Living at the Edge: A Large Deviations Approach to the Outage MIMO Capacity

Pavlos Kazakopoulos, Panayotis Mertikopoulos, Aris L. Moustakas and Giuseppe Caire

Abstract—A large deviations approach is introduced, which calculates the probability density and outage probability of the MIMO mutual information, and is valid for large antenna numbers N. In contrast to previous asymptotic methods that only focused on the distribution close to its *most probable* value, this methodology obtains the *full* distribution, including its non-Gaussian tails. The resulting distribution interpolates between the Gaussian approximation for rates R close its mean and the asymptotic distribution for large signal to noise ratios ρ **[1]**. For large enough N, this method provides the outage probability over the whole (R, ρ) parameter space. The presented analytic results agree very well with numerical simulations over a wide range of outage probabilities, even for small N. In addition, the outage probability thus obtained is more robust over a wide range of ρ and R than either the Gaussian or the large-ρ approximations, providing an attractive alternative in calculating the probability density of the MIMO mutual information. Interestingly, this method also yields the eigenvalue density constrained in the subset where the mutual information is fixed to R for given ρ. Quite remarkably, this eigenvalue density has the form of the Marčenko-Pastur distribution with square-root singularities.

Index Terms—Diversitymultiplexing tradeoff (DMT), Gaussian approximation, information capacity, large-system limit, multiple-input multiple-output (MIMO) channels.

I. INTRODUCTION

Considerable interest has arisen from the initial prediction [2], [3] that the use of multiple antennas in transmitting and receiving signals can lead to substantial gains in information throughput. To analyze the theoretical limits of such a MIMO (Multiple Input Multiple Output) system, it has been convenient to focus on the case of i.i.d. Gaussian noise and input. For the MIMO channel model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z} \tag{1}$$

with coherent detection and no channel state information at the transmitter [2], [3], the mutual information I_N for a given value of the channel matrix $\mathbf{H}$ takes the familiar form:

$$I_N = \log \det \left(\mathbf{I} + \rho \mathbf{H}^\dagger \mathbf{H} \right). \tag{2}$$

where "log" signifies the natural logarithm, ρ is the signal to noise ratio and $\mathbf{H}$ is the $M \times N$ channel matrix whose elements are independent $\mathcal{CN}(0, 1/N)$ random variables. This corresponds to the case of N transmitting and M receiving

antennas, which is captured by the ratio $\beta = M/N$. Without loss of generality we assume that $\beta \geq 1$; otherwise, if $\beta < 1$, we may simply replace ρ with $\rho_{\text{new}} = \rho\beta$ in (2) and interchange the roles of M and N.

If the channel matrix $\mathbf{H}$ varies in time according to a stationary ergodic process, and coding spans an arbitrarily large number of fading states, then the "ergodic" channel capacity is given by the mutual information expected value $\mathbb{E}[I_N]$ [3]. Initially, this quantity was calculated asymptotically for large N, with β remaining fixed and finite. In particular, in this case, $\mathbf{H}$ can be viewed as a large random matrix. Then, by applying ideas and methods from the theory of random matrices, it was shown in [4] that the value of the mutual information per antenna $I_N(\rho, \mathbf{H})/N$ "freezes" to a deterministic value in the large N limit, the so-called *ergodic average* $r_{\text{erg}}(\rho)$. Underlying this result is the fact that the very eigenvalue distribution of $\mathbf{H}^\dagger\mathbf{H}$ freezes to the celebrated Marčenko-Pastur distribution:

$$p(x) = \frac{\sqrt{(b-x)(x-a)}}{2\pi x} \tag{3}$$

where $a, b = (\sqrt{\beta} \pm 1)^2$ are the end-points of its support. Even though later the closed form solution of $\mathbb{E}[I_N]$ for general M, N was found [5], the asymptotic form of $r_{erg}(\rho)$ was particularly popular due to its simplicity and accuracy, even for small number of antennas.

Another more relevant regime is when the channel matrix is random, but varies in time much more slowly than the typical coding delay. In this case (usually referred to as the "quasi-static" fading channel) $\mathbf{H}$ can be considered as a random constant and the mutual information $I_N(\mathbf{H})$ is a random variable. In this regime, the relevant performance metric is the "rate versus outage probability" tradeoff [6], captured by the cumulative distribution function of $I_N(\mathbf{H})$. Various approaches [7]–[11] have shown that the mutual information $I_N(\mathbf{H})$ becomes asymptotically Gaussian for large N, with mean equal to the ergodic capacity $R_{\text{erg}} = Nr_{\text{erg}}(\rho)$ and a variance of order $\mathcal{O}(1)$ in N. This Gaussian variability of the mutual information is due to the fluctuations of the eigenvalues of the matrix around the most probable distribution described by the Marčenko-Pastur law. Since this Gaussian approximation is essentially a variation of the central limit theorem, it only applies within a small number of standard deviations away from the mean R_{erg}. As a result, this approximation fails to capture the tails of the distribution, e.g. the probability of the mutual information I_N falling below half its ergodic value $R_{\text{erg}}/2$, because this event only occurs $\mathcal{O}(N)$ standard deviations away from the mean.

Nevertheless, the tails of the distributions of the mutual information are important, because they correspond to regions with low outage probability, where one would want to operate a MIMO system. This is particularly important when, for large ρ, the slopes of the outage curves are large. The interplay between low outage and multiplexing gain was exemplified in the seminal paper [1] where the authors analyzed the asymptotics of the distribution of the mutual information in the limit of large ρ (keeping $R/\log\rho$ fixed). They found that the asymptotic form of the logarithm of the outage probability of the mutual information $P_{\text{out}}(R) \equiv \mathbb{P}(I_N(\mathbf{H}) \leq R)$ is a piecewise linear function of $R/\log\rho$, interpolating between the discrete set of values:

$$\log P_{\text{out}}(R_n) \sim -\log\rho\left(\frac{R_n}{\log\rho} - M\right)\left(\frac{R_n}{\log\rho} - N\right) \quad (4)$$

where $R_n = n\log\rho$ for integer $n \leq N \leq M$. When, in addition to ρ, N is also large, $\log P_{\text{out}}(R)$ in (4) becomes (to leading order) a continuous function of R/N. It should be pointed out that this approach generalizes the large N asymptotics discussed above, since it provides insight in the distribution of the mutual information quite far from its peak, which for large ρ (and large N) is situated at $I_N \approx N\log\rho$. More recently, in [12] the authors recast the DMT problem providing a formula to calculate $\log P_{out}$ as a function of R when R lies in each linear subsegment of (4). Nevertheless both approaches [1], [12] do not provide the offset to the leading, $O(\log\rho)$ behavior of (4). As a result, these approaches, while quite intuitive fail, often by a large margin, to provide an acceptable quantitative estimate of P_{out} unless $\log\rho$ is extremely large.

In the meantime, all variants [7]–[9] of the large N Gaussian approximation of the mutual information fail for large ρ. Specifically, they all predict that the outage probability is given asymptotically by:

$$\log P_{\text{out}}(R) \sim \frac{(\log\rho)^2}{2\log(1-\beta^{-1})}\left(\frac{R}{\log\rho} - N\right)^2 \quad (5)$$

where $\beta = M/N > 1$, an expression which is in striking disagreement with (4). Even though for $\beta = 1$ the asymptotic form of (4) is recovered within the Gaussian approximation [7], [9], the discrepancy for $\beta \neq 1$ indicates that the limits $N \to \infty$ and $\rho \to \infty$ cannot be naïvely interchanged. In the Gaussian approximation, one focuses on the most probable eigenvalue distribution, which converges vaguely to the Marčenko-Pastur distribution (3). However, as can be seen in (3), this distribution (almost surely) produces no eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$ close to zero when $\beta > 1$. Nevertheless, the analysis for large ρ focuses at the regime where the eigenvalues are of order $O(\rho^{-1})$. As a result, it is not surprising that the large-N Gaussian approximation of the mutual information distribution misses the correct behavior.

In summary, we have two methods, the large-N, fixed-ρ Gaussian approximation on the one hand and the large-ρ, fixed-N limit on the other, both having their own regions of validity, and both failing to produce quantitative results for the outage probability outside their respective regions. Thus, one still needs an approach that correctly describes the outage behavior of the mutual information distribution for arbitrary ρ and R.

In this paper, we introduce a large deviations approach to calculate the full asymptotic distribution of R. It is formally valid for large N, but works over the whole range of values of R and ρ. This method bridges the two regions of small/intermediate and large signal to noise ratios within a single framework and, in effect, it amounts to calculating the rate function of the logarithm of the average moment generating function of the mutual information. Our approach was first introduced in the context of random matrix theory by Dyson [13] and has been more recently applied in a variety of problems [14]–[17]. It is quite intuitive because it interprets the eigenvalues of $\mathbf{H}^\dagger\mathbf{H}$ as point charges on a line repelling each other logarithmically. This is the first time this approach has been applied in information theory and communications. As a byproduct of this approach, we obtain the most probable eigenvalue distribution constrained on the subset of channel matrices $\mathbf{H}^\dagger\mathbf{H}$ that have fixed total rate R and signal to noise ratio ρ. This is a generalized Marčenko-Pastur distribution that gives the constrained eigenvalue distribution for values of R even far from its ergodic value. It is worth pointing out that many of the results presented here could be set on a more formal mathematical footing using tools developed in [18]. However, we will follow the less formal but more intuitive approach developed by Dyson.

This generalized Marčenko-Pastur distribution can also be seen as the inverse of the so-called *Shannon transform* [19] in the following sense: while the Shannon transform produces the value of normalized mutual information I_N/N as a functional of the asymptotic eigenvalue distribution of $\mathbf{H}^\dagger\mathbf{H}$ (the Marčenko-Pastur distribution), the generalized Marčenko-Pastur distribution introduced here boils down to the asymptotic eigenvalue distribution of $\mathbf{H}^\dagger\mathbf{H}$ for a given value of the mutual information $R = Nr$, i.e., when $\mathbf{H}^\dagger\mathbf{H}$ is constrained on the subset defined by $r = I_N(\mathbf{H})/N$.

A. Outline

In the next section we will introduce the necessary mathematical methodology. In particular, Section II-A describes the mapping of the joint probability distribution of eigenvalues of the Wishart matrix to a Coulomb gas of charges with a continuous density (discussed in more detail in Appendix B) and the large-deviations analysis of the problem. Next, section II-B deals with the solution of the resulting integral equation that produces the most-likely eigenvalue distribution at the tails of the full distribution.

If one is not particularly interested in the details of our derivation, Section II may be skipped in favor of section III where we present our main results. Specifically, in Section III-A we rederive the Marčenko-Pastur distribution (that is, the most likely distribution without the mutual information constraint) to highlight the efficacy of our method. Subsequently, Sections III-B and III-C contain our results for the cases $\beta > 1$ and $\beta = 1$ respectively, while in Section III-D we show how to calculate the outage probability directly by means of the results of the previous sections. In Section IV

we analytically obtain previous results as limiting cases of this method, and also examine a number of different limiting cases. In Section V we provide numerical comparisons of our method to other approximations previously outlined and to Monte Carlo simulations.

The proofs of the properties of *tame* distributions (introduced in section II-A) are given in appendix A and we discuss Dyson's original construction of the Coulomb gas model in appendix B. Appendices C and D have been reserved for the exposition of some technical issues that cropped up during our calculations. Finally, Appendix E discusses higher order $O(1/N)$ corrections to our model and comparisons with Monte Carlo simulations.

II. METHODOLOGY

Our approach can roughly be divided in two main parts. First, in section II-A we reduce the original problem of finding the probability distribution of the mutual information to harvesting the minimum energy of a gas of charged particles (among other things we show here that the minimum energy configuration is unique). Then, in section II-B, we will solve the integral equation that comes up and actually obtain the minimum energy configuration of the charges.

A. Mapping the Problem to a Coulomb Gas

We begin by establishing the mathematical methodology, treading on the elegant footsteps of [15], [20]. Our overall aim will be to calculate the probability distribution of the mutual information (2), which can be written in terms of the eigenvalues λ_k of the Wishart matrix $\mathbf{H}^\dagger \mathbf{H}$ as:

$$I_N(\boldsymbol{\lambda}) = \sum_{k=1}^{N} \log\left(1 + \rho\lambda_k\right) \qquad (6)$$

Note that the aforementioned probability distribution of the mutual information thus depends on the joint probability distribution function of the eigenvalues $\lambda_1 \ldots \lambda_N$ of $\mathbf{H}^\dagger \mathbf{H}$. In its turn, this distribution takes the well-known form:

$$P_{\boldsymbol{\lambda}}(\lambda_1 \ldots \lambda_N) = A_N \Delta(\boldsymbol{\lambda})^2 \prod_{k=1}^{N} \lambda_k^{M-N} e^{-N\lambda_k} \qquad (7)$$

$$= A_N e^{-N^2 E(\boldsymbol{\lambda})} \qquad (8)$$

where A_N is a normalization constant and $\Delta(\boldsymbol{\lambda}) = \prod_{i>j}(\lambda_i - \lambda_j)$ is the Vandermonde determinant of the eigenvalues λ_k. The exponent $E(\boldsymbol{\lambda})$ is an energy function of the eigenvalues $\{\lambda_i\}$ that will become very useful later:

$$E(\boldsymbol{\lambda}) = \frac{1}{N} \sum_{k} \left(\lambda_k - (\beta - 1)\log \lambda_k\right) \qquad (9)$$
$$+ \frac{2}{N^2} \sum_{j>k} \log |\lambda_j - \lambda_k|$$

Note that the normalization we have chosen is such that $E(\boldsymbol{\lambda})$ corresponds roughly to the energy per eigenvalue.

The cumulative probability distribution (CDF) of the normalized mutual information I_N/N can then be written as a ratio of two volumes in $\boldsymbol{\lambda}$-space:

$$F_N(r) = \mathbb{P}(I_N/N \le r) = \frac{V_r}{V_{\text{tot}}} \qquad (10)$$
$$= \int P_{\boldsymbol{\lambda}}(\boldsymbol{\lambda})\,\Theta(r - I_N/N)\,d\boldsymbol{\lambda}$$

where I_N is given by (6), $\Theta(x)$ is the Heaviside step function ($\Theta(x) = 1$ if $x > 0$ and $\Theta(x) = 0$ if $x < 0$) and the integrals are taken with respect to the ordinary N-dimensional Lebesgue measure $d\boldsymbol{\lambda} = \prod_i d\lambda_i$. The above CDF is by definition the outage probability, i.e. the probability that the normalized mutual information falls below r. Its corresponding probability density (PDF) can be obtained from (10) by taking the derivative with respect to r [21]:

$$P_N(r) = F_N'(r) = \int P_{\boldsymbol{\lambda}}(\boldsymbol{\lambda})\,\delta(r - I_N/N)\,d\boldsymbol{\lambda} \qquad (11)$$

where we have used the fact that the (distributional) derivative of the step function is the Dirac δ-function: $\Theta'(x) = \delta(x)$.

Our primary goal will be to use (11) in order to obtain an analytic expression for the probability distribution function of the mutual information I_N. However, in general there is no standard way to evaluate integrals like V_r (except for some special cases [22]). Nevertheless, in the large-N limit it is possible to analyze such integrals in a systematic way. This so-called Coulomb-gas approach [23] is based on the intuitive idea to interpret the eigenvalues λ as the positions of N positive unit charges located on a line, a picture first proposed by Dyson [13]. Within this interpretation, the last term in the exponent $E(\boldsymbol{\lambda})$ in (9) corresponds to the logarithmic repulsion energy, while the first term is the potential due to a constant field and the second term is the repulsion of a point charge located at the origin.[1]

Now, it is instructive to look at the form of $E(\boldsymbol{\lambda})$ to get an intuitive understanding of the minimum energy configuration of $\boldsymbol{\lambda}$ in the absence of the constraint $I_N/N = r$. As discussed above, the first two terms in $E(\boldsymbol{\lambda})$ correspond to the external forces acting on the charges, while the last term represents the repulsion between charges. In the absence of the charge repulsion the minimum energy configuration will correspond to all charges settling at the minimum of the external potential, i.e. $\lambda_k = \beta - 1$ for all $k = 1, \ldots, N$. However, the repulsion between charges will make them move away from that point but still, from simple electrostatics considerations, the external forces will not allow this repulsion to carry charges too far away from the minimum. As a result, we expect that at the minimum of $E(\boldsymbol{\lambda})$ all charges will be concentrated in the neighborhood of $\beta - 1$. As the number of charges increases, it will make sense, at least for configurations with energy $E(\boldsymbol{\lambda})$ close to the minimum, to expect that the charge distribution will be approximately a *continuous* distribution. As a result, all sums over $\boldsymbol{\lambda}$ in $E(\boldsymbol{\lambda})$ may be replaced by integrals, and we expect that this will also be true in the presence of constraints as in (11).

---

[1]Note that these are simply the potentials that one obtains in classical two-dimensional electrostatics.

To make this continuum limit more precise, one begins by conditioning the probability law $\mathbb{P}$ of the eigenvalues of the Wishart matrix $\mathbf{H}^\dagger\mathbf{H}$ on the set $J_r = \{\lambda : I_N(\lambda)/N = r\}$, i.e. by considering the conditional probability law $\mathbb{P}(\cdot|I_N/N = r)$ and the corresponding PDF. As $N \to \infty$, large deviations theory suggests that this density function will be sharply concentrated around its most probable value, i.e. the minimum of the energy functional (9). Then, according to Dyson, this minimum can be asymptotically recovered by looking at the minimum of the *continuous* version of (9):

Conjecture 1 (Coulomb Gas Assumption). As $N \to \infty$, the empirical distribution of charges/eigenvalues under the rate constraint $I_N/N = r$ converges vaguely to an absolutely continuous density $p(x)$ which minimizes the continuous energy functional:

$$\mathcal{E}[p] = \int xp(x)\,dx - (\beta - 1)\int p(x)\log x\,dx \quad (12)$$
$$- \iint p(x)p(y)\log|x-y|\,dxdy$$

over the space of densities which satisfy the constraint $\int_0^\infty p(x)\log(1+\rho x)\,dx = r$. In other words, as $N \to \infty$, the total charge in any interval $I \subseteq \mathbb{R}$ will be given by:

$$\sigma(I) = \int_I p(x)dx, \quad (13)$$

with p as above.

This assumption is essentially identical to the one in Mehta's book [23] and has been extensively employed in the literature [13]–[15]. Unfortunately, despite its simple and intuitive nature, this assumption has resisted most attempts at a rigorous proof, thereby giving birth to different approaches, such as the one in [18]. Nevertheless, the results obtained there are in agreement with the ones obtained with the help of the Coulomb Gas assumption and, hence, we feel that our posit here is rather mild (see also appendix B for a more detailed discussion).

At any rate, to make proper use of the energy functional $\mathcal{E}$ (12) we must first make sure that it remains finite over a reasonably large class of densities $p(x)$. This leads us to the concept of "tameness":

Definition 2. An integrable function $p : \mathbb{R}_+ \to \mathbb{R}$ will be called ε-*tame* when:

(i) the "absolute mean" of p is finite:

$$\int_0^\infty x|p(x)|\,dx < \infty; \quad (14)$$

(ii) there exists some $\varepsilon > 0$ such that p is $L^{1+\varepsilon}$-integrable, i.e.

$$\int_0^\infty |p(x)|^{1+\varepsilon}\,dx < \infty. \quad (15)$$

Remark 2.1. The phrasing of condition (i) simply reflects our interest in tame functions $p \equiv p_X$ that are probability densities of random variables X with values in $\mathbb{R}_+$. In that case, condition (i) simply states that X has finite mean:

$$\mathbb{E}[X] = \int_0^\infty xp(x)\,dx < \infty. \quad (14')$$

Remark 2.2. Condition (ii) will be crucial to our analysis. At first, it might appear as a mere technical necessity (see e.g. section II-B and appendix C) but, in fact, it has a very deep physical interpretation: a probability density with finite mean might still fail to have finite energy, making it inadmissible on physical grounds. Condition (ii) ensures that $\mathcal{E}[p]$ will be finite (see lemma 3 below).

Remark 2.3. When it is not necessary to make explicit mention of the exponent ε, we will simply say that p is *tame*. Similarly, an absolutely continuous (signed) measure σ on $\mathbb{R}_+$ will be called *tame* when its Lebesgue derivative $p(x) = \frac{d\sigma(x)}{dx}$ is tame. Given this equivalence between continuous measures and Lebesgue derivatives, we will use the two terms interchangeably.

Going back to the energy functional $\mathcal{E}$ of (12), we can see that condition (i) guarantees that the first term in (12) is finite, while (ii) bounds the second and third terms. This is captured in the following:

Lemma 3 (Finiteness and Continuity of $\mathcal{E}$). *Let Ω be the space of tame functions on $\mathbb{R}_+$ and let $\mathcal{E}$ be defined as in (12). Then, $\mathcal{E}[p] < \infty$ for all $p \in \Omega$ and the restriction of $\mathcal{E}$ to any subspace of $L^{1+\varepsilon}$-integrable functions with finite mean is continuous (in the $L^{1+\varepsilon}$ norm). In other words, tame densities have finite energy and tame variations in density induce small variations in energy.*

We prove this lemma in Appendix A where we also give some background information on the L^r norms. For now, it will be more useful to express the probability density $P_N(r)$ as the ratio:

$$P_N(r) = \frac{\mathcal{Z}_r}{\mathcal{Z}} \quad (16)$$

where, in accordance with (8), (11) and (12), $\mathcal{Z}_r$ and $\mathcal{Z}$ are the (un-normalized) *partition functions*:[2]

$$\mathcal{Z}_r = \int_{\mathcal{X}_r} \mathcal{D}p\; e^{-N^2\mathcal{E}[p]} \quad (17)$$

$$\mathcal{Z} = \int_{\mathcal{X}} \mathcal{D}p\; e^{-N^2\mathcal{E}[p]} \quad (18)$$

and $\mathcal{D}p$ denotes the path-integral measure over the domains of tame densities $\mathcal{X}, \mathcal{X}_r \subseteq \Omega$:

$$\mathcal{X} = \left\{ p \in \Omega : p \geq 0 \text{ and } \int p(x)\,dx = 1 \right\} \quad (19)$$

$$\mathcal{X}_r = \left\{ p \in \mathcal{X} : \int p(x)\log(1+\rho x)\,dx = r \right\}. \quad (20)$$

Of course, from a mathematical point of view, constructing a measure $\mathcal{D}p$ over the infinite-dimensional space of functions is an intricate process which is far from trivial. Path integrals were first introduced by R. Feynman [24] in physics and have been used there extensively over the last 70 years. We prefer not to introduce them formally but, rather, to follow a more intuitive approach instead, in Appendix B.

---

[2]It is worth pointing out that the correction to the term $N^2\mathcal{E}[p]$ in the exponent is $\mathcal{O}(1)$ (see appendix B for more details). Also a nice analysis of the mapping from the $\boldsymbol{\lambda}$ integrals to path integrals over p can also be found in [20].

With all these considerations taken into account, we may take the large N limit and write:

$$\lim_{N \to \infty} \frac{1}{N^2} \log P_N(r) = \lim_{N \to \infty} \frac{1}{N^2} \left(\log \mathcal{Z}_r - \log \mathcal{Z} \right) \quad (21)$$

and, by invoking Varadhan's lemma [25], we obtain:

$$\lim_{N \to \infty} \frac{1}{N^2} \log P_N(r) = \mathcal{E}_0 - \mathcal{E}_1(r) \quad (22)$$

or, equivalently:

$$P_N(r) \sim e^{-N^2 (\mathcal{E}_1(r) - \mathcal{E}_0)} \quad (23)$$

where

$$\mathcal{E}_0 = \inf_{p \in \mathcal{X}} \mathcal{E}[p] \quad (24)$$

$$\mathcal{E}_1(r) = \inf_{p \in \mathcal{X}_r} \mathcal{E}[p] \quad (25)$$

In other words, we have reduced the problem of determining the asymptotic behavior of $P_N(r)$ to finding the minimum of the convex functional $\mathcal{E}$ over the two convex domains $\mathcal{X}$ and $\mathcal{X}_r$. To that end, we have:

Lemma 4 (Convexity of $\mathcal{E}$). *Let $\mathcal{X} \subseteq \Omega$ be the set of tame probability measures: $\mathcal{X} = \left\{ p \in \Omega : p \geq 0 \text{ and } \int p(x)\,dx = 1 \right\}$. Then, $\mathcal{X}$ is a convex subset of the topological vector space Ω and $\mathcal{E}$ is (strictly) convex on $\mathcal{X}$.*

Again, we will postpone the proof of this lemma until appendix A. However, an immediate corollary is that there exists a unique charge density p which minimizes (24) and (25). To find this unique solution - and the corresponding (global) minima $\mathcal{E}_0, \mathcal{E}_1(r)$ - it turns out to be more convenient to work over the whole space of tame measures Ω and introduce Lagrange multipliers for the two domains $\mathcal{X}$ and $\mathcal{X}_r$. This leads to the Lagrangian functions:

$$\mathcal{L}_0[p, \nu, c] = \mathcal{E}[p] - c \left(\int_0^\infty p(x)\,dx - 1 \right) \\ - \int_0^\infty \nu(x) p(x)\,dx \quad (26)$$

$$\mathcal{L}_1[p, \nu, c, k] = \mathcal{L}_0[p, \nu, c] \\ - k \left(\int_0^\infty p(x) \log(1 + \rho x)\,dx - r \right) \quad (27)$$

from which we obtain $\mathcal{E}_0$ and $\mathcal{E}_1(r)$ by maximizing over the dual parameters ν (non-negativity constraint), c (normalization constraint) and k (mutual information constraint):

$$\mathcal{E}_0 = \sup_{\nu \geq 0;\, c} \inf_p \mathcal{L}_0[p, \nu, c] \quad (28)$$

$$\mathcal{E}_1(r) = \sup_{\nu \geq 0;\, c, k} \inf_p \mathcal{L}_1[p, \nu, c, k] \quad (29)$$

The convexity of $\mathcal{L}_0$, $\mathcal{L}_1$ over p ensures that it suffices to find a local minimum $p(x)$ for the corresponding Lagrangian $\mathcal{L}$, for fixed ν, c, k. Then, any value of k, c that satisfies the constraints of p will be unique [26]. It is also worth pointing out that the only difference between $\mathcal{E}_0$ and $\mathcal{E}_1$ above is that the former can be seen as the maximum over $\mathcal{L}_1[p, \nu, c, k]$ keeping $k = 0$; this relation will come in handy later, because

it allows us to work with $\mathcal{L}_1$ and at the very last step set $k = 0$ to obtain $\mathcal{E}_0$.

We are now left to find a local minimum of $\mathcal{L}_1$ and the easiest way to do this is by looking at its functional derivative w.r.t. p. Indeed, recall that the functional derivative of $\mathcal{L}_1$ at $p \in \mathcal{X}_r$ is the distribution $\delta \mathcal{L}_1[p, \nu, c, k]$ whose action on test functions $\phi \in \Omega$ is given by:[3]

$$\langle \delta \mathcal{L}_1[p], \phi \rangle = \left. \frac{d}{dt} \right|_{t=0} \mathcal{L}_1[p + t\phi]. \quad (30)$$

Note now that the expression $\mathcal{L}_1[p + t\phi]$ is well-defined for all $p \in \mathcal{X}_r$, $\phi \in \Omega$, thanks to lemma 3 so that, at least, it makes sense to study its behavior as $t \to 0$. In addition to that, our convexity result (lemma 4) simplifies things even more because, if $\delta \mathcal{L}_1[p] = 0$ for some $p \in \mathcal{X}_r$, it immediately follows that $\mathcal{L}_1$ will be attaining its global minimum at p.[4] Then, maximizing the result with respect to k and c simply corresponds to enforcing the normalization and mutual information constraints that appear in (26) and (27):

$$\int_0^\infty p(x)\,dx = 1 \quad (31)$$

$$\int_0^\infty p(x) \log(1 + \rho x)\,dx = r \quad (32)$$

Furthermore, we must also maximize with respect to ν, in order to ensure that $p(x)$ be non-negative in $\mathbb{R}_+$. This optimization constraint can be enforced by observing that $\nu(x) = 0$ when $p(x) > 0$ and vice-versa, as we shall see below.

As a result, once we manage to find a solution to the above optimization problem, we will have:

Proposition 5 (Uniqueness of Solution). *Assume that the tame probability measure p satisfies the stationarity condition:*

$$\delta \mathcal{L}[p] = 0 \quad (\text{resp. } \delta \mathcal{L}_1[p] = 0) \quad (33)$$

along with the constraint (31) (resp. (31), (32)). Then, p is the unique global minimum point of (24) (resp. (25)).

This proposition stems directly from the convexity of $\mathcal{E}$ and will be of considerable help to us in what follows because it ensures that any stationary point of $\mathcal{L}, \mathcal{L}_1$ which satisfies the relevant constraints will be the (unique) solution to our original minimization problem.

B. Solving the Integral Equation

Our task now will be to actually *find* the solution of (30), subject to the constraints (31), (32). The solution for $\mathcal{E}_0$ in (28) can then be obtained by relaxing the constraint (32) and setting $k = 0$ in the final result. To that end, a brief calculation

---

[3]Since Ω is a locally convex space, this is just another guise of the Gâteaux/Fréchet derivative.

[4]Indeed, note that the function $w(t) = \mathcal{L}_1[p + t(q - p)], t \in [0, 1]$ is strictly convex in $[0, 1]$ for any choice of p and q in $\mathcal{X}_r$. Thus, if there were some $q \in \mathcal{X}_r$ with $\mathcal{L}_1[q] < \mathcal{L}_1[p]$, we would have $w'(0) = 0$ (on account of (30)) but also $w(0) > w(1)$, a contradiction.

(see appendix C) for the functional derivative for the functional derivative $\delta\mathcal{L}_1[p]$ of (30) yields the integral equation:

$$2\int_0^\infty p(x')\log|x-x'|\,dx' \;=\; x-(\beta-1)\log x \tag{34}$$
$$-\;c-k\log(1+\rho x)-\nu(x).$$

The role of $\nu(x)$ in the above equation is to enforce the inequality constraint $p(x)\geq 0$ for all $x\geq 0$. It is well known [26] that $\nu(x)>0$ only when the probability density $p(x)$ vanishes, while when the probability density is positive, $\nu(x)$ has to be zero.

The solution of the integral equation involves the inversion of the integral operator in the left-hand-side of (34), which is no simple task, because the inversion process depends on the support $\mathrm{supp}(p)$ of the density $p(x)$ [27]. As discussed in the previous subsection (and with a fair amount of hindsight gained from the Coulomb gas analogy), we will be looking for compactly supported solutions that are continuous in $(0,\infty)$; in other words, we will be assuming that $\mathrm{supp}(p)=[a,b]$ where $0\leq a<b<\infty$.

There is one important issue that must be mentioned here: when the dimensions of the channel matrix attain the critical value $\beta=1$, we will see that p exhibits two different behaviors depending on the values of r and ρ in constraint (32). On one hand, we could have $a>0$ which, by continuity, introduces the constraint $p(a)=0$; on the other hand, we could also have solutions with $a=0$ (which impose no extra constraints because p is assumed continuous only on $(0,\infty)$). If the rate r is less than some critical value $r_c(\rho)$, it turns out that solutions with $a>0$ must be rejected because they attain negative values. In that case, we are led to solutions with $a=0$ which have no such problems; the converse happens when $r>r_c$, while when $r=r_c$ the two solutions coincide.

Having said that, we may return to (34), where we have $\nu(x)>0$ if and only if $p(x)=0$. By restricting x to lie in the interval $[a,b]$, we may henceforth ignore $\nu(x)$ altogether. Furthermore, to eliminate c for the moment, a differentiation of (34) with respect to x yields:

$$2\mathcal{P}\int_a^b\frac{p(x')}{x-x'}dx' = 1-\frac{\beta-1}{x}-\frac{k\rho}{1+\rho x}\equiv f(x) \tag{35}$$

where $\mathcal{P}$ denotes the Cauchy principal value of the integral.[5]

The above equation has a straightforward physical meaning: it represents a balance of forces at every location $a\leq x<b$, because the repulsion from all other charges of the distribution located at x' (the LHS expression) is equal to the external forces (RHS). For $\beta>1$, we intuitively expect that $p(x)$ must vanish at $x=0$ because in this case the force from the finite charge density located at $x=0$ (the second term of (35)) would be infinite. As a result, we intuitively expect that $a>0$ for all $\beta>1$; this expectation will be vindicated shortly.

Indeed, the solution of this integral equation for general $f(x)$ can be obtained using standard methods from the theory of integral equations [27], [28]. So as not to interrupt the presentation, we will postpone the details until appendix C

[5]The principle value appears because of the absolute value $|x'-x|$ in (34).

and will only give the final result here:

$$p(x) = \frac{\mathcal{P}\int_a^b\frac{\sqrt{(y-a)(b-y)}f(y)}{y-x}dy+C'}{2\pi^2\sqrt{(x-a)(b-x)}} \tag{36}$$
$$= \frac{-x-\frac{k\sqrt{(1+a\rho)(1+b\rho)}}{1+\rho x}-\frac{(\beta-1)\sqrt{ab}}{x}+C}{2\pi\sqrt{(x-a)(b-x)}}$$

where C,C' are unknown constants to be determined by the condition $p(b)=0$.

As we explain in Appendix C, this formula is valid only when the function f is itself L^η-integrable for some $\eta>1$. This is always true if $\beta=1$, because the singular term proportional to $(\beta-1)$ is not present in the LHS of (35). However, as we have already mentioned, the case $\beta=1$ has its own set of subtleties, analyzed at length in section III-C. In particular, we obtain two different solutions depending on whether the support of p extends to 0 or not (imposing the constraints $a=0$ or $p(a)=0$ respectively), but only one of them is physically admissible (i.e. is a tame probability measure lying in the rate-constrained domain $\mathcal{X}_r$).

On the other hand, this dichotomy ceases to exist when $\beta>1$. Indeed, if $\beta>1$ and $a=0$, the LHS of (35) is no longer integrable. However, the RHS of (35) *is* $L^{1+\varepsilon}$-integrable whenever p is itself ε-tame, on account of the properties of the finite Hilbert transform [27] (see also appendix C). We thus conclude that any solution to (35) whose support extends to 0 cannot be tame and will thus have to be rejected. As a result, the support of p for $\beta>1$ has to be bounded away from 0, thus leading to the constraint $p(a)=0$ and proving our intuitive expectation above.

So, starting with the general case $a,b>0$, we find that the constraint of continuity requires that the distribution $p(x)$ vanish at the endpoints a,b of its support. The condition $p(b)=0$ determines the value of C in (36) resulting in the following form for $p(x)$:

$$p(x)=\frac{\sqrt{b-x}}{2\pi\sqrt{x-a}}\left(1-\frac{k\rho}{(1+\rho x)}\sqrt{\frac{1+a\rho}{1+b\rho}}-\frac{\beta-1}{x}\sqrt{\frac{a}{b}}\right) \tag{37}$$

The additional condition $p(a)=0$ (when $a>0$) results to

$$p(x) = \frac{1}{2\pi}\frac{\sqrt{(b-x)(x-a)}}{x(1+\rho x)}\left(\rho x+\frac{\beta-1}{\sqrt{ab}}\right) \tag{38}$$

with the value of a determined (as a function of b and k) by the equation:

$$\frac{k\rho}{\sqrt{(1+\rho a)(1+\rho b)}}+\frac{\beta-1}{\sqrt{ab}}=1. \tag{39}$$

Demanding that p be properly normalized as in (31), imposes the constraint:

$$\int_a^b p(x)dx = \frac{a+b-2k-2(\beta-1)}{4} \tag{40}$$
$$+\frac{k}{2\sqrt{(1+a\rho)(1+b\rho)}}=1.$$

In Appendix D we show that (39) and (40) admit a unique solution a,b for any given k and, as a result, Proposition 5

guarantees the existence of a (necessarily unique) density $p(x)$ that minimizes (29).

Now, given the resulting solution $p(x)$ we can readily calculate the minimum energy $\mathcal{E}[p]$ itself:

$$
\begin{aligned}
\mathcal{E}[p] &= \int_a^b x p(x)\, dx - (\beta - 1) \int_a^b p(x) \log x\, dx \\
&\quad - \int_a^b \int_a^b p(x) p(y) \log |x - y|\, dy\, dx \\
&= \frac{1}{2} \int_a^b x p(x)\, dx - \frac{\beta - 1}{2} \int_a^b p(x) \log x\, dx \\
&\quad + \frac{k}{2} \int_a^b p(x) \log(1 + \rho x)\, dx + \frac{c}{2}
\end{aligned}
\tag{41}
$$

where in the second line we eliminated the double integral by substituting it from (34) [15]. As for the value of c itself, it can be determined by evaluating (34) at a fixed value of x, say $x = a$:

$$
\begin{aligned}
c &= a - (\beta - 1) \log a - k \log(1 + \rho a) \\
&\quad - 2 \int_a^b \log(x - a) p(x)\, dx
\end{aligned}
\tag{42}
$$

Inserting this in (41) then yields:

$$
\begin{aligned}
\mathcal{E}[p] &= \frac{1}{2} \int_a^b x p(x)\, dx - \frac{\beta - 1}{2} \int_a^b p(x) \log x\, dx \\
&\quad - \int_a^b p(x) \log(x - a)\, dx \\
&\quad + \frac{1}{2} \left(k \left(r - \log(1 + \rho a) \right) + a - (\beta - 1) \log a \right)
\end{aligned}
\tag{43}
$$

III. PROBABILITY DISTRIBUTIONS $P_N(r)$, $P_{\text{OUT}}(r)$

The central aim of the paper is to evaluate the probability density of the rate r for large N, namely $P_N(r)$ given by (23)

$$
P_N(r) \approx B_N e^{-N^2(\mathcal{E}_1(r) - \mathcal{E}_0)}
\tag{44}
$$

where B_N is a normalization constant, while $\mathcal{E}_1(r)$ (25) and $\mathcal{E}_0$ (24) are the most probable values of the energy evaluated with and without the mutual information constraint (32), respectively. In this section we will calculate these values and derive the corresponding eigenvalue probability densities $p(x)$ that minimize the energy functional $\mathcal{E}[p]$. In Section III-A, we will derive $\mathcal{E}_0$ and we will show how the corresponding density $p(x)$ is the Marčenko-Pastur Distribution. In Sections III-B and III-C we will calculate $\mathcal{E}_1(r)$ for the cases $\beta > 1$ and $\beta = 1$ respectively. Finally, in Section III-D we will show how one can calculate the outage probability $P_{out}(r)$.

A. Evaluation of $\mathcal{E}_0$

As mentioned above, it is instructive to first calculate the most probable distribution of eigenvalues without the mutual information constraint (32), which will end up being the well-known Marčenko-Pastur distribution. This can be immediately extracted from the analysis in Section II-B by setting $k = 0$. Solving for a, b in (39), (40) gives

$$
a = \left(\sqrt{\beta} - 1 \right)^2
\tag{45}
$$

$$
b = \left(\sqrt{\beta} + 1 \right)^2
$$

and (38) then takes the well-known form (3). [6]

We may also evaluate the energy $\mathcal{E}_0$ by setting $k = 0$ in (43). Thus we get:

$$
\begin{aligned}
\mathcal{E}_0 &= \frac{1}{2} \int_a^b x p(x)\, dx + \frac{1}{2} \left(a - (\beta - 1) \log a \right) \\
&\quad - \frac{\beta - 1}{2} \int_a^b p(x) \log x\, dx - \int_a^b p(x) \log(x - a)\, dx
\end{aligned}
\tag{46}
$$

and, after some algebra, we can rewrite the above expression in the closed form:

$$
\begin{aligned}
\mathcal{E}_0 &= \frac{\Delta^2}{32} + \frac{a}{2} - \log \Delta - \frac{\beta - 1}{2} \log(a\Delta) \\
&\quad - \frac{\Delta}{2} \left[G\left(0, \frac{a}{\Delta}\right) + \frac{\beta - 1}{2} G\left(\frac{a}{\Delta}, \frac{a}{\Delta}\right) \right]
\end{aligned}
\tag{47}
$$

where $\Delta \equiv b - a$ and the function $G(x, y)$ is given by [29]:

$$
\begin{aligned}
G(x, y) &= \frac{1}{\pi} \int_0^1 \sqrt{t(1 - t)} \frac{\log(t + x)}{t + y}\, dt \\
&= -2\sqrt{y(1 + y)} \log \left[\frac{\sqrt{x(1 + y)} + \sqrt{y(1 + x)}}{\sqrt{1 + y} + \sqrt{y}} \right] \\
&\quad + (1 + 2y) \log \left[\frac{\sqrt{1 + x} + \sqrt{x}}{2} \right] \\
&\quad - \frac{1}{2} \left(\sqrt{1 + x} - \sqrt{x} \right)^2
\end{aligned}
\tag{48}
$$

When $\beta = 1$, a, b in (45) take the values $b = 4$ and $a = 0$, and hence (46) becomes $\mathcal{E}_0 = 3/2$.

B. Evaluation of $\mathcal{E}_1(r)$: $\beta > 1$

We will now calculate $\mathcal{E}_1$ for the case $\beta > 1$. To do so, we need to evaluate the constants a, b, k as a function of r and ρ using (39), (40) and (32). The values of these constants will determine the density of eigenvalues constrained on the subset with fixed total rate $R = Nr$ in the large N limit. After inserting (38) into the last equation and integrating, (32) can be expressed explicitly as

$$
\begin{aligned}
r &= \int_a^b p(x) \log(1 + \rho x)\, dx \\
&= \log \Delta\rho + \frac{\Delta k \rho}{2\sqrt{(1 + \rho a)(1 + \rho b)}} G\left(\frac{1 + \rho a}{\Delta\rho}, \frac{1 + \rho a}{\Delta\rho}\right) \\
&\quad + \frac{\Delta}{2} \left(1 - \frac{k\rho}{\sqrt{(1 + \rho a)(1 + \rho b)}} \right) G\left(\frac{1 + \rho a}{\Delta\rho}, \frac{a}{\Delta}\right)
\end{aligned}
\tag{49}
$$

where $G(x, y)$ is given in (48).

Based on the arguments discussed in the previous section, it suffices to show that there exists a distribution $p(x)$ in the form of (38) satisfying the constraints (31), (32). This corresponds to finding values of a, b, and k that satisfy (39), (40) and (49), while at the same time maintaining $p(x) \geq 0$ for all $x \in [a, b]$. If such a solution exists, then according to Theorem 5 it will be unique.

---

[6]Note that when $\beta = 1$, the lower endpoint vanishes ($a = 0$) and a square-root (integrable) singularity appears in $p(x)$ in (3).

In Appendix D we show that equations (39) and (40) admit a unique solution for any k. We therefore only need to show that (49) has a solution in k for any $r > 0$.

It suffices to show that the function defined solely as a function of k by the right-hand-side of (49) (with a and b expressed in terms of k) takes all values in $(0, \infty)$. Hence by continuity it will attain the value r for all positive rates $r > 0$. We first see that as $k \to -\infty$ the solution of (39), (40) is $a \approx (\sqrt{\beta} - 1)^2/(\rho|k|)$ and $b \approx (\sqrt{\beta} + 1)^2/(\rho|k|)$; then, inserting these solutions into (49), we see that it may be written in leading order as $r \approx \beta/|k|$. On the other hand, for $k \to \infty$ (39), (40) give $a \approx \sqrt{k + \beta - \rho^{-1}/2} - 1$ and $b \approx \sqrt{k + \beta - \rho^{-1}/2} + 1$, resulting to $r \approx \log k\rho$. This shows that the corresponding solution $p(x; r)$ is the unique minimizing distribution of $\mathcal{E}$ in $\mathcal{X}_r$.

In Fig. 1 we compare this distribution with the corresponding empirical probability distribution function obtained by numerical simulations. We see that the agreement is quite remarkable, indicating a quick convergence to the asymptotic distribution function of the eigenvalues constrained at the tails of the distribution of the mutual information. Furthermore, to get a feeling for the dependence of the eigenvalue distributions in terms of their parameters, in Fig. 2 we plot a few representative examples.

We may now calculate the value of $\mathcal{E}_1$. Inserting $p(x)$ from (38) into (43) and integrating finally gives us:

$$
\begin{aligned}
\mathcal{E}_1 &= \frac{\Delta^2}{32} + \frac{a}{2} - \log \Delta - \frac{\beta - 1}{2} \log(a\Delta) \quad (50) \\
&+ \frac{k}{2}\left(r - \log(1 + \rho a) - \frac{(\sqrt{1 + \rho b} - \sqrt{1 + \rho a})^2}{4\rho\sqrt{(1 + \rho a)(1 + \rho b)}}\right) \\
&- \frac{\Delta k\rho}{2\sqrt{(1 + \rho a)(1 + \rho b)}} \\
&\cdot \left[G\left(0, \frac{1 + \rho a}{\Delta\rho}\right) + \frac{\beta - 1}{2}G\left(\frac{a}{\Delta}, \frac{1 + \rho a}{\Delta\rho}\right)\right] \\
&- \frac{\Delta}{2}\left(1 - \frac{k\rho}{\sqrt{(1 + \rho a)(1 + \rho b)}}\right) \\
&\cdot \left[G\left(0, \frac{a}{\Delta}\right) + \frac{\beta - 1}{2}G\left(\frac{a}{\Delta}, \frac{a}{\Delta}\right)\right]
\end{aligned}
$$

where $G(x, y)$ is given by (48). Plugging this together with $\mathcal{E}_0$ into (44) we obtain $P_N(r)$, up to the normalization constant.

C. Evaluation of $\mathcal{E}_1(r)$: $\beta = 1$

The case $\beta = 1$ deserves special attention. In this case the logarithmic repulsion from the δ-function density of eigenvalues at the origin in (12) and (34) is no longer present. As discussed in Section II.B, depending on the parameters r and ρ there are two distinct types of solutions, which we treat here separately.

1) Case $\beta = 1$ and $r > r_c(\rho)$: We start by attempting to solve the problem as in the $\beta > 1$ case, namely by looking for solutions of $0 < a < b$ for the distribution's support. It is straightforward to show that the conditions (39) and (31) yield
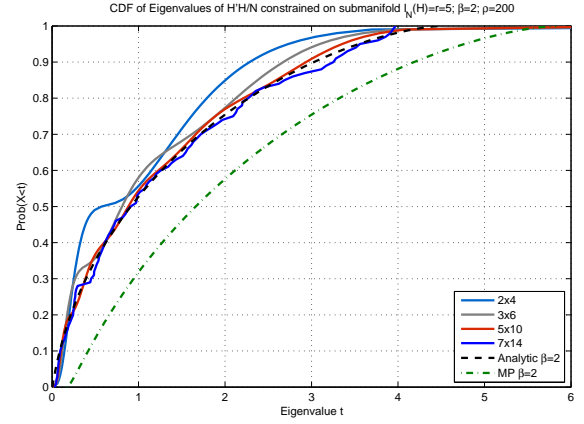

Fig. 1. Cumulative distribution function (CDF) of eigenvalues for a conventional and a generalized MP distribution with $\beta = 2$, $r = 5$ and $\rho = 200$. For $\beta = 2$ and $\rho = 100$, the value of the ergodic mutual information is $r_{erg} = 5.0014$. Thus, the generalized MP distribution with $\rho = 100$ would correspond to the conventional MP distribution above. Also plotted are the empirical CDFs for eigenvalues of $\mathbf{H}^\dagger\mathbf{H}/N$ conditioned on the subset $I_N \leq Nr$. It is remarkable that even for the not-too-large antenna array system 5×10 the empirical distribution converges to the analytic result.

the following values for a, b when $\beta = 1$:

$$
\begin{aligned}
a &= \left(\sqrt{k + 1} - 1\right)^2 - \rho^{-1} \\
b &= \left(\sqrt{k + 1} + 1\right)^2 - \rho^{-1}. \quad (51)
\end{aligned}
$$

As a result, the probality density function p becomes:

$$
p(x) = \frac{\rho}{2\pi}\frac{\sqrt{(b - x)(x - a)}}{1 + \rho x} \quad (52)
$$

The value of the parameter k can be obtained in a unique way from the mutual information condition, which now reads:

$$
r - \log \rho = (k + 1)\log(k + 1) - k\log k - 1. \quad (53)
$$

The monotonicity of the right-hand-side of this equation with respect to k implies a unique $k(r)$ satisfying (53) and hence a unique set of a,b in (51), guaranteeing uniqueness of (52).

In its turn, this can be used to evaluate the value of the outage exponent:

$$
\mathcal{E}_1 - \mathcal{E}_0 = \frac{k - 1}{2}(r - \log \rho) + k - \frac{1}{2} - \rho^{-1} - \frac{k\log k}{2}. \quad (54)
$$

From (51) we can see that this solution can only be valid for $k \geq k_c(z) \equiv \rho^{-1} + 2/\sqrt{\rho}$, or equivalently for $r > r_c(\rho)$ where

$$
\begin{aligned}
r_c(\rho) &\equiv \frac{1 + 2\sqrt{\rho}}{\rho}\log\left(1 + \frac{\rho}{1 + 2\rho}\right) \quad (55) \\
&+ 2\log\left(1 + \sqrt{\rho}\right) - 1 > r_{\text{erg}}
\end{aligned}
$$

The reason is that for $k < k_c(\rho)$ (or $r < r_c(\rho)$) the value of a becomes negative, which is unacceptable.

2) Case $\beta = 1$ and $r \leq r_c(\rho)$: In this case we can no longer treat a as a free variable. Instead, because $p(x) = 0$ for $x < 0$, the charge density becomes confined at the boundary $x = 0$. Thus, we need to look for solutions of (34) with $a = 0$, in which case the charge density has a square-root singularity
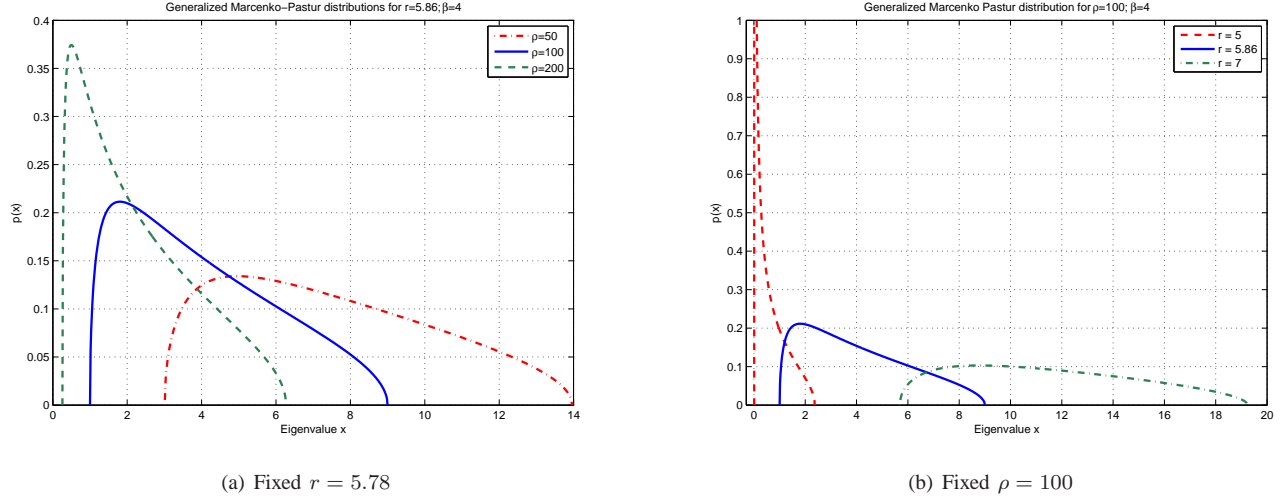
(a) Fixed $r = 5.78$

(b) Fixed $\rho = 100$

Fig. 2. Generalized MP distributions for $\beta = 4$ and different values of ρ and r. In (a) we plot the eigenvalue distributions for different values of ρ and fixed $r = 5.78$, which is the value of r_{erg} for the curve in the middle with $\rho = 100$. In (b) we plot the eigenvalue distributions for fixed $\rho = 100$ and different values r. We see that in the latter plot the distribution is more sensitive on r rather than ρ.

at $x = 0$ (instead of vanishing continuously). This is actually quite natural since we expect that, for $k = 0$ (or, equivalently, for $r = r_{\text{erg}}$), the charge distribution should take the form of the $\beta = 1$ Marčenko-Pastur density:

$$p(x) = \frac{\sqrt{4 - x}}{2\pi\sqrt{x}}. \tag{56}$$

Indeed, for general b, k, the distribution becomes:

$$p(x) = \frac{\sqrt{b - x}}{2\pi(1 + \rho x)\sqrt{x}}\left(\rho x + 1 - \frac{k\rho}{\sqrt{1 + \rho b}}\right), \tag{57}$$

and the normalization condition (31) implies

$$k = \frac{\frac{b}{2} - 2}{1 - \frac{1}{\sqrt{1 + \rho b}}} \tag{58}$$

It can easily be shown that the right-hand-side of (58) is increasing in b and, hence, (58) has a unique solution in b for all k.

In the last case ($a = 0$), the mutual information condition (32) can be integrated using (57) to give:

$$r = 2(k + 1)\log\frac{1 + \sqrt{1 + \rho b}}{2}$$
$$- \frac{1}{4\rho}\left(\sqrt{1 + \rho b} - 1\right)^2 - \frac{k}{2}\log\left(1 + \rho b\right). \tag{59}$$

We may use the same argument as in the previous subsection to show that this equation has at least one solution for any $0 < r < r_c(\rho)$. Indeed when $k = k_c$, the right-hand-side above takes the value of r_c. In contrast, when $k \to -\infty$, (58) gives $b \approx 4/(\rho|k|)$, in which case the right-hand-side of (59) becomes $\approx 1/|k|$. Thus all values between $(0, r_c(\rho))$ are taken when $k \in (-\infty, k_c(\rho))$. Hence by continuity it will attain the value $r \in (0, r_c)$.

After solving for b and k as a function of r and ρ, $\mathcal{E}_1$ can be calculated easily. Therefore, the exponent of the probability

distribution $P_N(r)$ becomes:

$$\mathcal{E}_1 - \mathcal{E}_0 = \frac{k}{2}\left(r - \frac{b}{4}\right) - \log\frac{b}{4} - k\log\frac{1 + \sqrt{1 + \rho b}}{2}$$
$$+ \frac{1}{32}(b - 4)\left(4\rho^{-1} + 3b + 12\right) \tag{60}$$

We should point out that just as the solution (57) is not valid for $r > r_c(\rho)$, the solution (57), which we found to be valid for $r > r_c(\rho)$ is not valid for $r < r_c(\rho)$. To see this, it is straightforward to show that in this case the constant term in the last parenthesis in (57) (namely $1 - k\rho/\sqrt{1 + \rho b}$) is negative. As a result, (57) cannot be valid for $k < k_c(\rho)$ because the charge density becomes negative at some point $x > 0$. As a result the solutions we found above are unique in their domains of validity. Interestingly there is a weak, third order discontinuity at the transition $r = r_c(\rho)$, in the sense that the first two derivatives of $\mathcal{E}_1(r)$ with respect to r evaluated at $r = r_c$ are continuous, while the third is discontinuous. This is analogous to the phase transition observed in [16].

D. Evaluation of the Outage Probability $P_{out}(r)$

In this section we will calculate the outage probability $P_{out}(r) = \mathbb{P}(I_N < Nr)$ from $\mathcal{E}_1(r)$. To do this we need to integrate $\exp\left[-N^2(\mathcal{E}_1(r) - \mathcal{E}_0)\right]$ over r. Generally it is impossible to evaluate this integral in closed form. Nevertheless, due to the presence of the factor N in the exponent, $P_N(r)$ falls rapidly away from its peak and thus we may use Watson's lemma [30] (a special case of Varadhan's lemma), to evaluate the asymptotic value of the integral. First, we will calculate the normalization factor of the distribution.

As we shall see in Section IV for r close to r_{erg}, $\mathcal{E}_1(r) - \mathcal{E}_0 \sim (r - r_{\text{erg}})^2/v_{\text{erg}}$, where v_{erg} is the ergodic variance (69) of the mutual information distribution. Therefore, we have

$$\int_0^\infty e^{-N^2(\mathcal{E}_1(r) - \mathcal{E}_0)}\, dr \approx \int_0^\infty e^{-\frac{N^2(r - r_{\text{erg}})^2}{2v_{\text{erg}}}}\, dr \approx \frac{\sqrt{2\pi v_{\text{erg}}}}{N} \tag{61}$$

which then gives

$$P_N(r) \approx \frac{N}{\sqrt{2\pi v_{\text{erg}}}} e^{-N^2(\mathcal{E}_1(r) - \mathcal{E}_0)} \qquad (62)$$

and fixes the normalization constant in (44). To calculate the outage probability $P_{\text{out}}(r) = \mathbb{P}(I_N < Nr)$ to leading order in N, we first note that for $r < r_{\text{erg}}$ ($r > r_{\text{erg}}$), $\mathcal{E}_1(r)$ is a decreasing (increasing) function of r. Therefore, to leading order, the behavior will be dominated by the value of the exponent at r. Using Watson's lemma once again we obtain the following expression for the outage probability:

$$P_{\text{out}}(r) \approx \frac{e^{-N^2\left[\mathcal{E}_1(r) - \mathcal{E}_0 - \frac{\mathcal{E}_1'(r)^2}{2\mathcal{E}_1''(r)}\right]} Q\left(\frac{N|\mathcal{E}_1'(r)|}{\sqrt{\mathcal{E}_1''(r)}}\right)}{\sqrt{\mathcal{E}_1''(r) v_{erg}}} \qquad (63)$$

when $r < r_{\text{erg}}$ and

$$P_{out}(r) \approx 1 - \frac{e^{-N^2\left[\mathcal{E}_1(r) - \mathcal{E}_0 - \frac{\mathcal{E}_1'(r)^2}{2\mathcal{E}_1''(r)}\right]} Q\left[\frac{N|\mathcal{E}_1'(r)|}{\sqrt{\mathcal{E}_1''(r)}}\right]}{\sqrt{\mathcal{E}_1''(r) v_{erg}}} \qquad (64)$$

when $r > r_{\text{erg}}$. In the above, $\mathcal{E}_1'(r)$ and $\mathcal{E}_1''(r)$ are the first and second derivatives of $\mathcal{E}_1(r)$ with respect to r and $Q(x)$ is given by

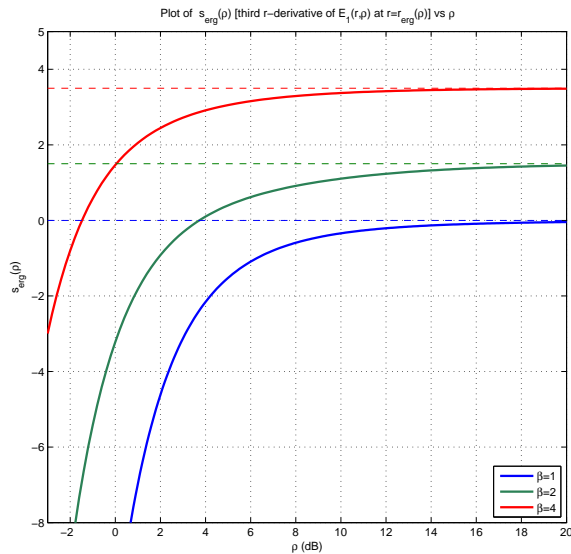$$Q(x) = \int_x^\infty \frac{dx}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \qquad (65)$$

Fig. 3. Dependence of $s_{erg} = \mathcal{E}_1'''(r_{erg})$ on ρ for different values of β. We see that for not too large ρ the behavior of s_{erg} quickly converges to the correct asymptotic limit (70), included here with dashed lines.

IV. ANALYSIS OF LIMITING CASES

We will now analyze the results of the previous section in specific limiting cases of the parameter space (ρ, r, β). We will thereby be able to connect with already existing results in specific regions, and also to describe the behavior of the probability density of $P_N(r)$ in other regions, which hitherto have defied asymptotic analysis.

A. Gaussian Region $r \approx r_{erg}(\rho)$

The most relevant limiting case is the Gaussian regime: after all, the Gaussian approximation, as well as the present approach assume that the number of antennas N is large. The difference is that our approach does not focus only in the region of $N|r - r_{\text{erg}}| = \mathcal{O}(1)$, where the Gaussian approximation should be valid. To reach that limit, we need to analyze the small k region of (49), (59) since, in the limit $k = 0$, both equations reduce to $r = r_{\text{erg}}(\rho)$, where the normalized ergodic mutual information r_{erg} is well known to be [4], [7], [31]:

$$r_{\text{erg}} = \log u + \beta \log\left[1 + \frac{\rho}{u}\right] - (1 - u^{-1}) \qquad (66)$$

with:

$$u = \frac{1}{2}\left(1 + \rho(\beta - 1) + \sqrt{(1 + \rho(\beta - 1))^2 + 4\rho}\right) \qquad (67)$$

By implicitly differentiating a, b, k with respect to r through the equations that define them, and expressing their values and the values of their derivatives at $r = r_{erg}$ we can obtain the following expansion

$$\mathcal{E}_1 - \mathcal{E}_0 = \frac{(r - r_{\text{erg}})^2}{2v_{\text{erg}}} + \frac{s_{erg}}{6}(r - r_{\text{erg}})^3 + \mathcal{O}\left((r - r_{\text{erg}})^4\right) \qquad (68)$$

where

$$v_{\text{erg}} = -\log\left[1 - \frac{(1 - u)^2}{\beta u^2}\right] \qquad (69)$$

coincides with the variance of the mutual information distribution as analyzed in [7], [9], and s_{erg} is the third *total* derivative of $\mathcal{E}_1$ with respect to r and evaluated at $r = r_{erg}(\rho)$. Without the cubic term, (68) is exactly the Gaussian limit of the mutual information distribution discussed in various papers in the past. This Gaussian limit is valid as long as the cubic (as well as all higher order) terms in the exponent of the probability are smaller than unity. Since this condition depends on s_{erg}, it is worth looking its behavior with ρ. In Fig. 3 we plot s_{erg} as a function of ρ. We see that it has a well-defined limit for large ρ. Specifically, it has the following asymptotic form

$$s_{erg}(\rho) \approx \begin{cases} -\frac{2}{\log(\rho)^3} & \beta = 1 \\ -\frac{1}{\beta(\beta-1)\log(1-\beta^{-1})^3} & \beta > 1 \end{cases} \qquad (70)$$

Also, for small $\rho \ll 1$ we can show that $s_{erg} \approx -c_\beta/\rho^3$, where $c_\beta > 0$ is a constant that depends on β. Thus the condition for validity of the Gaussian approximation is

$$|r - r_{erg}(\rho)| \ll \sqrt[3]{\frac{6}{|s_{erg}|}} N^{-2/3} \qquad (71)$$

We therefore see that the Gaussian approximation should not be valid for significant deviations from r_{erg}, e.g. $r = r_{erg}/2$. In contrast our large deviations (LD) approximation continues to be valid in that rate region as well.

B. Large ρ Approximation: $r < r_{erg}$

Next we analyze the behavior of the probability distribution of r in the large ρ limit, while keeping the ratio $r/\log \rho$ finite and less than 1.[7] Since in the large ρ limit $r_{erg} \sim \log \rho$, the region $q \leq 1$ with $\rho \gg 1$ corresponds to $k < 0$, equations (40), (39) will admit the following solutions for a, b:

$$a \quad \sim \quad \frac{(\beta - 1)^2}{4\rho(1-q)(\beta - q)} \tag{72}$$

$$b \quad \sim \quad 4q \tag{73}$$

where $q = r/\log \rho$ and we are assuming that $0 < q < 1$.

Now, note that the lower end of the spectrum has become of order $O(1/\rho)$, while the upper limit is still finite, just as expected. It is also interesting to calculate the proportion of eigenvalues that are in the neighborhood of $x = 1/\rho$ when $\rho \to \infty$. Indeed, by integrating the probability distribution $p(x)$ (38) from $a = \mathcal{O}(\rho^{-1})$ (72) to $L\rho^{-1}$ for some (arbitrarily) large L we get

$$\lim_{L \to \infty} \lim_{\rho \to \infty} \mathbb{P}(\rho x < L) = 1 - q \tag{74}$$

Thus, the proportion of "small" eigenvalues is simply $1 - q$, in agreement with [1]. Plugging (72), (73) into the equation for $\mathcal{E}_1$ then gives the expected result for the exponent:

$$\mathcal{E}_1 - \mathcal{E}_0 \sim \log \rho \left[(1-q)(\beta - q)\right] \tag{75}$$

which is exactly the diversity exponent (divided by N^2) of [1].

From the above, we see the difference between the two asymptotic analyses discussed above. In the previous section, the eigenvalue distribution did not deviate significantly from the most probable Marčenko-Pastur distribution, since k was assumed to be small. In contrast, here, k is finite, and in particular equal to $k = 2q - 1 - \beta$, In addition, a significant portion of the eigenvalues in this subset of fixed $r = q \log \rho$ is now to become very small, of order $1/\rho$.

In the above discussion, we see that generally the exponent $\mathcal{E}_1(r)$ is not only continuous, but also differentiable in r. This is in disagreement to the prediction by [1], [12] that when $\rho \to \infty$, the outage has a piecewise linear behavior. The length of these segments is $\Delta R \approx \log \rho$, or $\Delta r \approx \log \rho / N$. Thus for these segments to be pronounced we need

$$N \ll \log \rho \tag{76}$$

for large ρ. This provides a limit on the formal limitations of our large deviations (LD) approach. In particular, clearly the antenna number N has to be large, as in the Gaussian case. But, in contrast to the Gaussian approximation, there is no constraint here that the deviation of the rate from the ergodic rate has to be small, as in (71). Thus the *scale* of N at which the method should break down is given by $\log \rho$ for large ρ, rather than ρ itself. This is corroborated in the numerical results in the next section. Surprisingly, however, the analysis in this section shows that the *form* of the DMT exponent (4) is correctly predicted within the LD approach in (75).

[7]This is the region analyzed in the diversity-multiplexing trade-off [1].

C. Large ρ Approximation: $r > r_{erg}$

The regime of large ρ and fixed $q = r/\log \rho \leq 1$ is relevant in the analysis of the link-level outage probability. However, the opposite regime of $q > 1$ is also of interest in a cellular setting with many multi-antenna users receiving data in a TDMA fashion from a single multi-antenna base-station.[8] In this context, to analyze the system level throughput, it is the higher end of the probability distribution of the link-level mutual information that is important [8], [32]. Therefore, it is worthwhile to calculate the probability distribution of r for large ρ with $q > 1$.

Interestingly enough, the behavior here is quite different from the $q < 1$ case. Here $k \sim \rho^{q-1}$ and

$$a \quad \sim \quad \left(\sqrt{k + \beta} - 1\right)^2 \tag{77}$$

$$b \quad \sim \quad \left(\sqrt{k + \beta} + 1\right)^2 \tag{78}$$

resulting to

$$\mathcal{E}_1 - \mathcal{E}_0 \sim \rho^{q-1} = \frac{e^r}{\rho} \tag{79}$$

independent of β. The resulting probability distribution of r is

$$P(r) \sim e^{-N^2 e^r / \rho} \tag{80}$$

We see that when N is not too small, the probability of finding I_N significantly larger than its ergodic value is extremely small (in fact, doubly exponentially small in r). This is the manifestation of the fact that scheduling the best user in a MAC-layer in a multi-antenna setting does not seem to provide any clear advantage. Interestingly, in [8] the authors have the same conclusion, even though they assume a Gaussian distribution for I_N even for its tails. Here we see that the distribution of I_N goes to zero for $r > r_{erg}$ in a rate even faster than Gaussian, thereby making the above conclusion, to which they also reached even stronger.

This result has the following intuitive explanation. For large ρ and $r > r_{erg}$ all eigenvalues of the matrix $\mathbf{H}^\dagger \mathbf{H}$ will be large and the only constraint imposed upon them is (32). Thus, we may say that all of them are constrained by the condition $r \sim \log(1 + \rho \lambda_i) \sim \log \rho \lambda_i$ i.e. $\lambda_i \sim e^r / \rho$. In this limit, the exponent is roughly N times the sum of the eigenvalues.

D. Limit $r \to 0$

The final regime that is interesting to analyze is when $r \to 0$, independently of ρ. In this regime the solution of (49) (59) for small r is $r \sim \beta/|k|$ for $k \to -\infty$ and the corresponding values of a, b are:

$$a \quad \sim \quad \frac{r}{\rho \beta} \left(\sqrt{\beta} - 1\right)^2 \tag{81}$$

$$b \quad \sim \quad \frac{r}{\rho \beta} \left(\sqrt{\beta} + 1\right)^2 \tag{82}$$

resulting in:

$$\mathcal{E}_1 - \mathcal{E}_0 \sim -\beta \log \left[\frac{er}{\beta \rho}\right]. \tag{83}$$

[8]In that case a MAC-layer scheduler would be transmitting to the user with the best channel, for example.

where e is the Euler number. This means that the probability distribution $P_N(r)$ has a tail of the form

$$P(r) \sim \left(\frac{re}{\rho\beta}\right)^{MN} \tag{84}$$

The above behavior of $P_N(r)$ for small r is easy to understand: for r to be small, we need all matrix elements of the matrix $\mathbf{H}$ to be small. In fact, since $\mathbf{H}$ appears in a quadratic way in the mutual information equation (2) we need all MN elements of $\mathbf{H}$ to be less than $O(\sqrt{r/\rho})$. However, there are $2MN$ real degrees of freedom in the $M \times N$ complex matrix $\mathbf{H}$. Hence the allowed volume of space scales as $(r/\rho)^{MN}$ as above.

It should also be noted that the behavior $P(r) \sim \rho^{-MN}$ of the mutual information cumulative distribution function is precisely what is known as the "full diversity" of error probability, i.e., the SNR exponent of error probability for fixed but very small rate R while SNR ρ increases is ρ^{-MN}, which corresponds to the left extreme point of the Zheng-Tse exponent [1].

V. NUMERICAL SIMULATIONS

To test the applicability of this approach, we have performed a series of numerical simulations and have compared our large deviations (LD) approach to other popular approximations.

We start with the case of small rates r. In this limit the Gaussian approximation is guaranteed to give misleading results. For example, the Gaussian approximation predicts a finite outage probability at zero rate, while this is clearly wrong. The LD approximation, on the other hand, correctly predicts that the outage probability goes to zero at small r, as seen in (84). In Figs. 4 and 5 we plot the outage probability of the LD approach with the Gaussian and Monte Carlo simulations for low rates, small ρ and small square (2×2 and 3×3) antenna arrays. The comparison shows that while the Gaussian curves miss the correct outage, the LD curves remain close to the simulated ones, even for the 2×2 MIMO system. It is worthwhile to mention that the Gaussian outage probability is consistently greater than the correct (simulated) one. The reason for this can be traced to the fact that for all $\beta = 1$ and all values of ρ, the third derivative of the exponent $\mathcal{E}_1(r) - \mathcal{E}_0$ with respect to r evaluated at r_{erg}, i.e. $s_{erg}(\rho)$ in (68) is negative. Disturbing away from the peaks of the distribution we have

$$\log P_{out,Gaussian}(r) \approx -\frac{N^2(r - r_{erg})^2}{2v_{erg}} \tag{85}$$

while

$$\log P_{out}(r) \approx -\frac{N^2(r - r_{erg})^2}{2v_{erg}} - \frac{s_{erg}N^2(r - r_{erg})^3}{6} \tag{86}$$

We may thus conclude that when $r < r_{erg}$ and $s_{erg} < 0$ we should have $P_{out,Gaussian} > P_{out}$. From Fig. 3 we see that for increasing ρ, s_{erg} decreases in absolute size, which correctly predicts that the discrepancy between the Gaussian and the Monte-Carlo curves (and LD) decreases for larger ρ.

We have also analyzed the probability distribution for rates greater than the ergodic rate $r > r_{erg}$. Even though this region is not relevant for the outage probability evaluation, it is important in the analysis of the multiuser capacity for MIMO links in a multi-user setting with a greedy scheduler, such as a maximum rate scheduler. [8] In such a case, the multiuser diversity gain comes from the opportunity the scheduler has to schedule transmission to users when their fading rates are greater than their mean. Thus it is important to understand the tails of the distribution in this region. In Fig. 6 we obtained the complementary CDF (CCDF) of the mutual information, i.e. $1 - P_{out}(r)$, for a 3×3 setting. Here the probability of finding users with high rates falls faster than the Gaussian, especially in Fig. 6b for large ρ. We also find that the LD approximation follows the Monte Carlo simulations more accurately than the Gaussian curve, especially for lower outages. In this situation it is worth pointing out that the argument mentioned above regarding the sign of s_{erg} would make $1 - P_{out}(r)$ smaller in the Gaussian approximation compared to the correct result. We see that this only occurs for rates relatively close to the peak. In contrast, for rates greater than the critical rate $r_c(\rho)$ the behavior of the numerical and the LD outage probability changes markedly and they both become substantially smaller than the Gaussian curve. This is not surprising in view of the phase transition occurring at $r = r_c(\rho)$ as discussed in Section III-C2.

We next analyzed the outage probability as a function of the SNR. The outage has been analyzed in the large SNR limit for finite rates in [12], where they have dubbed this analysis as throughput reliability tradeoff (TRT). This model provides a piecewise linear function of the outage probability, which for completeness is provided below:

$$\begin{aligned}
\log_2 P_{out} &\approx c(k)R - g(k)\log_2 \rho \tag{87} \\
c(k) &= M + N - 2k - 1 \\
g(k) &= MN - k(k+1)
\end{aligned}$$

when ρ is large and $k\log_2 \rho < R < (k+1)\log_2 \rho$.

This piecewise linear behavior however is observable only at extremely high rates and SNRs, which may not necessarily be relevant for realistic MIMO systems. We analyzed the case of 3×6, 3×3 and 6×6 arrays in Figs. 7 and 8. In all three we have found that the LD approximation agrees with simulations over a wide region of rates r and SNR ρ. Characteristic is Fig. 7b, where the TRT curve is accurate in large SNR, the Gaussian is accurate in low SNR, but the LD curve is consistently closer to the correct outage. For the $N = M = 3$ case and extremely high SNRs and rates the piecewise linear behavior predicted by TRT starts becoming visible. Nevertheless, even in those high rates the TRT curve also fails to give quantitatively correct outage estimates and the LD curve is still closer to the correct outage.

It is sensible to point out that here the Gaussian outage probability is consistently less than the simulated and the LD values. In this case the argument made above for s_{erg} is reversed. As can be seen in Fig. 3 for $\beta = 2$ and large ρ the sign of s_{erg} is opposite, i.e. we have $s_{erg} > 0$ and hence indeed we should have $P_{out,Gaussian} < P_{out}$.

In Fig. 9, we plot the logarithm of the appropriately normalized probability density function (PDF) $P_N(r)$ as a function of the throughput r and we compare the result with the two

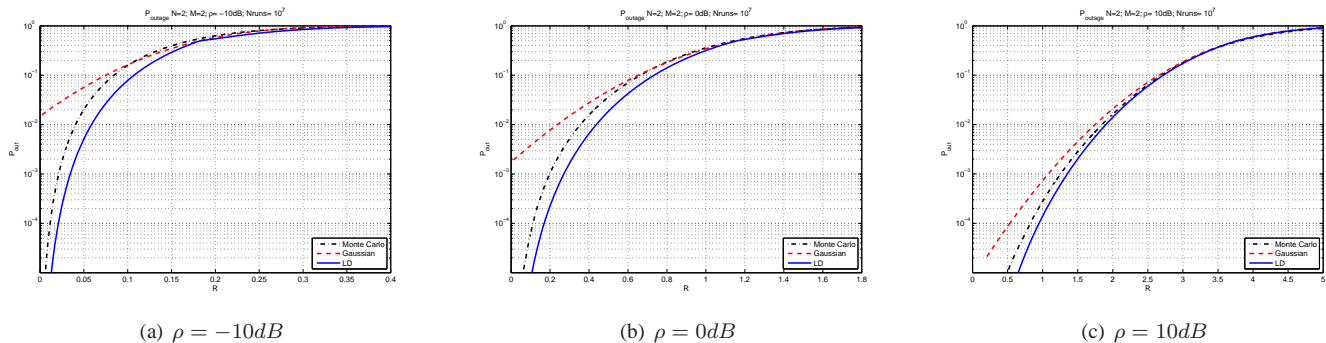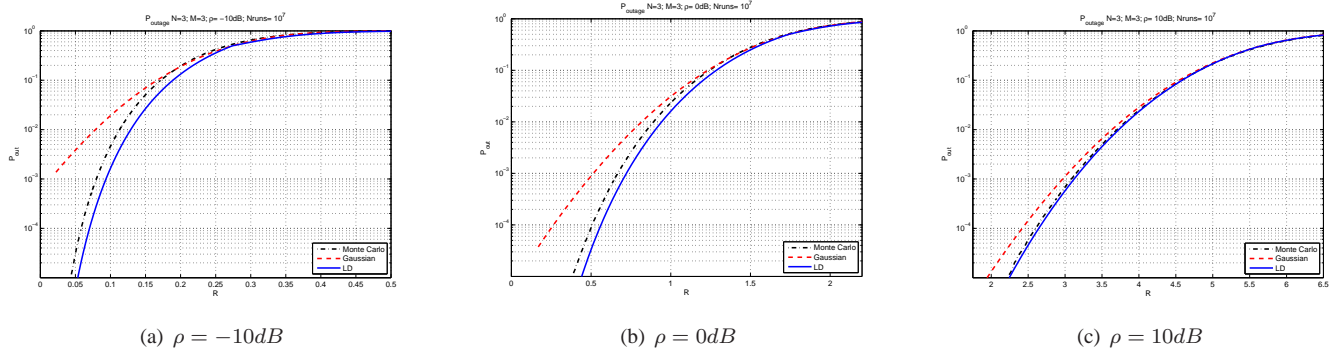(a) $\rho = -10dB$ (b) $\rho = 0dB$ (c) $\rho = 10dB$

Fig. 4. Comparison of the outage probability curves for $N = M = 2$ of the Large Deviation result with the Gaussian approach and Monte-Carlo simulations. The three subplots are for different SNR values: (a) with $\rho = -10dB$, (b) with $\rho = 0dB$ and (c) with $\rho = 10dB$. We see that for decreasing ρ the discrepancy between the Gaussian curve (dashed) and the other two, i.e. LD (solid)) and simulated (dash-dotted) is increasing.

(a) $\rho = -10dB$ (b) $\rho = 0dB$ (c) $\rho = 10dB$

Fig. 5. Comparison of the outage probability curves for $N = M = 3$ of the Large Deviation result with the Gaussian approach and Monte-Carlo simulations. The three subplots are for different SNR values: (a) with $\rho = -10dB$, (b) with $\rho = 0dB$ and (c) with $\rho = 10dB$. We see that for decreasing ρ the discrepancy between the Gaussian curve (dashed) and the other two, i.e. LD (solid)) and numerical (dash-dotted) is increasing. Comparing the $N = 3$ with the $N = 2$ results, we see that the former are generally closer to the simulated curve, nevertheless, the Gaussian curve is always clearly further away.
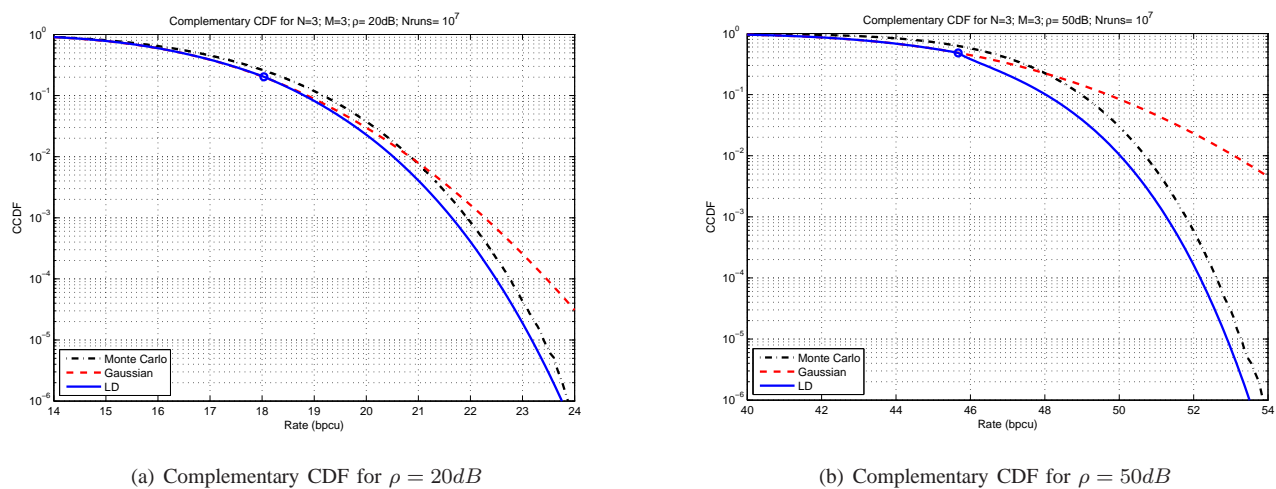

(a) Complementary CDF for $\rho = 20dB$ (b) Complementary CDF for $\rho = 50dB$

Fig. 6. In these figures we depict the complementary cumulative distribution function (CCDF) of the mutual information for the antenna array 3×3. In this region of parameters we compare the the current methodology (LD) (solid) with numerical Monte-Carlo simulations (dash-dotted) and the Gaussian approximation (dashed). We also depict the rate value r_c at which, for the given SNR, the exponent dependence on r changes from (60) to (54). We see that at that point the distribution starts deviating strongly from the Gaussian approximation. It should be pointed out that this point corresponds to a mild *phase transition* as discussed in Section III-C2 and also analyzed in a different context in [16]. Nevertheless, in both moderate and large SNRs the LD curve is consistently close to the simulated curves. (a) CCDF for $\rho = 20dB$ (b) CCDF for $\rho = 50dB$
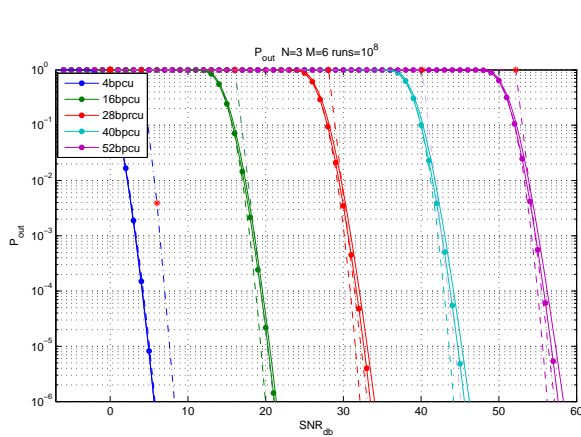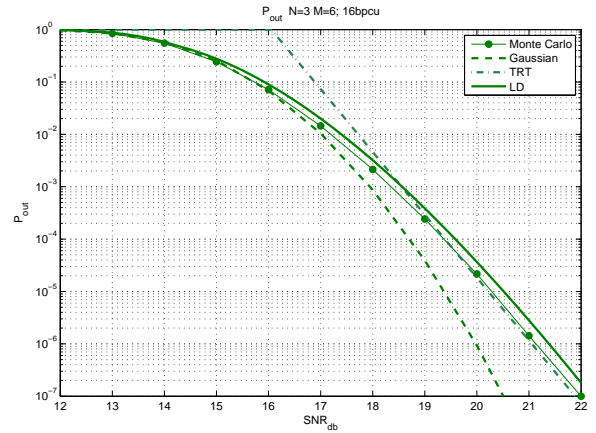
(a) Outage for $N = 3$, $M = 6$ and R=4, 16, 28, 40, 52 bpcu

(b) Outage probability only for R=16 bpcu

Fig. 7. In these figures we depict the outage probability as a function of SNR for the antenna array 3×6. The current methodology (LD) (solid) is compared with numerical Monte-Carlo simulations (10^8 runs, solid with dots) and two other approximations, the Gaussian (dashed) and the Throughput-Reliability-Tradeoff (TRT) approximation (dash-dot), analyzed in [12]. The red stars on the TRT curve depict the points at which the lines change slope. (a) In this figure, we collectively plot the curves at a number of bpcu values. At this scale all three candidates behave rather well, except perhaps for the TRT curve at the lowest bpcu value (R=4). (b) Nevertheless, zooming in for the R=16 bpcu case, we see that both the TRT and Gaussian approximations significantly depart from the numerical curve, at low and high SNRs correspondingly. In contrast, the LD curve is consistently closer to the numerics.
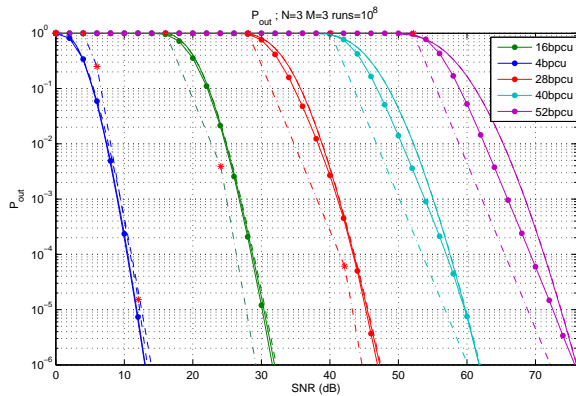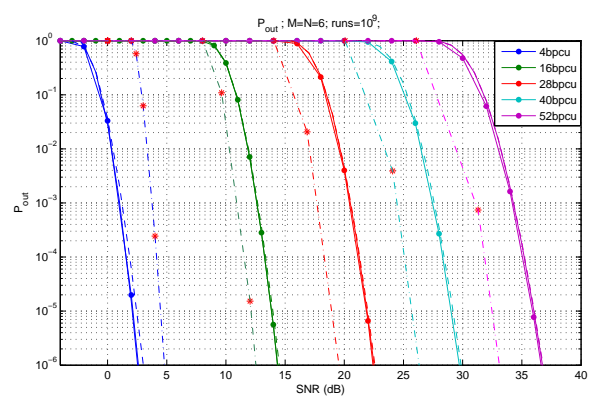

(a) Outage for $N = M = 3$, and R =4, 16, 28, 40, 52 bpcu

(b) Outage for $N = M = 6$, and R =4, 16, 28, 40, 52 bpcu

Fig. 8. In these figures we depict the outage probability as a function of SNR for the antenna arrays $N = M$. The current methodology (LD) (solid) is compared with numerical Monte-Carlo simulations (solid with dots) and two other approximation, the Gaussian (dashed) and the Throughput-Reliability-Tradeoff (TRT) approximation (dash-dot), analyzed in [12]. The red stars on the TRT curve depict the points at which the lines change slope. (a) Curves for outage probability versus SNR for the antenna array 3×3 for the same bpcu values as in Fig. 7. In contrast to that figure, for very large values of SNR ($\rho > 45dB$) both the LD and Gaussian approximations deviate from the numerics (10^8 runs), which exhibits a linear behavior (in a log-log plot). This deviation of the LD approximation is expected. Here the number of antennas is still quite small ($N = 3$), while the SNR is extremely large, making the LD approximation (in addition to the Gaussian) not valid. In these extreme SNRs the TRT approximation seems to have the correct slope, but also misses the exact value of the outage probability. For more reasonable SNR, the LD is quite close to the numerical plot. (b) Curves for outage probability versus SNR for the antenna array 6×6. In this case, the LD approximation works well even for such large SNRs.

other asymptotic forms, namely the Gaussian approximation of the mutual information [7] and the large-ρ asymptotic result given by (4) [1]. We see that our result performs much better at low outage, even at moderately large $\rho = 20dB$.

As discussed in the Introduction, the LD method is the correct generalization of the Gaussian approximation to capture the tails of the distribution of the mutual information. As a result, it is expected to give increasingly accurate results as the antenna number N increases. In the above comparisons we have compared the LD method with numerical simulations focusing on its tails (low outage P_{out} or low values of $1-P_{out}$) for small antenna numbers. We have found that the LD

approximation behaves well even at these values of N. The discrepancy between the LD approximation and Monte Carlo simulations becomes smaller for larger N as seen in Fig. 9.

In Appendix E we provide an improved estimate on the probability distribution close its center. This estimate is a result of the inclusion of the $O(1/N)$ corrections to the distribution derived in [7]. Fig. 10 shows the normalized probability distribution function of the Gaussian approximation as well as the LD approximation with and without the $O(1/N)$ corrections. We see that the improved estimate behaves extremely well when the antenna numbers are quite small, in which cases the leading approximation (without the $O(1/N)$ correction),
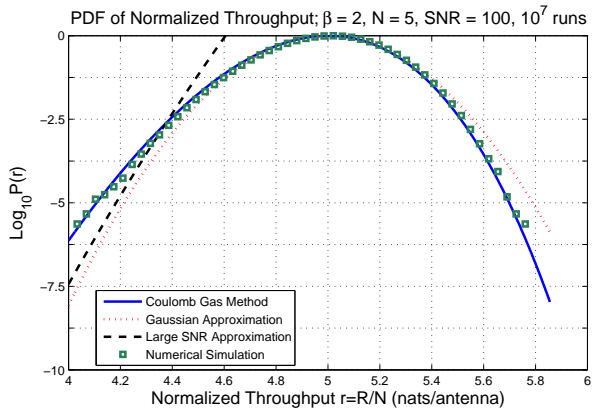
Fig. 9. Plot of the logarithm of the normalized probability distribution curve of the mutual information I_N/N for $\beta = 2$ and comparison to the Gaussian approximation and the large-ρ asymptotic result obtained by (4) [1]. The numerical result for $N = 5$ follows closely our result, even at large $\rho = 100$.

has some small discrepancies. (This should be contrasted with Fig. 9, where $N = 5$ and the $O(1/N)$ correction is no longer necessary to provide close agreement.)

VI. CONCLUSION

In this paper we have used a large deviation approach, first introduced in the context of statistical mechanics [13], [15], to calculate the probability distribution of the mutual information of MIMO channels in the limit of large antenna numbers. In contrast to previous approaches that focused only close to the mean of the distribution, [7]–[9], we also calculate the probability for *rare events* in the tails of the distribution, corresponding to instances where the observed mutual information differs by $\mathcal{O}(N)$ from the most probable value of the asymptotic distribution (where the Gaussian approximation for the mutual information is invalid). We find that the distribution in those tails is markedly different from what happens near the mean and our resulting probability distribution interpolates seamlessly between the Gaussian approximation for rates close to the ergodic mutual information and the results of [1] for large signal to noise ratios (where the outage probability is given asymptotically by (4)). Our method thus provides an analytic tool to calculate outage probabilities at any point in the (R, ρ, N) parameter space, as long as N is large enough. We performed numerical simulations that showed the robustness of our approximation over a wide range of parameters.

Additionally, this approach also provides the probability distribution of eigenvalues constrained in the subset where the mutual information is fixed to R for a given signal to noise ratio ρ. Interestingly, this eigenvalue density is of the form of the Marčenko-Pastur distribution with square-root singularities. Since the outage probability is an increasing function of the rate r for fixed ρ, we may use our approach to evaluate the transmission rate R for a required outage P_{out} and ρ. Thus, if the channel is known at the transmitter, we can optimize the transmitted rate by waterfilling on the known eigenvalue density that corresponds to the required outage probability [33]. This generalization is left for a future work.

Finally, it is worth pointing out that, to our knowledge, this is the first time this methodology has been applied to information theory and communications, and it is our belief that it may find other applications in this field. We can corroborate this belief by pointing out that this Coulomb gas methodology can be generalized to other channel distributions, as long as the resulting distribution can be written as a product of functions of the eigenvalues of $\mathbf{H}^\dagger \mathbf{H}$. Another related generalization is, for example, to include the correlations of the channel, a problem which is considerably more difficult compared to the present one. Some preliminary mathematical tools have already been developed in [34], and we will expand on this in the future.

APPENDIX A
PROPERTIES OF TAME PROBABILITY MEASURES

This appendix is largely devoted to the study of the energy functional $\mathcal{E}$:

$$\mathcal{E}[p] = \int xp(x)\,dx - (\beta - 1)\int p(x)\log x\,dx \quad (12)$$
$$- \iint p(x)p(y)\log|x - y|\,dx\,dy$$

where $p \in \Omega$ is a tame density. As evidenced by definition 2 where the concept of tameness was introduced, an extremely important part in our analysis will be played by the so-called L^r norm $\|\cdot\|_r$ defined by:

$$\|f\|_r \equiv \left(\int |f(x)|^r\,dx\right)^{1/r}. \quad (88)$$

If a function f has finite L^r norm it is called L^r-integrable and the space of such functions constitutes a complete vector space (also denoted by L^r). The completeness of this space follows from *Hölder's inequality* which we state without proof and which will be of great use to us [35]:

$$\|fg\|_1 \le \|f\|_r\|g\|_s \quad (89)$$

whenever the exponents $r, s > 1$ are *conjugate*, that is: $r^{-1} + s^{-1} = 1$.

We will also make heavy use of the convolution $f * g$ between two functions f and g:

$$(f * g)(x) = \int f(x - y)g(y)\,dy. \quad (90)$$

If $f \in L^1$ and $g \in L^r$, *Young's inequality* (pp. 240–241 in [35]) states that their convolution will be finite for almost every x and also that:

$$\|f * g\|_r \le \|f\|_1\|g\|_r. \quad (91)$$

We may now proceed with the proof of lemma 3 regarding the domain of $\mathcal{E}$ and its continuity properties:

Proof of Lemma 3: To show that $\mathcal{E}$ is finite for all tame functions $p \in \Omega$, we will study $\mathcal{E}[p]$ term by term. To that end, let $p : \mathbb{R}_+ \to \mathbb{R}$ be tame for some exponent $\varepsilon > 0$; that is, assume that $\int |p|^{1+\varepsilon} < \infty$ and that $\int xp(x)\,dx < \infty$. We then have:

- The first term of $\mathcal{E}[p]$ is finite by definition.

- The second term in (12) can be written as:

$$\left| \int p(x) \log x \, dx \right| \leq \int |p(x) \log x| \, dx$$

$$= \int_0^1 |p(x) \log x| \, dx + \int_1^\infty |p(x) \log x| \, dx.$$

Since $\log x < x$ for $x > 1$, the second integral will be bounded from above by $\int x |p(x)| \, dx < \infty$. As for the first integral, set $r = 1 + \varepsilon$ and $s = 1 + \frac{1}{\varepsilon}$ so that $r^{-1} + s^{-1} = 1$. Now, if $\chi_{[0,1]}$ is the indicator function of $[0,1]$, note that $\int |\chi_{[0,1]} \log x|^s \, dx = \int_0^1 |\log x|^s \, dx < \infty$ for all $s > -1$. As a result, Hölder's inequality yields:

$$\int_0^1 |p(x) \log x| \, dx = \| p \cdot \chi_{[0,1]} \log \|_1$$

$$\leq \|p\|_{1+\varepsilon} \cdot \|\chi_{[0,1]} \log \|_{1+1/\varepsilon} < \infty \tag{92}$$

on account of p being $L^{1+\varepsilon}$-integrable.

- For the last term of $\mathcal{E}$, let $D_+ = \{(x,y) \in \mathbb{R}^2 : y > x\}$ and note that:

$$\left| \iint p(x)p(y) \log |x - y| \, dy \, dx \right|$$

$$\leq 2 \iint_{D_+} |p(x)p(y) \log |x - y|| \, dy \, dx$$

$$= 2 \int_0^\infty |p(x)| \int_x^\infty |p(y) \cdot \log(y - x)| \, dy \, dx. \tag{93}$$

Now, the innermost integral can be written in the form:

$$\int_x^\infty |p(y)| \cdot |\log(y - x)| \, dy$$

$$= \int_0^\infty |p(x + w) \cdot \log w| \, dw$$

$$= \int_0^\infty |p(x+w) K(w)| \, dw + \int_1^\infty |p(x+w) \log w| \, dw$$

$$\leq \int_0^\infty |p(y)| K(y - x) \, dy$$

$$+ \int_0^\infty |p(1 + x + w) \log(1 + w)| \, dw. \tag{94}$$

where $K(w)$ is the kernel:

$$K(w) = \begin{cases} \log |w|, & 0 < w \leq 1 \\ 0, & \text{otherwise.} \end{cases} \tag{95}$$

As above, K will be L^s-integrable for all $s > -1$ and, in particular, for $s = 1 + \frac{1}{\varepsilon}$. Therefore, we will have:

$$\int_0^\infty |p(x)| \int_0^\infty |p(y)| K(x - y) \, dy \, dx$$

$$= \| |p| \cdot (|p| * |K|) \|_1$$

$$\leq \|p\|_{1+\varepsilon} \cdot \|p * K\|_{1+1/\varepsilon}$$

$$\leq \|p\|_{1+\varepsilon} \cdot \|p\|_1 \cdot \|K\|_{1+1/\varepsilon} < \infty \tag{96}$$

where the penultimate step is an application of Hölder's estimate and the last one follows from Young's inequality.

Finally, the second integral of (94) can be estimated by:

$$\int_0^\infty |p(1 + x + w) \log(1 + w)| \, dw$$

$$\leq \int_0^\infty |p(1 + x + w)| w \, dw$$

$$\leq Cx \int_0^\infty w |p(w)| \, dw \tag{97}$$

for some sufficiently large $C > 0$. Then, since p is tame (i.e. $\int w |p(w)| \, dw < \infty$), we may integrate (97) over x to finally obtain that $\mathcal{E}[p] < \infty$.

This completes the proof that $\mathcal{E}[p]$ is finite for all tame functions $p \in \Omega$. To show that $\mathcal{E}$ is continuous on all subspaces of $L^{1+\varepsilon}$-integrable functions with finite absolute mean, it simply suffices to note that all our estimates of $\mathcal{E}[p]$ are bounded by the $L^{1+\varepsilon}$ norm of p. ∎

Remark. If a function is in L^r for some $r > 1$ and has finite mean, it will necessarily be in L^1 as well; in this way, tame measures form a (dense) subspace Ω of $L^1(\mathbb{R}_+)$ that is similar to the union $\bigcup_{\varepsilon > 0} L^{1+\varepsilon}$.

We will now prove Lemma 4 showing that $\mathcal{E}$ is not only continuous but also convex over the (convex) domain $\mathcal{X}$ of tame *probability* measures.

Proof of Lemma 4: Let $p, q \in \mathcal{X}$ be two tame probability measures and introduce the bilinear pairing:

$$\langle p, q \rangle = - \int \int p(x) q(y) \log |x - y| \, dx \, dy \tag{98}$$

which is actually well-defined on the whole space Ω (as can be seen by the proof of lemma 3). Since the first two terms of $\mathcal{E}$ are linear (and hence convex), it will suffice to show that:

$$\langle (1 - t)p + tq, (1 - t)p + tq \rangle < (1 - t)\langle p, p \rangle + t \langle q, q \rangle \tag{99}$$

for all $t \in (0, 1)$. Indeed, if we let $\phi = p - q \in \Omega$, equation (99) reduces to showing that the pairing $\langle \cdot, \cdot \rangle$ is an inner product on the subspace of densities with zero total charge, i.e. that:

$$\langle \phi, \phi \rangle > 0 \tag{100}$$

for any nonzero tame $\phi \in \Omega$ with $\int \phi(x) \, dx = \int (p(x) - q(x)) \, dx = 0$.

From the point of view of electrostatics, this is plain to see: after all $\langle \phi, \phi \rangle$ is just the self-energy of the charge density ϕ. More specifically, let us define $D_+ = \{(x, y) : x < y\}$ as in the proof of lemma 3. Then we will have:

$$\langle \phi, \phi \rangle = -2 \int_{D_+} \phi(x)\phi(y) \log |x - y| \, dx \, dy$$

$$= -2 \int_0^\infty \phi(x) \int_0^x \phi(y) \log(x - y) \, dy \, dx$$

$$> 2 \int_0^\infty \phi(x) \int_0^x \phi(y)(y - x) \, dy \, dx \tag{101}$$

So, if we set $\Phi(x) = \int_0^x \phi(y) \, dy$ and integrate by parts, we

get:

$$
\begin{aligned}
\langle\phi,\phi\rangle \quad &> \quad \int_0^\infty \phi(x)\int_0^x y\phi(y)\,dy\,dx - \int_0^\infty x\phi(x)\Phi(x)\,dx \\
&= \quad -\int_0^\infty \phi(x)\left(\int_0^x \Phi(y)\,dy\right)dx \\
&= \quad \int_0^\infty \Phi^2(x)\,dx - \Phi(\infty)\int_0^\infty \Phi(y)\,dy \\
&> \quad 0
\end{aligned}
\tag{102}
$$

since $\Phi(\infty) \equiv \int_0^\infty \phi(x)\,dx = 0 = \Phi(0)$ on account of ϕ having zero total charge. ∎

APPENDIX B
CONSTRUCTION OF THE COULOMB GAS MODEL

In this appendix we will briefly show how the transition from discrete to continuous eigenvalue measures discussed in Section II-A occurs. As in the main text, we will not present any formal proof here either. However, we will argue that treating the formally discrete distribution of eigenvalues appearing in (9), as continuous in the large N limit is quite reasonable. A more formal method showing the same result appears in [15]. The main reasoning, also discussed in the main text, is that the external confining potentials defined by the first two terms in (9) or (12) are strong enough to overcome the logarithmic repulsion between eigenvalues (third term in (9)), and therefore guarantee that (with high probability) most of the eigenvalues will be confined in a finite width region near the minimum of the external potential. At the same time, this will mean that the eigenvalue density per unit length will be scaling with N if N is large enough. As a result, this can be seen as a high-density limit and therefore the continuous approximation for the measure will be valid, at least close to configurations whose energy is low enough.

In the remainder of this section we will motivate the transition from the discrete to continuous eigenvalue densities and show what kind of terms we expect to see. We start by focusing in a finite region of eigenvalues of length D. We then divide the integration over λ_k in (10) in L segments of length ℓ, such that $L\ell = D$. The length of each segment ℓ has to be small enough so that the energy (9) can be well approximated with all eigenvalues within a given segment being placed at the endpoint of the segment. At the same time, it has to be large enough so that there is a macroscopic (i.e. $\mathcal{O}(N)$) number of eigenvalues inside each segment. In principle, at the end of this exercise we need to take the limit $\ell \to 0$ as well, however we will discuss the subtleties of this limit later on. As a result, the integral over $\mathcal{D}\boldsymbol{\lambda}$ can be written as:

$$
\int \mathcal{D}\boldsymbol{\lambda} \sim \prod_{k=1}^N\left(\sum_{m_k=1}^L \ell\right) = \prod_{m=1}^L\left(\sum_{n_m=0}^N\right)\frac{N!\,\ell^N}{\prod_{m=1}^L n_m!}
\tag{103}
$$

$$
\sim \prod_{m=1}^L\left(\sum_{n_m=0}^N\right)\exp\left[-N\ell\sum_m p(m\ell)\log\left(p(m\ell)\right)\right]
\tag{104}
$$

where n_m are the number of λ_k's that appear in the mth segment, with constraint $\sum_m n_m = N$. The factorials appearing at the RHS of (103) are the number of ways the

N eigenvalues can be re-arranged in L segments. This factor constitutes the entropy term and, for large N and n_m, we can apply Stirling's formula to get the exponent in (104) (where $p(m\ell) = n_m/(N\ell)$ is the fraction of eigenvalues per unit length appearing in segment m).

We next look at the form of the energy in (9)

$$
\begin{aligned}
E(\boldsymbol{\lambda}) \quad \sim \quad & \ell\sum_m p(m\ell)\left(m\ell - (\beta-1)\log m\ell\right) \\
&+ \ell^2\sum_{m\neq m'} p(m\ell)p(m'\ell)\log|(m-m')\ell| \\
&+ \frac{\ell}{N}\sum_m p(m\ell)\log a_m\ell
\end{aligned}
\tag{105}
$$

The last term captures the repulsive interaction between eigenvalues in the same segment m. The value of a_m represents the typical distance between eigenvalues in segment m in units of ℓ and therefore is a number of order unity. We may now let $\ell \to 0$, which will make the sums converge to integrals $\ell\sum_m \to \int dx$ and $p(m\ell)$ can be written as a continuous function $p(x)$. Representing the sum over all possible states (i.e. the product of sums in (104)) by $\int \mathcal{D}p$ we can now get

$$
\mathcal{Z} \sim \int_\chi \mathcal{D}p\, e^{-N^2\mathcal{E}[p]}e^{-N\int dx\,p(x)\log p(x)}e^{N\int dx\,p(x)\log d(x)}
\tag{106}
$$

where $d(x) = a_m\ell$ is the average distance between eigenvalues at the position $x = m\ell$. One can estimate this average inter-eigenvalue distance to be

$$
d(x) \sim a_m\ell \sim \frac{1}{Np(x)}
\tag{107}
$$

This was first proposed by Dyson [13], [15], [23] and was shown explicitly more recently in [36]. It is remarkable that with this choice of $d(x)$ the $\mathcal{O}(N)$ dependence on $p(x)$ in the exponent of (106) vanishes. This surprising fact is true only for complex matrices [23] in which, up to uninteresting constants, the leading correction to the $N^2\mathcal{E}[p]$ term in the exponent is $\mathcal{O}(1)$.

APPENDIX C
SOLUTION OF THE VARIATIONAL EQUATION

In this appendix, we give a more detailed account of the solution of the variational equation:

$$
\delta\mathcal{L}_1[p] = 0
$$

where $\mathcal{L}_1$ is the Lagrangian function of (27). To that end, if $\phi \in \Omega$ is tame, we get:

$$
\begin{aligned}
\mathcal{L}_1[p+t\phi] = \mathcal{L}_1[p] + t\mathcal{L}_1[\phi] \\
- 2t\iint \phi(x)p(y)\log|x-y|\,dy\,dx + \mathcal{O}(t^2)
\end{aligned}
\tag{108}
$$

and a simple differentiation at $t = 0$ yields:

$$\langle \delta \mathcal{L}_1[p], \phi \rangle = \frac{d}{dt}\Big|_{t=0} \mathcal{L}_1[p + t\phi]$$

$$= \mathcal{L}_1[\phi] - 2 \iint \phi(x) p(y) \log |x - y| \, dy \, dx$$

$$= \int \phi(x) \Psi[p, x] \, dx, \quad (109)$$

where the expression $\Psi[p, x]$ is given by:

$$\Psi[p, x] = 2 \int p(y) \log |x - y| \, dy - x$$
$$+ (\beta - 1) \log x + c + k \log(1 + \rho x) + \nu(x). \quad (110)$$

Thus, for the above expression to vanish identically for all $\phi \in \Omega$, we must have $\Psi[p, x] = 0$, and this is precisely (34), repaeted below:

$$2 \int_0^\infty p(x') \log |x - x'| \, dx' = x - (\beta - 1) \log x$$
$$- c - k \log(1 + \rho x) - \nu(x). \quad (111)$$

Having derived this stationarity equation in terms of p, we will devote the rest of this appendix to the expression (35), also repeated below for convenience, that is obtained after differentiating (34) above:

$$2 \mathcal{P} \int_a^b \frac{p(y)}{x - y} dy = 1 - \frac{\beta - 1}{x} - \frac{k\rho}{1 + \rho x} \equiv f(x) \quad (112)$$

for all $x \in [a, b]$ (cf. section II-B). This integral equation is known as the *airfoil equation* and can be studied with the help of the *finite* Hilbert transform [27]:

$$\mathcal{T}[\phi](x) = \mathcal{P} \int_{-1}^1 \frac{\phi(y)}{y - x} \, dy. \quad (113)$$

If $r > 1$, the $\mathcal{T}$-transform maps L^r to L^r but, nevertheless, it lacks a unique inverse.[9] Indeed, the kernel of $\mathcal{T}$ is spanned by the function $\omega(x) = (1 - x^2)^{-\frac{1}{2}}$: $\mathcal{T}[\omega](x) = 0$ for all $x \in (-1, 1)$. Outside this kernel, the solutions ϕ to the airfoil equation $\mathcal{T}[\phi] = g$ with $\phi, g \in L^r[-1, 1]$ will satisfy [27]:

$$\phi(x) = -\frac{1}{\pi} \mathcal{P} \int_{-1}^1 \sqrt{\frac{1 - y^2}{1 - x^2}} \frac{g(y)}{y - x} \, dy + \frac{c}{\sqrt{1 - x^2}} \quad (114)$$

where c is an arbitrary constant that stems from the fact that any two solutions of the airfoil equation differ by a multiple of $\omega(x) = (1 - x^2)^{-\frac{1}{2}}$.

Hence, after rescaling the interval $[-1, 1]$ to $[a, b]$, the solution of the stationarity equation (35) will be given by:

$$p(x) = \frac{\mathcal{P} \int_a^b \frac{\sqrt{(y-a)(b-y)} f(y)}{y-x} dy + C'}{2\pi^2 \sqrt{(x - a)(b - x)}} \quad (115)$$

whenever f is itself $L^{1+\varepsilon}$-integrable. So, by substituting $f(x) = 1 - \frac{\beta-1}{x} - \frac{k}{x+z}$ from (35) and performing one last integration, we obtain the final result (36).

It is worthwhile to mention here again how this procedure breaks down if we allow the support of p to extend to $a = 0$

---

[9]This is a remarkable difference from the case of the *infinite* Hilbert transform which integrates over all $\mathbb{R}$ and which *is* invertible [27].

for $\beta > 1$: in that case, the function f also extends all the way to $a = 0$ and the term $\frac{\beta-1}{x}$ makes it non-integrable. However, since the Hilbert transform preserves L^r-integrability for $r > 1$ and p is assumed tame (and hence $L^{1+\varepsilon}$-integrable), equation (35) would equate an integrable function with a non-integrable one, thus yielding a contradiction. Therefore, as we stated in section II-B, solutions with $a = 0$ are physically inadmissible when $\beta > 1$.

APPENDIX D
PROOF OF UNIQUENESS OF SOLUTION OF (39),(40)

In order to show that (39), (40) admit a unique solution, we start by observing that for fixed k, β and z, (39) has a unique positive solution $a \leq b$. Then, from the implicit function theorem, this solution can be captured in terms of b by a smooth function $a(b)$ whose derivative can be obtained implicitly from (39) (and which is negative). With this in mind, the normalization integral $g(b) = \int_{a(b)}^b p(x) \, dx$ takes the form:

$$g(b) = \frac{a(b) + b}{4} + \frac{1}{2} \left(\rho^{-1} - k - (\beta - 1) \left(1 + \frac{1}{\rho \sqrt{a(b)b}} \right) \right)$$

and this is actually an increasing function of b. Indeed, after a somewhat painful calculation, one obtains:

$$g'(b) = \frac{\rho}{4} \left[1 + \frac{(\beta - 1)}{\rho \sqrt{a(b) \, b^3}} \right] \frac{b - a(b)}{1 + \rho b} > 0 \quad (116)$$

However, with $a(b)$ decreasing and bounded below by 0, this last equation yields $g'(b) > 1/8$ for large enough b, i.e. $\lim_{b \to \infty} g(b) = +\infty$. So, by continuity, there will be a (necessarily) unique b^* such that $g(b^*) = 1$. Hence, the pair $a^* = a(b^*), b = b^*$ will be the unique solution to (39), (40).

APPENDIX E
$O(1/N)$ CORRECTION TO THE LD APPROXIMATION

Here we provide an improved estimate on the probability distribution close to the center of the distribution. This estimate is a result of the inclusion the $O(1/N)$ higher moment corrections to the distribution derived in [7].

It is well known [37] that to provide asymptotic corrections to the limiting Gaussian distribution due to the presence of a small (but finite) skewness we need to change the distribution as follows:

$$P_N(x) = \frac{e^{-\frac{x^2}{2v}}}{\sqrt{2\pi v}} \left(1 - \frac{s}{2v^2} \left(x - \frac{x^3}{3v} \right) \right) \quad (117)$$

where v is the variance of the asymptotically Gaussian distribution and s is the third moment of the distribution. Clearly, the above distribution cannot be valid over the entire support of x since the cubic polynomial will become negative for some value of x. Nevertheless, since the third moment is small for large N this value of x will become asymptotically large.

We may therefore apply the above formula to our model. The value of the third moment $s = s_3/N$ has been calculated in [(60) in [7]] and it is of order $O(1/N)$. As a result,

(a) PDF $N = 2$, $M = 4$ for $\rho = 20dB$ (b) PDF $N = 2$, $M = 4$ for $\rho = 50dB$
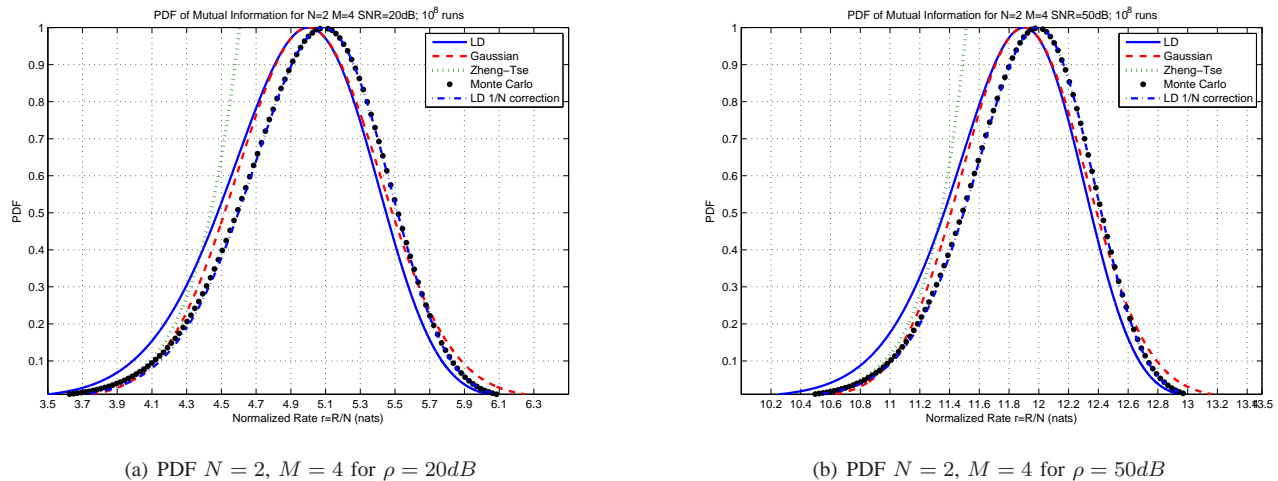
Fig. 10. Normalized probability distribution curves for the PDF of the mutual information for the antenna array $N = 2$, $M = 4$ for $\rho = 20dB$ (a) and $\rho = 50dB$ (b). In addition to the LD and Gaussian approximations and the Monte Carlo-generated curves, we have plotted the LD approximation including the $O(1/N)$ correction analyzed in Appendix E. We see that the latter curve agrees very well with the numerical one.

the correction to the Gaussian approximation of the mutual information is given by

$$P_N(R) = \frac{e^{-\frac{(R - Nr_{erg})^2}{2v_{erg}}}}{\sqrt{2\pi v_{erg}}} \cdot \qquad (118)$$
$$\left(1 - \frac{s_3}{2Nv_{erg}^2}(R - Nr_{erg}) + \frac{(R - Nr_{erg})^3}{3v_{erg}}\right)$$

To order $O(1/N)$, there is also the correction to the mean of the mutual information [7], which needs to be subtracted off from I_N.

Now, to obtain the correction to the LD approximation, we need to take into account that the large deviations function $\mathcal{E}_1$ also has a cubic term for $r \approx r_{erg}$, which needs to be balanced. This can be done by adding a cubic term that cancels this term for $r \approx r_{erg}$. Thus we obtain

$$P_N(r) = \frac{Ne^{-N^2(\mathcal{E}_1(r) - \mathcal{E}_0)}}{\sqrt{2\pi v_{erg}}}\left(1 - \frac{s_3}{2v_{erg}^2}(r - r_{erg})\right.$$
$$\left. + \frac{N^2}{6}\left(\frac{s_3}{v_{erg}^3} + s_{erg}\right)(r - r_{erg})^3\right) \qquad (119)$$

where s_{erg} is given by (70).

REFERENCES

[1] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[2] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, pp. 311–335, 1998.

[3] I. E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Transactions on Telecommunications and Related Technologies*, vol. 10, no. 6, pp. 585–596, Nov. 1999.

[4] P. B. Rapajic and D. Popescu, "Information capacity of a random signature multiple-input multiple-output chanel," *IEEE Trans. Commun.*, vol. 48, no. 8, p. 1245, Aug. 2000.

[5] Z. Wang and G. B. Giannakis, "Outage mutual information of space-time MIMO channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 4, pp. 657–662, Apr. 2004.

[6] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, no. 6, p. 2619, Oct. 1998.

[7] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis," *IEEE Trans. Inform. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.

[8] B. M. Hochwald, T. L. Marzetta, and V. Tarokh, "Multi-antenna channel hardening and its implications for rate feedback and scheduling," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1893–1909, Sept. 2004.

[9] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for capacity analysis of large dimensional multi-antenna channels," *IEEE Trans. Inform. Theory*, vol. 54, pp. 3987–4004, Sep. 2008.

[10] G. Taricco, "On the capacity of separately-correlated MIMO Rician fading channels," *Proc. IEEE Globecom 2006*, Dec. 2006.

[11] ——, "Asymptotic mutual information statistics of separately-correlated MIMO Rician fading channels," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, p. 3490, Aug. 2008.

[12] K. Azarian and H. El-Gamal, "The Throughput Reliability Tradeoff in block-fading MIMO channels," *IEEE Trans. Inform. Theory*, vol. 53, no. 2, p. 488, Feb. 2007.

[13] F. Dyson, "Statistical theory of the energy levels of complex systems. I," *J. Math. Phys.*, vol. 3, p. 140, 1962.

[14] S. N. Majumdar, *Random Matrices, the Ulam Problem, Directed Polymers & Growth Models, and Sequence Matching*, ser. Les Houches, M. Mézard and J. P. Bouchaud, Eds. Elsevier, July 2006, vol. Complex Systems.

[15] P. Vivo, S. N. Majumdar, and O. Bohigas, "Large deviations of the maximum eigenvalue in Wishart random matrices," *J. Phys. A*, vol. 40, pp. 4317–4337, 2007.

[16] ——, "Distributions of conductance and shot noise and associated phase transitions," *Phys. Rev. Lett.*, vol. 101, p. 216809, 2008.

[17] C. Nadal and S. N. Majumdar, "Nonintersecting brownian interfaces and wishart random matrices," *Phys. Rev. E*, vol. 79, p. 061117, 2009.

[18] K. Johansson, "On fluctuations of eigenvalues of random hermitian matrices," *Duke Math. J.*, vol. 91, no. 1, pp. 151–204, 1998.

[19] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.

[20] D. S. Dean and S. N. Majumdar, "Extreme value statistics of eigenvalues of Gaussian random matrices," *Phys. Rev E*, vol. 77, p. 041108, 2008.

[21] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. Singapore: McGraw-Hill, 1991.

[22] S. H. Simon and A. L. Moustakas, "Optimizing MIMO systems with channel covariance feedback," *IEEE J. Select. Areas Commun.*, vol. 21, no. 3, Apr. 2003.

[23] M. L. Mehta, *Random Matrices*, 2nd ed. San Diego, CA: Academic Press, 1991.

[24] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals*. New York: McGraw-Hill, 1965.

[25] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. New York, USA: Springer-Verlag Inc., 1998.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge Univ. Press, 2004.

[27] F. G. Tricomi, *Integral Equations*, ser. Pure Appl. Math V. London: Interscience, 1957.

[28] S. G. Mikhlin, *Integral Equations*. New York: Pergamon, 1964.

[29] Y. Chen and S. M. Manning, "Some eigenvalue distribution functions of the laguerre ensemble," *J. Phys. A: Math. Gen.*, vol. 29, pp. 7561–7579, 1996.

[30] C. M. Bender and S. A. Orszag, *Advanced Mathematical Methods for Scientists and Engineers*. New York, NY: McGraw-Hill, 1978.

[31] S. Verdú and S. Shamai, "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inform. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.

[32] P. Bender, P. Black, M. Glob, R. Padovani, N. Sindhushayaba, and A. Viterbi, "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users," *IEEE Communications Magazine*, pp. 70–77, Jul. 2000.

[33] L. G. Ordónez, D. P. Palomar, and J. R. Fonollosa, "Ordered eigenvalues of a general class of hermitian random matrices with application to the performance analysis of mimo systems," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 672–689, 2009.

[34] A. Matytsin, "On the large-N limit of the Itzykson-Zuber integral," *Nuclear Physics B411*, pp. 805–820, 1994.

[35] G. B. Folland, *Real Analysis*, 2nd ed. Wiley-Interscience, 1999.

[36] E. Brézin and A. Zee, "Universality of the correlations between eigenvalues of large random matrices," *Nuclear Physics B (FS)*, vol. 402, pp. 613–627, 1993.

[37] J.-P. Bouchaud and M. Potters, *Theory of Financial Risk and Derivative Pricing*, 2nd ed. Cambridge, UK: Cambridge, 2003.