not an unreasonable model for a mobile communications environment, with reasonable antenna spacing, and small delay spread. In [5], the authors use Eq. (4) to derive bounds on the capacity. Specifically, they derive a useful lower bound of the form:

$$C/N > (1 + \rho^{-1}) \log_2(1 + \rho) - \log_2 e + \epsilon_n + o(n^{-1}).$$

(5)

The closed form formula obtained in [1] is given by (for unequal number of transmit/receive antennas)

$$C = N(\log_2(\rho + 1)).$$

(6)

In this paper, we borrow some analytical techniques from theoretical physics to derive an alternative expression for the expression in Eq. (4) for the case of equal number of antennas.
4. INTRODUCTION TO GRASSMANN VARIABLES AND ALGEBRA

One mathematical technique which can assist in finding an approximate expression for the RHS of Eq. (4) is the use of Grassmann variables. In this section, we will first provide an extremely abbreviated introduction to the concept and use of Grassmann variables. A more complete overview of Grassman algebra can be found in [4].

At their simplest, Grassmann variables are mathematical objects which obey the following anti-commutation rule:

$$\{\theta_i, \theta_j\} = \theta_i\theta_j + \theta_j\theta_i = 0, \text{ for all } i, j.$$  \hspace{1cm} (7)

The anti-commuting rule of Eq. (7) holds in particular for $i = j$, and hence we see that the square of an arbitrary variable $\theta_i$ is zero in this algebra,

$$\theta_i^2 = 0.$$ \hspace{1cm} (8)

Note that these defining properties of Grassmann variables are mathematical conventions, and we do not attempt to ascribe any physical interpretation to such variables. Because of the property expressed in Eq. (8), any function $f$ of the anti-commuting variables is a finite polynomial. A further convention in Grassmann algebra is that the complex conjugate of a product of variables $\theta_1, \theta_2, \ldots, \theta_n$, is taken to be the product of the complex conjugates of these variables.

$$(\theta_1\theta_2\ldots\theta_n)^* = \theta_1^*\theta_2^*\ldots\theta_n^*.$$ \hspace{1cm} (9)

We shall also need to define the complex conjugate of a complex conjugate $(\theta_i^*)^*$. In principle, one has different options and it is matter of taste which one is chosen. In this paper, we shall define it thus:

$$(\theta_i^*)^* = -\theta_i.$$ \hspace{1cm} (10)

At first glance, the definition in Eq. (10) is counter-intuitive, since it differs in sign on the right-hand side from the corresponding definition in conventional algebra. However, it is quite convenient for anti-commuting variables. For example, for the quantity $\theta_i^*\theta_i$, we have

$$(\theta_i^*\theta_i)^* = -\theta_i\theta_i^* = \theta_i^*\theta_i.$$ \hspace{1cm} (11)

We see from this equation that the quantity $\theta^*\theta$ does not change under complex conjugation and therefore can be considered as "real", which is consistent with our expectations.

It is also straightforward to define linear combinations of the set of anti-commuting variables $\gamma_i, i = 1, \ldots, n$, to produce new variables $\theta_i$:

$$\theta_i = \sum_{k=1}^{n} a_{ik}\gamma_k.$$ \hspace{1cm} (12)

where the $a_{ik}$ represent conventional (non-Grassmann) algebraic scalars. By calculating the product of different $\theta_i$ we can verify the following identity:

$$\theta_1\theta_2\ldots\theta_n = \gamma_1\gamma_2\ldots\gamma_n \det A,$$ \hspace{1cm} (13)

where $A_{i,k} = a_{ik}$. The definition of derivatives of Grassmann variables [4] also proceeds in an expected fashion. However, the definition of integrals over Grassmann variables requires a little care, as explained further in the next section.
4.1. Integrals over Grassmann variables

Let us consider integrals over Grassmann variables, as first introduced by Berezin [9]. They are defined formally as follows:

\[ \int d\theta_i = \int d\theta_i^* = 0, \]
\[ \int \theta_i d\theta_i = \int \theta_i^* d\theta_i^* = 1. \]

The definition in Eqs. (14), (15) is completely formal. The notation \( \int \) is only a symbol, and one should not try to imagine this integral as representing an infinite sum. This definition implies that the "differentials" \( d\theta_i, d\theta_i^* \) anti-commute with each other and with the variables \( \theta_i, \theta_i^* \).

\[ \{ d\theta_i, d\theta_i \} = \{ d\theta_i, d\theta_i^* \} = \{ d\theta_i^*, d\theta_i^* \} = 0 \]  
\[ \{ d\theta_i, \theta_i \} = \{ d\theta_i, \theta_i^* \} = \{ d\theta_i^*, \theta_i^* \} = 0. \]  

The definition in Eq. (16) is sufficient to introduce integrals of an arbitrary function. If a function depends on one variable \( \theta_i \) it must be linear in \( \theta_i \) because already \( \theta_i^2 = 0 \) (and higher order powers). Assuming that the integral of the sum of two functions equals the sum of the integrals we can calculate the integral of the function with Eq. (15). Repeated integrals are considered to be integrals over several variables. This enables us to calculate the integral of a function of an arbitrary number of variables. Note that the choice of unity on the right-hand side of Eq. (15) is completely arbitrary; one could write any finite number.

An important result for our later calculations is the formula for the integral of the Gaussian exponential of multiple Grassmann variables [4] given by,

\[ \int \exp(-\theta^* A \theta) \prod_{i=1}^{n} d\theta_i d\theta_i^* = \det A, \]

where \( A \) is an \( n \times n \) matrix whose entries are conventional (non-Grassmann) algebraic scalars and \( \theta \) is the column vector composed of Grassmann variables. We also note that

\[ \int \exp(-\theta^* b \theta_1) d\theta_i d\theta_i = b. \]

5. APPLICATION OF REPLICA ANALYSIS AND GRASSMANN VARIABLES TO CAPACITY EVALUATION

Before returning to the subject of MIMO capacity, we will also briefly introduce another technique found in theoretical physics termed replica analysis [6, 7]. In 1975, Edwards and Anderson, when studying disordered systems of spins, proposed a new method for the investigation of disordered systems—the so-called replica method. In this method, one replaces a single disordered system by \( n \) systems which are identical to the original. Then, for example, instead of calculating the free energy, \( F = T \log Z \), one calculates the quantity, \( F_n = T(\frac{Z^n - 1}{n}) \). The limit of \( F_n \) as \( n \to 0 \) coincides with the free energy \( F \). Mathematically this is represented by

\[ F = T \log Z. \]

The above expression can be calculated as

\[ F = T \left( \lim_{n \to 0} \frac{Z^n - 1}{n} \right). \]
We shall now motivate our discussion of Grassmann variables and replica analysis, by showing how they allow us to provide an alternative approximation of the ergodic capacity of the MIMO communication system discussed earlier. Recall, that the capacity is given by

\[ C = E_H \{ \log \det(I_N + \frac{\rho}{N} H^H H) \}. \]

At relatively high SNR we have

\[ C = E_H \{ \log \det(\frac{\rho}{N} H^H H) \}. \tag{22} \]

Firstly, borrowing from replica analysis, we can rewrite the capacity expression (at relatively high SNR) as

\[ C = \lim_{n \to 0} \frac{d}{dn} E_H \{ \det(\frac{\rho}{N} H^H H)^n \}. \tag{23} \]

We can write

\[ \det(\frac{\rho}{N} H^H H) = \det(\sqrt{\frac{\rho}{N}} H) \det(\sqrt{\frac{\rho}{N}} H^H) \] \tag{24} \]

We will now use Grassmann algebra to evaluate the determinant term. Introducing two sets of Grassmann variables \( \psi_{ai}, \bar{\psi}_{ai}, \chi_{ai}, \bar{\chi}_{ai} \). Using Eqs. (18), (23), (24), we have

\[ G = E_H \{ \det(\frac{\rho}{N} H^H H)^n \} = \int d\mu(H) \]

\[ \int d\mu(\psi, \chi) \exp\left( -\sqrt{\frac{\rho}{N}} \bar{\chi}_a H^H \chi_a - \sqrt{\frac{\rho}{N}} \bar{\psi}_a H \psi_a \right) \tag{25} \]

where

\[ d\mu(H) = (\pi)^{-N^2} \exp(-Tr(H H^H)) \prod_{m=1}^2 \prod_{i,j=1}^N dH^{(m)} \] \tag{26} \]

where \( d\mu(\psi, \chi) = \prod_{a=1}^N d\psi_a d\chi_a \), is the integration measure, the \( \psi_a \) is a vector of Grassmann variables and \( \bar{\psi}_a \) is the (Hermitian) complex conjugate of \( \psi_a \).

By carrying out integration with respect to \( H \), we arrive at

\[ G = \int d\mu(\psi, \chi) \exp\left\{ \frac{\rho}{N} \bar{\psi}_{ai} \psi_{aj} \bar{\chi}_{bi} \chi_{bi} \right\} \] \tag{27} \]

Note that by carrying out integration with respect to \( H \) we generate quartic terms in Grassmann variables. We introduce an auxiliary matrix \( Q_{ab} \) to decouple the quartic term in Grassmann variables (a standard technique used in statistical physics) as

\[ G = \lim_{\epsilon \to 0} \int_{C^{n \times n}} dQ \exp(-Tr(Q Q^H)) \int d\bar{\psi}_a d\chi_a \]

\[ \exp \left\{ (\bar{\psi}_{ai} \bar{\chi}_{ai}) \begin{pmatrix} \epsilon \delta_{ab} - \sqrt{\frac{\rho}{N}} Q_{ab} & \sqrt{\frac{\rho}{N}} Q_{ab} H^H \epsilon \delta_{ab} \\ \sqrt{\frac{\rho}{N}} Q_{ab} & \epsilon \delta_{ab} \end{pmatrix} \begin{pmatrix} \psi_{bi} \\ \chi_{bi} \end{pmatrix} \right\} \]

where \( \delta_{ab} \) is the Kronecker delta. Carrying out trivial Grassmann integration [4], we obtain

\[ G = \lim_{\epsilon \to 0} \int_{C^{n \times n}} dQ \exp(-Tr(Q Q^H)) \det \left( \begin{pmatrix} \epsilon I & -\sqrt{\frac{\rho}{N}} Q \\ \sqrt{\frac{\rho}{N}} Q^H & \epsilon I \end{pmatrix} \right)^N \]
which can be further written as (after simplification and ignoring a term which approaches unity in the replica limit) and also taking limit of $\epsilon$,

$$G = \int_{C^{n\times n}} dQ \exp(-\text{Tr} QQ^H) \det \left( \frac{\rho}{N} QQ^H \right)^N$$

(28)

where we have used the following property of determinants:

$$\det \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(AD - BC) \text{ if } CD = DC$$

(29)

which can be further written as (after simplification and ignoring a term which approaches unity in the replica limit)

$$G = \int_{C^{n\times n}} dQ \exp(-\text{Tr} QQ^H) \det \left( \frac{\rho}{N} QQ^H \right)^N$$

(30)

where $A, B, C, D$ are block matrices of appropriate dimensions, and where $dQ = \pi^{-n^2} \prod_{a,b} d^2 Q_{ab}$ and $d\psi = \prod_{a,i} d\psi_{ai}$, with a similar measure defined for $\chi$. A complex matrix $Q \in C^{n\times n}$ can be uniquely written using the singular value decomposition as $Q = U \Sigma V^H$, where $U \in \text{U}(n)$ and $V \in \text{V}(n)$ are unitary matrices and $\lambda_a$ are the eigenvalues of $QQ^H$. The Euclidean measure $dQ$ on $C^{n\times n}$ is related to the normalized Haar measures $dU$ on $\text{U}(n)$ and $dV$ on $\text{V}(n)$ by

$$dQ = dU dV \Delta(\lambda)^2 \prod_{a=1}^n d\lambda_a,$$

(31)

where $\Delta \lambda = \prod_{a>b}(\lambda_a - \lambda_b)$ is Vandermonde determinant. Carrying out the integration over $U$ and $V$ and using the property of Haar measures, i.e., $\int dU = \int dV = 1$, we obtain

$$G = \int_0^\infty \prod_{a=1}^n d\lambda_a \exp(-\lambda_a) \left( \frac{\rho}{N} \lambda_a \right)^N \Delta(\lambda)^2$$

(32)

up to an irrelevant constant factor that approaches unity in the replica limit. The above equation can further be written as

$$G = \int_0^\infty \prod_{a=1}^n d\lambda_a \exp\{-\lambda_a + N \ln \left( \frac{\rho}{N} \lambda_a \right)\} \Delta(\lambda)^2.$$

(33)

We now consider the large-$N$ limit using saddle point analysis (where large $N$ refers to the number of antennas). In the large-$N$ limit, the integral can be evaluated using a saddle-point approximation [8]. The saddle point method is essentially the generalization of the Laplace’s method to integrals in the complex planes, and is used for integrals which can be expressed in the form:

$$f(z) = \int_C \exp(N h(z)) dz,$$

where $C$ is some closed contour in the complex plane, $h(z)$, is analytic function of $z$ in some domain of the complex plane which contains $C$, and $N$ is a positive number. The problem is to find the asymptotic approximation for $f(z)$ with large $N$ (see [8] for more details). Returning to our problem, the saddle point equation is

$$\frac{d(N h(\lambda))}{d\lambda} = 0,$$

where

$$N h(\lambda) = -\lambda_a + N \ln \left( \frac{\rho}{N} \lambda_a \right).$$
For $\sqrt{\rho} > 1$, we can determine the point about which $Nh(\lambda)$ is expanded to first order. Differentiating $Nh(\lambda)$ with respect to $\lambda$ and putting it to zero, we obtain

$$\lambda_0 = N. \quad (34)$$

Thus integral can be approximated as

$$\int d\lambda \exp(Nh(\lambda)) = \exp((Nh(\lambda_0)) + (d Nh(\lambda))_{\lambda=\lambda_0}). \quad (35)$$

The second term (in the exponential term) on the right hand side of the above equation is zero because the first derivative is zero in the saddle point approximation. Substituting this result into the capacity expression in Eq. (23), and taking the limit gives

$$C = N \log_2 e \cdot (\ln \rho - 1) \text{ bps/Hz} \quad (36)$$

which provides a useful approximation to the ergodic capacity under the chosen assumptions. Note the striking resemblance between this expression and the expression given in [1] and also lower bound calculated by Foschini [5]. In [1] the equation was derived for the case of uncorrelated and unequal number of antennas at the transmitter and at the receiver with number of antennas going to infinity on one of the end while the other end is kept fixed. The above equation shows that for large (but finite) number of antennas, the capacity scales linearly with the number of antennas. The above result can be thought of as generalization of [1] to large (but finite) equal number of transmit/receive antennas case.

5.1. Correlated MIMO system capacity using replica analysis

Now we consider to come up with the closed form expression for ergodic capacity for correlated MIMO case. We assume that the correlation is on one side of the transmission system. The ergodic capacity of the MIMO channel is given by

$$C = E_H \{ \log \det(I_N + \frac{\rho}{N}H^H H) \}, \quad (37)$$

where the expectation is with respect to probability given by

$$P(H) = C_1 \exp(-Tr(\Sigma^{-1} H^H H)), \quad (38)$$

where $C_1$ is irrelevant constant. $\Sigma$ is the correlation matrix. Carrying out transformation $H \Sigma^{-\frac{1}{2}} \rightarrow H$, we have for capacity expression (at relatively high SNR)

$$C = E_H \{ \log \det(\frac{\rho}{N} \Sigma H^H H) \},$$

with probability measure given by

$$P(H) = C_1 \exp(-Tr(H^H H)).$$

We can take out $\Sigma$ from the determinant and we end up with the capacity expression as in previous section. It is straightforward to show that $\det(\Sigma) < 1$. Carrying out similar analysis as is done in previous section gives

$$C = \log_2 e (\ln(\det \Sigma) + N(\ln \rho - 1)) \text{ bps/Hz.} \quad (38)$$
Figure 1. Capacity results for N=M=4 antenna system. Solid curve indicates capacity using Monte-Carlo simulations. Star-solid line is capacity evaluation using Eq. (36), solid-plus curve indicates capacity evaluation using closed form result in [1], and dashed line is the Foschini’s lower bound.

Figure 2. Capacity results for N=M=2 antenna system. Solid curve indicates capacity using Monte-Carlo simulations. Star-solid line is capacity evaluation using Eq. (36), solid-plus curve indicates capacity evaluation using closed form result in [1], and dashed line is the Foschini’s lower bound.
The above equation shows that the capacity degrades as the correlation increases between antennas. We can also extend the above result to correlations at both ends. Here we give the result directly for correlations on
both ends as

\[ C = \log_2 e (\ln(\det \Sigma_1) + \ln(\det \Sigma_2) + N(\log \rho - 1)) \text{ bps/Hz} \]  

(39)

where \( \Sigma_1, \Sigma_2 \) are correlation matrices.

To illustrate the accuracy of this approximation, in Fig. 1 we have capacity results for four transmit and four receive antennas. The solid line represents the capacity evaluation using Monte Carlo type simulations. The dashed line represents Foschini’s lower bound. Monte Carlo simulation is carried by averaging over 10000 independent channel realizations. The star-solid line gives the capacity evaluation using Eq. (36). Solid-plus curve is the capacity evaluated using closed form formula in [1]. It is clear from the figure that our result is in close agreement even for four transmit and the same number of receive antennas and gives accurate results for relatively high SNR. Figure 2 shows similar results for two transmit and two receive antennas case. The solid line represents the capacity evaluation using Monte Carlo type simulations. The dashed line represents Foschini’s lower bound. Monte Carlo simulation is carried by averaging over 10000 independent channel realizations. The star-solid line gives the capacity evaluation using Eq. (36). Solid-plus curve is the capacity evaluated using closed form formula in [1]. Figure 3 gives the results for the capacity for correlated system. We consider \( 2 \times 2 \) antenna case. Solid curve is the capacity evaluation using Monte Carlo simulations and star-solid line is capacity evaluation using Eq. (38). In Fig. 4 we have \( 2 \times 2 \) system with correlations at both ends of the transmission system with the same correlation matrix. Solid line is capacity using Monte Carlo simulations and star-solid line is capacity calculation using Eq. (39). We can see from Fig. 3 and Fig. 4 that correlation degrades the capacity of the system.

6. CONCLUSION

In this paper, we derived a closed form expression for the ergodic channel capacity of MIMO systems with equal numbers of transmit and receive antennas, flat fading, and independently distributed antenna gains for correlated and uncorrelated case. The expression was obtained by using replica analysis and non-commuting (Grassmann) variables, with the key step being writing a determinant expression in the form of a Gaussian integral over Grassmann variables. It is shown that in large (but finite) system too, the capacity scales linearly with the number of antennas. Our expression has very close resemblance with the expression obtained in [1] and [5]. Although the analysis was derived for the case of a large number of antennas, it gives satisfactory results for even \( 2 \times 2 \) and \( 4 \times 4 \) antenna system. In the correlated case the capacity of the system decreases rapidly as the correlation increases.

REFERENCES
