On the Capacity of Vector Gaussian Channels With Bounded Inputs

Borzoo Rassouli and Bruno Clerckx

Abstract—The capacity of a deterministic multiple-input multiple-output (MIMO) channel under the peak and average power constraints is investigated. For the identity channel matrix, the approach of Shamai et al. is generalized to the higher dimension settings to derive the necessary and sufficient conditions for the optimal input probability density function. This approach prevents the usage of the identity theorem of the holomorphic functions of several complex variables which seems to fail in the multi-dimensional scenarios. It is proved that the support of the capacity-achieving distribution is a finite set of hyper-spheres with mutual independent phases and amplitude in the spherical domain. Subsequently, it is shown that when the average power constraint is relaxed, if the number of antennas is large enough, the capacity has a closed form solution and constant amplitude signaling at the peak power achieves it. Moreover, it will be observed that in a discrete-time memoryless Gaussian channel, the average power constrained capacity, which results from a Gaussian input distribution, can be closely obtained by an input where the support of its magnitude is a discrete finite set. Finally, we investigate some upper and lower bounds for the capacity of the non-identity channel matrix and evaluate their performance as a function of the condition number of the channel.

Index Terms—Vector Gaussian channel, peak power constraint, discrete magnitude, spherical symmetry

I. INTRODUCTION

The capacity of a point-to-point communication system subject to peak and average power constraints was investigated in [1] for the scalar Gaussian channel where it was shown that the capacity-achieving distribution is unique and has a probability mass function with a finite number of mass points. In [2], Shamai and Bar-David gave a full account on the capacity of a quadrature Gaussian channel under the aforementioned constraints and proved that the optimal input distribution has a discrete amplitude and a uniform independent phase. This discreteness in the optimal input distribution was surprisingly shown in [3] to be true even without a peak power constraint for the Rayleigh-fading channel when no channel state information (CSI) is assumed either at the receiver or the transmitter. Following this work, the authors in [4] and [5] investigated the capacity of noncoherent AWGN and Rician-fading channels, respectively. In [6], a point to

Borzoo Rassouli is with the Intelligent Systems and Networks group of Department of Electrical and Electronics, Imperial College London, United Kingdom. email: b.rassouli12@imperial.ac.uk

Bruno Clerckx is with the Communication and Signal Processing group of Department of Electrical and Electronics, Imperial College London and the School of Electrical Engineering, Korea University, Korea. email: b.clerckx@imperial.ac.uk

This paper was presented in part at the IEEE International Conference on Communications (ICC) 2015, London, UK.

point real scalar channel is considered in which sufficient conditions for the additive noise are provided such that the support of the optimal bounded input has a finite number of mass points. These sufficient conditions are also useful in multi-user settings as shown in [7] for the MAC channel under bounded inputs.

The analysis of the MIMO channel under the peak power constraints per antenna is a straightforward problem after changing the vector channel into parallel AWGN channels and applying the results of [1] or [2]. Recently, the vector Gaussian channel under the peak and average power constraints has become more practical by the new scheme proposed in [8]. More specifically, this scheme enables multiple antenna transmission using only one RF chain and the peak power constraint (i.e., a peak constraint on the norm of the input vector rather than on each antenna separately) is the very result of this single RF chain. The capacity of the vector Gaussian channel under the peak and average power constraints has been explored in [9] and [10]. However, according to [11], it seems that the results in the higher dimension settings are not rigorous due to the usage of the identity theorem for holomorphic functions of several complex variables without fulfilling its conditions. As shown by an example in section IV of [11], a holomorphic function of several complex variables can be zero on $\mathbb{R}^n$, but not necessarily zero on $\mathbb{C}^n$. Since $\mathbb{R}^n$ is not an open subset of $\mathbb{C}^n$, the identity theorem cannot be applied. To address this problem, the contributions of this paper are as follows.

- For the identity channel matrix, the approach of [2] is generalized to the vector Gaussian channel in which the complex extension will be done only on a single variable which is the amplitude of the input in the spherical coordinates. The necessary and sufficient conditions for the optimality of the input distribution are derived and it is proved that the magnitude of the capacity-achieving distribution has a probability mass function over a finite number of mass points which determines a finite number of hyper spheres in the spherical coordinates. Further, the magnitude and the phases of the capacity-achieving distribution are mutually independent and the phases are distributed in a way that the points are uniformly distributed on each of the hyper spheres.
- It is shown that if the average power constraint is relaxed, when the ratio of peak power to the number of dimensions remains below a certain threshold (≈ 3.4), the constant amplitude signaling at the peak power achieves the capacity.
- It is also shown that for a fixed SNR, the gap between the Shannon capacity and the constant amplitude signaling

decreases as $O(\frac{1}{n})$ for large values of n, where n denotes the number of dimensions.

- Finally, the case of the non-identity channel matrix is considered where we start from the MISO channel and show that the support of the optimal input does not necessarily have discrete amplitude. Afterwards, several upper bounds and lower bounds are provided for the general n by m MIMO channel capacity. The performance of these bounds are evaluated numerically as a function of the condition number of the channel.

The paper is organized as follows. The system model and some preliminaries are provided in section II, respectively. The main result of the paper is given in section III for the identity channel. The general case of the non-identity channel matrix is briefly investigated in section IV. Numerical results and the conclusion are given in sections V and VI, respectively. Some of the calculations are provided in the appendices at the end of the paper.

II. System Model and preliminaries

In a discrete-time memoryless vector Gaussian channel, the input-output relationship for the identity channel is given by

$$\mathbf{Y}(t) = \mathbf{X}(t) + \mathbf{W}(t), \tag{1}$$

where $\mathbf{X}(t)$, $\mathbf{Y}(t)$ $(\in \mathbb{R}^n)$ denote the input and output of the channel, respectively. t denotes the channel use and $\{\mathbf{W(t)}\}$ is an i.i.d. noise vector process with $\mathbf{W}(t) \sim N(\mathbf{0}, \mathbf{I}_n)$ which is independent of $\mathbf{X}(t)$ for every transmission t. [1]

The capacity of the channel in (1) under the peak and the average power constraints is

$$C(u_p, u_a) = \sup_{F_{\mathbf{X}}(\mathbf{x}): \|\mathbf{X}\|^2 \leq u_p, \ E(\|\mathbf{X}\|^2) \leq u_a} I(\mathbf{X}; \mathbf{Y}), \tag{2}$$

where $F_{\mathbf{X}}(\mathbf{x})$ denotes the input cumulative distribution function (CDF) of the input vector, and u_p, u_a are the upper bounds for the peak and the average power, respectively. Throughout the paper, any operator that involves a random variable reads with the term *almost-surely* (e.g. $\|\mathbf{X}\|^2 \overset{a.s.}{\leq} u_p)^2$.

It is obvious that

$$\sup_{F_{\mathbf{X}}(\mathbf{x}): \|\mathbf{X}\|^2 \leq u_p, \ E(\|\mathbf{X}\|^2) \leq u_a} I(\mathbf{X}; \mathbf{Y}) \leq \sup_{F_{\mathbf{X}}(\mathbf{x}): E(\|\mathbf{X}\|^2) \leq \min(u_p, u_a)} I(\mathbf{X}; \mathbf{Y}).$$

Therefore, a trivial upper bound for the capacity is given by

$$C(u_p, u_a) \leq C_G = \frac{n}{2} \ln\left(1 + \frac{\min(u_p, u_a)}{n}\right), \tag{3}$$

where C_G is achieved by a Gaussian input vector distributed as $N\left(\mathbf{0}, \frac{\min(u_p, u_a)}{n}\mathbf{I}_n\right)$.

We formulate the optimization problem in the spherical domain. The rational behind this change of coordinates is due to the spherical symmetry of the white Gaussian noise and the constraints which, as it will be clear, enables us to perform

[1]It is obvious that the m-dimensional complex AWGN channel can be mapped to the channel in (1) with $n = 2m$.

[2]More precisely, let Ω be the sample space of the probability model over which the random vector $\mathbf{X}$ is defined. $\|\mathbf{X}\|^2 \overset{a.s.}{\leq} u_p$ is equivalent to $\Pr\{\omega \in \Omega | \|\mathbf{X}(\omega)\|^2 \leq u_p\} = 1$.

the optimization problem only on the magnitude of the input. By writing the mutual information in terms of the differential entropies, we have

$$I(\mathbf{X}; \mathbf{Y}) = h(\mathbf{Y}) - h(\mathbf{Y}|\mathbf{X}) = h(\mathbf{Y}) - \frac{n}{2} \ln 2\pi e,$$

where the entropies are in nats. Motivated by the spherical symmetry of the white Gaussian noise and the constraints, $\mathbf{Y}$ and $\mathbf{X}$ can be written in spherical coordinates as

$$\mathbf{Y} = R\mathbf{a}(\mathbf{\Psi}) \quad , \quad \mathbf{X} = P\mathbf{a}(\mathbf{\Theta}),$$

where R and P denote the magnitude of the output and the input, respectively. $\mathbf{\Psi} = [\Psi_1, \Psi_2, \ldots, \Psi_{n-1}]^T$ and $\mathbf{\Theta} = [\Theta_1, \Theta_2, \ldots, \Theta_{n-1}]^T$ are, respectively, the phase vectors of the output and the input, in which $\Psi_i, \Theta_i \in [0, \pi](i \in [1 : n-2])$ and $\Psi_{n-1}, \Theta_{n-1} \in [0, 2\pi)$. $\mathbf{a}(\phi) = [a_1(\phi), \ldots, a_n(\phi)]^T$ is a unit vector in which

$$a_k(\phi) = \begin{cases} \cos\phi_k \prod_{i=1}^{k-1} \sin\phi_i & k \in [1 : n-1] \\ \prod_{i=1}^{k-1} \sin\phi_i & k = n \end{cases}. \tag{4}$$

As it will become clear later, this change of coordinates prevents the usage of the identity theorem for holomorphic functions of several complex variables. The optimization problem in (2) is equivalent to

$$C(u_p, u_a) = \sup_{F_{P,\mathbf{\Theta}}(\rho,\theta): P^2 \leq u_p, \ E(P^2) \leq u_a} h(\mathbf{Y}) - \frac{n}{2} \ln 2\pi e. \tag{5}$$

The differential entropy of the output is given by

$$h(\mathbf{Y}) = -\int_{\mathbb{R}^n} f_{\mathbf{Y}}(\mathbf{y}) \ln f_{\mathbf{Y}}(\mathbf{y}) d\mathbf{y}$$

$$= -\int_0^\infty \underbrace{\int_0^\pi \cdots \int_0^\pi}_{n-2 \text{ times}} \int_0^{2\pi} f_{\mathbf{Y}}(\mathbf{y}(r,\psi)) \ln f_{\mathbf{Y}}(\mathbf{y}(r,\psi)) |\frac{\partial \mathbf{y}}{\partial(r,\psi)}| d\psi dr$$

$$= -\int_0^\infty \underbrace{\int_0^\pi \cdots \int_0^\pi}_{n-2 \text{ times}} \int_0^{2\pi} f_{R,\mathbf{\Psi}}(r,\psi) \ln \frac{f_{R,\mathbf{\Psi}}(r,\psi)}{|\frac{\partial \mathbf{y}}{\partial(r,\psi)}|} d\psi dr$$

$$= h(R, \mathbf{\Psi}) + \int_0^\infty f_R(r) \ln r^{n-1} dr$$

$$+ \sum_{i=1}^{n-2} \int_0^\pi f_{\Psi_i}(\psi_i) \ln \sin^{n-i-1} \psi_i d\psi_i, \tag{6}$$

where $|\frac{\partial \mathbf{y}}{\partial(r,\psi)}| (= r^{n-1} \prod_{i=1}^{n-2} \sin^{n-i-1} \psi_i)$ is the Jacobian of the transform. The conditional pdf of $R, \mathbf{\Psi}$ conditioned on $P, \mathbf{\Theta}$ is given by

$$f_{R,\mathbf{\Psi}|P,\mathbf{\Theta}}(r,\psi|\rho,\theta) = \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{r^2+\rho^2-2r\rho\mathbf{a}^T(\theta)\mathbf{a}(\psi)}{2}} r^{n-1}$$

$$\times \prod_{i=1}^{n-2} \sin^{n-i-1} \psi_i. \tag{7}$$

From (7), the joint pdf of the magnitude and phases of the output is

$$f_{R,\mathbf{\Psi}}(r,\psi) = \int_0^\infty \underbrace{\int_0^\pi \cdots \int_0^\pi}_{n-2 \text{ times}} \int_0^{2\pi} f_{R,\mathbf{\Psi}|P,\mathbf{\Theta}}(r,\psi|\rho,\theta) d^n F_{P,\mathbf{\Theta}}(\rho,\theta), \tag{8}$$

in which $F_{P,\Theta}(\rho,\theta)$ denotes the joint CDF of (P,Θ). By integrating (8) over the phase vector ψ, we have

$$f_R(r) = \int_0^\infty L(r,\rho) f_P(\rho) d\rho, \qquad (9)$$

where [3]

$$L(r,\rho) = \underbrace{\int_0^\pi \cdots \int_0^\pi}_{n-2 \text{ times}} \int_0^{2\pi} f_{R,\Psi|P,\Theta}(r,\psi|\rho,\theta) d\psi_{n-1} \ldots d\psi_1.$$

It is obvious that

$$h(R,\Psi) \le h(R) + \sum_{i=1}^{n-1} h(\Psi_i) \le h(R) + \sum_{i=1}^{n-2} h(\Psi_i) + \ln 2\pi, \tag{10}$$

where the first inequality is tight iff the elements of $\{R, \Psi_1, \ldots, \Psi_{n-1}\}$ are mutually independent, and the second inequality becomes tight iff Ψ_{n-1} is uniformly distributed over $[0, 2\pi)$. From (6) and (10),

$$h(\mathbf{Y}) \le h(R) + \sum_{i=1}^{n-2} h(\Psi_i) + \int_0^\infty f_R(r) \ln r^{n-1} dr$$
$$+ \sum_{i=1}^{n-2} \int_0^\pi f_{\Psi_i}(\psi_i) \ln \sin^{n-i-1}\psi_i d\psi_i + \ln 2\pi. \tag{11}$$

For the sake of readability, the following change of variables is helpful

$$V = \frac{R^n}{n} \;,\; U_i = \int_0^{\Psi_i} \sin^{n-i-1}\delta d\delta \;,\; i \in [1:n-2]. \tag{12}$$

Since $R \ge 0$ and $\Psi_i \in [0,\pi](i \in [1:n-2])$, it is easy to show that the two mappings $R \to V$ and $\Psi_i \to U_i$ (defined in (12)) are invertible. Also, the support set of U_i is $S_{U_i} = [0, \alpha_i]$ where $\alpha_i = \frac{\sqrt{\pi}\Gamma(\frac{n-i}{2})}{\Gamma(\frac{n-i+1}{2})}$ (the Gamma function is defined as $\Gamma(t) = \int_0^\infty x^{t-1}e^{-x}dx$.) From (9), the pdf of V is [4]

$$f_V(v) = f_V(v; F_P) = \int_0^\infty K_n(v,\rho) dF_P(\rho), \tag{13}$$

where the notation $; F_P$ in $f_V(v; F_P)$ is to emphasize that V has been induced by F_P. Not that the integral transform in (13)

---

[3] The reason that $L(r,\rho)$ is not a function of the phase vector θ is due to the spherically symmetric distribution of the white Gaussian noise. In other words, $L(r,\rho)$ is the integral of the Gaussian pdf $N(\mathbf{x}, \mathbf{I})$ over the surface of an n-sphere with radius r which is invariant to the position of $\mathbf{x}$ as long as $\|\mathbf{x}\| = \rho$, i.e.

$$L(r, \|\mathbf{x}\|) = \int_{\|\mathbf{y}\|=r} \frac{e^{-\frac{\|\mathbf{y}-\mathbf{x}\|^2}{2}}}{(\sqrt{2\pi})^n} d\mathbf{y} = \frac{e^{-\frac{r^2+\|\mathbf{x}\|^2}{2}}}{(\sqrt{2\pi})^n} \int_{\|\mathbf{y}\|=r} e^{\mathbf{x}^t\mathbf{y}} d\mathbf{y}$$

which is constant on $\|\mathbf{x}\| = \rho$. (9) implies that in the AWGN channel in (1), $f_R(r)$ is induced only by $f_P(\rho)$ and not $f_\Theta(\theta)$.

[4] The existence of $f_V(v)$ is guaranteed by the Gaussian distribution of the additive noise.

---

is invertible as shown in Appendix D. The kernel $K_n(v,\rho)$ is given by

$$K_n(v,\rho) = \frac{L(\sqrt[n]{nv}, \rho)}{(\sqrt[n]{nv})^{n-1}}$$
$$= \underbrace{\int_0^\pi \cdots \int_0^\pi}_{n-2 \text{ times}} \int_0^{2\pi} \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{(\sqrt[n]{nv})^2+\rho^2-2\sqrt[n]{nv}\rho \mathbf{a}^T(\theta)\mathbf{a}(\psi)}{2}}$$
$$\cdot \prod_{i=1}^{n-2} \sin^{n-i-1}\psi_i d\psi_{n-1}\ldots d\psi_1 \tag{14}$$
$$= e^{-\frac{(\sqrt[n]{nv})^2+\rho^2}{2}} \begin{cases} \frac{I_{\frac{n}{2}-1}(\rho\sqrt[n]{nv})}{(\rho\sqrt[n]{nv})^{\frac{n}{2}-1}} & \rho v \ne 0 \\ \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}-1}} & \rho v = 0 \end{cases}, \forall n \ge 2, \tag{15}$$

where $I_\alpha(.)$ is the modified bessel function of the first kind and order α. The calculations are provided in Appendix A. Note that $K_n(v,\rho)$ is continuous on its domain. The differential entropy of V is

$$h(V) = h(V; F_P)$$
$$= -\int_0^\infty f_V(v; F_P) \ln f_V(v; F_P) dv$$
$$= -\int_0^\infty f_R(r) \ln \frac{f_R(r)}{r^{n-1}} dr. \tag{16}$$

The differential entropy of U_i is given by

$$h(U_i) = -\int_{S_{U_i}} f_{U_i}(u) \ln f_{U_i}(u) du$$
$$= -\int_0^\pi f_{\Psi_i}(\psi_i) \ln \frac{f_{\Psi_i}(\psi_i)}{\sin^{n-i-1}\psi_i} d\psi_i \;,\; i \in [1:n-2]. \tag{17}$$

Rewriting (5), we have

$$C(u_p, u_a) = \sup_{F_{P,\Theta}(\rho,\theta): P^2 \le u_p, E[P^2] \le u_a} h(\mathbf{Y}) - \frac{n}{2}\ln 2\pi e$$

$$\le \sup_{F_{P,\Theta}(\rho,\theta): P^2 \le u_p, E[P^2] \le u_a} h(V; F_P) + \sum_{i=1}^{n-2} h(U_i)$$
$$+ \left(1 - \frac{n}{2}\right)\ln 2\pi - \frac{n}{2} \tag{18}$$

$$\le \sup_{F_P(\rho): P^2 \le u_p, E[P^2] \le u_a} h(V; F_P) + \sum_{i=1}^{n-2} \ln \alpha_i$$
$$+ \left(1 - \frac{n}{2}\right)\ln 2\pi - \frac{n}{2}, \tag{19}$$

where (18) results from (11), (16) and (17). (19) is due to the fact that since S_{U_i} (the support of U_i) is bounded, $h(U_i)$ is maximized when U_i is uniformly distributed. It is easy to verify that if the magnitude and phases of the input are mutually independent with the phases having the distributions as

$$\Theta_{n-1} \sim U[0, 2\pi] \;,\; f_{\Theta_i}(\theta_i) = \alpha_i^{-1}\sin^{n-i-1}\theta_i \;,\; i \in [1:n-2], \tag{20}$$

the magnitude and phases of the output become mutually independent with the phases having the distributions as

$$\Psi_{n-1} \sim U[0, 2\pi] \;,\; f_{\Psi_i}(\psi_i) = \alpha_i^{-1}\sin^{n-i-1}\psi_i \;,\; i \in [1:n-2], \tag{21}$$

where $\alpha_i = \frac{\sqrt{\pi}\Gamma(\frac{n-i}{2})}{\Gamma(\frac{n-i+1}{2})}$. In other words, having the input distribution

$$F_{P,\Theta}(\rho,\theta) = \frac{\theta_{n-1}}{2\pi} F_P(\rho) \prod_{i=1}^{n-2} \int_0^{\theta_i} \alpha_i^{-1} \sin^{n-i-1}\theta d\theta \quad (22)$$

results in

$$F_{R,\Psi}(r,\psi) = \frac{\psi_{n-1}}{2\pi} F_R(r) \prod_{i=1}^{n-2} \int_0^{\psi_i} \alpha_i^{-1} \sin^{n-i-1}\psi d\psi. \quad (23)$$

The above result can be easily checked either by solving for $f_{R,\Psi}(r,\psi)$ in (8) or by the fact that the summation of two independent spherically symmetric random vectors is still spherically symmetric.[5] Also, note that having Ψ_i ($i = 1, \ldots, n-2$) distributed as in (21) implies uniform U_i on $[0, \alpha_i]$ ($i = 1, \ldots, n-2$). It can be observed that the input pdf in (22) makes the inequalities in (18) and (19) tight. Since the constraint is only on the magnitude of the input and $f_V(v)$ is induced only by $f_P(\rho)$, it is concluded that the optimal input distribution must have mutually independent phases and magnitude with the phases being distributed as in (20). Therefore,

$$C(u_p, u_a) = \sup_{F_P(\rho): P^2 \leq u_p, E[P^2] \leq u_a} h(V; F_P)$$
$$+ \sum_{i=1}^{n-2} \ln\alpha_i + (1 - \frac{n}{2})\ln 2\pi - \frac{n}{2}. \quad (24)$$

Before proceeding further, it is interesting to check whether the problem in (24) boils down to the classical results when the peak power constraint is relaxed (i.e., $u_p \to \infty$). From the definition of V,

$$E[V^{\frac{2}{n}}] = \frac{1}{\sqrt[n]{n^2}} E[n + P^2].$$

This can be verified by a change of variable (i.e., $V = \frac{R^n}{n}$) and using the derivative of (112) (in Appendix D) with respect to β. Therefore, when $u_p \to \infty$, the problem in (24) becomes maximization of the differential entropy over all the distributions having a bounded moment of order $\frac{2}{n}$ which is addressed in Appendix B for an arbitrary moment. Substituting m with $\frac{2}{n}$ and A with $\frac{n+u_a}{\sqrt[n]{n^2}}$ in (92), the optimal distribution for V is obatined and from (13), the corresponding $f_{P^*}(\rho)$ has the general Rayleigh distribution as

$$f_{P^*}(\rho) = \frac{n^{\frac{n}{2}} \rho^{n-1} e^{-\frac{n\rho^2}{2u_a}}}{2^{\frac{n-2}{2}} u_a^{\frac{n}{2}} \Gamma(\frac{n}{2})},$$

which is the only solution, since (13) is an invertible transform (see Appendix D). Furthermore, it can be verified that the maximum is

$$C(\infty, u_a) = \frac{n}{2}\ln(1 + \frac{u_a}{n}), \quad (25)$$

which coincides with the classical results for the identity channel matrix [12].

[5]The magnitude and the unit vector of a spherically symmetric random vector are independent and the unit vector is uniformly distributed on the unit ball. It can be verified that this property is equivalent to the vector having the distribution of (23) in spherical coordinates.

Similar to [1] and [2], we define the marginal entropy density of V as

$$\tilde{h}_V(x; F_P) = -\int_0^\infty K_n(v, x) \ln f_V(v; F_P) dv, \quad (26)$$

which satisfies

$$h(V; F_P) = \int_0^\infty \tilde{h}_V(\rho; F_P) dF_P(\rho).$$

(26) is shown to be an invertible transform in Appendix D and this property will become useful later on.

III. MAIN RESULTS

Let ϵ_P denote the set of points of increase[6] of $F_P(\rho)$ in the interval $[0, \sqrt{u_p}]$. The main result of the paper is given in the following theorem.

Theorem. The supremization in (24), which is for the identity channel matrix, has a unique solution and the optimal input achieving the supremum (and therefore the maximum) has the following distribution in the spherical coordinates,

$$F_{P,\Theta}^*(\rho,\theta) = \frac{\theta_{n-1}}{2\pi} F_P^*(\rho) \prod_{i=1}^{n-2} \int_0^{\theta_i} \alpha_i^{-1} \sin^{n-i-1}\theta d\theta, \quad (27)$$

where $F_P^*(\rho)$ has a finite number of points of increase (i.e., ϵ_{P^*} has a finite cardinality). Further, the necessary and sufficient condition for $F_P^*(\rho)$ to be optimal is the existence of a $\lambda(\geq 0)$ for which

$$\tilde{h}_V(\rho; F_P^*) \leq h(V; F_P^*) + \lambda(\rho^2 - u_a), \ \forall\rho \in [0, \sqrt{u_p}] \quad (28)$$
$$\tilde{h}_V(\rho; F_P^*) = h(V; F_P^*) + \lambda(\rho^2 - u_a), \ \forall\rho \in \epsilon_{P^*}. \quad (29)$$

Note that when the average power constraint is relaxed (i.e., $u_a \geq u_p$), $\lambda = 0$.

Proof. The phases of the optimal input distribution have already been shown to be mutually independent and have the distribution in (20) being independent of the magnitude. Therefore, it is sufficient to show the optimal distribution of the input magnitude. This is proved by reductio ad absurdum. In other words, it is shown that having an infinite number of points of increase results in a contradiction. The detailed proof is given in Appendix C. □

Remark 1. When the average power constraint is relaxed (i.e. $u_a \geq u_p$), the following input distribution is asymptotically ($\frac{u_p}{n} \to 0$) optimal

$$F_{P,\Theta}^{**}(\rho,\theta) = \frac{\theta_{n-1}}{2\pi} u(\rho - \sqrt{u_p}) \prod_{i=1}^{n-2} \int_0^{\theta_i} \alpha_i^{-1} \sin^{n-i-1}\theta d\theta, \quad (30)$$

where $u(.)$ is the unit step function. Further, the resulting capacity is given by

$$C(u_p, u_p) \approx \frac{u_p}{2} \quad \text{when} \quad \frac{u_p}{n} \ll 1.$$

Later, in the numerical results section, we observe that the density in (30) remains optimal for the non-vanishing ratio $\frac{u_p}{n}$ when it is below a certain threshold.

[6]A point Z is said to be a point of increase of a distribution if for any open set Γ containing Z, we have $\Pr\{\Gamma\} > 0$.

Proof. Since the density in (30) has spherical symmetry, it is sufficient to show that $F_P^{**}(\rho) = u(\rho - \sqrt{u_p})$ is optimal when $\frac{u_p}{n} \to 0$. From (3), we have

$$\lim_{\frac{u_p}{n} \to 0} C(u_p, u_a) \leq \frac{u_p}{2}. \tag{31}$$

The CDF $F_P^{**}(\rho) = u(\rho - \sqrt{u_p})$ induces the following output pdf

$$f_V(v; F_P^{**}) = K_n(v, \sqrt{u_p}) = e^{-\frac{(\sqrt[n]{nv})^2 + u_p}{2}} \frac{I_{\frac{n}{2}-1}(\sqrt{u_p}\sqrt[n]{nv})}{(\sqrt{u_p}\sqrt[n]{nv})^{\frac{n}{2}-1}}. \tag{32}$$

When $\frac{u_p}{n}$ is small, the entropy of V is given by (35) on top of the next page. In (33), we have approximated the modified bessel function with the first two terms in its power series expansion as follows

$$I_n(x) \approx \frac{x^n}{\Gamma(n+1)2^n}(1 + \frac{x^2}{4(n+1)}) \quad , \quad \frac{x}{n} \to 0.$$

In (34), we use the approximation $\ln(1+x) \approx x$ ($x \ll 1$) and in (35), the higher order term is neglected. Given the input distribution F_P^{**}, the achievable rate with small ratio $\frac{u_p}{n}$ is given by (see (24))

$$\lim_{\frac{u_p}{n} \to 0} h(V; F_P^{**}) + \sum_{i=1}^{n-2} \ln \alpha_i + (1 - \frac{n}{2})\ln 2\pi - \frac{n}{2} = \frac{u_p}{2}, \tag{36}$$

where we have used the fact that

$$\sum_{i=1}^{n-2} \ln \alpha_i = -\ln \Gamma(\frac{n}{2}) + \frac{n-2}{2}\ln \pi.$$

From (36) and (31), it is concluded that the pdf in (30) is asymptotically optimal for $\frac{u_p}{n} \ll 1$ when $u_p \leq u_a$. Note that the distribution in (30) is not the only asymptotically optimal distribution. There are many possible alternatives, one of which, for example, is the binary PAM in each dimension with the points $-\sqrt{\frac{u_p}{n}}$ and $\sqrt{\frac{u_p}{n}}$ which can be verified to have an achievable rate of $\frac{u_p}{2}$ when $\frac{u_p}{n} \ll 1$. Specifically, in the low peak power regime ($u_p \ll 1$), a sufficient condition for the input distribution to be asymptotically optimal is as follows. First, it has a constant magnitude at $\sqrt{u_p}$. Second, its Θ_1 is independent of $(P, \Theta_2, \ldots, \Theta_{n-1})$ and has a zero first Fourier coefficient i.e.,

$$\int_0^\pi e^{j\theta} f_{\Theta_1}(\theta) d\theta = 0. \tag{37}$$

The claim is justified by noting that fulfilling the second condition results in the spherical symmetric output distribution of (23) as follows. Using the approximation $e^x \approx 1+x$ ($x \ll 1$), at small values of u_p, (7) can be approximated as

$$f_{R,\Psi|P,\Theta}(r, \psi|\rho, \theta) \approx \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{r^2+\rho^2}{2}}(1 + r\rho\mathbf{a}^T(\theta)\mathbf{a}(\psi))$$

$$\times r^{n-1} \prod_{i=1}^{n-2} \sin^{n-i-1}\psi_i. \tag{38}$$

If Θ_1 is independent of $(\Theta_2, \ldots, \Theta_{n-1}, P)$, substituting (38) in (8) results in

$$f_{R,\Psi}(r, \psi) \approx \int_0^\infty \underbrace{\int_0^\pi \cdots \int_0^\pi}_{n-3 \text{ times}} \int_0^{2\pi} \int_0^\pi \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{r^2+\rho^2}{2}} r^{n-1}$$

$$\times (1 + r\rho\mathbf{a}^T(\theta)\mathbf{a}(\psi)) \prod_{i=1}^{n-2} \sin^{n-i-1}\psi_i$$

$$dF_{\Theta_1}(\theta_1) d^{n-1} F_{P,\Theta_2^{n-1}}(\rho, \theta_2^{n-1}) \tag{39}$$

where $\theta_2^{n-1} = (\theta_2, \theta_3, \ldots, \theta_{n-1})$. If Θ_1 has a zero first Fourier coefficient, due to the structure of $\mathbf{a}(\theta)$ (see (4)), we have

$$\int_0^\pi \mathbf{a}^T(\theta)\mathbf{a}(\psi) dF_{\Theta_1}(\theta_1) = 0.$$

Therefore, (39) simplifies as

$$f_{R,\Psi}(r, \psi) \approx \int_0^\infty \frac{1}{(\sqrt{2\pi})^n} e^{-\frac{r^2+\rho^2}{2}} r^{n-1} \prod_{i=1}^{n-2} \sin^{n-i-1}\psi_i dF_P(\rho)$$

which implies that when $u_p \to 0$, having Θ_1 independent of all other spherical variables with a zero first Fourier coefficient results in the output distribution in (23) which makes the inequalities (18) and (19) tight. Finally, fulfilling the first condition (i.e., having a constant magnitude at $\sqrt{u_p}$) validates the previous reasoning starting from (32).

The asymptotic optimality of the constant-magnitude signaling in (30) can alternatively be proved by inspecting the behavior of the marginal entropy density $\tilde{h}_V(\rho; F_P)$ when $\frac{u_p}{n}$ is sufficiently small. From (13)

$$f_V(v; F_P) \to \frac{e^{-\frac{(\sqrt[n]{nv})^2}{2}}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}-1}} \underbrace{\int_0^\infty e^{-\frac{\rho^2}{2}} dF_P(\rho)}_{\text{constant} = C} \quad \text{when} \quad \frac{u_p}{n} \to 0.$$

Therefore,

$$\tilde{h}_V(\rho; F_P) = -\int_0^\infty e^{-\frac{(\sqrt[n]{nv})^2 + \rho^2}{2}} \frac{I_{\frac{n}{2}-1}(\rho\sqrt[n]{nv})}{(\rho\sqrt[n]{nv})^{\frac{n}{2}-1}} \ln f_V(v; F_P) dv$$

$$\to \int_0^\infty e^{-\frac{(\sqrt[n]{nv})^2 + \rho^2}{2}} \frac{I_{\frac{n}{2}-1}(\rho\sqrt[n]{nv})}{(\rho\sqrt[n]{nv})^{\frac{n}{2}-1}} \left[\frac{(\sqrt[n]{nv})^2}{2} \right.$$

$$\left. + \ln\left(\frac{\Gamma(\frac{n}{2})2^{\frac{n}{2}-1}}{C}\right) \right] dv$$

$$= \frac{\rho^2 + n}{2} + \ln\left(\frac{\Gamma(\frac{n}{2})2^{\frac{n}{2}-1}}{C}\right) \tag{40}$$

It is obvious that (40) is a (strictly) convex (strictly) increasing function. Hence, the necessary and sufficient conditions in (28) and (29) are satisfied if and only if the input has only one point of increase at $\sqrt{u_p}$ which proves the asymptotic optimality of (30) for $\frac{u_p}{n} \ll 1$ and $u_a \geq u_p$. $\square$

Remark 2. For a fixed SNR, the gap between Shannon capacity and the constant amplitude signaling decreases as $O(\frac{1}{n})$ for large values of n.

Proof. By writing the first two terms of the Taylor series expansion of the logarithm (i.e., $\ln(1+x) \approx x - \frac{x^2}{2}, x \ll 1$), we have

$$\text{when } n \to \infty \quad , \quad \frac{n}{2}\ln(1 + \frac{u_p}{n}) \approx \frac{u_p}{2} - \frac{u_p^2}{4n}.$$

$$\lim_{\frac{u_p}{n}\to 0} h(V; F_P^{**}) = \lim_{\frac{u_p}{n}\to 0} - \int_0^\infty f_V(v; F_P^{**}) \ln f_V(v; F_P^{**}) dv$$

$$= \lim_{\frac{u_p}{n}\to 0} \int_0^\infty e^{-\frac{(\sqrt[n]{nv})^2 + u_p}{2}} \frac{I_{\frac{n}{2}-1}(\sqrt{u_p}\sqrt[n]{nv})}{(\sqrt{u_p}\sqrt[n]{nv})^{\frac{n}{2}-1}} \left[\frac{(\sqrt[n]{nv})^2 + u_p}{2} - \ln\left(\frac{I_{\frac{n}{2}-1}(\sqrt{u_p}\sqrt[n]{nv})}{(\sqrt{u_p}\sqrt[n]{nv})^{\frac{n}{2}-1}} \right) \right] dv$$

$$= \frac{n}{2} + \ln\left(\Gamma(\frac{n}{2})2^{\frac{n}{2}-1} \right)$$

$$+ \lim_{\frac{u_p}{n}\to 0} \left\{ u_p - \int_0^\infty e^{-\frac{(\sqrt[n]{nv})^2 + u_p}{2}} \frac{I_{\frac{n}{2}-1}(\sqrt{u_p}\sqrt[n]{nv})}{(\sqrt{u_p}\sqrt[n]{nv})^{\frac{n}{2}-1}} \ln\left(1 + \frac{u_p(\sqrt[n]{nv})^2}{2n} \right) dv \right\} \tag{33}$$

$$= \frac{n}{2} + \ln\left(\Gamma(\frac{n}{2})2^{\frac{n}{2}-1} \right) + \lim_{\frac{u_p}{n}\to 0} \left\{ u_p - \frac{u_p}{n}(\frac{n+u_p}{2}) \right\} \tag{34}$$

$$= \frac{n}{2} + \ln\left(\Gamma(\frac{n}{2})2^{\frac{n}{2}-1} \right) + \frac{u_p}{2}. \tag{35}$$

From (34), the achievable rate obtained by the constant envelope signaling is

$$\text{when } n \to \infty \quad , \quad I(\mathbf{X};\mathbf{Y}) \approx \frac{u_p}{2} - \frac{u_p^2}{2n}.$$

This shows that the gap between achievable rate and the Shannon capacity decreases as $\frac{u_p^2}{4n}(= O(\frac{1}{n}))$, when n goes to infinity. $\square$

While remark 2 shows an asymptotic behavior of the gap, the following remark provides an analytical lower bound for any values of n.

Remark 3. The following lower bound holds for the capacity of constant amplitude signaling.

$$\sup_{F_\mathbf{X}(\mathbf{x}):\|\mathbf{X}\|^2 = u_p} I(\mathbf{X};\mathbf{Y}) \geq \frac{n-1}{2} \log\left(1 + \frac{2^{\frac{2}{n-1}-1}u_p}{e\left[(n-1)\Gamma(\frac{n-1}{2})\right]^{\frac{2}{n-1}}} \right) \tag{41}$$

Proof. Let $\mathbf{X}'$ and $\mathbf{X}'$ be defined as

$$\mathbf{X}' = [X_1, X_2, \ldots, X_{N-1}, 0]^T \quad , \quad \mathbf{Y}' = [Y_1, Y_2, \ldots, Y_{N-1}, 0]^T. \tag{42}$$

Due to the Markov chain $\mathbf{X}' \leftrightarrow \mathbf{X} \leftrightarrow \mathbf{Y} \leftrightarrow \mathbf{Y}'$ and the fact that $\|\mathbf{X}\|^2 = u_p$ implies $\|\mathbf{X}'\|^2 \leq u_p$, we can write

$$\sup_{\substack{F_\mathbf{X}(\mathbf{x}) \\ \|\mathbf{X}\|^2 = u_p}} I(\mathbf{X};\mathbf{Y}) \geq \sup_{F_{\mathbf{X}'}(\mathbf{x}'):\|\mathbf{X}'\|^2 \leq u_p} I(\mathbf{X}';\mathbf{Y}')$$

$$= \sup_{F_{\mathbf{X}'}(\mathbf{x}'):\|\mathbf{X}'\|^2 \leq u_p} h(\mathbf{Y}'; F_{\mathbf{X}'})$$

$$- \frac{n-1}{2} \log 2\pi e$$

$$\geq \sup_{F_{\mathbf{X}'}(\mathbf{x}'):\|\mathbf{X}'\|^2 \leq u_p} \frac{n-1}{2} \log\left(2^{\frac{2}{n-1}h(\mathbf{X}')} + 2\pi e \right)$$

$$- \frac{n-1}{2} \log 2\pi e \tag{43}$$

$$= \frac{n-1}{2} \log\left(1 + \frac{2^{\frac{2}{n-1}-1}u_p}{e\left[(n-1)\Gamma(\frac{n-1}{2})\right]^{\frac{2}{n-1}}} \right) \tag{44}$$

where in (43), the $(n-1)$-dimensional EPI has been used[7] and (44) is due to the fact that for the $(n-1)$-dimensional vector $\mathbf{X}'$, we can write

$$\sup_{F_{\mathbf{X}'}(\mathbf{x}'):\|\mathbf{X}'\|^2 \leq u_p} h(\mathbf{X}') = \log\left(\frac{2(\pi u_p)^{\frac{n-1}{2}}}{(n-1)\Gamma(\frac{n-1}{2})} \right) \quad , \quad n \geq 2 \tag{45}$$

whose proof follows the same steps from (72) to (77) with $\lambda = 0$ and $a = \frac{n}{(\sqrt{u_p})^n}$. $\square$

The asymptotic decrease of the gap in remark 2 can be alternatively proved by using the lower bound in (44) which is provided in Appendix E.

Remark 4. When $u_a < u_p$, the following input distribution is asymptotically ($u_a \to 0$) optimal

$$F_{P,\Theta}^{**}(\rho,\theta) = \left[(1 - \frac{u_a}{u_p})u(\rho) + \frac{u_a}{u_p}u(\rho - \sqrt{u_p}) \right]$$

$$\times \frac{\theta_{n-1}}{2\pi} \prod_{i=1}^{n-2} \int_0^{\theta_i} \alpha_i^{-1} \sin^{n-i-1}\theta d\theta \tag{46}$$

and the resulting capacity is given by

$$C(u_p, u_p) \approx \frac{u_a}{2} \quad \text{when } u_a \ll 1.$$

Proof. The proof is given in Appendix F. $\square$

Remarks 1 and 4 are essential for the initial stage of the simulation results when either u_a or u_p are assumed to be very small at first and afterwards they are increased gradually by a step size.

Remark 5. The fact that the magnitude of the optimal input distribution has a finite number of mass points remains unchanged if the average constraint in (2) is generalized as

$$E(g(P)) \leq u_a \tag{47}$$

in which $g(z)$ is holomorphic on an open subset $\mathbb{D}(\subseteq \mathbb{C})$ which includes the non-negative real line (i.e., $\mathbb{R}_{\geq 0} \subset \mathbb{D}$).

---

[7]Note that the reduction of dimensions from n to $n-1$ in (42) is necessary. The reason is that the usage of the n-dimensional EPI is not permissible for the constant amplitude vector, since an n-dimensional vector with a fixed norm has at most $(n-1)$ degrees of freedom (or equivalently at most $(n-1)$-dimensional support).

Proof. The proof is given in Appendix G. □

Remark 6. The peak power constraint in (2) can be generalized to

$$\|P\|^2 \overset{a.s.}{\in} \mathbb{D}_{u_p} \subseteq [0, u_p].$$

Proof. Since all the conditions (compactness, continuity, etc.) remain unchanged, the support of the optimal input distribution will be some concentric shells having the mass points of the magnitude in $\mathbb{D}_{u_p}$. □

IV. THE MIMO CASE WITH DETERMINISTIC CHANNEL

First, we consider the multiple-input single-output (MISO) channel in which (1) changes to

$$Y(t) = \mathbf{h}^T \mathbf{X}(t) + W(t) \tag{48}$$

where $\mathbf{h}(\in \mathbb{R}^{n \times 1})$ is the deterministic channel vector and $W \sim N(0, 1)$. Let $X_{new} = \mathbf{h}^T \mathbf{X}$. The capacity of this channel under the peak and average power constraints is given by

$$
\begin{aligned}
C(u_p, u_a) &= \sup_{\substack{F_\mathbf{X}(\mathbf{x}): \|\mathbf{X}\|^2 \le u_p, \\ E[\|\mathbf{X}\|^2] \le u_a}} I(\mathbf{X}; Y) \\
&= \sup_{\substack{F_\mathbf{X}(\mathbf{x}): \|\mathbf{X}\|^2 \le u_p, \\ E[\|\mathbf{X}\|^2] \le u_a}} I(\mathbf{X}, X_{new}; Y) \tag{49} \\
&= \sup_{\substack{F_\mathbf{X}(\mathbf{x}): \|\mathbf{X}\|^2 \le u_p, \\ E[\|\mathbf{X}\|^2] \le u_a}} I(X_{new}; Y) + \underbrace{I(\mathbf{X}; Y | X_{new})}_{=0} \tag{50} \\
&= \sup_{\substack{F_\mathbf{X}(\mathbf{x}): \|\mathbf{X}\|^2 \le u_p, \\ E[\|\mathbf{X}\|^2] \le u_a}} I(X_{new}; Y) \\
&\le \sup_{\substack{F_{X_{new}}(x): |X_{new}| \le \sqrt{u_p}\|\mathbf{h}\|, \\ E[|X_{new}|^2] \le u_a \|\mathbf{h}\|^2}} I(X_{new}; Y) \tag{51}
\end{aligned}
$$

where (49) is due to the fact that X_{new} is a function of $\mathbf{X}$ and (50) is a result of the following Markov chain $\mathbf{X} \longrightarrow X_{new} \longrightarrow Y$. (51) is due to the fact that any input cdf having the support $\|\mathbf{X}\|^2 \le u_p$ and satisfying $E[\|\mathbf{X}\|^2] \le u_a$ induces a cdf for X_{new} with the support in $[-\sqrt{u_p}\|\mathbf{h}\|, \sqrt{u_p}\|\mathbf{h}\|]$ and satisfying $E[|X_{new}|^2] \le u_a \|\mathbf{h}\|^2$. This could be readily verified by the following convex optimization problem

$$\max_{\mathbf{x}} \mathbf{h}^T \mathbf{x}$$
$$\text{S.t. } \|\mathbf{x}\|^2 \le u_p$$

where the maximum is $\sqrt{u_p}\|\mathbf{h}\|$ and it is achieved when $\mathbf{x}$ is matched to the channel (i.e., $\mathbf{x} = \sqrt{u_p}\frac{\mathbf{h}}{\|\mathbf{h}\|}$). Further, from Cauchy-Shwartz inequality, we have

$$E[|X_{new}|^2] = E[|\mathbf{h}^t \mathbf{X}|^2] \le E[\|\mathbf{h}\|^2]E[\|\mathbf{X}\|^2] \le u_a \|\mathbf{h}\|^2 \tag{52}$$

where the inequalities change to equality iff $\mathbf{X}$ is in the direction of $\mathbf{h}$ and $E[\|\mathbf{X}\|^2] = u_a$.

The supremization in (51) is the same problem of finding the capacity of a scalar Gaussian channel which has been addressed in [1] where it was shown that the optimal input distribution is a pmf over a finite set of points in the interval defined by the peak power constraint and also it satisfies the average power inequality with equality. It is obvious that

having $\mathbf{X}$ located on the hyperplane $\mathbf{h}^T \mathbf{X} = e_i$ (confined in the ball $\|\mathbf{X}\|^2 \le u_p$) with probability p_i results in having X_{new} equal to the mass point $e_i \in [-\sqrt{u_p}\|\mathbf{h}\|, \sqrt{u_p}\|\mathbf{h}\|]$ with probability p_i. If the average power constraint is relaxed (i.e., $u_a \ge u_p$), the support of the capacity-achieving distribution of the MISO channel with the input bounded in a ball becomes a finite number of hyper planes confined in that ball (all of these hyperplane have the normal vector $\mathbf{h}$). Note that the discrete amplitude property is no longer a necessity for the optimal input distribution in contrast to the MIMO with identity channel. In other words, the necessary and sufficient condition for the optimality is that $\mathbf{X}$ is located on each of these hyperplanes with the corresponding probabilities. There is a common characteristic of the optimal input distribution in both the MIMO (with identity channel) and MISO scenarios which is the fact that the support of the optimal input distribution does not include any open set in $\mathbb{R}^n$. Finally, if the average power constraint is active (i.e., $u_a < u_p$), the support of the optimal input becomes a finite number of mass points in the direction of $\mathbf{h}$ (from (52) and the fact that $E[|X_{new}|^2] = u_a\|\mathbf{h}\|^2$) and confined in the ball $\|\mathbf{X}\|^2 \le u_p$.

For the general deterministic MIMO channel, we have

$$\mathbf{Y}'(t) = \mathbf{H}\mathbf{X}'(t) + \mathbf{W}'(t) \tag{53}$$

where $\mathbf{H} \in \mathbb{R}^{n_r \times n_t}$ denotes the deterministic channel. By an SVD (i.e., $\mathbf{H} = \mathbf{D}\mathbf{\Lambda}\mathbf{N}^T$ where $\mathbf{D} \in \mathbb{R}^{n_r \times n_r}$, $\mathbf{\Lambda} \in \mathbb{R}^{n_r \times n_t}$, $\mathbf{N} \in \mathbb{R}^{n_t \times n_t}$), we get

$$\tilde{\mathbf{Y}}'(t) = \mathbf{D}^T \mathbf{Y}'(t) = \mathbf{\Lambda} \underbrace{\mathbf{N}^T \mathbf{X}'(t)}_{\tilde{\mathbf{X}}'(t)} + \underbrace{\mathbf{D}^T \mathbf{W}'(t)}_{\tilde{\mathbf{W}}'(t)}. \tag{54}$$

Let $n = \text{rank}(\mathbf{H})$ and $\mathbf{Q}(t)$ be the first n elements of the vector $\tilde{\mathbf{Q}}'(t)$ (for $\mathbf{Q} = \mathbf{Y}, \mathbf{X}$ and $\mathbf{W}$). It is obvious that (54) is equivalent to the following

$$\mathbf{Y}(t) = \mathbf{X}(t) + \mathbf{N}(t) \tag{55}$$

with the noise distributed as $N(\mathbf{0}, \mathbf{\Sigma})$ where $\mathbf{\Sigma} = \text{diag}\{\lambda_1^{-2}, \lambda_2^{-2}, \ldots, \lambda_n^{-2}\}$ and λ_i ($i \in [1:n]$) is the ith singular value of $\mathbf{H}$. Therefore, the capacity of the deterministic channel in (53) is the same as the capacity of the additive non-white Gaussian noise channel in (55). It is assumed that the condition number of $\mathbf{H}$ is not unity, since in that case, it becomes equivalent to the scenario with identity channel matrix discussed in section II. From now on, we consider $n = 2$.

Two possible changes of coordinates are as follows. Motivated by the elliptical symmetry of the noise, $\mathbf{X}$ and $\mathbf{Y}$ could be written in the following elliptical coordinates

$$\mathbf{Y} = R\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{a}(\mathbf{\Psi}) \quad , \quad \mathbf{X} = P\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{a}(\mathbf{\Theta}) \tag{56}$$

and using a similar approach as in section II, the optimization problem becomes

$$C(u_p, u_a) = \sup_{\substack{F_{P,\Theta(\rho,\theta)}: P^2 \mathbf{a}^T(\theta)\mathbf{\Sigma}\mathbf{a}(\theta) \le u_p, \\ E[P^2 \mathbf{a}^T(\theta)\mathbf{\Sigma}\mathbf{a}(\theta)] \le u_a}} h(V, \Psi; F_{P,\Theta}) - \ln 2\pi e, \tag{57}$$

where $V = \frac{R^2}{2}$. The joint entropy of the output variables is given by

$$h(V, \Psi; F_{P,\Theta}) = \int_0^\infty \int_0^{2\pi} \tilde{h}_{V,\Psi}(\rho, \theta; F_{P,\Theta}) d^2 F_{P,\Theta}(\rho, \theta), \quad (58)$$

where the joint marginal entropy density writes as

$$\tilde{h}_{V,\Psi}(\rho, \theta; F_{P,\Theta}) = -\int_0^\infty \int_0^{2\pi} K(v, \psi, \rho, \theta) \times \ln f_{V,\Psi}(v, \psi; F_{P,\Theta}) d\psi dv, \quad (59)$$

in which

$$f_{V,\Psi}(v, \psi; F_{P,\Theta}) = \int_0^\infty \int_0^{2\pi} K(v, \psi, \rho, \theta) d^2 F_{P,\Theta}(\rho, \theta), \quad (60)$$

and

$$K(v, \psi, \rho, \theta) = \frac{1}{2\pi} e^{-v - \frac{\rho^2}{2} + \rho\sqrt{2v}\cos(\psi - \theta)}. \quad (61)$$

Alternatively, due to the spherical symmetry of the constraint, the input and the output could be written in the spherical coordinates in which

$$C(u_p, u_a) = \sup_{F_{P,\Theta(\rho,\theta)}: P^2 \leq u_p, E[P^2] \leq u_a} h(V, \Psi; F_{P,\Theta}) - \ln(2\pi e\sqrt{|\Sigma|}). \quad (62)$$

(58) to (60) remain unchanged, while the kernel is given by

$$K(v, \psi, \rho, \theta) = \frac{e^{-\frac{1}{2}\left[\sqrt{2v}\mathbf{a}(\psi) - \rho\mathbf{a}(\theta)\right]^T \Sigma^{-1}\left[\sqrt{2v}\mathbf{a}(\psi) - \rho\mathbf{a}(\theta)\right]}}{2\pi\sqrt{|\Sigma|}}. \quad (63)$$

Using neither of the above coordinates makes the separation of the magnitude and the phases possible as done in (10). This is due to the different symmetries of the noise (elliptical) and the peak power constraint (spherical). Since the conditions of compactness, convexity and continuity remain unchanged, we can only proceed up to the point of writing the necessary and sufficient conditions for the joint cdf $F_{P,\Theta}(\rho, \theta)$ to be the optimal solution. By using the spherical coordinates, the necessary and sufficient conditions for the optimal input distribution is given by

$$\tilde{h}_{V,\Psi}(\rho, \theta; F_{P,\Theta}^*) \leq h(V, \Psi; F_{P,\Theta}^*) + \lambda(\rho^2 - u_a)$$
$$, \forall \rho \in [0, \sqrt{u_p}], \forall \theta \in [0, 2\pi), \quad (64)$$

$$\tilde{h}_{V,\Psi}(\rho, \theta; F_{P,\Theta}^*) = h(V, \Psi; F_{P,\Theta}^*) + \lambda(\rho^2 - u_a)$$
$$, \forall (\rho, \theta) \in \epsilon_{P,\Theta}^*, \quad (65)$$

where $\epsilon_{P,\Theta}^*$ is the set of points of increase in $F_{P,\Theta}^*$.

To make the problem caused by the different symmetries of the noise and the constraint more clear, let's assume $\lambda_1 = \lambda_2$ (i.e., as in the previous section with identity channel.) In this case, we rewrite the optimization problem as

$$C(u_p, u_a) = \sup_{F_{P,\Theta(\rho,\theta)}: P^2 \leq u_p, \ E[P^2] \leq u_a} h(V, \Psi; F_{P,\Theta}) - \ln(2\pi e\lambda_1^2). \quad (66)$$

It is already known that the optimal solution must have independent phase and magnitude with the former being uniformly distributed on $[0, 2\pi)$. This can alternatively be inferred from the above necessary and sufficient conditions as follows. Let $f_{P,\Theta}^*(\rho, \theta) = f_P^*(\rho) f_{\Theta|P}^*(\theta|\rho)$ denote the (unique) solution of

(66) with $\epsilon_{P,\Theta}^*$ as its points of increase. Let the pdf $l_{P,\Theta}^\epsilon$ be defined as

$$l_{P,\Theta}^\epsilon(\rho, \theta) = f_P^*(\rho) f_{\Theta|P}^*(\theta - \epsilon|\rho),$$

where ϵ is a constant arbitrarily chosen from $(0, 2\pi)$. Let $L_{P,\Theta}^\epsilon$ be the corresponding CDF. It can be easily verified that

$$f_{V,\Psi}(v, \psi; L_{P,\Theta}^\epsilon) = f_{V,\Psi}(v, \psi - \epsilon; F_{P,\Theta}^*) \quad (67)$$

and therefore,

$$h(V, \Psi; L_{P,\Theta}^\epsilon) = h(V, \Psi; F_{P,\Theta}^*).$$

Since $L_{P,\Theta}^\epsilon$ satisfies the constraints and the optimal solution is unique, it is concluded that

$$f_{P,\Theta}^*(\rho, \theta) = l_{P,\Theta}^\epsilon(\rho, \theta)$$

which in turn results in

$$f_{\Theta|P}(\theta|\rho) = f_{\Theta|P}(\theta - \epsilon|\rho).$$

Since $\epsilon \in (0, 2\pi)$ was chosen arbitrarily, we conclude that $f_{\Theta|P}(\theta|\rho) = f_\Theta(\theta) = \frac{1}{2\pi}$. The problem in the case when $\lambda_1 \neq \lambda_2$ is that if the elliptical domain is used, (67) remains true, but $L_{P,\Theta}^\epsilon$ does not satisfy the spherical constraints any more, and if the spherical domain is considered, $L_{P,\Theta}^\epsilon$ satisfies the constraints, but (67) does not hold any longer. Therefore, in what follows, we provide some upper bounds and lower bounds for the capacity of the deterministic channel.

1) Bounds based on the cubic constraints: For brevity, let

$$\mathbb{F}(\mathbf{a}, \mathbf{b}) = \{F_\mathbf{X}(\mathbf{x})|F_{X_i}(x_i) = 0 \text{ for } x_i < 0,$$
$$F_{X_i}(x_i) = 1 \text{ for } x_i^2 \geq a_i,$$
$$\int_{\mathbb{R}^n} x_i^2 d^n F_\mathbf{X}(\mathbf{x}) \leq b_i, \forall i \in [1 : n]\}$$

be the set of all CDFs with the cubic constraints defined by the vectors $\mathbf{a}$ and $\mathbf{b}$, respectively. By strengthening or weakening the constraints of (2), we have

$$\sup_{F_\mathbf{X}(\mathbf{x}) \in \mathbb{F}_1} I(\mathbf{X}; \mathbf{Y}) \leq C(u_p, u_a) \leq \sup_{F_\mathbf{X}(\mathbf{x}) \in \mathbb{F}_2} I(\mathbf{X}; \mathbf{Y}) \quad (68)$$

as long as $\mathbb{F}_1 \subseteq \{F_\mathbf{X}(\mathbf{x})|F_\mathbf{X}(\mathbf{x}) = 1 \text{ for } \|\mathbf{x}\|^2 \geq u_p, \int_{\mathbb{R}^n} \|\mathbf{x}\|^2 d^n F_\mathbf{X}(\mathbf{x}) \leq u_a\} \subseteq \mathbb{F}_2$. One possible choice for $\mathbb{F}_2$ is obtained with the enhanced cubic constraints as follows

$$\mathbb{F}_2 = \mathbb{F}(u_p \mathbf{1}, u_a \mathbf{1})$$

where $\mathbf{1}$ is the n-dimensional all-one vector. Also, a trivial option for $\mathbb{F}_1$ would be

$$\mathbb{F}_1 = \mathbb{F}(\frac{u_p}{n}\mathbf{1}, \frac{u_a}{n}\mathbf{1}).$$

Since the noise elements are independent, we have

$$\sum_{i=1}^n \sup_{\substack{F_{X_i}(x_i): |X_i|^2 \leq \frac{u_p}{n} \\ E[|X_i|^2] \leq \frac{u_a}{n}}} I(X_i; Y_i) \leq C(u_p, u_a)$$

$$\leq \sum_{i=1}^n \sup_{\substack{F_{X_i}(x_i): |X_i|^2 \leq u_p \\ E[|X_i|^2] \leq u_a}} I(X_i; Y_i)$$

which leads to

$$\sum_{i=1}^{n} C_S(\frac{\lambda_i^2 u_p}{n}, \frac{\lambda_i^2 u_a}{n}) \leq C(u_p, u_a)$$

$$\leq \sum_{i=1}^{n} C_S(\lambda_i^2 u_p, \lambda_i^2 u_a),$$

in which $C_S(.,.)$ is the capacity of a scalar AWGN channel under peak and average power constraints defined in [1]. The resources could alternatively be allocated according to the noise covariance matrix Σ such that the resource of each component is inversely proportional to its noise variance. Therefore, another possible set for obtaining a lower bound is

$$\mathbb{F}_1 = \mathbb{F}(u_p \mathbf{v}, u_a \mathbf{v})$$

in which $v_i = \frac{\lambda_i^2}{\sum_{j=1}^{n} \lambda_j^2}$. We name this last set of constraints as modified cubic constraints.

2) Bounds based on the elliptical constraints: Another possible set of lower and upper bounds is obtained by strengthening or weakening the constraints in (57). By noting that

$$\min\{\lambda_1^{-2}, \lambda_2^{-2}, \ldots, \lambda_n^{-2}\} \leq \mathbf{a}^T(\theta)\Sigma\mathbf{a}(\theta)$$
$$\leq \max\{\lambda_1^{-2}, \lambda_2^{-2}, \ldots, \lambda_n^{-2}\}, \quad (69)$$

we get the two following sets of constraints for the lower and the upper bounds of (57), respectively.

$$\mathbb{F}_1 = \{F_{P,\Theta}(\rho, \theta)|P^2 \leq \min\{\lambda_1^2, \ldots, \lambda_n^2\}u_p ,$$
$$E[P^2] \leq \min\{\lambda_1^2, \ldots, \lambda_n^2\}u_a\},$$
$$\mathbb{F}_2 = \{F_{P,\Theta}(\rho, \theta)|P^2 \leq \max\{\lambda_1^2, \ldots, \lambda_n^2\}u_p ,$$
$$E[P^2] \leq \max\{\lambda_1^2, \ldots, \lambda_n^2\}u_a\}.$$

Following the same approach as in the proof of the theorem, it can be verified that with these sets of constraints, the lower and the upper bounds results from the input distributions that have finite number of concentric hyper-ellipsoids as their support.

3) Bounds based on whitening the noise: Another trivial set of upper and lower bounds is obtained by whitening the noise and therefore, making it spherically symmetric. It is obvious that

$$\sup_{\Sigma=\max\{\lambda_1^{-2},\ldots,\lambda_n^{-2}\}\mathbf{I}} I(\mathbf{X};\mathbf{Y}) \leq C(u_p, u_a)$$

$$\leq \sup_{\Sigma=\min\{\lambda_1^{-2},\ldots,\lambda_n^{-2}\}\mathbf{I}} I(\mathbf{X};\mathbf{Y}), \quad (70)$$

where the bounds are obtained by distributions that have finite number of concentric hyper-spheres as their support as in section III. It can be easily verified that the bounds in 2) and 3) are actually the same, although the former is based on weakening or strengthening the constraint and the latter is based on whitening the noise.

4) Lower bound based on Entropy Power Inequality (EPI): The mutual information can be lower bounded as

$$I(\mathbf{X};\mathbf{Y}) = h(\mathbf{Y}) - \frac{1}{2}\ln((2\pi e)^n|\Sigma|)$$
$$\geq \frac{n}{2}\ln\left(e^{\frac{2}{n}h(\mathbf{X})} + e^{\frac{1}{n}\ln((2\pi e)^n|\Sigma|)}\right)$$
$$- \frac{1}{2}\ln((2\pi e)^n|\Sigma|) \quad (71)$$

where in (71), vector EPI [13] has been used. In order to get a lower bound for the capacity, we notice that the maximization of $h(\mathbf{X})$ under the peak and average constraints could be written as

$$\sup_{\substack{F_{\mathbf{X}}(\mathbf{x}):\|\mathbf{X}\|^2 \leq u_p \\ E(\|\mathbf{X}\|^2) \leq u_a}} h(\mathbf{X}) = \sup_{\substack{F_P(\rho):P^2 \leq u_p \\ E(P^2) \leq u_a}} -\int_0^{\infty} f_P(\rho)\ln\frac{f_P(\rho)}{\rho^{n-1}}d\rho$$
$$+ \sum_{i=1}^{n-2} \ln\alpha_i + \ln 2\pi. \quad (72)$$

By the change of variable $T = \frac{P^n}{n}$, we have

$$\sup_{\substack{F_P(\rho):P^2 \leq u_p \\ E(P^2) \leq u_a}} -\int_0^{\infty} f_P(\rho)\ln\frac{f_P(\rho)}{\rho^{n-1}}d\rho = \sup_{\substack{F_T(t):T \leq \frac{u_p^{\frac{n}{2}}}{n} \\ E(T^{\frac{2}{n}}) \leq \frac{u_a}{n^{\frac{2}{n}}}}} h(T).$$
$$(73)$$

It can be verified that optimization theory guarantees a unique solution for (73) and the necessary and sufficient conditions for f_T^* to be the optimal pdf is the existence of a $\lambda \geq 0$ for which the following inequality holds for any $f_T(t)$ that has its support inside the interval $[0, \frac{u_p^{\frac{n}{2}}}{n}]$

$$\int_0^{\frac{u_p^{\frac{n}{2}}}{n}} (\ln f_T^*(t) + \lambda t^{\frac{2}{n}})(f_T^*(t) - f_T(t))dt \leq 0. \quad (74)$$

It is obvious that when $u_a \geq \frac{nu_p}{n+2}$, $\lambda = 0$ and the optimal distribution will be uniform. In the case $u_a < \frac{nu_p}{n+2}$, $\lambda \neq 0$ and the optimal distribution is given by

$$f_T^*(t) = ae^{-\lambda t^{\frac{2}{n}}} , \ t \in [0, \frac{u_p^{\frac{n}{2}}}{n}],$$

or equivalently

$$f_P^*(\rho) = a\rho^{n-1}e^{-\frac{\lambda\rho^2}{(\sqrt[n]{n})^2}} , \ \rho \in [0, \sqrt{u_p}],$$

since it satisfies (74) with equality. The two degrees of freedom a, λ are uniquely obtained by solving the two following equations:

$$\frac{\int_0^{\frac{u_p^{\frac{n}{2}}}{n}} t^{\frac{2}{n}}e^{-\lambda t^{\frac{2}{n}}} dt}{\int_0^{\frac{u_p^{\frac{n}{2}}}{n}} e^{-\lambda t^{\frac{2}{n}}} dt} = \frac{u_a}{n^{\frac{2}{n}}} \quad (75)$$

$$a = \left(\int_0^{\frac{u_p^{\frac{n}{2}}}{n}} e^{-\lambda t^{\frac{2}{n}}} dt\right)^{-1}. \quad (76)$$

It can be verified that the left-hand side of (75) is a strictly decreasing function of λ having the range

$(0, \frac{nu_p}{(n+2)n^{\frac{2}{n}}}]$ and by continuity, there exists a unique $\lambda > 0$ that satisfies (75). Substituting this λ in (76) gives the value of a which results in

$$h(\mathbf{X}) = \frac{\lambda u_a}{(\sqrt[n]{n})^2} + \ln\left(\frac{2(\sqrt{\pi})^n}{a\Gamma(\frac{n}{2})}\right). \quad (77)$$

Substituting (77) in (71), we get the following lower bound for the capacity

$$C(u_p, u_a) \geq \frac{n}{2}\ln\left(\frac{2^{\frac{2}{n}}\pi}{(a\Gamma(\frac{n}{2}))^{\frac{2}{n}}}e^{\frac{2\lambda u_a}{n(\sqrt[n]{n})^2}} + 2\pi e\sqrt[n]{|\Sigma|}\right)$$
$$- \frac{1}{2}\ln((2\pi e)^n|\Sigma|) \quad (78)$$

A visual representation of some of the bounds is shown in figure 1 for $n = 2$, $\lambda_1^2 = 2\lambda_2^2$ and $u_a \geq u_p$. It is obvious that the figures inside the circle (which shows the peak power constraint for the 2-dimensional channel) strengthen the constraint and those outside the circle weaken it. In figure 1(a), the two ellipsoids are obtained from (69). In other words, the inner and the outer ellipsoids are given by

$$a^T(\theta)\mathbf{\Sigma}\mathbf{a}(\theta) = \min\{\lambda_1^{-2}, \lambda_2^{-2}\}$$

and

$$a^T(\theta)\mathbf{\Sigma}\mathbf{a}(\theta) = \max\{\lambda_1^{-2}, \lambda_2^{-2}\},$$

respectively. The inner and outer squares in figure 1(b) are $[-\sqrt{\frac{u_p}{2}}, \sqrt{\frac{u_p}{2}}]^2$ and $[-\sqrt{u_p}, \sqrt{u_p}]^2$, respectively. The modified cubic constraint in figure 1(c) is based on resource allocation according to the channel gains (i.e., λ_1 and λ_2). Channels 1 and 2 have the peak power of $\frac{\lambda_1^2}{\lambda_1^2+\lambda_2^2}u_p(= \frac{2}{3}u_p$ in this example) and $\frac{\lambda_2^2}{\lambda_1^2+\lambda_2^2}u_p(= \frac{1}{3}u_p$ in this example), respectively.

V. NUMERICAL RESULTS

As stated in the Theorem, the magnitude of the optimal input has a finite number of mass points and the phases are distributed according to (20). The algorithm [8] for finding the number, the positions and the probabilities of the optimal mass points is exactly the same as that explained in [2]. When the average power constraint is relaxed, figures 2 to 6 show the capacity of the channel in (2) along with the capacity-achieving input distribution for different values of n. In these figures, black, red and green points have their probabilities in the intervals $[0.7, 1]$, $[0.3, 0.7]$ and $[0, 0.3]$, respectively.

Figure 7 shows the capacity of the four dimensional channel versus u_p along with the optimal input for a fixed average power $u_a = 10$. It is obvious that the capacity saturates at its conventional value given in (25). This saturation shows the near-optimal performance of the discrete input for the conventional unbounded scenario. For example, when $n = 4$ and $u_a = 10$, the capacity of the channel with unbounded input (i.e., $C_G = 2.5055$), which is achieved by a generalized Rayleigh distributed P, can also be achieved with good approximation (i.e., $I(\mathbf{X}; \mathbf{Y}) = 2.5052$) by a pmf having only three mass points below $\sqrt{30}$.

[8]The codes for this section are available at http://www.ee.ic.ac.uk/bruno.clerckx/Research.html .

Figure 8 shows the capacity versus the average power constraint for a fixed value of the peak power ($u_p = 20$). It is obvious that for $u_a \geq u_p$, the average constraint becomes inactive and the capacity is determined only by u_p. We have already shown that when the peak power is very small (i.e., $u_p \ll 1$) and $u_a \geq u_p$, the optimal input has only one mass point at $\rho = \sqrt{u_p}$. Let F_{P_1} denote the cdf of this optimal input. Therefore,

$$f_V(v; F_{P_1}) = K_n(v, \sqrt{u_p}),$$
$$\tilde{h}_V(\rho; F_{P_1}) = -\int_0^\infty K_n(v, \rho)\ln(K_n(v, \sqrt{u_p}))dv.$$

When $u_p \ll 1$, the above marginal entropy density is a convex and increasing function of ρ and satisfies the equality of (29) (with $\lambda = 0$) at $\rho = \sqrt{u_p}$ and the inequality of (28) at all other points. As u_p increases, F_{P_1} remains optimal until it violates the necessary and sufficient conditions. By observing the behavior of $\tilde{h}_V(\rho, F_{P_1})$, it is concluded that as u_p increases, the first point to violate the necessary and sufficient conditions will happen at $\rho = 0$. Therefore, the peak power threshold u_p^t for which F_{P_1} remains optimal (when $u_a \geq u_p$) is obtained by solving the following equation for u_p^t

$$\tilde{h}_V(0; F_{P_1}) = h(V; F_{P_1}). \quad (79)$$

By solving (79) numerically, the values of the peak power threshold are obtained for different values of n as shown in figure 9. For example, for $n = 4$, $u_p^t \approx 12.81$ which means that when the peak power is below 12.81, the support of the optimal input has only one hyper-sphere, and at this threshold it gets another mass point at zero as already shown in figure 4. For $n = 20$, when $u_p \leq 66$, constant amplitude signaling is optimal which is consistent with figure 6. From figure 9, it can be observed that the ratio $\frac{u_p}{n}$ does not necessarily need to be vanishingly small to guarantee the optimality of F_{P_1}. Specifically, for the ratios of $\frac{u_p}{n}$ below (approximately) 3.4, F_{P_1} remains optimal.

It has already been shown that when the number of antennas is above a certain threshold, constant amplitude signaling at the peak power (i.e., $\|\mathbf{X}\| = \sqrt{u_p}$) becomes optimal. Figure 10 compares the achievable rate of the constant amplitude signaling [9]at the peak power with the capacity of the channel (with the constraint $\|\mathbf{X}\|^2 \leq u_p$) and the unbounded Gaussian input having an average power of u_p. As it can be observed, when the number of antennas is sufficiently large, constant amplitude signaling is not only optimal but also it has a performance close to that of the unbounded Gaussian signaling.
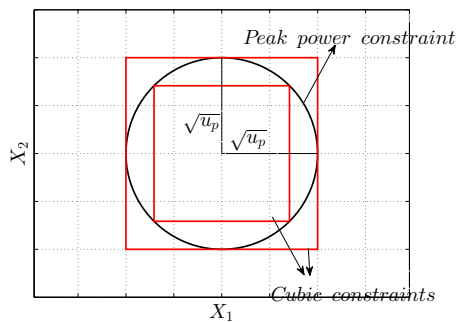
Figures 11 and 12 demonstrate the bounds for the deterministic MIMO channel in (53) for two values of the condition

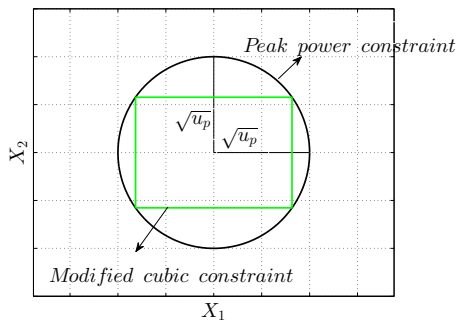[9]The rate has been obtained by numerical evaluation of

$$\sup_{F_{\mathbf{X}}(\mathbf{x}):\|\mathbf{X}\|^2=u_p} I(\mathbf{X}; \mathbf{Y}) = -\int_0^\infty e^{-\frac{(\sqrt[n]{nv})^2+u_p}{2}}\frac{I_{\frac{n}{2}-1}(\sqrt{u_p}\sqrt[n]{nv})}{(\sqrt{u_p}\sqrt[n]{nv})^{\frac{n}{2}-1}} \times$$
$$\ln\left(e^{-\frac{(\sqrt[n]{nv})^2+u_p}{2}}\frac{I_{\frac{n}{2}-1}(\sqrt{u_p}\sqrt[n]{nv})}{(\sqrt{u_p}\sqrt[n]{nv})^{\frac{n}{2}-1}}\right)dv$$
$$- \frac{n}{2}\ln(2e) + \ln 2.$$
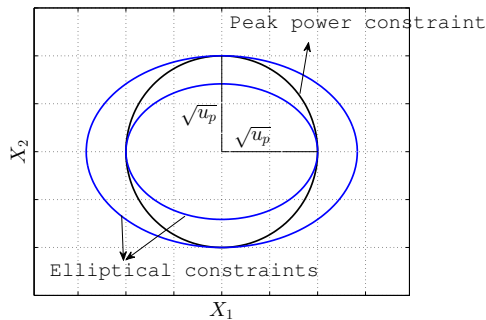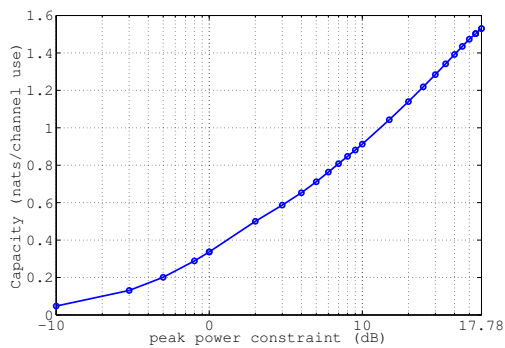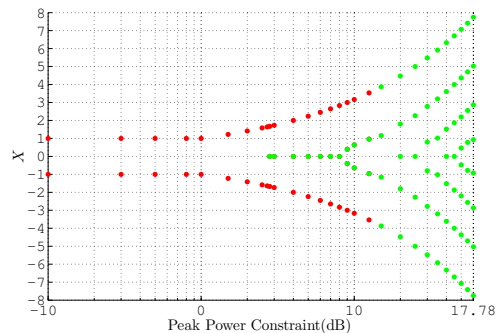
(a) Elliptical Constraints

(b) Cubic Constraints

(c) Modified Cubic Constraint

Fig. 1: Weakening or strengthening the peak power constraint for $n = 2$ and $\lambda_1^2 = 2\lambda_2^2$.
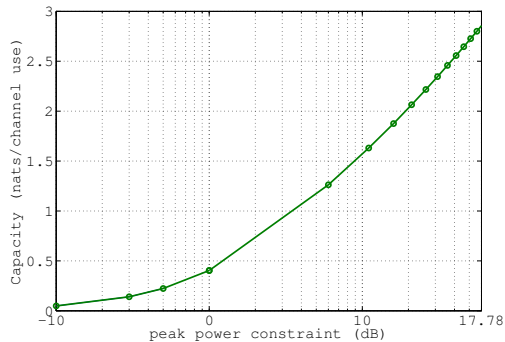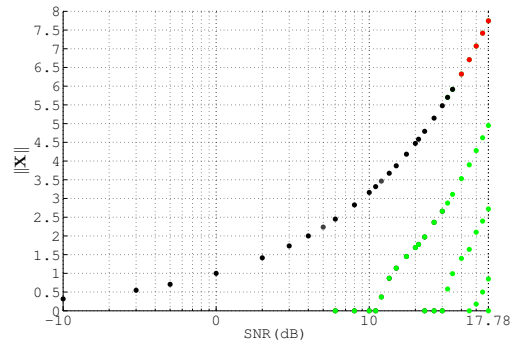

(a) Capacity
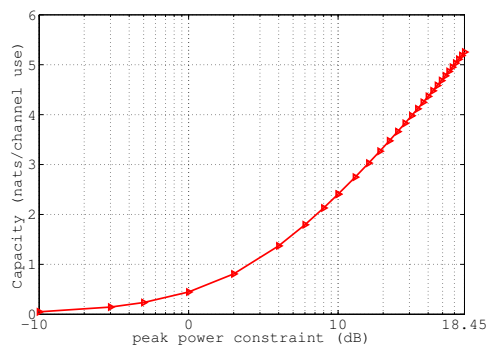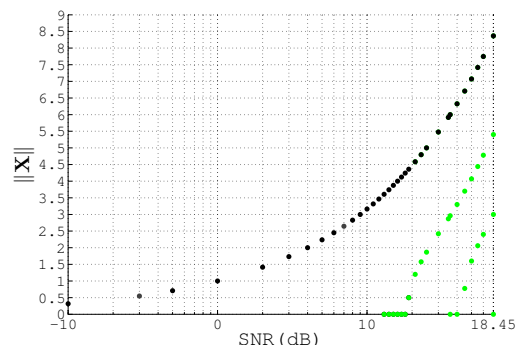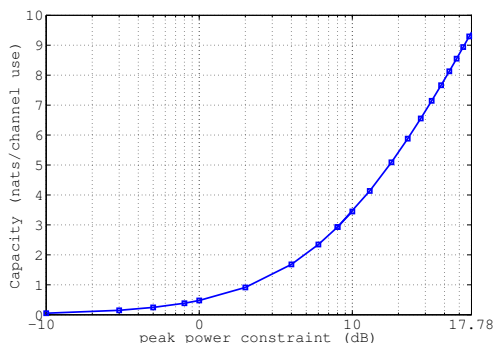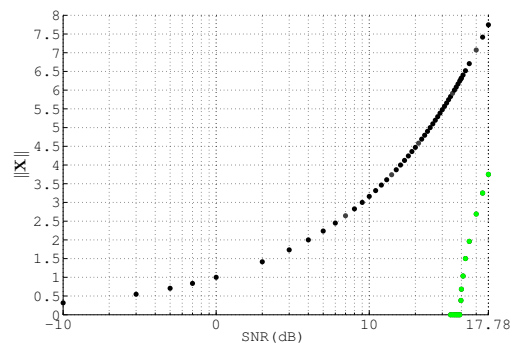

(b) Optimal Input Distribution

Fig. 2: Capacity vs. u_p for $n = 1$ ($u_a \geq u_p$), and the optimal input mass points.
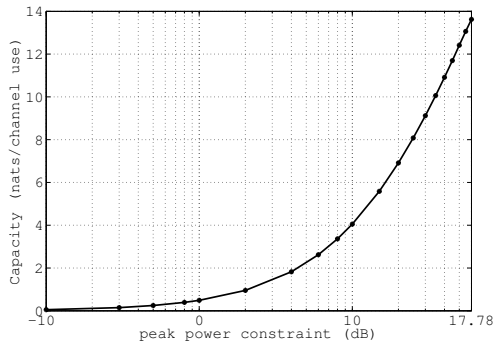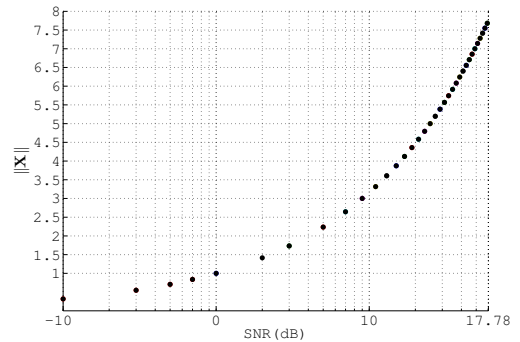
(a) Capacity

(b) Optimal Input Distribution

Fig. 3: Capacity vs. u_p for $n = 2$ ($u_a \geq u_p$), and the optimal input mass points.

(a) Capacity

(b) Optimal Input Distribution

Fig. 4: Capacity vs. u_p for $n = 4$ ($u_a \geq u_p$), and the optimal input mass points.

(a) Capacity

(b) Optimal Input Distribution

Fig. 5: Capacity vs. u_p for $n = 10$ ($u_a \geq u_p$), and the optimal input mass points.

number of the channel. It can be observed that the gap between the elliptical lower and upper bound increases with the condition number. This is intuitively justified by noting that the elliptical bounds converge to the actual capacity of the channel when the condition number approaches unity. For large values of the condition number, the lower bound obtained by modified cubic constraints performs better than the equal resource allocation at small values of the peak power. Finally, it is important to note that although the lower bound obtained by EPI is loose in these two figures, it becomes asymptotically tight for large values of u_p. It can be easily verified by the fact that when the average power constraint is relaxed, we have $\lambda = 0$ and $a = \frac{n}{u_p^{\frac{n}{2}}}$ in (77). When $u_p \rightarrow \infty$ the lower bound in (78) gets arbitrarily close to $h(\mathbf{X})$ in (77) which is obviously an upper bound for the capacity. This justifies the asymptotic tightness of the bound resulted from EPI at large values of u_p.
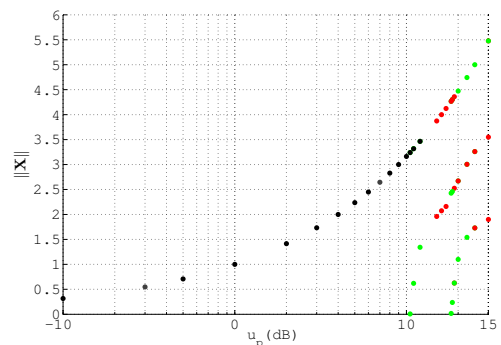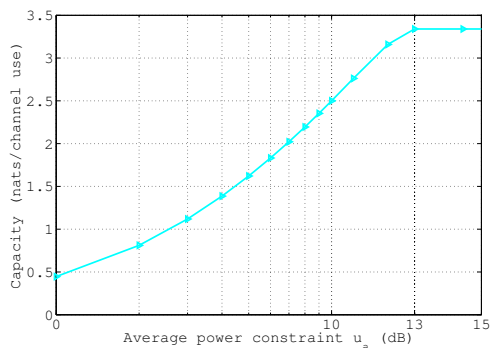
(a) Capacity

(b) Optimal Input Distribution

Fig. 6: Capacity vs. u_p for $n = 20$ ($u_a \geq u_p$), and the optimal input mass points.
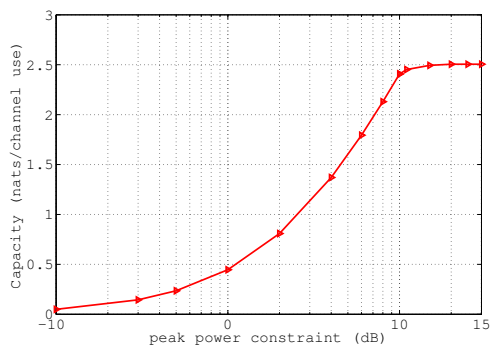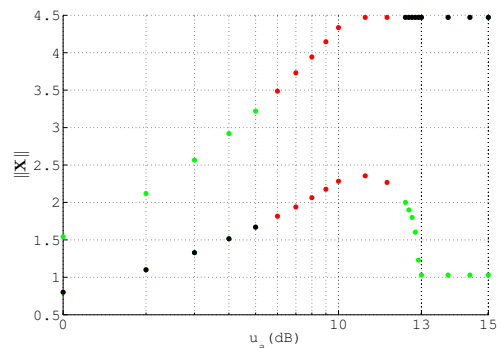

(a) Capacity

(b) Optimal Input Distribution

Fig. 7: Capacity vs. u_p for $n = 4$ ($u_a = 10$), and the optimal input mass points.

(a) Capacity

(b) Optimal Input Distribution

Fig. 8: Capacity vs. u_a for $n = 4$ ($u_p = 20$), and the optimal input mass points.

VI. CONCLUSION

We have shown that the capacity-achieving distribution of the vector Gaussian channel with identity channel matrix under the peak and average power constraints 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. It was shown that when the peak power is the only active constraint, constant amplitude signaling at the peak power is optimal when the number of dimensions is above a

threshold. Finally, some upper and lower bounds were given for the general deterministic channel and their performance was evaluated numerically as a function of the condition number of the channel.

The results of the paper could be applied to the MIMO communication systems with only one single RF chain at the transmitter which is of great interest and necessitate the peak power constraint. The importance of the results becomes more pronounced in the massive MIMO settings, where it
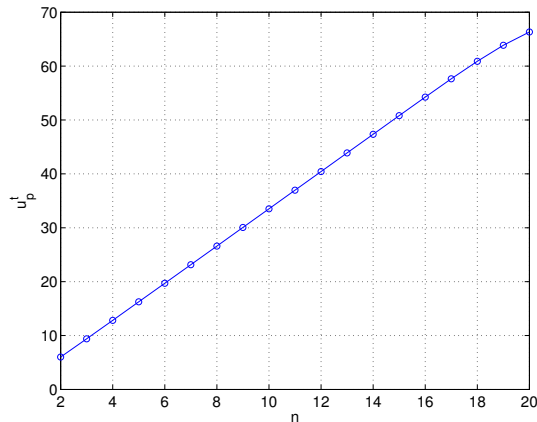
Fig. 9: The peak power threshold for which F_{P_1} remains optimal versus n ($u_a \geq u_p$).

Fig. 11: Bounds for the capacity of the deterministic MIMO channel ($\lambda_2^2 = 2\lambda_1^2 = 1$).
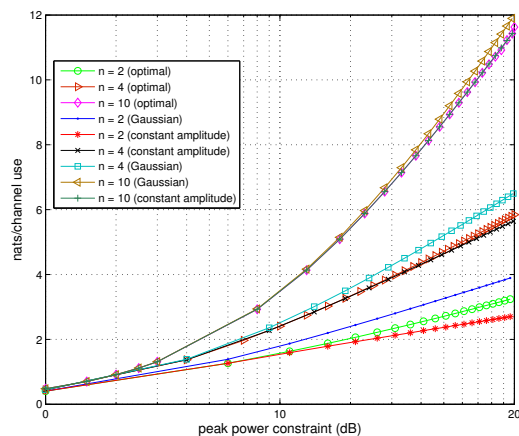

Fig. 10: Achievable rate by the constant amplitude signaling at the peak power (i.e., $\|\mathbf{X}\| = \sqrt{u_p}$) when the average power constraint is relaxed.
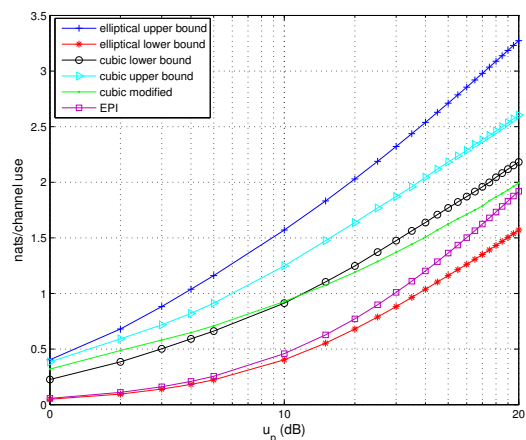
Fig. 12: Bounds for the capacity of the deterministic MIMO channel ($\lambda_2^2 = 10\lambda_1^2 = 1$).

was shown that the capacity has a closed form solution and no computer program is needed to find the optimal input distribution.

APPENDIX A
DERIVATION OF (15)

The following lemma is useful in the sequel.

Lemma 1. Let a and b be two real numbers with $a > 0$. Also, let $\mathbb{N}_0$ be the set of non-negative integers. Then,

$$\int_{-1}^{1} I_n(a\sqrt{1-u^2})(\sqrt{1-u^2})^n e^{-bu} du$$
$$= \sqrt{2\pi} a^n \frac{I_{n+\frac{1}{2}}(\sqrt{a^2+b^2})}{(\sqrt{a^2+b^2})^{n+\frac{1}{2}}}, \; n = \frac{k}{2}, \; \forall k \in \mathbb{N}_0. \tag{80}$$

Proof. By using [14, pp. 698], (80) could be shown for $n = 0$. Also, by some manipulation, (80) holds true for $n = \frac{1}{2}, 1, \frac{3}{2}$.
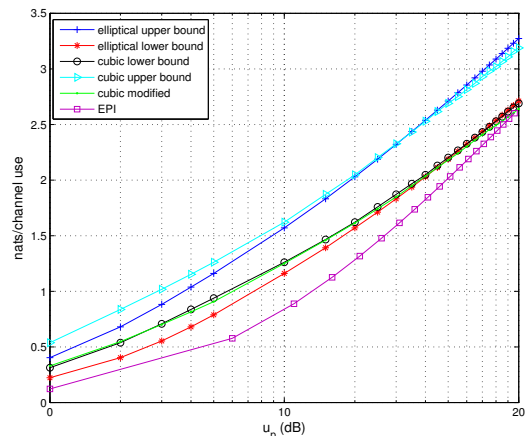
For general n, we use induction as follows. Denote the left-hand side of (80) by Q_n. It is shown that if (80) is true for n, it will also be true for $n + \frac{1}{2}$. In other words, if

$$Q_n = \sqrt{2\pi} a^n \frac{I_{n+\frac{1}{2}}(\sqrt{a^2+b^2})}{(\sqrt{a^2+b^2})^{n+\frac{1}{2}}} \quad (n \geq \frac{3}{2}), \tag{81}$$

then

$$Q_{n+\frac{1}{2}} = \sqrt{2\pi} a^{n+\frac{1}{2}} \frac{I_{n+1}(\sqrt{a^2+b^2})}{(\sqrt{a^2+b^2})^{n+1}}. \tag{82}$$

By using the recursive identity for the bessel function (i.e., $I_\alpha(z) = I_{\alpha-2}(z) - \frac{2(\alpha-1)}{z} I_{\alpha-1}(z)$), we have

$$Q_{n+\frac{1}{2}} = \int_{-1}^{1} I_{n-\frac{3}{2}}(a\sqrt{1-u^2})(\sqrt{1-u^2})^{n+\frac{1}{2}} e^{-bu} du$$
$$- \frac{2(n-\frac{1}{2})}{a}$$

$$\times \int_{-1}^{1} I_{n-\frac{1}{2}}(a\sqrt{1-u^2})(\sqrt{1-u^2})^{n-\frac{1}{2}}e^{-bu}du$$

$$= \int_{-1}^{1} I_{n-\frac{3}{2}}(a\sqrt{1-u^2})(\sqrt{1-u^2})^{n+\frac{1}{2}}e^{-bu}du$$

$$- 2(n-\frac{1}{2})\sqrt{2\pi}a^{n-\frac{3}{2}}\frac{I_n(\sqrt{a^2+b^2})}{(\sqrt{a^2+b^2})^n}, \qquad (83)$$

where in (83), we have used (81). From (81), we have

$$Q_{n-\frac{1}{2}} = \sqrt{2\pi}a^{n-\frac{1}{2}}\frac{I_n(\sqrt{a^2+b^2})}{(\sqrt{a^2+b^2})^n}. \qquad (84)$$

By taking the derivative of (84) with respect to a and using the identity $I'_\alpha(z) = \frac{1}{2}(I_{\alpha-1}(z) + I_{\alpha+1}(z))$ for $\alpha \neq 0$, we have

$$\int_{-1}^{1} I_{n-\frac{3}{2}}(a\sqrt{1-u^2})(\sqrt{1-u^2})^{n+\frac{1}{2}}e^{-bu}du + Q_{n+\frac{1}{2}}$$

$$= 2\sqrt{2\pi}\frac{\partial}{\partial a}\left\{a^{n-\frac{1}{2}}\frac{I_n(\sqrt{a^2+b^2})}{(\sqrt{a^2+b^2})^n}\right\}. \qquad (85)$$

Solving for $Q_{n+\frac{1}{2}}$ in (83) and (85) results in

$$Q_{n+\frac{1}{2}} = \sqrt{2\pi}a^{n-\frac{1}{2}}\frac{\partial}{\partial a}\left\{\frac{I_n(\sqrt{a^2+b^2})}{(\sqrt{a^2+b^2})^n}\right\}$$

$$= \sqrt{2\pi}a^{n+\frac{1}{2}}\frac{I_{n+1}(\sqrt{a^2+b^2})}{(\sqrt{a^2+b^2})^{n+1}}, \qquad (86)$$

where in (86), we have used the identity $\frac{d}{dx}\{\frac{I_n(x)}{x^n}\} = \frac{I_{n+1}(x)}{x^n}$. This completes the proof of lemma. $\qquad \square$

(15) is equivalent to

$$\underbrace{\int_0^\pi \dots \int_0^\pi}_{n-2 \text{ times}} \int_0^{2\pi} \frac{1}{(\sqrt{2\pi})^n}e^{x\mathbf{a}^T(\theta)\mathbf{a}(\psi)}\prod_{i=1}^{n-2}\sin^{n-i-1}\psi_i d\psi_{n-1}\dots d\psi_1$$

$$= \begin{cases} \frac{I_{\frac{n}{2}-1}(x)}{(x)^{\frac{n}{2}-1}} & x \neq 0 \\ \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}-1}} & x = 0 \end{cases} \quad \forall n \geq 2. \qquad (87)$$

If $x = 0$, it is obvious that the left-hand side of (87) is the hyper-surface area of an n-sphere with unit radius ($= \frac{2\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2})}$) divided by $(\sqrt{2\pi})^n$ which results in the value shown on the right-hand side. Therefore, we consider $x \neq 0$. It is obvious that (87) is valid for $n = 2$. Denote the left-hand side of (87) by W_n and assume it is valid for $n \geq 2$. It can be verified that

$$W_{n+1} = \int_0^\pi \frac{I_{\frac{n}{2}-1}(x\sin\theta\sin\psi)}{\sqrt{2\pi}(x\sin\theta\sin\psi)^{\frac{n}{2}-1}}\sin^{n-1}\psi e^{x\cos\theta\cos\psi}d\psi$$

$$= \int_{-1}^{1} \frac{I_{\frac{n}{2}-1}(x\sin\theta\sqrt{1-u^2})}{\sqrt{2\pi}(x\sin\theta)^{\frac{n}{2}-1}}(\sqrt{1-u^2})^{\frac{n}{2}-1}e^{-x\cos\theta u}du \qquad (88)$$

$$= \frac{I_{\frac{n-1}{2}}(x)}{(x)^{\frac{n-1}{2}}} \qquad (89)$$

where in (88), $u = -\cos\psi$ and in (89), we have used lemma 1. This completes the proof of (87).

APPENDIX B

Proposition. Let X be a non-negative random variable and $m \in \mathbb{R}^+$. The following optimization problem

$$\sup_{F_X(x):E[X^m]\leq A} h(X), \qquad (90)$$

has a unique solution. Further, the maximum is

$$\frac{\Gamma(\frac{m+1}{m})}{\Gamma(\frac{1}{m})} - \ln\left(\frac{m\sqrt[m]{\frac{\Gamma(\frac{m+1}{m})}{\Gamma(\frac{1}{m})A}}}{\Gamma(\frac{1}{m})}\right), \qquad (91)$$

and is achieved by the following distribution

$$f_{X^*}(x) = \frac{m\sqrt[m]{\frac{\Gamma(\frac{m+1}{m})}{\Gamma(\frac{1}{m})A}}}{\Gamma(\frac{1}{m})}e^{-\frac{\Gamma(\frac{m+1}{m})}{A\Gamma(\frac{1}{m})}x^m}. \qquad (92)$$

Proof. Let Ω denote the set of all probability density functions on the non-negative real line. It can be shown that Ω is convex and compact in the Levy metric. Further, the following function

$$L(f_X(x)) = h(X) - \lambda(\int_0^\infty x^m f_X(x)dx - A)$$

is for $\lambda \geq 0$, a continuous, weakly differentiable and strictly concave function of $f_X(x)$ having the weak derivative at $f_X^0(x)$ as

$$L'_{f_X^0(x)}(f_X(x)) = \int_0^\infty (\ln f_X^0(x) + \lambda x^m)(f_X^0(x) - f_X(x))dx.$$

Therefore, the Lagrangian optimization guarantees a unique solution for (90) and the necessary and sufficient condition for $f_{X^*}(x)$ to be the optimal solution is the existence of a $\lambda \geq 0$ for which $L'_{f_{X^*}(x)}(f_X(x)) \leq 0 \quad \forall f_X(x) \in \Omega$. It can be verified that for $\lambda = \frac{\Gamma(\frac{m+1}{m})}{A\Gamma(\frac{1}{m})}$, the distribution in (92) results in $L'_{f_{X^*}(x)}(f_X(x)) = 0$ which satisfies the necessary and sufficient conditions. Hence, the pdf in (92), which has the differential entropy in (91), is the unique solution of (90). $\qquad \square$

APPENDIX C
PROOF OF THE THEOREM

Let $\mathbb{F}_{u_p}$ denote the space of all cumulative distribution functions satisfying the peak power constraint, i.e.

$$\mathbb{F}_{u_p} = \{F_P(\rho)|F_P(\rho) = 0 \ \forall \rho < 0 \ , \ F_P(\rho) = 1 \ \forall \rho \geq \sqrt{u_p}\}.$$

The metric space $(\mathbb{F}_{u_p}, d_L)$ is convex and compact ([15], [3, Appendix I]) where d_L denotes the Levy metric [16] (note that the proof of the compactness in [3] relies only on the average power constraint). The differential entropy $h(V; F_P) : \mathbb{F}_{u_p} \to \mathbb{R}$ is continuous ([15], [2, Proposition 3], [3, Appendix I], [6, Proposition 1]) (note that the proof of continuity in [6] is more general in the sense that it does not rely on the Schwartz properties), strictly concave and weakly differentiable ([15],

[2, Proposition 4], [3, Appendix II], [6, Proposition 2]) and has the weak derivative at F_P^0 given by

$$h'_{F_P^0}(V; F_P) = \lim_{\zeta \to 0} \frac{h(V; (1-\zeta)F_P^0 + \zeta F_P) - h(V; F_P^0)}{\zeta}$$

$$= \int_0^{\sqrt{u_p}} \tilde{h}_V(\rho; F_P^0) dF_P(\rho) - h(V; F_P^0), \forall F_P \in \mathbb{F}_{u_p}.$$

The average power constraint is denoted by

$$G(F_P) = \int_0^{\sqrt{u_p}} \rho^2 dF_P(\rho) - u_a \leq 0.$$

It is obvious that $G : \mathbb{F}_{u_p} \to \mathbb{R}$ is linear and weakly differentiable having the weak derivative at F_P^0 given by

$$G'_{F_P^0}(F_P) = G(F_P) - G(F_P^0) \quad , \forall F_P \in \mathbb{F}_{u_p}.$$

Since $h(V; F_P)$ and $G(F_P)$ are concave maps from $\mathbb{F}_{u_p}$ to $\mathbb{R}$, Lagrangian optimization [17] guarantees a unique solution for (24) and the necessary and sufficient condition for F_{P*} to be the optimal solution is the existence of a $\lambda(\geq 0)$ such that

$$\int_0^{\sqrt{u_p}} (\tilde{h}_V(\rho; F_{P*}) - \lambda \rho^2) dF_P(\rho) \leq h(V; F_{P*}) - \lambda u_a,$$

$$\forall F_P \in \mathbb{F}_{u_p}. \qquad (93)$$

It can be shown that (93) is equivalent to (28) and (29) ([1, Corollary 1]). In order to show the finiteness of the cardinality of ϵ_{P*}, we extend the marginal entropy density in (26) to the complex domain i.e.,

$$\tilde{h}_V(z; F_P) = -\int_0^\infty K_n(v, z) \ln f_V(v; F_P) dv \quad , \quad z \in \mathbb{C}. \qquad (94)$$

Proposition 1. The kernel $K_n(v, z)$ is an entire function in z for every v.

Proof. This can be verified by the fact that the real and imaginary parts of $K(v, z = x + jy)$ have continuous partial derivatives and satisfy the Cauchy-Riemann equations which leads to its holomorphy over the complex plane. As a result, by Cauchy's theorem, for every rectifiable closed curve γ in $\mathbb{C}$,

$$\int_\gamma K_n(v, z) dz = 0. \qquad (95)$$

$\square$

Proposition 2. The marginal entropy density $\tilde{h}_V(z; F_P)$ is an entire function.

Proof. First, we show the continuity of $\tilde{h}_V(z; F_P)$. Let $\{z_m\}_1^\infty$ be a sequence of complex numbers converging to z_0. Since $K_n(v, z)$ is holomorphic (see Proposition 1), it is continuous. Therefore,

$$\lim_{m \to \infty} K_n(v, z_m) \ln f_V(v; F_P) = K_n(v, z_0) \ln f_V(v; F_P).$$

Because the kernel is continuous and $K_n(v, +\infty) = 0$, it is also bounded (i.e., $0 \leq K_n(v, \rho) < \infty$ for all $\rho \in \mathbb{R}_{\geq 0}$.) The continuity and boundedness of the kernel guarantees the continuity of $f_V(v; F_P)$ given in (13) by the application of

Lebesgue's dominated convergence theorem. This allows us to write

$$0 < e^{-\frac{(\sqrt[n]{nv})^2 + u_p}{2}} \frac{1}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}-1}}$$

$$\leq \min_{\rho \in [0, \sqrt{u_p}]} K_n(v, \rho)$$

$$\leq f_V(v; F_P)$$

$$\leq \max_{\rho \in [0, \sqrt{u_p}]} K_n(v, \rho)$$

$$\leq e^{-\frac{(\sqrt[n]{nv})^2}{2}} \frac{I_{\frac{n}{2}-1}(u_p \sqrt[n]{nv})}{(u_p \sqrt[n]{nv})^{\frac{n}{2}-1}}$$

$$< \infty, \qquad (96)$$

since $\frac{I_n(x)}{x^n}(x > 0)$ is a strictly increasing function. Therefore,

$$|\ln f_V(v; F_P)| \leq \frac{(\sqrt[n]{nv})^2 + u_p}{2} + |\ln(\frac{I_{\frac{n}{2}-1}(u_p \sqrt[n]{nv})}{(u_p \sqrt[n]{nv})^{\frac{n}{2}-1}})|$$

$$\leq \frac{(\sqrt[n]{nv})^2 + u_p}{2} + u_p \sqrt[n]{nv} + \ln(\Gamma(\frac{n}{2}) 2^{\frac{n}{2}-1}) \qquad (97)$$

$$\leq \frac{(\sqrt[n]{nv})^2}{2}(1 + u_p) + u_p + \ln(\Gamma(\frac{n}{2}) 2^{\frac{n}{2}-1}) \qquad (98)$$

where in (97), we have used the inequality

$$\frac{I_\nu(x)}{x^\nu} < \frac{\cosh x}{2^\nu \Gamma(\nu+1)} \overset{x > 0}{<} \frac{e^x}{2^\nu \Gamma(\nu+1)}, \qquad (99)$$

which was proved in [18]. From (98), it can be verified that

$$|\tilde{h}_V(z_m; F_P)|$$

$$\leq \int_0^\infty |e^{-\frac{(\sqrt[n]{nv})^2 + z_m^2}{2}}| \frac{I_{\frac{n}{2}-1}(z_m \sqrt[n]{nv})}{(z_m \sqrt[n]{nv})^{\frac{n}{2}-1}}| |\ln f_V(v; F_P)| dv$$

$$\leq |e^{-\frac{z_m^2}{2}}| \int_0^\infty e^{-\frac{(\sqrt[n]{nv})^2}{2}} \frac{I_{\frac{n}{2}-1}(|z_m| \sqrt[n]{nv})}{(|z_m| \sqrt[n]{nv})^{\frac{n}{2}-1}} |\ln f_V(v; F_P)| dv \qquad (100)$$

$$\leq |e^{\frac{|z_m|^2 - z_m^2}{2}}| \left(\frac{(|z_m|^2 + n)}{2}(1 + u_p) + u_p \right.$$

$$\left. + \ln(\Gamma(\frac{n}{2}) 2^{\frac{n}{2}-1}) \right) \qquad (101)$$

$$< \infty$$

where in (100), we have used the fact that $|I_n(z)| \leq I_n(|z|)$ and in (101) the upper bound in (98) has been used. Since the absolute value of the integrand of $\tilde{h}_V(z_n; F_P)$ is integrable, by Lebesgue's dominated convergence theorem, we have

$$\lim_{m \to \infty} \tilde{h}_V(z_m; F_P) = \lim_{m \to \infty} \int_0^\infty K_n(v, z_m) \ln f_V(v; F_P) dv$$

$$= \int_0^\infty \lim_{m \to \infty} K_n(v, z_m) \ln f_V(v; F_P) dv$$

$$= \int_0^\infty K_n(v, z_0) \ln f_V(v; F_P) dv$$

$$= \tilde{h}_V(z_0; F_P)$$

which proves the continuity of $\tilde{h}_V(z, F_P)$. Let ∂T denote an arbitrary triangle in the complex plane. We can write,

$$\int_{\partial T} \tilde{h}_V(z; F_P)dz = -\int_{\partial T}\int_0^\infty K_n(v, z)\ln f_V(v; F_P)dvdz$$

$$= -\int_0^\infty \int_{\partial T} K_n(v, z)dz \ln f_V(v; F_P)dv \tag{102}$$

$$= 0 \tag{103}$$

where (102) is allowed by Fubini's theorem, because for a given rectifiable triangle ∂T

$$\int_{\partial T} |\tilde{h}_V(z; F_P)|dz < \infty.$$

(103) is due to the holomorphy of $K_n(v, z)$ (see (95)). Therefore, by Morera's theorem (with weakened hypothesis), it is concluded that $\tilde{h}_V(z; F_P)$ is holomorphic on the entire complex plane.

Alternatively, the holomorphy of the marginal entropy density can be proved as follows. The following integral

$$\tilde{h}_V(z; F_P) = -\int_0^\infty K_n(v, z)\ln f_V(v; F_P)dv$$

is uniformly convergent for all $z \in \mathbb{K}$ (where $\mathbb{K}$ is a compact subset of $\mathbb{C}$) in the sense that for $\forall \delta > 0$, there exists some real number L_0 such that

$$|-\int_{L_1}^{L_2} K_n(v, z)\ln f_V(v; F_P)dv| < \delta,$$

for $\forall L_1, L_2$ satisfying $L_0 < L_1 < L_2$. Therefore, by the differentiation lemma [19], $\tilde{h}_V(z; F_P)$ is holomorphic on the complex plane. $\square$

If ϵ_{P*} has infinite number of points, since it is a bounded subset of the real line ($\subseteq [0, \sqrt{u_p}]$), it has an accumulation point in $\mathbb{R}$ by Bolzano-Weierstrass theorem [20]. Hence, according to (29), the two holomorphic functions $\tilde{h}_V(z; F_{P*})$ and $h(V; F_{P*}) + \lambda(z^2 - u_a)$ become equal on an infinite set that has an accumulation point in $\mathbb{C}$. Therefore, by the identity theorem for holomorphic functions of one complex variable [19], the two functions are equal on the whole complex plane, i.e.

$$\tilde{h}_V(z; F_{P*}) = h(V; F_{P*}) + \lambda(z^2 - u_a) \quad , \quad \forall z \in \mathbb{C},$$

which results in

$$\tilde{h}_V(\rho; F_{P*}) = h(V; F_{P*}) + \lambda(\rho^2 - u_a) \quad , \quad \forall \rho \in \mathbb{R}. \tag{104}$$

In the following, we show that (104) leads to a contradiction.

1) $\lambda = 0$. In this case, in which the average power constraint is relaxed, (104) results in

$$f_V(v; F_P) = e^{-h(V; F_{P*})}, \tag{105}$$

which is a constant and is guaranteed by the invertibility of (26) to be the only solution. The uniform distribution in (105) cannot be a legitimate pdf for V on the non-negative real line. This contradiction can be observed in

an alternative way. By noting that from (105) and (13), if $f_V(v; F_P)$ is to be constant (shown by C), then

$$f_P(\rho) = C\rho^{n-1}, \quad \rho \geq 0,$$

which is the only solution for $f_P(\rho)$ by the invertibility of (13). Again, it is not a legitimate pdf for ρ and obviously violates the peak power constraint.

2) $\lambda > 0$. In this case (104) holds iff

$$f_V(v; F_P) = \frac{2(\sqrt{\lambda})^n}{\Gamma(\frac{n}{2})}e^{-\lambda(\sqrt[n]{nv})^2}, \tag{106}$$

which also holds iff

$$f_P(\rho) = (\sqrt{\frac{\lambda}{1-2\lambda}})^n \frac{\rho^{n-1}e^{-\frac{\lambda}{1-2\lambda}\rho^2}}{\Gamma(\frac{n}{2})}, \tag{107}$$

with $\lambda = \frac{\Gamma(\frac{n}{2}+1)}{\Gamma(\frac{n}{2})(u_a+\frac{n}{2})}$. It is obvious that for $0 < \lambda < \frac{1}{2}$, the solution in (107) violates the peak power constraint and for $\lambda > \frac{1}{2}$, no legitimate $f_P(\rho)$ results in (106). For $\lambda = \frac{1}{2}$, $f_P(\rho) = \delta(\rho)$ which implies a unit mass point at zero. This, of course, contradicts the first assumption of F_{P*} having infinite points of increase and also results in $C(u_p, u_a) = 0$.

Therefore, the magnitude of the optimal input has a finite number of mass points. This completes the proof of the theorem.

APPENDIX D
TWO INVERTIBLE TRANSFORMS

In this section, we show that the two following integral transforms are invertible (i.e., one-to-one),

$$q(v) = \int_0^\infty K_n(v, \rho)t(\rho)d\rho, \tag{108}$$

$$w(\rho) = \int_0^\infty K_n(v, \rho)g(v)dv, \tag{109}$$

where t is allowed to have at most an exponential order and g a polynomial with a finite degree, so that the transforms exist. The invertibility of (108) and (109) is equivalent to the invertibility of (13) and (26), respectively. The following lemma will be helpful in the sequel.

Lemma 2. The kernel function $K_n(v, \rho)$ satisfies the two following equations,

$$\int_0^\infty K_n(v, \rho)\rho^{n-1}e^{-s\rho^2}d\rho = \frac{e^{-\frac{s}{2s+1}(\sqrt[n]{nv})^2}}{(\sqrt{2s+1})^n}, \tag{110}$$

$$\int_0^\infty K_n(v, \rho)e^{-s(\sqrt[n]{nv})^2}dv = \frac{e^{-\frac{s}{2s+1}\rho^2}}{(\sqrt{2s+1})^n}, \tag{111}$$

where $s \geq 0$.

Proof. From the properties of probability density functions,

$$\int_{R^n} \frac{1}{(\sqrt{2\pi\sigma^2})^n}e^{-\frac{\|\mathbf{y}-\mathbf{x}\|^2}{2\sigma^2}}d\mathbf{y} = 1.$$

By writing $\mathbf{y}$ and $\mathbf{x}$ in spherical coordinates (i.e., $\mathbf{y} \equiv (r, \psi)$ and $\mathbf{x} \equiv (\rho, \theta)$), and by substituting $\beta = \frac{1}{2\sigma^2}$ and $\alpha = \frac{\rho}{\sigma^2}$, we get (112) on top of the next page.

$$\int_0^\infty \underbrace{\int_0^\pi \cdots \int_0^\pi}_{n-2 \text{ times}} \int_0^{2\pi} e^{-\beta r^2 + \alpha r \mathbf{a}^T(\theta)\mathbf{a}(\psi)} r^{n-1} \prod_{i=1}^{n-2} \sin^{n-i-1}\psi_i \, d\psi_{n-1} d\psi_{n-2} \ldots d\psi_1 dr = \left(\sqrt{\frac{\pi}{\beta}}\right)^n e^{\frac{\alpha^2}{4\beta}}. \tag{112}$$

By using (112) and by change of variables, (110) and (111) are obtained. $\qquad\square$

In order to show the invertibility of (108), it is sufficient to show that the following

$$\int_0^\infty K_n(v,\rho)t(\rho)d\rho = 0, \tag{113}$$

results in $t(\rho) = 0$. From (113), we have

$$\int_0^\infty \int_0^\infty K_n(v,\rho)t(\rho)d\rho e^{-s(\sqrt[n]{nv})^2} dv = 0 \quad s \geq 0.$$

By changing the order of integration, which is allowed here by Fubini's theorem, and by (111),

$$\int_0^\infty t(\rho)\frac{e^{-\frac{s}{2s+1}\rho^2}}{(\sqrt{2s+1})^n}d\rho = 0, \quad s \geq 0,$$

which results in

$$\int_0^\infty \frac{t(\sqrt{x})}{\sqrt{x}}e^{-\mu x}dx = 0, \quad \mu \in [0,\tfrac{1}{2}). \tag{114}$$

Again, by extending μ to the complex domain, it is easy to verify that the left-hand side of (114) is holomorphic on the complex plane. Since this holomorphic function is zero on an infinite set ($[0,\frac{1}{2})$) which has an accumulation point in $\mathbb{C}$, it is zero on the whole complex plane and consequently the real line by the identity theorem. Therefore,

$$\int_0^\infty \frac{t(\sqrt{x})}{\sqrt{x}}e^{-\mu x}dx = 0, \quad \mu \in \mathbb{R},$$

which results in $t(\rho) = 0$. The uniqueness of this solution results from the invertibility of Laplace transform (by considering the non-negative values for μ). It is obvious that the same approach can be carried out to show the invertibility of the transform (109). Alternatively, the following property of the kernel function

$$K_n(v,\rho) = K_n(\frac{\rho^n}{n}, \sqrt[n]{nv})$$

could be used in (108) to show the invertibility of (109).

APPENDIX E
ALTERNATIVE PROOF FOR REMARK 2

From [21] and [22], we have[10]

$$\Gamma(x+1) < \sqrt{\pi}(\frac{x}{e})^x(8x^3 + 4x^2 + x + \frac{1}{30})^{\frac{1}{6}}. \tag{115}$$

Let $f(n) \triangleq 2e\left[\frac{(n-1)}{2}\Gamma(\frac{n-1}{2})\right]^{\frac{2}{n-1}}$. From (44), we can write

$$C_G \geq \frac{n-1}{2}\log\left(1+\frac{u_p}{f(n)}\right)$$
$$\geq \frac{n-1}{2}\log\left(1+\frac{u_p}{F(n)}\right),$$

[10]Tighter bounds for Gamma function can be found in [23].

in which $F(n)$ is an upper bound for $f(n)$ and is obtained from (115) as

$$F(n) = 2e\left[\frac{(n-1)}{2}\sqrt{\pi}\left(\frac{n-3}{2e}\right)^{\frac{n-3}{2}}\left(8(\frac{n-3}{2})^3\right.\right.$$
$$\left.\left.+4(\frac{n-3}{2})^2 + \frac{n-3}{2} + \frac{1}{30}\right)^{\frac{1}{6}}\right]^{\frac{2}{n-1}}.$$

The behavior of $F(n)$ as n goes to infinity can be obtained as follows.

$$\lim_{n\to\infty}\ln\frac{F(n)}{2e} = \lim_{n\to\infty}\frac{n-3}{n-1}\ln(\frac{n-3}{2e})$$
$$+\lim_{n\to\infty}\frac{2}{n-1}\ln\left[\frac{(n-1)}{2}\sqrt{\pi}\left(8(\frac{n-3}{2})^3\right.\right.$$
$$\left.\left.+4(\frac{n-3}{2})^2 + \frac{n-3}{2} + \frac{1}{30}\right)^{\frac{1}{6}}\right]$$
$$= +\infty.$$

Therefore, $\frac{u_p}{F(n)}$ goes to zero with n, and from the expansion of $\ln(1+x)$ when $x \ll 1$, we can write

$$\lim_{n\to\infty}\frac{n-1}{2}\ln\left(1+\frac{u_p}{F(n)}\right) = \lim_{n\to\infty}\frac{u_p(n-1)}{2F(n)}$$
$$\geq \lim_{n\to\infty}\frac{u_p(n-1)}{2(n+25)}, \tag{116}$$

where in (116), we have used the fact that for $n \leq 10^{10}$, it can be verified that $n < F(n) < n+25$. The gap between C_G and constant amplitude signaling can be written as

$$\lim_{n\to\infty}\{C_G - \sup_{F_\mathbf{X}(\mathbf{x}):\|\mathbf{X}\|^2 = u_p} I(\mathbf{X};\mathbf{Y})\} \leq \lim_{n\to\infty}\frac{u_p}{2}\left(1 - \frac{n-1}{n+25}\right)$$
$$= \frac{13u_p}{n+25},$$

which completes the proof.

APPENDIX F
PROOF OF REMARK 3

We have

$$C(u_p, u_a) \leq C(\infty, u_a) = \frac{n}{2}\ln(1+\frac{u_a}{n}),$$

and

$$\lim_{u_a\to 0} C(u_p, u_a) \leq \frac{u_a}{2}. \tag{117}$$

The CDF $F_P^{**}(\rho) = (1 - \frac{u_a}{u_p})u(\rho) + \frac{u_a}{u_p}u(\rho - \sqrt{u_p})$ induces the following output pdf

$$f_V(v; F_P^{**}) = (1 - \frac{u_a}{u_p})K_n(v, 0) + \frac{u_a}{u_p}K_n(v, \sqrt{u_p})$$

$$= (1 - \frac{u_a}{u_p})\frac{e^{-\frac{(\sqrt[n]{nv})^2}{2}}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}-1}}$$

$$+ \frac{u_a}{u_p}e^{-\frac{(\sqrt[n]{nv})^2+u_p}{2}}\frac{I_{\frac{n}{2}-1}(\sqrt{u_p}\sqrt[n]{nv})}{(\sqrt{u_p}\sqrt[n]{nv})^{\frac{n}{2}-1}}$$

$$= (1 - \frac{u_a}{u_p})\frac{e^{-\frac{(\sqrt[n]{nv})^2}{2}}}{\Gamma(\frac{n}{2})2^{\frac{n}{2}-1}}\left[1 + \frac{u_a}{u_p - u_a}\right.$$

$$\left.\times \frac{e^{-\frac{u_p}{2}}\Gamma(\frac{n}{2})2^{\frac{n}{2}-1}I_{\frac{n}{2}-1}(\sqrt{u_p}\sqrt[n]{nv})}{(\sqrt{u_p}\sqrt[n]{nv})^{\frac{n}{2}-1}}\right].$$

When u_a is small, the entropy of V is given by (121) on the next page.

The six terms in (120) are obtained by multiplying the terms in the brackets of (119) in order. In (121), we have neglected the last higher order term in (120) and have used the approximation $\ln(1 - x) \approx -x$ when $x \ll 1$. Therefore,

$$\lim_{u_a \to 0} h(V; F_P^{**}) + \sum_{i=1}^{n-2} \ln \alpha_i + (1 - \frac{n}{2})\ln 2\pi - \frac{n}{2} = \frac{u_a}{2}. \quad (122)$$

(122) and (117) show the asymptotic optimality of the distribution in (46).

APPENDIX G
PROOF OF REMARK 4

Since $\sqrt{z}$ is holomorphic on the complex plane excluding the non-positive real line (i.e., the domain where the principal branch of the complex logarithm function is holomorphic), $g(\sqrt{x})$ has the following power series expansion about $\epsilon > 0$

$$g(\sqrt{x}) = \sum_{m=0}^{\infty} g_m(x - \epsilon)^m = \sum_{m=0}^{\infty} \tilde{g}_m x^m, \quad (123)$$

where its interval of convergence is $(0, \infty)$. Assuming infinite number of mass points, with the constraint in (47), (104) changes to

$$\tilde{h}_V(\rho; F_{P^*}) = h(V; F_{P^*}) + \lambda(g(\rho) - u_a) \ , \ \ \forall \rho \in \mathbb{R} \quad (124)$$

or equivalently

$$-\int_0^{\infty} K_n(v, \rho)\ln f_V(v; F_P^*)dv = \lambda g(\rho) + h(V; F_P^*) - \lambda u_a,$$

$$\forall \rho \in \mathbb{R}. \quad (125)$$

Multiplying both sides of (125) by $\rho^{n-1}e^{-s\rho^2}$ ($s \geq 0$) and integrating with respect to ρ gives

$$-\int_0^{\infty} \ln f_V(v; F_P^*)\frac{e^{-\frac{s}{2s+1}(\sqrt[n]{nv})^2}}{(\sqrt{2s+1})^n}dv$$

$$= \int_0^{\infty} [\lambda g(\rho) + h(V; F_P^*) - \lambda u_a]\rho^{n-1}e^{-s\rho^2}d\rho,$$

where we have used the transform in (110). By a change of variables as $v = \frac{t^{\frac{n}{2}}}{n}$ and $x = \rho^2$, we have

$$-\int_0^{\infty} \ln f_V(\frac{t^{\frac{n}{2}}}{n}; F_P^*)t^{\frac{n}{2}-1}\frac{e^{-\frac{s}{2s+1}t}}{(\sqrt{2s+1})^n}dt$$

$$= \int_0^{\infty} [\lambda g(\sqrt{x}) + h(V; F_P^*) - \lambda u_a]x^{\frac{n}{2}-1}e^{-sx}dx. \quad (126)$$

By substituting (123) in (126), we get

$$-\int_0^{\infty} \ln f_V(\frac{t^{\frac{n}{2}}}{n}; F_P^*)t^{\frac{n}{2}-1}\frac{e^{-\frac{s}{2s+1}t}}{(\sqrt{2s+1})^n}dt$$

$$= \sum_{m=1}^{\infty} \frac{\tilde{g}_m\Gamma(\frac{n}{2}+m)}{s^{\frac{n}{2}+m}} + \frac{[h(V; F_P^*) - \lambda u_a + \lambda\tilde{g}_0]\Gamma(\frac{n}{2})}{s^{\frac{n}{2}}}.$$

Taking the inverse transform gives the unique solution as

$$\ln f_V(\frac{t^{\frac{n}{2}}}{n}; F_P^*) = \sum_{m=0}^{\infty} c_m t^m,$$

where the coefficients are obtained from the set of equations in (127) on the next page.

If there is no solution satisfying (127), (124) does not hold, which is the desired contradiction. However, in the case of having a solution for the coefficients in (127), we have

$$f_V(v; F_P^*) = e^{\sum_{m=0}^{\infty} c_m(\sqrt[n]{nv})^{2m}}. \quad (128)$$

In the case $c_m = 0$ ($m \geq 1$), f_V becomes a constant on the non-negative real line which cannot be a probability density function. The case $c_m = 0$ ($m \geq 2$) does not result in a legitimate pdf, either (see (106) and its following discussion.) For the remaining case of having at least one non-zero $c_m(m \geq 3)$, (128) leads to a contradiction as follows. Let $m^* = \max_m\{m|c_m \neq 0\}$. If $c_{m^*} > 0$, (128) is not integrable over the non-negative real line, hence, it is not a pdf. However, if $c_{m^*} < 0$, no $F_P(\rho)$ can result in f_V, since from (96),

$$f_V^{-1}(v; F_P) = O(e^{\frac{(\sqrt[n]{nv})^2}{2}}), \quad (129)$$

while the behavior of the inverse of (128) is different from (129) as v goes to infinity. Therefore, it is concluded that (128) cannot be resulted by any $F_P(\rho)$ due to its behavior at large v. This implies that the discrete nature of the magnitude of the optimal input distribution does not change when the average constraint is generalized to (47).

REFERENCES

[1] J. Smith, "The information capacity of amplitude and variance con-strained scalar gaussian channels," *Inform. Contr.*, vol. 18, pp. 203–219, 1971.

[2] S. Shamai 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, July 1995.

[3] I. Abou-Faycal, M. Trott, and S. Shamai, "The capacity of discrete-time memoryless rayleigh-fading channels," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1290–1300, May 2001.

$$\lim_{u_a \to 0} h(V; F_P^{**}) = \lim_{u_a \to 0} - \int_0^\infty f_V(v; F_P^{**}) \ln f_V(v; F_P^{**}) dv \tag{118}$$

$$= \lim_{u_a \to 0} \int_0^\infty \left\{ (1 - \frac{u_a}{u_p}) \frac{e^{-\frac{(\sqrt[n]{nv})^2}{2}}}{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}-1}} + \frac{u_a}{u_p} e^{-\frac{(\sqrt[n]{nv})^2 + u_p}{2}} \frac{I_{\frac{n}{2}-1}(\sqrt{u_p}\sqrt[n]{nv})}{(\sqrt{u_p}\sqrt[n]{nv})^{\frac{n}{2}-1}} \right\} \times$$

$$\left\{ \frac{(\sqrt[n]{nv})^2}{2} + \ln\left(\frac{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}-1}}{(1 - \frac{u_a}{u_p})}\right) - \frac{u_a}{u_p - u_a} \frac{e^{-\frac{u_p}{2}} \Gamma(\frac{n}{2}) 2^{\frac{n}{2}-1} I_{\frac{n}{2}-1}(\sqrt{u_p}\sqrt[n]{nv})}{(\sqrt{u_p}\sqrt[n]{nv})^{\frac{n}{2}-1}} \right\} dv \tag{119}$$

$$= \lim_{u_a \to 0} \frac{n}{2}(1 - \frac{u_a}{u_p}) + (1 - \frac{u_a}{u_p}) \ln\left(\frac{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}-1}}{(1 - \frac{u_a}{u_p})}\right) - \frac{u_a}{u_p} + \frac{u_a}{u_p}(\frac{n + u_p}{2})$$

$$+ \frac{u_a}{u_p} \ln\left(\frac{\Gamma(\frac{n}{2}) 2^{\frac{n}{2}-1}}{(1 - \frac{u_a}{u_p})}\right) - \frac{u_a^2}{u_p - u_a} \underbrace{\gamma(u_p)}_{\text{constant}} \tag{120}$$

$$= \frac{n}{2} + \frac{u_a}{2} + \ln\left(\Gamma(\frac{n}{2}) 2^{\frac{n}{2}-1}\right) \tag{121}$$

$$\left\{ \begin{array}{l} -\sum_{m=0}^\infty \frac{c_m \Gamma(\frac{n}{2}+m)(2s+1)^m}{s^{\frac{n}{2}+m}} \equiv \sum_{m=1}^\infty \frac{\tilde{g}_m \Gamma(\frac{n}{2}+m)}{s^{\frac{n}{2}+m}} + \frac{[h(V;F_P^*) - \lambda u_a + \lambda \tilde{g}_0]\Gamma(\frac{n}{2})}{s^{\frac{n}{2}}} \\ h(V; F_P^*) = -\int_0^\infty f_V(v; F_P^*) \ln f_V(v; F_P^*) dv \end{array} \right. . \tag{127}$$

---

[4] M. Katz and S. Shamai(Shitz), "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent awgn channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2257–2270, October 2004.

[5] M. Gursoy, H. Poor, and S. Verdú, "The noncoherent rician fading channel-parti: Structure of the capacity-achieving input," *IEEE Trans. Wireless Comm.*, vol. 4, no. 5, pp. 2193–2206, September 2005.

[6] A. Tchamkerten, "On the discreteness of capacity-achieving distributions," *IEEE Trans. Inf. Theory*, vol. 50, no. 11, pp. 2773–2778, November 2004.

[7] B. Mamandipoor, K. Moshkar, and A. Khandani, "Capacity-achieving distributions in Gaussian multiple access channel with peak power constraints," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, pp. 6080–6092, October 2014.

[8] M. A. Sedaghat, R. R. Mueller, and G. Fischer, "A novel single-rf transmitter for massive mimo," in *18th International ITG Workshop on Smart Antennas (WSA)*. VDE, 2014, pp. 1–8.

[9] R. Palanki, "On the capacity achieving distributions of some fading channels," in *Proc., 40th Annu. Allerton Conf. Communication, Control, and Computing*, Monticello, IL, Oct. 2002, pp. 337–346.

[10] S. Chan, S. Hranilovic, and F. Kschischang, "Capacity-achieving probability measure for conditionally gaussian channels with bounded inputs," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 2073–2088, June 2005.

[11] J. Sommerfeld, I. Bjelaković, and H. Boche, "On the boundedness of the support of optimal input measures for rayleigh fading channels," in *IEEE International Symposium on Information Theory (ISIT)*, July 2008, pp. 1208–1212.

[12] E. Telatar, "Capacity of multi-antenna gaussian channels," *European Trans. on Telecommunications*, vol. 10, no. 6, pp. 585 –595, 1999.

[13] A. E. Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2012.

[14] I. S. Gradshteyn and I. M. Ryzhik, *Table of integrals, series, and products*. Academic Press, 2007.

[15] J.G.Smith, *On the information capacity of peak and average power constrained Gaussian channels*. Ph.D. dissertation., Dep. Elec. Eng., Univ. of California, Berkeley, CA, Dec. 1969.

[16] M. Loéve, *Probability Theory*. New York:Van Nostrand, 1960.

[17] D. G. Luenberger, *Optimization by Vector Space Methods*. New York:Wiley, 1969.

[18] Y. Luke, "Inequalities for generalized hypergeometric functions," *Journal of Approximation Theory*, vol. 5, no. 1, pp. 41–65, January 1972.

[19] S. Lang, *Complex Analysis*. 4th ed. Springer, 1999.

[20] R. Bartle, *Elements of Real Analysis*. New York:Wiley, 1964.

[21] E. Karatsuba, "On the asymptotic representation of the Euler gamma function by Ramanujan," *J. Computat. Appl. Math*, vol. 135, pp. 225–240, 2001.

[22] H. Alzer, "On Ramanujans double-inequality for the gamma function," *Bull. Lond. Math. Soc.*, vol. 35, pp. 601–607, 2003.

[23] ——, "Sharp upper and lower bounds for the gamma function," *Royal Society of Edinburgh*, vol. 139A, p. 709718, 2009.

Borzoo Rassouli received his M.Sc. and Ph.D. degrees in communication systems engineering from University of Tehran, Iran, in 2012 and Imperial College, London, UK in 2016, respectively. He is currently a research associate in the Intelligent Systems and Networks Group at Imperial College. His research interests are in the general area of information theory, wireless communications, detection and estimation theory.

Bruno Clerckx received the M.S. and Ph.D. degrees in applied science from Universite catholique de Louvain, Belgium. He is now a Senior Lecturer (Associate Professor) at Imperial College London. He held visiting research positions at Stanford University and EURECOM and was with Samsung Electronics from 2006 to 2011. He actively contributed to 3GPP LTE/ LTE-A and IEEE802.16m. He is the author or coauthor of two books on MIMO wireless communications and networks and numerous research papers, standard contributions and patents. He received the Best Student Paper Award at the IEEE Symposium on Communications and Vehicular Technology in 2002 and several awards from Samsung in recognition of special achievements. Dr. Clerckx has served as an Editor for IEEE TRANSACTIONS ON COMMUNICATIONS and is currently an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.