

On the Capacity of Multiple-Antenna Systems and Parallel Gaussian Channels With Amplitude-Limited Inputs

Ahmad ElMoslmany, *Student Member, IEEE*, and Tolga M. Duman, *Fellow, IEEE*

Abstract—We propose upper and lower bounds on the capacity of multiple-input multiple-output (MIMO) systems with amplitude-limited inputs. The results are derived by considering an equivalent channel via singular value decomposition, and by enlarging and reducing the corresponding feasible region of the channel input vector, for the upper and lower bounds, respectively. We analytically characterize the asymptotic behavior of the derived bounds for high and low noise levels, and study the gap between them. We also consider parallel Gaussian channels with peak and average power-constrained inputs. For such channels, the capacity-achieving distribution has been reported in the literature to be discrete, which can be computed using numerical optimization techniques. However, there is no closed-form expression and finding the capacity-achieving distribution is computationally tedious. With this motivation, we derive approximate expressions for the capacity at low and high noise variance levels. We illustrate our findings on both MIMO channels and parallel Gaussian channels via several numerical examples.

Index Terms—MIMO systems, parallel Gaussian channels, amplitude-limited inputs, channel capacity.

I. INTRODUCTION

CAPACITIES of various single-user memoryless continuous-alphabet channels with different constraints on the channel input have been extensively studied. The most commonly used constraint is the average power constraint for which the capacity of the Gaussian channel is first derived by Shannon. Although transmissions subject to both average and peak power constraints is of utmost importance as it is a better representative of the limitations in practical communication systems, channel capacity studies under these assumptions have been very limited.

Capacity of a Gaussian channel with peak and average power constraints was first studied by Smith [1] where he

Manuscript received October 4, 2015; revised February 12, 2016 and April 7, 2016; accepted May 4, 2016. Date of publication May 16, 2016; date of current version July 12, 2016. This work is funded by National Science Foundation under the contract NSF-ECCS 1102357 and by the EC Marie Curie Career Integration Grant PCIG12-GA-2012-334213. This paper was presented at the IEEE Black Sea Conference on Communications and Networking, May 2014. The associate editor coordinating the review of this paper and approving it for publication was V. Stankovic.

A. ElMoslmany was with the School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287-5706 USA. He is now with Polaris Wireless, Mountain View, CA 94043, USA (email: aelmoslmany@polariswireless.com).

T. M. Duman is with the Department of Electrical and Electronics Engineering, Bilkent University, Ankara 06800, Turkey (e-mail: duman@ee.bilkent.edu.tr).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TCOMM.2016.2569085

showed that under these constraints on the input, the capacity-achieving distribution is discrete. Finding the mass point locations of this discrete distribution and the associated probabilities are enabled through a numerical optimization algorithm, which is feasible since the problem is reduced to a finite-dimensional convex optimization problem. Tchamkerten [2] extended Smith's results on channel capacity with amplitude limited inputs to general additive noise channels and he derived sufficient conditions on the noise probability density function that guarantee that the capacity-achieving input has a finite number of mass points. Capacity of multi-input channels with amplitude-limited inputs has been considered recently, but in a different communication scenario. Specifically, the authors in [3] and [4] study the capacity of a multiple access channel (MAC) with amplitude-constrained inputs, and they show that the sum-capacity achieving distribution is discrete.

Our interest in the first part of the paper is the capacity of multiple-input multiple-output (MIMO) systems under amplitude limited inputs. Although closely related to the MAC scenario, our set-up is significantly different as there is a single message being transmitted over two antennas as opposed to two independent messages being transmitted separately from two transmitters. Unfortunately, the existing results for amplitude limited inputs do not generalize to the MIMO scenario fully. In particular, proving that the capacity achieving distribution is discrete (as in the point to point scenario and the limited results for the MAC case) does not seem feasible. The reason behind this claim is that one of the main ingredients in all the proofs (e.g., [2], [3], [5], [6]) is what is called the Identity Theorem [7] in complex analysis which holds for one dimensional functions, however, no counterpart exists for the multi-dimensional case (as was also noted by Smith in his classical work [8]).

We highlight that the capacity of multiple-antenna systems is explored in [6] (along with other channel models) in the context of conditionally Gaussian channels, and the discreteness of the capacity-achieving distribution is claimed. However, as clarified in [9], the results in the higher dimensional scenario are not fully rigorous since the Identity Theorem (known as the uniqueness theorem) is used without fulfilling its hypothesis. Since proving the discreteness of the capacity achieving input distribution for amplitude limited inputs (for all the specific set-ups studied in the literature to date) requires a proper application of the Identity Theorem (to arrive at

a contradiction), the nonexistence of such a result for the case of multi-dimensional functions makes similar set of ideas inapplicable in the MIMO scenario motivating part of the work in this paper.

We direct our attention to development of upper and lower bounds on the capacity of MIMO systems with amplitude limited inputs. The proposed bounds rely on transforming the coupled MIMO channel into decoupled (independent) channels with coupled inputs, and then relaxing the constraint imposed on these inputs or restricting them further. We obtain the relaxed and constrained regions in such a way that the problem decouples into capacity computation of independent point to point channels for which the channel capacity results are available. We detail the approach for the case of 2×2 MIMO channels, and describe how these results can be generalized for an arbitrary number of antennas. We note that for the special case of multiple-input single-output (MISO) channels the problem is still “one-dimensional,” and it is easy to argue that the capacity achieving distribution discrete. However, there is no uniqueness result as the capacity-optimization problem boils down to an optimization problem that aims to choose a distribution for the sum of the channel inputs which maximizes the mutual information. Hence, different marginal distributions for the inputs can be chosen such that the distribution of their sum is the same as the distribution that results from the solution of the optimization problem. Some of our results along these lines have been presented in [10]. Our contributions in the context of general MIMO channels can be summarized as follows: 1) we provide an upper (and lower bound) on the capacity by relaxing (and enlarging) the feasible region in such a way that the MIMO channel can be decomposed into parallel (scalar) channels with amplitude limited inputs whose capacity characterization is possible, 2) we develop tight approximations to the channel capacity for high and low noise variance levels analytically.

In the second part of the paper, we investigate the capacity of parallel Gaussian channels under peak and average power constraints. It is straightforward to show that the capacity achieving distribution for this set-up is discrete which is a result that can be implied directly from Smith’s original work. Also, the authors in [6] show (through a different approach) that the capacity-achieving distribution is discrete. The authors in [11] study the capacity of Gaussian vector channel under peak and average power constraints and extend the results of [5] to show that the capacity achieving distribution has a finite number of mass points for its amplitude and the points are uniformly distributed on the hyper-spheres determined by the amplitude mass points. They also, via relaxing the constraints on the problem, derive bounds on the capacity of the channel. However, there is no-closed form expression and finding the capacity achieving distribution requires extensive numerical computations in determining the optimal power assignment for each of the N -parallel channels. With this motivation, we consider the behavior of these channels asymptotically in the low and high signal-to-noise ratio (SNR) regimes. Our approach is based on the following: in the very high noise variance regime, the capacity-achieving distribution consists only of two mass points [1], hence using the data

processing inequality the problem boils down to be finding the capacity of (parallel) binary input channels. In the low noise variance regime, the capacity-achieving distribution can be approximated by a continuous distribution [8]. We utilize these approximations to derive different optimal power assignment policies for different values of the noise variance. For the case of high noise variance, we use the Karush-Kuhn-Tucker (KKT) conditions to find the optimal power assignment policy. For the low noise variance case, we first adopt a technique from calculus of variations to show that the capacity-achieving distribution is a Gaussian-like distribution with a truncated domain, and then we find the optimal power assignment policy.

The paper is organized as follows. In Section II, we describe the channel model for MIMO and parallel Gaussian noise channels. In Section III, we compute the capacity of MISO systems, and we show that the capacity-achieving distribution is discrete. In Section IV, we use singular value decomposition to decouple the system into independent parallel channels. Then, we propose the upper and lower bounds by relaxing the input constraints of the new optimization problem. Finally, we describe an alternative approach providing a lower bound. In Section V, we study the capacity of the parallel Gaussian channels under peak and average power constraints. In Section VI, we present numerical examples to demonstrate our findings, and conclude the paper in Section VII.

Notation: We denote scalars by lower-case letters, e.g., x , vectors of scalars by lower-case boldface letters, e.g., \mathbf{x} . Matrices and random variables are denoted by upper-case letters, e.g., X . Vectors of random variables are denoted by upper-case boldface letters, e.g., \mathbf{X} . Superscript H denotes Hermitian (conjugate transpose) operator. We reserve \mathbf{E} for the statistical expectation and A_{x_i} to represent the peak power constraint on the i th channel. The determinant of the matrix X is denoted by $\det(X)$.

II. SYSTEM MODELS

We consider a MIMO system where the received signal \mathbf{Y} is written as

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

where H is an $N_r \times N_t$ channel matrix, N_r is the number of receive elements, and N_t is the number of transmit elements. The channel inputs $\mathbf{X} = [X_1, X_2, \dots, X_{N_t}]^T$ where $|X_i| \leq A_{x_i}$, the channel output is $\mathbf{Y} = [Y_1, Y_2, \dots, Y_{N_r}]^T$. The channel matrix H is assumed to be deterministic. The vector \mathbf{Z} denotes an additive white Gaussian noise (AWGN) vector whose elements are independent and identically distributed (i.i.d.) and each is $Z_i \sim \mathcal{N}(0, \sigma_i^2)$, where σ_i^2 is the noise variance at the i th receiver, $i = 1, 2, \dots, N_r$. We assume that the channel inputs and outputs, the channel matrix and noise terms are all real valued.

For most of our results on MIMO communications, we consider the case of a 2×2 MIMO system with $N_r = 2$ and $N_t = 2$ where we assume that the channel matrix H is non-singular. Hence, the channel input \mathbf{X} has a two-dimensional

joint distribution function $F(x_1, x_2)$ and the channel inputs are amplitude limited as $|X_1| \leq A_{x_1}$ and $|X_2| \leq A_{x_2}$. The capacity of the 2×2 MIMO system is given by

$$C = \max_{F(x_1, x_2): |X_1| \leq A_{x_1}, |X_2| \leq A_{x_2}} I(Y_1, Y_2; X_1, X_2). \quad (2)$$

For the second scenario, we have a set of N parallel Gaussian channels with outputs given by

$$Y_i = X_i + Z_i, \quad \forall i = 1, 2, \dots, N, \quad (3)$$

which can be written in the vector form as

$$\mathbf{Y} = \mathbf{X} + \mathbf{Z}, \quad (4)$$

where $\mathbf{X} = [X_1, X_2, \dots, X_N]^T$, $\mathbf{Y} = [Y_1, Y_2, \dots, Y_N]^T$, $|X_i| \leq A_{x_i}$, and $\mathbb{E}[\mathbf{X}^T \mathbf{X}] \leq P_0$. The Gaussian noise vector \mathbf{Z} is given by $\mathbf{Z} = [Z_1, Z_2, \dots, Z_N]^T$, where Z_i is an AWGN with zero mean and variance σ_i^2 , $i = 1, 2, \dots, N$. The mutual information between the input vector \mathbf{X} and the output vector \mathbf{Y} is given by

$$I(\mathbf{X}; \mathbf{Y}) = h(\mathbf{Y}) - \sum_{i=1}^N h(Z_i), \quad (5)$$

where $h(Z_i)$ is the differential entropy of the Gaussian noise random variable Z_i which is given by $h(Z_i) = \frac{1}{2} \log(2\pi e \sigma_i^2)$. The optimal input distribution for the N parallel Gaussian channels under peak and average power constraints has been shown in to be discrete and independent of each other [6]. Hence,

$$\begin{aligned} I(X_1, X_2, \dots, X_N; Y_1, Y_2, \dots, Y_N) \\ = \sum_{i=1}^N h(Y_i) - \sum_{i=1}^N \frac{1}{2} \log(2\pi e \sigma_i^2). \end{aligned} \quad (6)$$

Therefore, the channel capacity is given by

$$C = \max_{\substack{F(x_1, x_2, \dots, x_N): |X_i| \leq A_{x_i}, \\ \mathbb{E}[\mathbf{X}^T \mathbf{X}] \leq P_0}} \sum_{i=1}^N h(Y_i) - \sum_{i=1}^N \frac{1}{2} \log(2\pi e \sigma_i^2). \quad (7)$$

Each of the terms in the summation is maximized by a discrete input distribution with a finite number of mass points, and the overall problem boils down to a power assignment one.

III. CAPACITY OF MISO SYSTEMS WITH AMPLITUDE-LIMITED INPUTS

We first consider multiple-input single-output systems. Since there is only one receive antenna in this case, the received signal Y can be written as

$$Y = h_1 X_1 + h_2 X_2 + Z, \quad (8)$$

where h_1, h_2 are the channel coefficients from the two transmit antennas to the receive antenna. Define an auxiliary variable U such that $U = h_1 X_1 + h_2 X_2$. Since X_1 and X_2 are amplitude-limited, U will also be amplitude limited, i.e.,

$$-|h_1|A_{x_1} - |h_2|A_{x_2} \leq U \leq |h_1|A_{x_1} + |h_2|A_{x_2}. \quad (9)$$

Thus, the received signal Y can be written as $Y = U + Z$, and the problem boils down to the classical point-to-point scalar

problem that has been investigated by Smith. Hence, the probability density function of the auxiliary random variable U that achieves the capacity is discrete, i.e.,

$$f_U(u) = \sum_{i=0}^{N_u-1} p(u_i) \delta(u - u_i) \quad (10)$$

where $p(u_i)$ is the probability associated with the mass point u_i , $\delta(\cdot)$ is the Dirac delta function. the number of mass points N_u are to be determined numerically by solving the capacity optimization problem using the algorithm given in [1]. The specific channel inputs X_1 and X_2 can be arbitrarily generated such that their weighted sum (weighted by the channel coefficients) follows the optimal probability mass function of the random variable U .

IV. BOUNDS ON THE CAPACITY OF 2×2 MIMO SYSTEMS WITH AMPLITUDE-LIMITED INPUTS

For a 2×2 MIMO system, we obtain an equivalent model via singular value decomposition of the channel matrix H , i.e., $H = U \Omega W^H$. That is,

$$\tilde{\mathbf{Y}} = \Omega \tilde{\mathbf{X}} + \tilde{\mathbf{Z}}, \quad (11)$$

where the equivalent channel outputs and inputs are $\tilde{\mathbf{Y}} = U^H \mathbf{Y}$, $\tilde{\mathbf{X}} = W^H \mathbf{X}$, respectively. The equivalent noise vector is $\tilde{\mathbf{Z}} = U^H \mathbf{Z}$, where U and W are unitary matrices. Define $V = W^H$. Since the amplitude of the first channel input is constrained by A_{x_1} and the amplitude of the second input is constrained by A_{x_2} , the domain of \mathbf{X} is a rectangular region. However, after applying the singular value decomposition, in the equivalent formulation, the region defining the input constraint turns out to be a parallelogram. Further, this region will be centered at origin (since the original rectangular region is symmetric around origin).

Define the following terms that characterize the new input constraint

$$a = \frac{\det(V)}{v_{22}}, \quad b = \frac{v_{12}}{v_{22}}, \quad c = -\frac{\det(V)}{v_{21}}, \quad d = \frac{v_{11}}{v_{21}},$$

where v_{ij} is the ij th element of the matrix V . If the vector $\tilde{\mathbf{X}} = [X, Y]^T$, then the feasible region of the equivalent channel in (11) is,

$$-\frac{1}{a}X + \frac{b}{a}Y \leq A_{x_1}, \quad \frac{1}{a}X - \frac{b}{a}Y \leq A_{x_1},$$

and

$$-\frac{1}{c}X + \frac{d}{c}Y \leq A_{x_2}, \quad \frac{1}{c}X - \frac{d}{c}Y \leq A_{x_2}.$$

We derive upper and lower bounds on the capacity using this new formulation. We obtain the lower bound by looking for a smaller feasible region inside the parallelogram, i.e., we consider a rectangle, and we compute the corresponding mutual information between the input and output with the channel input vector constrained to be inside this rectangular region. For the upper bound, we follow a similar approach, i.e., we look for the smallest rectangle that inscribes the parallelogram. This geometrical interpretation of the approach is illustrated in Fig. 1. This approach enables us to separate the two-dimensional problem into two one-dimensional problems whose solutions are readily available.

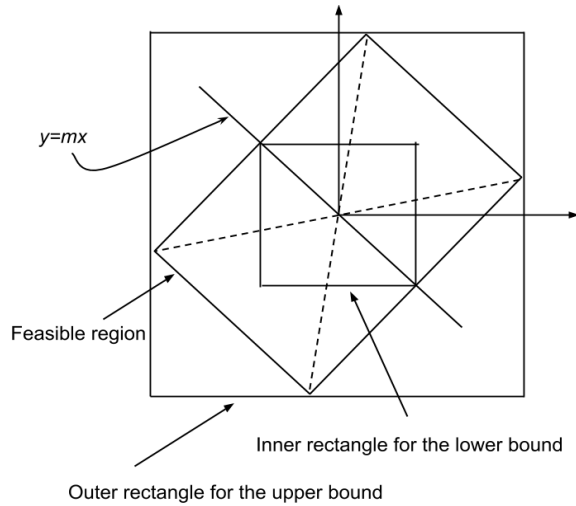


Fig. 1. Feasible region, and the rectangular regions used for the lower and upper bounds.

A. An Upper Bound on the Capacity of 2×2 MIMO Systems With Amplitude-Limited Inputs

A capacity upper bound is derived by solving the capacity optimization problem over the smallest rectangle that inscribes the original feasible region. This rectangle is constructed from the intersection points of every pair of lines forming the feasible region. The four intersection points are denoted by the pairs $\{(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4)\}$. The first intersection point is

$$x_1 = \frac{bc}{d-b}A_{x_2} - \frac{ad}{d-b}A_{x_1}, \quad y_1 = \frac{c}{d-b}A_{x_2} - \frac{a}{d-b}A_{x_1},$$

the second point is

$$x_2 = -\frac{bc}{d-b}A_{x_2} - \frac{ad}{d-b}A_{x_1}, \quad y_2 = -\frac{c}{d-b}A_{x_2} - \frac{a}{d-b}A_{x_1},$$

the third point is

$$x_3 = \frac{bc}{d-b}A_{x_2} + \frac{ad}{d-b}A_{x_1}, \quad y_3 = \frac{c}{d-b}A_{x_2} + \frac{a}{d-b}A_{x_1},$$

and the fourth point is

$$x_4 = -\frac{bc}{d-b}A_{x_2} + \frac{ad}{d-b}A_{x_1}, \quad y_4 = -\frac{c}{d-b}A_{x_2} + \frac{a}{d-b}A_{x_1}.$$

Using the geometric interpretation of the feasible region, it is easy to show that in the equivalent formulation of the 2×2 MIMO system, the amplitude limits on the two inputs are

$$\begin{aligned} \Delta x_{upp} &= \left| \frac{bc}{d-b}A_{x_2} \right| + \left| \frac{ad}{d-b}A_{x_1} \right|, \\ \Delta y_{upp} &= \left| \frac{c}{d-b}A_{x_2} \right| + \left| \frac{a}{d-b}A_{x_1} \right|. \end{aligned} \quad (12)$$

These new input constraints can be used to compute an upper bound on the channel capacity of the original MIMO system. Namely, the upper bound of the channel capacity is given by

$$C \leq C_0(\Delta x_{upp}) + C_0(\Delta y_{upp}), \quad (13)$$

where $C_0(A)$ is the capacity of the point-to-point AWGN channel for a given amplitude constraint A (computed using Smith's approach in [1]).

B. A Lower Bound on the Capacity of MIMO Systems With Amplitude-Limited Inputs

A lower bound on the capacity of the channel can be found by optimizing the mutual information over a smaller rectangular region inside the feasible region (parallelogram). To find such a rectangle, we determine the intersection of a straight line, $y = mx$, that passes through the origin (as the region is centered at the origin) and the boundary of the feasible region. In this case, it is easy to show that

$$\Delta x_{low} = \min \left(\left| \frac{aA_{x_1}}{1+bm} \right|, \left| \frac{aA_{x_1}}{1-bm} \right|, \left| \frac{cA_{x_2}}{1+dm} \right|, \left| \frac{cA_{x_2}}{1-dm} \right| \right), \quad (14)$$

$$\Delta y_{low} = \min \left(\left| \frac{amA_{x_1}}{1+bm} \right|, \left| \frac{amA_{x_1}}{1-bm} \right|, \left| \frac{cmA_{x_2}}{1+dm} \right|, \left| \frac{cmA_{x_2}}{1-dm} \right| \right), \quad (15)$$

for some arbitrary values for the slope m such that the set of points

$$\{(l\Delta x_{low}, k\Delta y_{low}) \in \mathcal{R} : l, k \in \{1, -1\}\},$$

where \mathcal{R} is the feasible region. Thus, the lower bound on the channel capacity is given by

$$C \geq C_0(\Delta x_{low}) + C_0(\Delta y_{low}). \quad (16)$$

C. Capacity of 2×2 MIMO Systems With Amplitude- and Power-Limited Inputs

Smith in [1] showed that for any amplitude-limited and power-limited point-to-point Gaussian channel, a unique capacity-achieving distribution exists and it is discrete. Again, extending these results to the case of MIMO systems does not seem feasible since there is no result corresponding to the Identity Theorem used in Smith's proof for multi-dimensional functions [8]. (See also the results in [9].) However, we can follow a similar approach as in the previous subsection to find upper and lower bounds on the capacity of MIMO systems with amplitude-limited inputs by relaxing the constraints on the amplitude and solving the capacity optimization problem over rectangular regions that inscribe and are inscribed by the original feasible region, respectively. We do not pursue this problem formulation any further in this work.

D. Asymptotic Bounds on the Capacity of the 2×2 MIMO Systems With Amplitude-Limited Inputs

In this section we study the asymptotic behavior of the upper and lower bounds on the capacity of MIMO systems at very high and low noise levels.

1) *Very Low Noise Levels:* For very low noise variances, the entropy of the noise is very small compared to the entropy of the input and hence its contribution to the output entropy can be neglected. Thus, the following approximations are valid [1],

$$h(\tilde{\mathbf{Y}}) \gg h(\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}) \text{ and } h(\tilde{\mathbf{X}}) \gg h(\tilde{\mathbf{X}}|\tilde{\mathbf{Y}}).$$

Since $I(\tilde{\mathbf{X}}; \tilde{\mathbf{Y}}) = h(\tilde{\mathbf{X}}) - h(\tilde{\mathbf{X}}|\tilde{\mathbf{Y}}) = h(\tilde{\mathbf{Y}}) - h(\tilde{\mathbf{Y}}|\tilde{\mathbf{X}})$, we have $h(\tilde{\mathbf{X}}) \approx h(\tilde{\mathbf{Y}})$. That is, the capacity can be approximated as

$$\begin{aligned} C &= \max_{F(\tilde{\mathbf{X}})} I(\tilde{\mathbf{X}}; \tilde{\mathbf{Y}}), \\ &\approx \max_{F(\tilde{\mathbf{X}})} h(\tilde{\mathbf{X}}) - h(\tilde{\mathbf{Y}}|\tilde{\mathbf{X}}), \\ &\stackrel{(a)}{=} \log(4A_{x_1}A_{x_2}) - \frac{1}{2} \log((2\pi e \sigma_1^2) - \frac{1}{2} \log(2\pi e \sigma_2^2)), \end{aligned}$$

where (a) is justified by the fact that the entropy of a random variable with finite support is maximized by a uniform distribution. Therefore, the capacity of the 2×2 MIMO system can be upper and lower bounded for low noise variances as

$$C \leq \log(4\Delta x_{upp} \Delta y_{upp}) - \frac{1}{2} \log(2\pi e \sigma_1^2) - \frac{1}{2} \log(2\pi e \sigma_2^2), \quad (17)$$

$$C \geq \log(4\Delta x_{low} \Delta y_{low}) - \frac{1}{2} \log(2\pi e \sigma_1^2) - \frac{1}{2} \log(2\pi e \sigma_2^2). \quad (18)$$

The lower bound on the capacity can be optimized by choosing the slope m (as defined in the previous section) that maximizes the mutual information between the input and the output. We have

$$\begin{aligned} C \geq \max_m \log(4|\Delta x_{low} \Delta y_{low}|) - \frac{1}{2} \log(2\pi e \sigma_1^2) \\ - \frac{1}{2} \log(2\pi e \sigma_2^2), \quad (19) \end{aligned}$$

where the set of points

$$\{(l\Delta x_{low}, k\Delta y_{low}) \in \mathcal{R} : l, k \in \{1, -1\}\},$$

where Δx_{low} , Δy_{low} , and \mathcal{R} are as defined in the previous section.

2) *Very High Noise Levels:* For very high noise variances, the optimal distribution consists of only two mass points with the same probabilities [12]. The capacity of this discrete-time binary-input AWGN is well known [13], and the upper and lower bounds on the capacity are

$$C \leq g\left(\frac{\Delta x_{upp}}{\sigma_1}\right) + g\left(\frac{\Delta y_{upp}}{\sigma_2}\right), \quad (20)$$

$$C \geq g\left(\frac{\Delta x_{low}}{\sigma_1}\right) + g\left(\frac{\Delta y_{low}}{\sigma_2}\right), \quad (21)$$

where

$$g(x) = 1 - \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(u-x)^2}{2}} \log_2(1 + e^{-2ux}) du. \quad (22)$$

3) *The Gap Between the Upper and Lower Bounds:* For very low noise variances, it is easy to see that the gap between the upper and lower bounds does not depend on the amplitude constraint if $A_{x_1} = A_{x_2}$. From (12), and if $A_{x_1} = A_{x_2} = A_0$ we have,

$$\Delta x_{upp} = G_{upp}A_0, \quad \Delta y_{upp} = H_{upp}A_0,$$

where G_{upp} and H_{upp} are only function of the channel coefficients. Also from (14) and (15),

$$\Delta x_{low} = G_{low}A_0, \quad \Delta y_{low} = H_{low}A_0,$$

where G_{low} and H_{low} are only functions of the channel coefficients.

Thus, the gap between the upper and lower bounds ΔC can be written as

$$\begin{aligned} \Delta C &= \log(4G_{upp}H_{upp}A_0^2) - \log(4G_{low}H_{low}A_0^2), \\ &= \log\left(\frac{G_{upp}H_{upp}}{G_{low}H_{low}}\right), \end{aligned}$$

which is independent of the amplitude constraints imposed on the inputs.

E. Alternative Lower Bounds on the Capacity

It is also possible to obtain achievable rates by simply considering an input distribution (satisfying the input constraints) without paying attention to its optimality and computing the mutual information between the input and the output. For instance, by considering a discrete input distribution constructed from the cartesian product of the input distributions corresponding to the upper bound with mass points falling within the feasible region one can compute achievable rates. While it is clear that other input distributions can also be adopted, in the numerical results section VI, we will illustrate that this may be a good selection (at least for the examples considered) and the achievable rates computed approach the upper bounds, hence allowing for a more accurate channel capacity characterization. For a given MIMO capacity calculation problem, this approach should be considered along with the approach in the previous sections as we cannot definitively say that one is superior to the other.

F. Bounds on the Capacity of General MIMO Systems With Amplitude-Limited Inputs

For the case of MIMO systems with larger number of transmit and receive elements, a similar approach can be followed to derive upper and lower bounds on the capacity of the channel with amplitude constraints. However, the feasible region of the capacity optimization problem will not be a simple rectangle in the two-dimensional space as in the case of 2×2 systems.

Consider an $N_t \times N_r$ MIMO channel. We first apply the singular value decomposition, then the new channel input vector is

$$\tilde{\mathbf{X}} = W^H \mathbf{X}, \quad (23)$$

where W is a unitary matrix. We also have $\mathbf{X} = W \tilde{\mathbf{X}}$. Define the vector $\mathbf{a} = [A_{x_1}, A_{x_2}, \dots, A_{x_{N_t}}]^T$, where A_{x_i} represents the amplitude constraint on the i th antenna element. As a result, the new feasible region is constructed as

$$\begin{bmatrix} W & 0 \\ 0 & -W \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{X}} \\ \tilde{\mathbf{X}} \end{bmatrix} \leq_K \begin{bmatrix} \mathbf{a} \\ \mathbf{a} \end{bmatrix} \quad (24)$$

where $\tilde{\mathbf{X}} = [\tilde{X}_1, \tilde{X}_2, \dots, \tilde{X}_{N_t}]^T$, and $K = \mathbb{R}^+$ is the nonnegative orthant (i.e., the inequality is componentwise). Hence, as in the case of two transmit antennas, we can find smaller and larger feasible regions, and decouple the capacity determination problem to complete the solution.

We note that these arguments can also be extended to the case of transmission over bandpass channels for which the channel coefficients and the inputs can be taken as complex variables, and hence the inputs are complex variables. However, with MIMO systems, the dimensionality of the problem grows further.

1) *Upper Bound on the Capacity of $N_t \times N_r$ MIMO Channels:* For the upper bound, we solve the capacity optimization problem over a larger feasible region that inscribes the original feasible region of the problem. First, we assume that the $N_r \times N_t$ channel matrix H has M' nonzero eigen values. In this case, the new feasible region is an N_t -dimensional hyperrectangle which is constructed from the vertex points of the feasible region in (24). For N_t -dimensional hyperrectangle there are 2^{N_t} vertices, the coordinates of the i th vertex is denoted as \hat{x}_i

$$\begin{aligned} \hat{x}_1 &= (z_{11}, z_{21}, \dots, z_{N_t 1}), \\ \hat{x}_2 &= (z_{12}, z_{22}, \dots, z_{N_t 2}), \\ &\vdots \\ \hat{x}_i &= (z_{1i}, z_{2i}, \dots, z_{N_t i}), \end{aligned}$$

for all $i = 1, 2, \dots, 2^{N_t}$. Thus, the new relaxed constraints on the new channel inputs are $\tilde{\mathbf{a}}$ which is given by

$$\tilde{\mathbf{a}} = [\tilde{A}_1, \tilde{A}_2, \dots, \tilde{A}_{N_t}]^T \quad (25)$$

where \tilde{A}_i is given by

$$\tilde{A}_i = \max_j |z_{ij}|, \quad (26)$$

we assume that the vector $\tilde{\mathbf{A}}$ is an ordered vector. As a result, the upper bound on the capacity is

$$C \leq \sum_{i=1}^{M'} \log(2\tilde{A}_i) - \frac{1}{2} \sum_{i=1}^{M'} \log(2\pi e \sigma_i^2). \quad (27)$$

2) *Lower Bound on the Capacity of $N_t \times N_r$ MIMO Channels:* For the lower bound, we solve the capacity optimization problem over smaller hyperrectangle that is inscribed by the feasible region of the original problem. To facilitate this, for instance, we can assume that this hyperrectangle is just a scaled version of the hyperrectangle derived for the upper bound and then a feasibility problem is solved to find the scaling coefficients. As another approach, we can consider mass points within the feasible region (at certain locations, e.g., determined by the locations of the mass

points from the upper bound falling in the feasible region), and run standard optimization algorithms to find the optimal masses to obtain an achievable rate for the original system.

3) *Asymptotic Bounds on the Capacity of $N_t \times N_r$ MIMO Systems With Amplitude-Limited Inputs:* The asymptotic results derived in Section IV-D can be extended to the case of $N_t \times N_r$ MIMO systems. At high values of noise variances, the capacity is achieved by a discrete input distribution with two mass points and hence the channel capacity C can be bounded as

$$C \leq \sum_{i=1}^{M'} g\left(\frac{\tilde{A}_i}{\sigma_i}\right), \quad (28)$$

where $g(\cdot)$ is defined in (22).

However, for low values of noise variance, the capacity is achieved by a continuous distribution and it can be approximated as

$$C \approx \sum_{i=1}^{M'} \log(2\tilde{A}_i) - \sum_{i=1}^{M'} \frac{1}{2} \log(2\pi e \sigma_i^2). \quad (29)$$

V. CAPACITY OF INDEPENDENT PARALLEL GAUSSIAN CHANNELS WITH AMPLITUDE AND POWER-LIMITED INPUTS

In this section, we turn our attention to a more tractable problem compared to the case of MIMO channels, namely to parallel Gaussian channels, and we study the capacity of N independent parallel Gaussian channels with peak and average power-limited inputs. Specifically, we consider analytical approximations to the capacity at low and high noise level regimes.

A. Capacity of Parallel Gaussian Channels for Very High Noise Levels

At very high noise levels, the optimal input distribution that maximizes the mutual information between the input and the output (for amplitude-limited and power-limited inputs) is discrete with only two mass points and each one has a probability of one half [1]. Using the data processing inequality, one can easily argue that the capacity of each parallel channel is lower bounded by the capacity of a binary symmetric channel denoted by C_{BSC} . Recall that $C_{BSC} = 1 - H(p)$, where p is the cross-over probability which can be computed for the problem at hand as

$$p = Q\left(\sqrt{\frac{P}{\sigma^2}}\right), \quad (30)$$

where P is the average transmit power. We consider N parallel channels where the power assigned to each is P_1, P_2, \dots, P_N , respectively. Thus, the capacity of the channel is determined by the power assigned to each channel, and the capacity optimization problem boils down to a power assignment problem, i.e., choosing the optimal power for each channel so that the mutual information is maximized.

Since the input distribution for each channel input consists of two mass points with equal probabilities, we also have

$$P_i \leq A_{x_i}^2, \quad \forall i = 1, 2, \dots, N. \quad (31)$$

Define a function $J(P_i)$, which is basically the binary entropy function, as

$$J(P_i) = -Q\left(\sqrt{\frac{P_i}{\sigma_i^2}}\right) \log\left(Q\left(\sqrt{\frac{P_i}{\sigma_i^2}}\right)\right) - \left(1 - Q\left(\sqrt{\frac{P_i}{\sigma_i^2}}\right)\right) \log\left(1 - Q\left(\sqrt{\frac{P_i}{\sigma_i^2}}\right)\right). \quad (32)$$

Then, the channel capacity of the parallel Gaussian channel is lower bounded by,

$$C \geq \max_{P_i, \forall i=1,2,\dots,N, 0 \leq P_i \leq A_{x_i}^2} N - \sum_{i=1}^N J(P_i). \quad (33)$$

Solving this optimization problem results in the following power assignment policy (the proof is detailed in Appendix A). Let us assume that the power assigned to the i th channel is P_i^* . Then, we consider three cases $P_i^* = 0$, $0 < P_i^* < A_{x_i}^2$, and $P_i^* = A_{x_i}^2$ separately, for $i = 1, 2, \dots, N$, and solve for the candidate power assignments for the resulting $3^N - 2$ cases. For each case, considering only the non-zero power assignments with $P_i^* < A_{x_i}^2$, the water-filling parameter ν is chosen such that $P_i^* = g_i^{-1}(\nu)$, where $g_i(P_i)$ is defined as follows

$$g_i(P_i) = \frac{1}{2\sqrt{2\pi} P_i \sigma_i^2} \exp\left(-\frac{P_i}{2\sigma_i^2}\right) \log\left(\frac{1}{Q_i} - 1\right), \quad (34)$$

where Q_i is defined as

$$Q_i \triangleq Q\left(\sqrt{\frac{P_i}{\sigma_i^2}}\right) \quad \forall i = 1, 2, \dots, N. \quad (35)$$

We also have $\sum_{i=1}^N P_i^* = P_0$. The optimal power assignment policy is chosen such that the KKT conditions are satisfied and the channel capacity is maximized.

There is an analogy between this solution and the classical water-filling solution used in assigning transmitted signal powers in the standard parallel Gaussian channel problem. The term ‘‘water-filling’’ arises from the similarity between the curve $g_i(P_i)$ and a bowl into which water (power) is poured, filling the bowl until there is no more power to use. The amount of water/energy in any subchannel is the depth of the water at the corresponding point in the bowl. There are some bowls that will be left empty. For the non-empty bowls a water (power) level is chosen such that there is no power and the mutual information is maximized.

B. Capacity of Parallel Gaussian Channels for Very Low Noise Levels

For a point-to-point scalar Gaussian channel where the noise variance is low, the entropy of the noise is very small compared to the entropy of the input. Thus, the following approximations are valid as noted in [1],

$$h(\mathbf{Y}) \gg h(\mathbf{Y}|\mathbf{X}) \quad \text{and} \quad h(\mathbf{X}) \gg h(\mathbf{X}|\mathbf{Y}),$$

hence

$$h(\mathbf{Y}) \approx h(\mathbf{X}). \quad (36)$$

That is, the capacity can be approximated as [1]

$$\begin{aligned} C &\triangleq \max_{F(\mathbf{x}): |\mathbf{X}| \leq \mathbf{K} \mathbf{a}} I(\mathbf{X}; \mathbf{Y}), \\ &= \max_{F(\mathbf{x}): |\mathbf{X}| \leq \mathbf{K} \mathbf{a}} h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}), \\ &\approx \max_{F(\mathbf{x}): |\mathbf{X}| \leq \mathbf{K} \mathbf{a}} h(\mathbf{X}) - h(\mathbf{Y}|\mathbf{X}), \end{aligned} \quad (37)$$

where $F(\mathbf{x})$ is the input probability distribution function. The intuition behind this approximation is that at very low noise levels the output is not highly affected by the imposed noise. Thus the capacity optimization problem boils down to the maximization of the input entropy instead of maximization of the output entropy (which applies assuming independent additive noise).

Lemma 1: Consider a random variable X with a probability density function $f(x) \in \mathcal{F}_X$ where $|X| \leq A$, $\mathbf{E}[X^2] \leq P$, and \mathcal{F}_X denotes the corresponding class of probability distribution functions such that $P(X > A) = 0$ and $P(X < -A) = 0$. The probability density function that maximizes its entropy is $f(x) = c_1 \exp(-c_2 x^2)$, where c_1 and c_2 are the solutions of

$$c_1 = \frac{1 - 2c_2 P}{2A \exp(-c_2 A^2)}, \quad (37)$$

and

$$\frac{1 - 2c_2 P}{2A \exp(-c_2 A^2)} \left[\sqrt{\frac{\pi}{c_2}} \operatorname{erf}(\sqrt{c_2} A) \right] = 1. \quad (38)$$

Proof: See Appendix B. ■

The resulting differential entropy is given by

$$h(X) = - \int_{-A}^A f(x) \log(f(x)) dx, \quad (39)$$

$$= h(X) = -\log(c_1) + c_2 P. \quad (40)$$

Based on this Lemma, the mutual information between \mathbf{X} and \mathbf{Y} is approximated as

$$\begin{aligned} I(X_1, X_2, \dots, X_N; Y_1, Y_2, \dots, Y_N) \\ \approx \sum_{i=1}^N h(X_i) - \sum_{i=1}^N \frac{1}{2} \log(2\pi e \sigma_i^2). \end{aligned} \quad (41)$$

Similar to the previous subsection, the optimal power assignment for each channel can be performed through water-filling (details of the power assignment policy are given in Appendix C). For the i th channel, let us assume that the power assigned is P_i^* . Then, there are three possible cases, i.e., $P_i^* = 0$ or $P_i^* = A_{x_i}^2$ or $0 < P_i^* < A_{x_i}^2$. One can consider $3^N - 2$ cases separately, and when $0 < P_i < A_{x_i}^2$, the water-filling parameter ν is chosen such that $P_i^* = w_i^{-1}(\nu)$, where the function and $w_i(P_i)$ is the derivative of the objective function which is defined in Appendix C. The optimal power assignment policy is chosen by examining the necessary conditions for optimality and selecting the case with largest information rate.

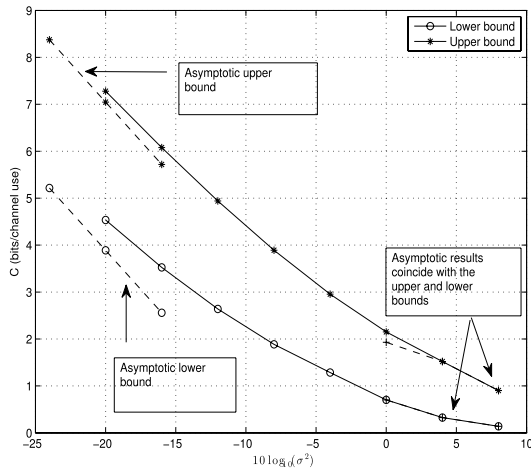


Fig. 2. Upper and lower bounds on the capacity for H_1 , along with the asymptotic capacity at low and high noise variances, with an amplitude constraint of 2 (for both inputs).

We note an important difference between the power assignment policy derived for the low and low noise levels cases, the high noise levels policy depends on the noise variance of the channel besides the amplitude constraints. However, the low noise levels policy does not depend on the noise variance, i.e., the same power assignment policy applies for the entire SNR range.

VI. NUMERICAL EXAMPLES

In this section, we present numerical examples that show the derived upper and lower bounds on the capacity of general MIMO systems for different channels and different input constraints.

A. 2×2 MIMO Systems

In this subsection, we present numerical examples that show the upper and lower bounds on the capacity of 2×2 MIMO systems for different channel coefficient matrices and different amplitude constraints. We consider two arbitrarily picked channel matrices given by

$$H_1 = \begin{bmatrix} 0.177 & 0.28 \\ 1 & 0.31 \end{bmatrix}, \quad H_2 = \begin{bmatrix} 0.997 & 0.295 \\ 1 & 0.232 \end{bmatrix}.$$

We assume that the amplitude constraints imposed on the inputs are identical, and both channels have the same noise variances.

Figs. 2 and 3 show the upper and lower bounds on the capacity and their asymptotic behavior for the two channels considered, we observe that the asymptotic characterizations of the bounds are tight with the bounds at low and high SNR, respectively. The gap between the upper and lower bounds indicate that there is more work to be done for a tighter characterization of the MIMO channel capacity with amplitude constraints.

Fig. 4 shows the upper and lower bounds on the capacity of the second channel for different values of amplitude constraints. Clearly, when the amplitude constraint is increased, the capacity upper and lower bounds are also increased. Also, we observe that the gap between the upper and lower bounds does not depend on the amplitude constraint for very low noise

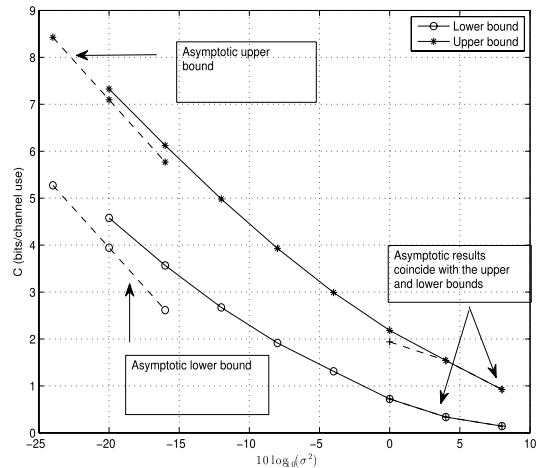


Fig. 3. Upper and lower bounds on the capacity for H_2 , along with the asymptotic capacity at low and high noise variances, with an amplitude constraint of 2 (for both inputs).

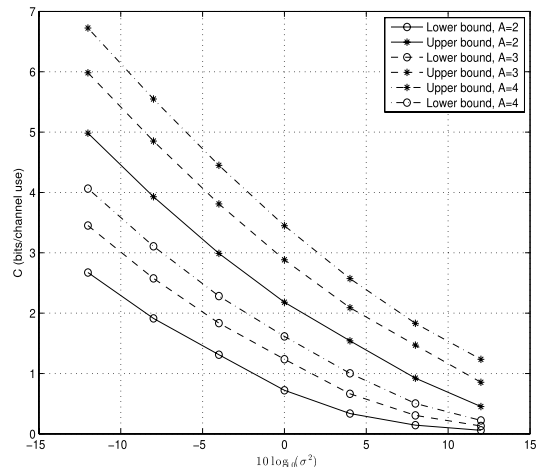


Fig. 4. Illustration of capacity upper and lower bounds of the capacity of H_2 for different amplitude constraints on the inputs.

variance values given that the same amplitude constraints are imposed on both antenna elements. As the value of the noise variance increases, i.e., the SNR decreases, the gap between the upper and lower bound decreases as the number of mass points for the optimal input distribution decreases (eventually it converges to only two mass points). Fig. 5 shows the upper and lower bounds derived from the approach described before along with the alternative achievable lower bound at which we consider a suboptimal input distribution inspired by the optimal input distributions for the bounds, the alternative input distribution shows a better performance, yet the approach of constructing this input distribution is heuristic.

We note that the gap between the upper and lower bounds developed depends highly on the structure of the channel matrix. That is, when the channel matrix is “close” to a diagonal, then the gap between the upper and the lower bounds is small. More, precisely if after the application of the SVD, if the resulting feasible region is closer to a hyper-rectangle, then the restricted region and the relaxation result nearly in the same feasible region, hence the upper and lower bounds are computed using almost the same channels — making the results tight. However, if after the SVD, the region

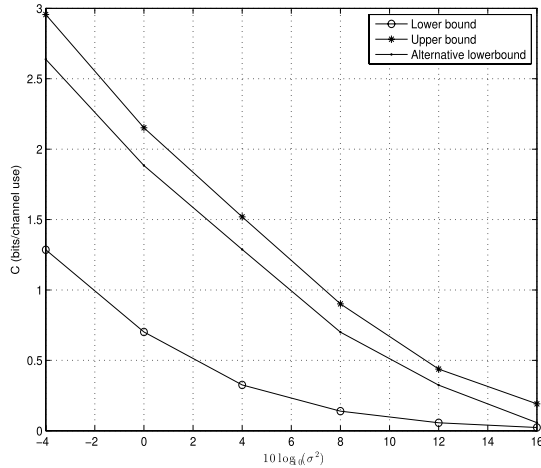


Fig. 5. Upper and lower bounds on the capacity along with the alternative lower bound on the capacity of a MIMO channel.

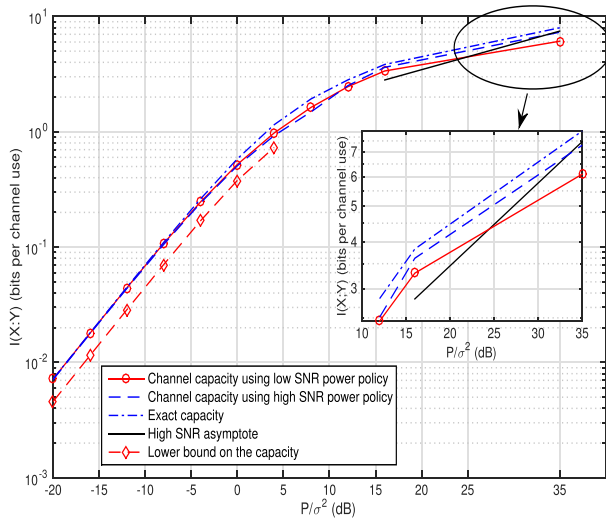


Fig. 6. The channel capacity evaluated using low noise levels policy and high noise levels policy for two parallel Gaussian channel.

is far from a hyper-rectangle, then there is no guarantee of the bounds being very tight. For typical cases, however, as the examples illustrate the results are reasonably tight allowing for a good characterization of the channel capacity.

B. Parallel Gaussian Channels

In this subsection, we present three numerical examples that demonstrate the behavior of the capacity and the bounds for different noise levels regimes. In the first example, we consider two parallel Gaussian channels where the peak power constraint on each channel is $|A_1| \leq 2$, and $|A_2| \leq 4$. The total power constraint is 4. We assume that the variance of the two parallel channels is the same. Fig. 6 shows the capacity of the two parallel Gaussian channel along with the upper and lower bounds derived in the Section V. We also show the evaluation for the channel capacity using the low noise levels policy, high noise levels policy, and uniform power assignment, i.e., for each subchannel the optimal solution is computed using numerical techniques as in [1], given the particular power and amplitude limits. We note that low noise

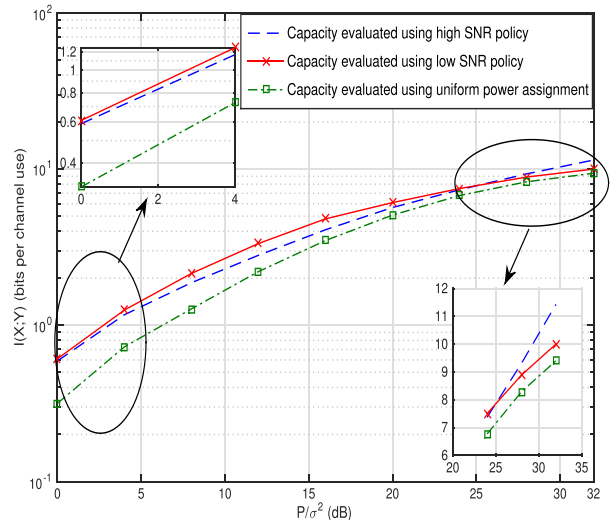


Fig. 7. The channel capacity evaluated using low noise levels policy and high noise levels policy for six parallel Gaussian channel.

(high noise) levels power assignment policy works better at low noise (high noise) levels as expected.

For the second example we consider six parallel Gaussian channels with total power constraint $P_0 = 15$ and with different amplitude constraints for each channel, i.e., $|A_1| \leq 0.1$, $|A_2| \leq 0.1$, $|A_3| \leq 1$, $|A_4| \leq 1$, $|A_5| \leq 10$, $|A_6| \leq 10$. Fig. 7 shows the capacity of the channel evaluated using power assignments resulting from the policy developed for high noise levels, the policy for the low noise variance case, and with uniform power assignments. From the figure we notice that at high noise levels there is a gap between the exact capacity and the bound with the BSC approximation which is due to the fact that the BSC acts as a quantized version of the Gaussian channel. The benefits of the proposed bounds appear when the number of parallel channels is very large as in this case computing the exact channel capacity is not viable as the brute force calculation of the optimal power assignments requires extensive computations. However, by using the proposed bounds we can characterize the channel capacity at different noise levels with a low computational complexity and at the same time obtain tight bounds.

VII. CONCLUSIONS

We study the capacity of multiple-input multiple-output systems with amplitude-limited inputs. Computing the channel capacity of such systems does not seem viable, however, we are able to compute analytical upper and lower bounds. We start by transforming the MIMO channel with coupled inputs to an equivalent channel with decoupled inputs using singular value decomposition. Then, we derive the upper bound by solving the capacity optimization problem over a larger region inscribing the original one. For the lower bound, we solve the problem by constraining the feasible region to a smaller one inscribed within the original. Furthermore, we study the capacity of the parallel Gaussian channels, and provide an analytical characterization at low and high noise levels. For high noise levels, the capacity-achieving distribution consists of only two points for each

channel. For low noise levels, for each channel, the capacity-achieving distribution can be approximated by a continuous distribution whose exact form is determined using techniques from calculus of variations. In both cases, the corresponding optimal power assignment policy is derived by solving the capacity optimization problem by applying the KKT conditions to complete the solution.

APPENDIX A

OPTIMAL POWER ASSIGNMENT FOR THE PARALLEL GAUSSIAN CHANNELS FOR HIGH NOISE LEVELS

The channel capacity is the sum of the capacity of the individual channels, i.e., we can write (33) as

$$C \geq \max_{\substack{P_i, \\ \mathbf{1}^T \mathbf{P} \leq P_0 \\ 0 \leq P_i \leq A_{x_i}^2}} \sum_{i=1}^N J(P_i). \quad (\text{A.1})$$

We first assume that $P_0 < \sum_{i=1}^N A_{x_i}^2$, since for $P_0 \geq \sum_{i=1}^N A_{x_i}^2$, it is straightforward to see that the optimal power assignment policy is $P_i = A_{x_i}^2$, $\forall i = 1, 2, \dots, N$ where Q_i is defined in (35), and $J(P_i)$ is defined as

$$J(P_i) = -Q_i \log Q_i - (1 - Q_i) \log(1 - Q_i), \quad \forall i = 1, 2, \dots, N. \quad (\text{A.2})$$

Clearly, for the optimal power assignment $\sum_{i=1}^N P_i = P_0$. Hence, the Lagrangian of the optimization problem as a minimization of negative of the RHS in (A.1) is

$$L(P_i, \lambda, \nu) = -N + \sum_{i=1}^N J(P_i) + \sum_{i=1}^N \lambda_i (P_i - A_{x_i}^2) - \sum_{i=1}^N \lambda'_i P_i + \nu (\mathbf{1}^T \mathbf{P} - P_0). \quad (\text{A.3})$$

The derivative of the entropy term with respect to the power assigned to the channel is

$$\frac{dJ(P_i)}{dP_i} = Q'_i \log \left(\frac{1}{Q_i} - 1 \right) \quad \forall i = 1, 2, \dots, N, \quad (\text{A.4})$$

where

$$\frac{dQ_i}{dP_i} = \frac{-1}{2\sqrt{2\pi P_i \sigma_i^2}} \exp \left(-\frac{P_i}{2\sigma_i^2} \right) \quad \forall i = 1, 2, \dots, N. \quad (\text{A.5})$$

From the KKT conditions, necessary conditions for optimality are

$$\begin{aligned} \mathbf{1}^T \mathbf{P} &= P_0, \quad \lambda_i \geq 0, \quad P_i \geq 0, \\ \lambda_i (P_i - A_{x_i}^2) &= 0, \quad \lambda'_i \geq 0, \quad \lambda'_i P_i = 0, \\ \nu (\mathbf{1}^T \mathbf{P} - P_0) &= 0, \\ \frac{-1}{2\sqrt{2\pi P_i \sigma_i^2}} \exp \left(-\frac{P_i}{2\sigma_i^2} \right) \log \left(\frac{1}{Q_i} - 1 \right) &+ \lambda_i - \lambda'_i + \nu = 0, \\ &i = 1, \dots, N. \end{aligned} \quad (\text{A.6})$$

For simplicity, we define the function $g(P_i)$ as

$$g_i(P_i) = \frac{1}{2\sqrt{2\pi P_i \sigma_i^2}} \exp \left(-\frac{P_i}{2\sigma_i^2} \right) \log \left(\frac{1}{Q_i} - 1 \right), \quad (\text{A.7})$$

and write the last condition as

$$g_i(P_i) + \lambda_i - \lambda'_i + \nu = 0, \quad i = 1, \dots, N. \quad (\text{A.8})$$

We note that

$$\lim_{P_i \rightarrow 0} g_i(P_i) = \frac{1}{\pi \sigma_i^2}, \quad (\text{A.9})$$

and

$$\lim_{P_i \rightarrow \infty} g_i(P_i) = 0. \quad (\text{A.10})$$

Moreover, it is easy to show that the function $g_i(P_i)$ is monotonically decreasing. Let us assume that the power assigned to the i th channel is P_i^* . Then, there are three possible power assignments, i.e., $P_i^* = 0$ or $P_i^* = A_{x_i}^2$ or $0 < P_i^* < A_{x_i}^2$. Thus, we have $3^N - 2$ cases to consider (we exclude $P_i^* = 0 \quad \forall i = 1, 2, \dots, N$ and $P_i^* = A_{x_i}^2 \quad \forall i = 1, 2, \dots, N$). For the non-zero power assignments if $P_i^* < A_{x_i}^2$, we have $P_i^* = g_i^{-1}(\nu)$.

APPENDIX B

ASYMPTOTIC CAPACITY ACHIEVING DISTRIBUTION OF AMPLITUDE-LIMITED POWER-LIMITED GAUSSIAN CHANNEL IS CONTINUOUS WITH A TRUNCATED GAUSSIAN-LIKE PDF

For a random variable X with a probability density function $f(x)$, $|X| \leq A$ almost surely, and $\mathbf{E}[X^2] \leq P$, we are interested in the probability density function that maximizes the entropy $h(X)$, i.e., we want to

$$\begin{aligned} \text{maximize } h(X) &= - \int_{-A}^A f(x) \log(f(x)) dx, \\ \text{subject to } \mathbf{E}[X^2] &\leq P, \quad \int_{-A}^A f(x) dx = 1. \end{aligned} \quad (\text{B.1})$$

The Lagrangian can be written as,

$$\begin{aligned} u(x, f(x)) &= \int_{-A}^A f(x) \log(f(x)) dx \\ &+ \lambda_1 \int_{-A}^A f(x) (x^2 - P) dx \\ &+ \lambda_2 \int_{-A}^A \left(f(x) - \frac{1}{2A} \right) dx, \end{aligned} \quad (\text{B.2})$$

where λ_1 and λ_2 are the Lagrange multipliers where $\lambda_1 \geq 0$. We can rewrite $u(x, f(x))$ as

$$u(x, f(x)) = \int_{-A}^A g(x, f(x)) dx, \quad (\text{B.3})$$

and

$$\begin{aligned} g(x, f(x)) &= f(x) \log(f(x)) + \lambda_1 x^2 f(x) \\ &- \lambda_1 P f(x) + \lambda_2 f(x) - \frac{\lambda_2}{2A}. \end{aligned} \quad (\text{B.4})$$

Using the Euler-Lagrange equation from calculus of variations [15], i.e.,

$$\frac{\partial g(x, f(x))}{\partial f(x)} = 0, \quad (\text{B.5})$$

results in

$$\log(f(x)) + 1 + \lambda_1 x^2 - \lambda_1 P + \lambda_2 = 0, \quad \text{for } x \in [-A, A]. \quad (\text{B.6})$$

Then, for the optimal probability density function we obtain

$$f(x) = \exp(-\lambda_1 x^2 + \lambda_1 P - \lambda_2 - 1), \quad (\text{B.7})$$

$$= c_1 \exp(-c_2 x^2), \quad (\text{B.8})$$

for some constants c_1, c_2 . We have

$$c_1 \int_{-A}^A \exp(-c_2 x^2) dx = 1, \\ c_1 \int_{-A}^A x^2 \exp(-c_2 x^2) dx = P,$$

hence we can solve for c_1 and c_2 using

$$c_1 = \frac{1 - 2c_2 P}{2A \exp(-c_2 A^2)}, \quad (\text{B.9})$$

and

$$c_1 \left[\sqrt{\frac{\pi}{c_2}} \operatorname{erf}(\sqrt{c_2} A) \right] = 1, \quad (\text{B.10})$$

where the error function $\operatorname{erf}(\cdot)$ is defined as $\operatorname{erf}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$. By combining (B.9) and (B.10), the solution for c_2 can be obtained from

$$\frac{1 - 2c_2 P}{2A \exp(-c_2 A^2)} \left[\sqrt{\frac{\pi}{c_2}} \operatorname{erf}(\sqrt{c_2} A) \right] = 1. \quad (\text{B.11})$$

We can then use in (B.9) to solve for c_1 .

APPENDIX C

OPTIMAL POWER ASSIGNMENT FOR THE PARALLEL GAUSSIAN CHANNELS AT LOW NOISE LEVELS

We have shown in Appendix B that the input distribution $f(x)$ that maximizes the input entropy $h(X)$ is given by (B.8), (B.9), and (B.11). The corresponding differential entropy is given by

$$h(X) = -\log(c_1) + c_2 P. \quad (\text{C.1})$$

Using (B.9), the entropy of X can be written as

$$h(X) = -\log(1 - 2c_2 P) + \log(2A) - c_2 A^2 + c_2 P, \quad (\text{C.2})$$

where c_2 satisfies (B.11). For simplicity, we refer to the single constant c_2 as c_i for the i th channel in the following.

To find the optimal power assignment at low noise levels, we solve the following capacity optimization problem (assuming that $\sum_{i=1}^N A_{x_i}^2 \geq P_0$).

$$\max_{\substack{P_i, \forall i=1,2,\dots,N \\ \mathbf{1}^T \mathbf{P} = P_0 \\ 0 \leq P_i \leq A_{x_i}^2}} \sum_{i=1}^N -\log(1 - 2c_i P_i) + \log(2A_{x_i}) - c_i A_{x_i}^2 + c_i P_i. \quad (\text{C.3})$$

The Lagrangian of the optimization problem is

$$L(P_i, \boldsymbol{\omega}, \nu) = \sum_{i=1}^N \log(1 - 2c_i P_i) - \log(2A_{x_i}) + c_i A_{x_i}^2 \\ - c_i P_i - \sum_{i=1}^N \omega'_i P_i + \sum_{i=1}^N \omega_i (P_i - A_{x_i}^2) \\ + \nu (\mathbf{1}^T \mathbf{P} - P_0). \quad (\text{C.4})$$

The derivative of the Lagrangian is given by

$$\frac{dL(P_i, \boldsymbol{\omega}, \nu)}{dP_i} = -\frac{2c_i}{1 - 2c_i P_i} - \frac{2P_i \frac{dc_i}{dP_i}}{1 - 2c_i P_i} + A_{x_i}^2 \frac{dc_i}{dP_i} \\ - P_i \frac{dc_i}{dP_i} - c_i - \omega'_i + \omega_i + \nu \quad (\text{C.5})$$

where $\frac{dc_i}{dP_i}$ can be found by from (B.11) as

$$\frac{dc_i}{dP_i} = \frac{\pi c_i (\operatorname{erf}(\sqrt{c_i} A_{x_i}))^2}{\beta}, \quad (\text{C.6})$$

where

$$\beta = -\frac{\pi}{c_i} (\operatorname{erf}(\sqrt{c_i} A_{x_i}))^2 + A_{x_i}^3 \sqrt{\pi c_i} \operatorname{erf}(\sqrt{c_i} A_{x_i}) \exp(-c_i A_{x_i}^2) \\ + \frac{A}{2} \sqrt{\frac{\pi}{c_i}} \operatorname{erf}(\sqrt{c_i} A_{x_i}) \exp(-c_i A_{x_i}^2) + A_{x_i}^2 \exp(-2c_i A_{x_i}^2). \quad (\text{C.7})$$

Let us denote $\frac{dc_i}{dP_i}$ by $r(c_i)$. Therefore, we can replace $g(c_i, P_i)$ by $r(c_i)$. Thus,

$$\frac{dL(P_i, \boldsymbol{\omega}, \nu)}{dP_i} = \frac{-2c_i}{1 - 2c_i P_i} - \frac{2P_i r(c_i)}{1 - 2c_i P_i} - A_{x_i}^2 r(c_i) \\ + P_i r(c_i) + c_i - \omega'_i + \omega_i + \nu. \quad (\text{C.8})$$

By applying the KKT conditions, we obtain the following necessary conditions for optimality

$$\omega_i \geq 0, \quad \omega'_i \geq 0, \quad P_i \geq 0, \\ \omega'_i P_i = 0, \quad \omega_i (P_i - A_{x_i}^2) = 0 \quad i = 1, 2, \dots, N, \\ \nu (\mathbf{1}^T \mathbf{P} - P_0) = 0, \quad (\text{C.9})$$

and

$$\frac{dL(P_i, \boldsymbol{\omega}, \nu)}{dP_i} = \frac{-2c_i}{1 - 2c_i P_i} - \frac{2P_i r(c_i)}{1 - 2c_i P_i} - A_{x_i}^2 r(c_i) \\ + P_i r(c_i) + c_i - \omega'_i + \omega_i + \nu = 0. \quad (\text{C.10})$$

For simplicity we define the function $w_i(P_i)$ as

$$w_i(P_i) = \frac{2c_i}{1 - 2u^{-1}(P_i)P_i} + \frac{2P_i r(u^{-1}(P_i))}{1 - 2P_i u^{-1}(P_i)} \\ + A_{x_i}^2 r(u^{-1}(P_i)) - P_i r(u^{-1}(P_i)) + u^{-1}(P_i). \quad (\text{C.11})$$

For the i th channel, let's assume that the power assigned to this channel is P_i^* . Then, there are three possible power assignments, i.e., $P_i^* = 0$ or $P_i^* = A_{x_i}^2$ or $0 < P_i^* < A_{x_i}^2$. For the non-zero power assignments, we have $P_i^* = w_i^{-1}(\nu)$, and $\sum_{i=1}^N P_i^* = P_0$. The optimal power assignment policy is chosen such that the KKT conditions are satisfied and the channel capacity is maximized.¹

¹We note that the KKT conditions in this case are necessary but not sufficient.

REFERENCES

- [1] J. G. Smith, "The information capacity of amplitude- and variance-constrained scalar Gaussian channels," *Inf. Control*, vol. 18, no. 3, pp. 203–219, 1971.
- [2] A. Tchamkerten, "On the discreteness of capacity-achieving distributions," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2773–2778, Nov. 2004.
- [3] B. Mamandipoor, K. Moshksar, and A. K. Khandani, "On the sum-capacity of Gaussian MAC with peak constraint," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2012, pp. 26–30.
- [4] O. Ozel and S. Ulukus, "On the capacity region of the Gaussian MAC with batteryless energy harvesting transmitters," in *Proc. IEEE Global Commun. Conf.*, Dec. 2012, pp. 2385–2390.
- [5] S. Shamai (Shitz) and I. Bar-David, "The capacity of average and peak-power-limited quadrature Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 4, pp. 1060–1071, Jul. 1995.
- [6] T. H. Chan, S. Hranilovic, and F. R. Kschischang, "Capacity-achieving probability measure for conditionally Gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, Jun. 2005.
- [7] K. Knopp and F. Bagemihl, *Theory of Functions*, vol. 1. New York, NY, USA: Dover Publications, 1996.
- [8] J. G. Smith, "On the information capacity of peak and average power constrained Gaussian channels," Ph.D. dissertation, Dept. Elect. Eng., Univ. California, Berkeley, Berkeley, CA, USA, 1969.
- [9] J. Sommerfeld, I. Bjelaković, and H. Boche, "On the boundedness of the support of optimal input measures for Rayleigh fading channels," in *Proc. IEEE Int. Symp. Inf. Theory*, Jul. 2008, pp. 1208–1212.
- [10] A. ElMoslimary and T. M. Duman, "On the capacity of MIMO systems with amplitude-limited inputs," in *Proc. IEEE Int. Black Sea Conf. Commun. Netw.*, May 2014, pp. 97–101.
- [11] B. Rassouli and B. Clerckx, "On the capacity of vector Gaussian channels with bounded inputs," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 4030–4035.
- [12] M. Raginsky, "On the information capacity of Gaussian channels under small peak power constraints," in *Proc. 46th Annu. Allerton Conf. Commun., Control, Comput.*, Sep. 2008, pp. 286–293.
- [13] T. M. Duman and A. Ghrayeb, *Coding for MIMO Communication Systems*. New York, NY, USA: Wiley, 2007.
- [14] J. G. Proakis and M. Salehi, *Digital Communications*, 5th ed. New York, NY, USA: McGraw-Hill, 2008, p. 1150.
- [15] R. Weinstock, *Calculus of Variations: With Applications to Physics and Engineering*. New York, NY, USA: Dover Publications, 1952.



Ahmad ElMoslimary received the B.S. degree from Ain Shams University, Cairo, Egypt, in 2009, the M.S. degree from Nile University, Cairo, in 2011, and the Ph.D. degree from Arizona State University, Tempe, AZ, USA, in 2015, all in electrical engineering. He is currently with Polaris Wireless, Mountain View, CA, USA.

He was a Research Assistant with the Wireless Intelligent Network Center, Nile University, from 2009 to 2011. From 2011 to 2015, he was a Graduate Research Associate with Arizona State University.

His research interest includes signal processing, information theory, channel coding theory, communications theory, wireless channel modeling, MIMO communications, and channel estimation.



Tolga M. Duman (S'95–M'98–SM'03–F'11) received the B.S. degree from Bilkent University, Turkey, in 1993, and the M.S. and Ph.D. degrees from Northeastern University, Boston, in 1995 and 1998, respectively, all in electrical engineering. He was with the Electrical Engineering Department, Arizona State University, first as an Assistant Professor (1998–2004), an Associate Professor (2004–2008), and a Professor (2008–2015). He joined Bilkent University, Turkey, in 2012. He is currently a Professor with

the Electrical and Electronics Engineering Department, Bilkent University, and an Adjunct Professor with the School of ECEE, Arizona State University. His current research interests are in systems, with particular focus on communication and signal processing, including wireless and mobile communications, coding/modulation, coding for wireless communications, data storage systems, and underwater acoustic communications.

Dr. Duman is a recipient of the National Science Foundation CAREER Award and the IEEE Third Millennium Medal. He served as an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2003–2008), the IEEE TRANSACTIONS ON COMMUNICATIONS (2007–2012), and the IEEE ONLINE JOURNAL OF SURVEYS AND TUTORIALS (2002–2007). He has been the Coding and Communication Theory Area Editor of the IEEE TRANSACTIONS ON COMMUNICATIONS (2011–present) and an Editor of the *Physical Communications Journal* (Elsevier; 2010–present).