On the Distribution of MIMO Mutual Information: An In-Depth Painlevé-Based Characterization

Shang Li, Matthew R. McKay, *Member, IEEE*, and Yang Chen

***Abstract*—This paper builds upon our recent work which computed the moment generating function of the multiple-input multiple-output mutual information exactly in terms of a Painlevé V differential equation. By exploiting this key analytical tool, we provide an in-depth characterization of the mutual information distribution for sufficiently large (but finite) antenna numbers. In particular, we derive systematic closed-form expansions for the high-order cumulants. These results yield considerable new insight, such as providing a technical explanation as to why the well-known Gaussian approximation is quite robust to large signal-to-noise ratio for the case of unequal antenna arrays, while it deviates strongly for equal antenna arrays. In addition, by drawing upon our high-order cumulant expansions, we employ the Edgeworth expansion technique to propose a *refined* Gaussian approximation which is shown to give a very accurate closed-form characterization of the mutual information distribution, both around the mean and for moderate deviations into the tails (where the Gaussian approximation fails remarkably). For stronger deviations where the Edgeworth expansion becomes unwieldy, we employ the saddle point method and asymptotic integration tools to establish new analytical characterizations which are shown to be very simple and accurate. Based on these results, we also recover key well-established properties of the tail distribution, including the diversity-multiplexing-tradeoff.**

***Index Terms*—Channel capacity, multiple-input multiple-output (MIMO) systems, random matrix theory.**

I. Introduction

MULTIPLE-INPUT multiple-output (MIMO) technologies form a key component of emerging broadband wireless communication systems due to their ability to provide substantial capacity growth over power and bandwidth constrained channels. Such technologies have received huge attention for over a decade now, with recent trends focusing mainly on incorporating MIMO into complicated system configurations, e.g., those employing relaying [1]–[5], cooperative multicell processing [6]–[8], information-theoretically secure systems [9]–[11], and ad hoc networking [12]–[15]. However, despite the huge progress, some basic questions regarding the information-theoretic limits of MIMO systems still remain unclear, even for the simplest point-to-point communication scenarios. This is particularly true for the *outage capacity*. In effect, this quantity gives an achievable transmission rate over quasi-static channels, which have been considered as a ubiquitous model for various systems, such as those lacking adequate dynamics, diversity, or with stringent constraints (e.g., on allowable delays). On the other hand, as discussed in the recent article [16], with many contemporary wireless systems employing technologies such as link adaptation, hybrid automatic repeat request, and orthogonal frequency division multiplexing, one may rightly question the relevance of the "classical" notion of outage. Nonetheless, as discussed in [16], even under this contemporary viewpoint, the outage capacity and the associated outage probability still remain as meaningful metrics, but just interpreted differently. In particular, under perfect link adaptation, for a given rate, the complement of the outage probability implies the percentage of time that transmission can reliably take place at or above that rate (a similar analogy also holds for the outage capacity). As such, whether adopting a classical or more conventional viewpoint, theoretical characterization of outage measures such as the outage capacity and outage probability still remains of interest.

Characterizing outage measures requires solving for the entire distribution of the channel mutual information, which has been the topic of considerable research interest since the early works of Telatar [17] and Foschini and Gans [18]. Despite the huge literature on this topic, the distribution is still partially understood. For example, in [19], assuming independent and identically distributed (IID) Rayleigh fading with perfect channel state information (CSI) at the receiver, the mutual information distribution was characterized via an exact expression for the moment generating function (MGF). This result was given in terms of a determinant of a certain Hankel matrix which yields little insight and becomes unwieldy when the number of antennas are not small.

A similar determinant representation for the MGF of a more general model, i.e., including spatial correlation, was adopted in [20]–[22] to provide a Gram–Charlier and saddle point approximation for the cumulative distribution function (CDF). However, though some asymptotic results of the determinant-based approach was given in [23], the solution was quite complicated and did not readily lend itself to intuitive interpretation. An alternative result was presented in [24] which considered the MGF of the mutual information at high signal-to-noise ratios

Manuscript received June 27, 2012; revised March 07, 2013; accepted May 13, 2013. Date of publication May 21, 2013; date of current version August 14, 2013. This work was supported by the Hong Kong Research Grants Council under Grant 616911. The material in this paper was presented in part at the 2012 Asilomar Conference on Signals, Systems and Computers.

S. Li is with the Department of Electrical Engineering, Columbia University, New York, NY 10027 USA (e-mail: shang@ee.columbia.edu).

M. R. McKay is with the Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Kowloon, Hong Kong (e-mail: eemckay@ust.hk).

Y. Chen is with the Department of Mathematics, University of Macau, Taipa, China (e-mail: yangbrookchen@yahoo.co.uk).

Communicated by A. Lozano, Associate Editor for Communications.

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

Digital Object Identifier 10.1109/TIT.2013.2264505

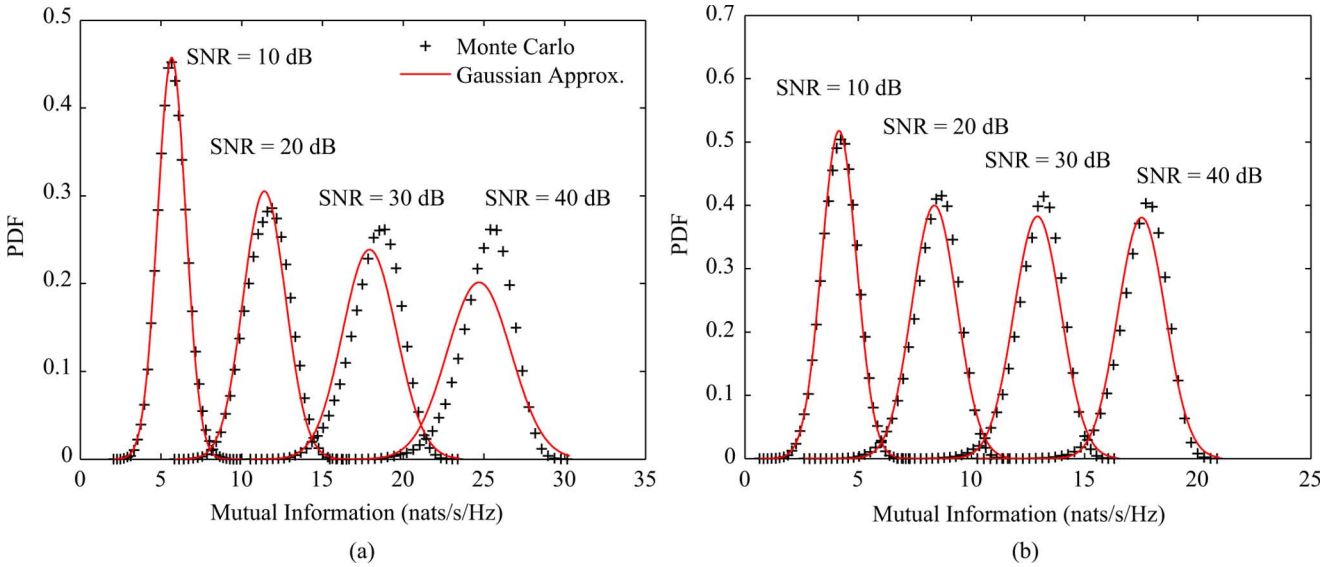

Fig. 1. PDF of the mutual information of MIMO Rayleigh fading channels for different antenna configurations and different SNRs. (a) $n_t = n_r = 3$; (b) $n_t = 3,\ n_r = 2$.

(SNRs) and used this to establish a Chernoff bound on the CDF. While this approach avoided dealing with complicated determinants, simulations demonstrated that the bound was not particularly tight, particularly when considering the CDF region representing outage probabilities of practical interest. As an alternative method to characterizing the mutual information distribution through its MGF, Smith *et al.* [25], [26] took the more direct approach of using classical transformation theory to derive exact expressions for the probability density function (PDF) and CDF for MIMO systems with small numbers of antennas. It was shown however, that even for the simplest MIMO configuration with dual antennas (i.e., having two transmit or two receive antennas), closed-form solutions were not forthcoming and one must rely on numerically evaluating complicated integrals.

To overcome the complexities of finite-antenna characterizations, another major line of work has focused on giving a large-antenna asymptotic analysis, which provides more intuitive results; see, e.g., [27]–[30]. In such analyzes, the most well-known conclusion is that the mutual information distribution approaches a Gaussian as the number of antennas become sufficiently large. Different approaches have also been employed to derive closed-form expressions for the asymptotic mean and variance [27]–[32]. Quoting [30] as an example, for a MIMO system subjected to IID Rayleigh fading with perfect receiver CSI, with n_t transmit antennas, n_r receive antennas, and SNR P, if n_t and n_r are both sufficiently large, then the mutual information distribution is approximated by a Gaussian with mean μ_0 and variance σ_0^2 given as

$$\mu_0 = n\left[\frac{a+b}{2}\ln\left(\frac{\sqrt{\beta+aP}+\sqrt{\beta+bP}}{2\sqrt{\beta}}\right) - \sqrt{ab}\ln\left(\frac{\sqrt{a\,(\beta+bP)}+\sqrt{b\,(\beta+aP)}}{\sqrt{a\beta}+\sqrt{b\beta}}\right) - \frac{\left(\sqrt{\beta+aP}-\sqrt{\beta+bP}\right)^2}{4P}\right], \quad (1)$$

$$\sigma_0^2 = 2\ln\left[\frac{1}{2}\left(\frac{\beta+aP}{\beta+bP}\right)^{1/4} + \frac{1}{2}\left(\frac{\beta+bP}{\beta+aP}\right)^{1/4}\right] \quad (2)$$

where

$$n := \min\{n_t, n_r\},\ m := \max\{n_t, n_r\},\ \beta := m/n,$$
$$a := \left(\sqrt{\beta+1}-1\right)^2,\ b := \left(\sqrt{\beta+1}+1\right)^2.$$

The Gaussian approximation has been considered extensively due to its relative simplicity compared to the exact characterizations. However, in practice, the number of antennas in MIMO systems is typically not huge, and it turns out that the Gaussian approximation may sometimes be very inaccurate. These deviations have been reported in several previous contributions [20], [30], [33], and here we give two concrete examples to demonstrate them.

As a first example, as shown in Fig. 1(a) and (b), the PDF of the Gaussian approximation deviates significantly from the true (simulated) PDF of the mutual information for finite-antenna arrays when the SNR becomes large. This phenomena was emphasized and investigated in our recent work [30] for the specific case of equal-antenna arrays (i.e., $n_t = n_r$) by looking at the cumulants of the mutual information. Quite interestingly, the figures indicate that the deviation from Gaussian is significantly *stronger* when $n_r = n_t$ compared with the alternative case, despite the fact that there are more antennas. Thus, when $n_t \neq n_r$, it appears that the Gaussian approximation is more robust to increasing SNR. This key observation is unexpected, and to the best of our knowledge, it hitherto lacks any rigorous explanation.

As a second example, aside from the deviation observed in the bulk of the distribution for large SNR, it turns out that the Gaussian approximation is typically very inaccurate in the tail of the distribution when n is finite and the SNR is either small or large. This is observed in Fig. 2(a) and (b), where in both cases the Gaussian curve fails markedly in tracking the simulations

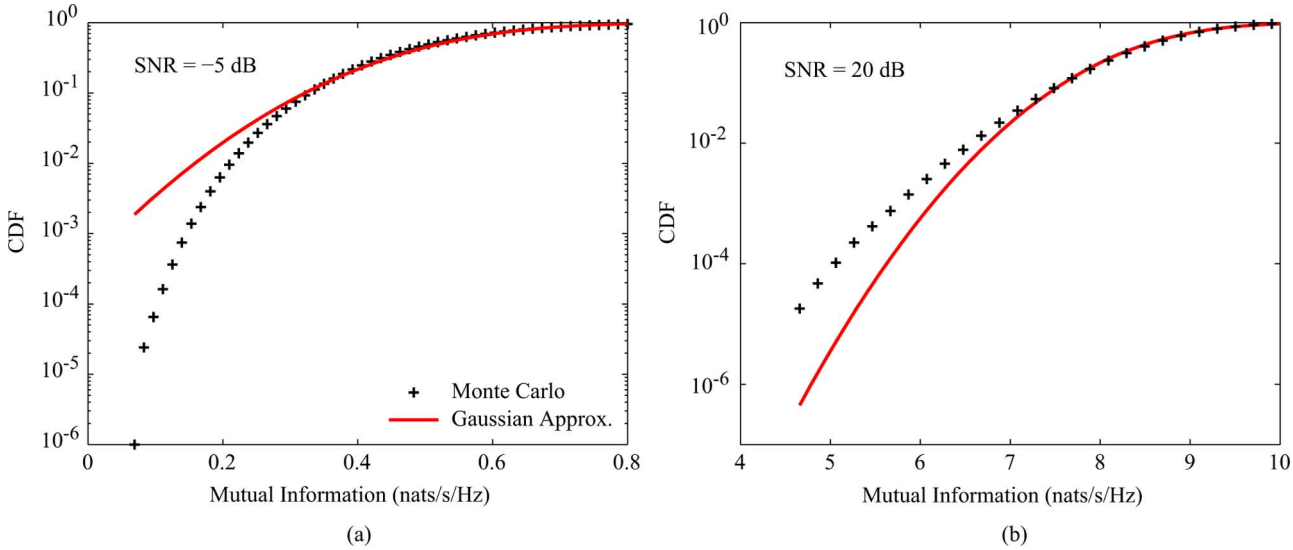

Fig. 2. CDF of mutual information of MIMO Rayleigh fading channels for small and large SNRs. (a) $n_t = 4,\ n_r = 2$; (b) $n_t = 4,\ n_r = 2$.

for small but practical outage probabilities. This strong deviation from Gaussian was also discussed in [33], where a refined approximation was presented based on adopting an intuitive large-n Coulomb fluid interpretation from statistical physics, leading to a set of coupled nonlinear equations requiring numerical computation. As discussed therein, the main utility of the large deviation approach is that it allows one to capture the tail behavior in the regime of $O(n)$ deviations away from the mean, while the Gaussian is restricted to capturing deviations which are close to $O(1)$.

Obtaining a clear and rigorous understanding of the distributional behavior indicated above appears difficult with random matrix theoretic tools which are currently well known to information and communication theorists, such as those based on the Stieltjes transform [29] and the replica method [27], [32]. To address this problem, in our recent work [30], we introduced a powerful methodology involving orthogonal polynomials and their so-called ladder operators which led to a new and convenient exact characterization of the MGF of the mutual information in terms of a Painlevé differential equation. This valuable representation provides the machinery for systematically capturing the finite-n corrections to the asymptotic Gaussian results, and thereby investigating deviations from Gaussian and correcting for them. Some related subsequent work for interference-limited multiuser MIMO systems was also presented in [34]. However, while the Painlevé MGF representation in [30] laid the platform for further analysis, the main focus of the analysis presented therein was restricted to the case $n_t = n_r$. Moreover, even for this case, a rigorous treatment of the large-deviation region considered in [33] was not pursued.

In this paper, we significantly expand upon our existing studies in [30] to provide a much broader characterization and understanding of the MIMO mutual information distribution. Starting with the exact Painlevé representation for the mutual information MGF derived in [30], for sufficiently large n (but not assuming that $n_t = n_r$), we systematically compute new closed-form expansions for the higher order cumulants of the mutual information distribution. These series expansions reveal interesting fundamental differences between the two cases, $n_t \neq n_r$ and $n_t = n_r$. Notably, it is demonstrated that the typical approach of considering only the leading-order terms in the large-n series expansions for each cumulant (e.g., (1) and (2), respectively, for the mean and variance) is relatively stable for the asymmetric case $n_t \neq n_r$ compared with the symmetric case $n_t = n_r$, when the SNR P becomes large. This is because the correction series for each cumulant (i.e., comprising all terms other than the leading n term) is shown to converge to a bounded constant as P increases, and therefore becomes quite small relative to the leading term when P is large. For the symmetric case, on the other hand, the situation is very different—the correction series for each cumulant not only grows with P, but also at a faster rate than the leading term, a phenomenon which was discussed at length in [30]. These results allow us to provide a technical explanation for the intriguing large-SNR phenomena observed in Fig. 1(a) and (b).

In addition to gaining fundamental insight into the behavior of the mutual information distribution, we also provide new accurate analytical characterizations for the distribution by employing two different approaches, each being useful in their own region of interest. First, we draw upon the Edgeworth expansion technique along with our derived cumulant expressions to give a simple refined closed-form distribution approximation which is shown to be very accurate around the mean and also for certain moderate deviations into the tails. These results generalize previous Edgeworth expansions which were derived for the specific case of $n_t = n_r$ in [30]. It is shown that in contrast to the Gaussian approximation, which is only capable of successfully characterizing near $O(1)$ deviations around the mean (or the *bulk*) as n increases, the Edgeworth expansion captures deviations of up to $O(n^{\epsilon})$, where $0 < \epsilon \leq 1$. Here, ϵ is a key parameter which, as it increases, allows the approximation to capture the correct distribution further into the tails; however, it also requires the addition of more cumulants to be included in the Edgeworth series, thereby increasing the complexity. In the extreme case, as $\epsilon \to 1$, *all* cumulants must be included; thus, the Edgeworth technique becomes unwieldy and alterna-

tive methods are required. In this scenario, in order to capture such "large deviations" into the tail, we exploit the saddle point method along with asymptotic integration tools to give further analytical representations. Very simple formulas are obtained for the cases of high and low SNR which, taken together, are shown to be very accurate over almost the entire range of SNR values. We point out that our saddle point results provide, in effect, an alternative characterization to the results proposed in [33], which also considered the "large deviation" regime, focusing on $O(n)$ deviations from the mean. Our results, however, are based on a more rigorous footing, stemming from the exact Painlevé representation for the mutual information MGF, as opposed to intuitive statistical physics analogies. Quite surprisingly, they are also found to be simpler.

As a final step, to further emphasize the utility of our results and methodology, we use our analytical framework to extract the well-known diversity-multiplexing-tradeoff (DMT) formula of Zheng and Tse [35] which relates to the large deviation region in the left tail at large SNR, while also deriving similar results for the right tail. These results are found to be consistent with those obtained via the Coulomb fluid analogy in [33].

The rest of this paper is structured as follows. Section II describes the system model under consideration and introduces the Painlevé representation for the MGF of the mutual information, which is the key tool underpinning our analysis. Then, in Section III, we provide a systematic derivation of the mutual information cumulants to leading order in n, as well as finite-n correction terms, giving closed-form expressions in both cases. Based on these new results, Section IV analyzes the n-asymptotic Gaussian approximation at large SNR, revealing fundamental differences between the two scenarios: $n_t = n_r$ and $n_t \neq n_r$. In Section V, we draw upon our cumulant expressions and the Edgeworth expansion technique to provide a refined approximation to the mutual information distribution, both around the mean and for moderate deviations into the tails. Subsequently, Section VI exploits the saddle point method and asymptotic integration tools to characterize the "large deviation" region, extracting key behavior such as the DMT. In Appendix I, we describe some additional details of the machinery to derive the cumulants expressions, while in Appendix II, correspondences are drawn between our analytical framework and the Coulomb fluid large deviation method used in [33].

II. System Model

Considering a point-to-point communication system with n_t transmit and n_r receive antennas, under flat-fading, the linear MIMO channel model takes the form:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n} \tag{3}$$

where $\mathbf{y} \in \mathbb{C}^{n_r}$ and $\mathbf{x} \in \mathbb{C}^{n_t}$ denote the received and transmitted vector, respectively, while $\mathbf{H} \in \mathbb{C}^{n_r \times n_t}$ represents the channel matrix, and $\mathbf{n}_{n_r \times 1} \in \mathbb{C}^{n_r}$ represents noise. Assuming rich scattering, $\mathbf{H}$ is modeled as Rayleigh fading, having IID entries $h_{i,j} \sim \mathcal{CN}(0,1)$, known to the receiver only. The noise is assumed $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{n_r})$. The input is selected to be the ergodic-capacity-achieving input distribution $\mathbf{x} \sim \mathcal{CN}\left(0, \frac{P}{n_t}\mathbf{I}_{n_t}\right)$, where P is the transmitted power constraint and also represents the SNR due to the normalized noise. For the model under consideration, the mutual information between the channel input and output is [17]

$$\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) = \begin{cases} \ln\det\left(\mathbf{I}_n + \frac{P}{m}\mathbf{H}\mathbf{H}^\dagger\right), & n_t \geq n_r \\ \ln\det\left(\mathbf{I}_n + \frac{P}{n}\mathbf{H}^\dagger\mathbf{H}\right), & n_t < n_r. \end{cases} \tag{4}$$

According to (4), we can assume $n_t \geq n_r$ without loss of generality; otherwise, if $n_t < n_r$, we only need replace P with βP.

Throughout this paper, we make the well-known assumption that the channel exhibits block-fading, such that the fading coefficients vary independently from one coding block to another, but remain constant for the duration of each block. In this case, the outage probability becomes an important performance indicator, and the "capacity-versus-outage" tradeoff comes into play [36]. Quantifying this tradeoff requires the entire distribution of the mutual information (4). Consequently, a common approach is to investigate the MGF of $\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})$:

$$\begin{aligned}\mathcal{M}(\lambda) &:= E\left[\exp\left(\lambda\mathcal{I}(\mathbf{x};\boldsymbol{y}|\boldsymbol{H})\right)\right] \\ &= E_{\mathbf{H}}\left\{\left[\det\left(\mathbf{I}_n + \frac{P}{m}\mathbf{H}\mathbf{H}^\dagger\right)\right]^\lambda\right\}. \end{aligned} \tag{5}$$

With this, the cumulant generating function (CGF) can be expressed as a power series about $\lambda = 0$:

$$\mathcal{K}(\lambda) := \ln\mathcal{M}(\lambda) = \sum_{\ell=1}^{\infty}\kappa_\ell\frac{\lambda^\ell}{\ell!} \tag{6}$$

where the coefficient κ_ℓ is the ℓth cumulant of $\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})$.

To lay the foundation of our analysis, we first quote the following exact representation for the MGF (5) (or equivalently, the CGF) in [30]; a key result derived by the authors by drawing upon methods from random matrix theory (see, e.g., [37]–[39]).

Proposition 1: The MGF (5) admits the following compact representation:

$$\mathcal{M}(\lambda) = \exp\left(\int_{\infty}^{\beta/P}\frac{G_n(x)}{x}dx\right) \tag{7}$$

where $G_n(x)$ satisfies a version of the Painlevé V continuous σ–form:

$$\begin{aligned}\left(xG_n''\right)^2 =& n^2\left(xG_n' + G_n'\left(n+m+\lambda\right)/n - G_n\right)^2 \\ &- 4\left(xG_n' - G_n + nm\right)\left(G_n'^2 + \lambda n G_n'\right), \end{aligned} \tag{8}$$

with $'$ denoting the derivative with respect to (w.r.t.) x.

While an explicit solution to the Painlevé V differential equation does not exist in general, we will show in the next section that it can be used to great effect to extract deep insight into the behavior of the mutual information distribution. In particular, we will use it to provide a systematic method for computing closed-form expressions for the leading-order and correction terms to the mean, variance, and higher order cumulants, as the numbers of antennas grow large. This will allow us to obtain more accurate characterizations than the asymptotic Gaussian approximation, which is based on only the leading-order terms of the mean and variance.

III. Systematic Derivation of the Mutual Information Statistics

In this section, we further elaborate upon our work in [30, Sec. IV], by making use of the Painlevé representation (i.e., Proposition 1) to obtain closed-form expressions for the higher order cumulants to leading order in n, as well as first-order correction terms, which apply for *arbitrary* β. Through these cumulants, we may then obtain insights into the shape of the distribution, for example, studying its "Gaussianity." Furthermore, the cumulant information may also be employed to give closed-form approximations for the outage probability, avoiding explicit computation of the inverse MGF (i.e., an inverse Laplace Transform). A key contribution in this section is the generalization of the results in [30, Sec. IV], beyond the case $\beta = 1$. Based on these results, we will see that the mutual information distribution exhibits some stark differences in the two scenarios: $\beta = 1$ and $\beta > 1$.

For a preliminary study, one may start by assuming that P is small, and use standard methods to obtain an asymptotic series solution to (8). This, in turn, directly gives an expression for the CGF $\mathcal{K}(\lambda)$ as a power series of P, which upon rearranging in terms of $\lambda^\ell, \ell = 1, 2, \ldots$ allows us to simply read off the asymptotic (i.e., small-P) cumulants. This procedure is explained in detail in [30, Sec. IV], which considered the case $\beta = 1$; thus, we do not further elaborate upon it here. A key consequence of the small-P study is that it reveals that the cumulants, when expanded in n, admit the interesting structure:

$$\begin{aligned}
\mu &= n\,C_{1,0} + \frac{1}{n}C_{1,1} + \frac{1}{n^3}C_{1,2} + \frac{1}{n^5}C_{1,3} \cdots \\
\sigma^2 &= C_{2,0} + \frac{1}{n^2}C_{2,1} + \frac{1}{n^4}C_{2,2} + \cdots \\
\kappa_3 &= \frac{1}{n}C_{3,0} + \frac{1}{n^3}C_{3,1} + \frac{1}{n^5}C_{3,2} \cdots \\
\kappa_4 &= \frac{1}{n^2}C_{4,0} + \frac{1}{n^4}C_{4,1} + \cdots \\
&\vdots \qquad\qquad \ddots \qquad\qquad \ddots
\end{aligned} \tag{9}$$

Here, $C_{i,j}$ denotes the term of order $1/n^{i-2+2j}$ for the ith cumulant, which is dependent on P and β (but not n). This structure gives an important general understanding of the mutual information distribution in terms of cumulants, and also paves the way for our arbitrary P analysis. As seen from (9), as n grows large with other parameters fixed, only the leading terms of the mean and variance survive; hence, the mutual information distribution approaches Gaussian with increasing n. This agrees with previous results; see, e.g., [19], [28], and [29].

We now present a general approach to computing the coefficients $C_{i,j}$ in closed form for arbitrary P. We first present a generalization of [30, Sec. IV-D], in order to establish a recursive computation for all cumulants to leading order in n [i.e., generating $C_{\ell,0}, \ell = 1, 2, \ldots$ along the leading diagonal in (9)]. Then, we obtain a recursion for systematically computing the first-order correction terms for all cumulants, allowing us to derive closed-form expressions for $C_{\ell,1}, \ell = 1, 2, \ldots$ along the second main diagonal in (9). This approach was not adopted in [30], and thus, our analysis presents new results for both $\beta = 1$ and $\beta > 1$. This, in turn, allows us to capture improved accuracy of each cumulant for finite n.

In addition to computing the terms along the diagonals of (9), we also demonstrate the recursion for systematically computing the horizontal power series expansion in (9) for a given cumulant, to arbitrary degree of accuracy. This, in essence, is a generalization of [30, Sec. IV-B], to arbitrary β. We present results for the mean by way of example, with the other cumulant series' being computable in a similar way.

A. Cumulants to Leading Order in n

Recalling the definition of the cumulants in (6) and their large-n series structure in (9), we start by making the replacement $\lambda \to n\lambda$ and taking n large, such that $G_n(x)$ admits

$$G_n(x) = n^2 Y(x) + O(1) \tag{10}$$

where $Y(x) = \lambda y_1(x) + \lambda^2 y_2(x) + \cdots$, independent of n. Here, $y_\ell(x)$ corresponds to the leading-order term (in n) for the ℓth cumulant by the following integral[1]:

$$C_{\ell,0} = \ell! \int_\infty^{\beta/P} \frac{y_\ell(x)}{x} dx, \qquad \ell = 1, 2, \ldots. \tag{11}$$

The remaining challenge is to compute closed-form formulas for $y_\ell(x)$, $\ell = 1, 2, \ldots$, in order to obtain the corresponding $C_{\ell,0}$. Substituting (10) for $G_n(x)$ into (8) (with $\lambda \to n\lambda$) and keeping only the leading-order terms in n (i.e., $O(n^6)$), we obtain the following equation involving[2] $Y(x)$:

$$\left[(x + \beta + \lambda + 1)\,Y' - Y\right]^2 = 4\,(xY' - Y + \beta)\,(Y' + \lambda)\,Y'. \tag{12}$$

Note that this equation completely characterizes all of the cumulants to leading order in n [i.e., based on this, we may compute $C_{\ell,0}, \ell = 1, 2, \ldots$ in the structure (9)]. This can be done upon substituting $Y(x) = \lambda y_1(x) + \lambda^2 y_2(x) + \cdots$ into (12) and matching the coefficients of λ^ℓ on the left and right-hand side. Using this approach, we systematically obtain equations involving $y_\ell(x), \ell = 1, 2, \ldots$ Full details are provided in Appendix I-A, and the result is the recursive expression (13), at the bottom of the next page, for $\ell > 2$, with the initial values

$$y_1(x) = \frac{\sqrt{x^2 + 2\,(\beta+1)\,x + (\beta-1)^2} - x - (\beta+1)}{2}, \tag{14}$$

$$\begin{aligned}
y_2(x) = {} & \frac{x}{2\sqrt{x^2 + 2\,(\beta+1)\,x + (\beta-1)^2}} \\
& - \frac{1}{2}\,\frac{x^2 + (\beta+1)\,x}{x^2 + 2\,(\beta+1)\,x + (\beta-1)^2}.
\end{aligned} \tag{15}$$

[1]Note that here and in the following, we implicitly assume that the summation and integration may be exchanged. Since we are effectively dealing with asymptotic series in n with unknown coefficients, it seems difficult to formally prove the validity of this assumption. Nonetheless, as we will see in the next section, our numerical results agree very well with simulations, providing strong empirical justification.

[2]Here and henceforth, without causing confusion, we will omit the argument x when presenting differential or difference (recursion) equations.

Here, $y_1(x)$ and $y_2(x)$ correspond to the leading terms of the mean and variance. Integrating these two terms via (11), the closed-form expressions for $C_{1,0}$ and $C_{2,0}$ can be obtained, respectively. With these expressions, we find that $nC_{1,0}$ and $C_{2,0}$ agree precisely with the n-asymptotic mean and variance of the mutual information [i.e., μ_0 in (1) and σ_0^2 in (2)], which correspond to the large-n Gaussian approximation.

Beyond the mean and variance, the higher order cumulants are characterized by the case $\ell > 2$. Evaluating and investigating these cumulants is important, since it allows one to study deviations from Gaussian for finite dimensions. As we will see, these higher order cumulants may also be used to "refine" the Gaussian approximation to provide increased accuracy. Note that the right-hand side of (13) depends only on the previously computed $y_i(x),\ i < \ell$; thus, this recursion equation gives us $y_\ell(x)$ for any value of ℓ in closed form. Here, we simply write down $y_3(x)$ and $y_4(x)$ in (16) and (17), at the bottom of the page, respectively, by way of example. Integrating $y_3(x), y_4(x)$ via (11), we obtain $C_{3,0}$ and $C_{4,0}$ in (9) in closed-form, which are given by (18) and (19), at the bottom of the page, respectively. As this procedure continues, we can systematically compute the leading-order expressions (in n) for any desired number of higher order cumulants, in sequence.

B. Cumulants to Second Order in n

In addition to computing the cumulants to leading order in n, it is also of interest to compute the correction terms (in n), to capture deviations and achieve higher accuracy at finite n. To this end, similar to before, we consider

$$G_n(x) = n^2 Y(x) + Z(x) + O\left(n^{-2}\right) \tag{20}$$

where $Y(x)$ is defined as in (10) while $Z(x) = \lambda z_1(x) + \lambda^2 z_2(x) + \cdots$, independent of n. Here, $z_\ell(x)$ corresponds to

$$y_\ell = \frac{1}{2\left[(x+\beta+1)\,y_1' - y_1\right]}\Bigg\{2\sum_{i=1}^{\ell-1} y_i'\left[\sum_{j=1}^{\ell-i}\left(2xy_j'y_{\ell-i-j+1}' - 2y_j'y_{\ell-i-j+1}\right) + (x-\beta-1)\,y_{\ell-i}' - y_{\ell-i}\right]$$
$$- \sum_{i=1}^{\ell-2} y_i'y_{\ell-i-1}' - \sum_{i=2}^{\ell-1}\left\{\left[x^2+2(\beta+1)x+(\beta-1)^2\right]y_i'y_{\ell-i+1}' - 2(x+\beta+1)\,y_i'y_{\ell-i+1} + y_iy_{\ell-i+1}\right\}\Bigg\}, \quad \ell > 2 \tag{13}$$

$$y_3(x) = -\frac{1}{2}\,x\,(x+1-\beta)\,(x-1+\beta)\left(\frac{x+\beta+1-\sqrt{x^2+2(\beta+1)x+(\beta-1)^2}}{\left(x^2+2(\beta+1)x+(\beta-1)^2\right)^{5/2}}\right) \tag{16}$$

$$y_4(x) = \frac{x}{2}\frac{3x^4+(\beta+1)x^3-(\beta-1)^2\left[6x^2+3(\beta+1)x-(\beta-1)^2\right]}{\left(x^2+2(\beta+1)x+(\beta-1)^2\right)^{7/2}}$$
$$+\frac{x}{2}\frac{-3x^5-4(1+\beta)x^4+(\beta-1)^2\left[5x^3+9(\beta+1)x^2+2(\beta+1)^2x-(\beta+1)(\beta-1)^2\right]}{\left(x^2+2(\beta+1)x+(\beta-1)^2\right)^4} \tag{17}$$

$$C_{3,0} = \frac{\beta+1}{2\beta} - \frac{3\beta P}{\beta^2+2(\beta+1)\beta P+(\beta-1)^2P^2} - \frac{1}{2}\frac{(\beta-1)^2\left[\left(\beta^2+1\right)P^3+3\beta(\beta+1)P^2\right]+3\beta^2\left(\beta^2+1\right)P+\beta^3(\beta+1)}{\beta\left(\beta^2+2(\beta+1)\beta P+(\beta-1)^2P^2\right)^{3/2}} \tag{18}$$

$$C_{4,0} = -\frac{\beta^2+1}{2\beta^2} + \frac{18\beta^4+28\beta^3(\beta+1)P+(\beta-1)^2\left[3\beta^2P^2-6\beta(\beta+1)P^3+(\beta-1)^2P^4\right]}{\left(\beta^2+2(\beta+1)\beta P+(\beta-1)^2P^2\right)^3}P^2 + \Big[\left(\beta^2+1\right)\beta^5$$
$$+5\beta^4(\beta+1)\left(\beta^2+1\right)P+\left(10\beta^4+10\beta^3-16\beta^2+10\beta+10\right)\beta^3P^2+10\beta^2(\beta+1)\left(\beta^2+\beta+1\right)(\beta-1)^2P^3$$
$$+\beta\left(5\beta^4+14\beta^2+5\right)(\beta-1)^2P^4+(\beta+1)(\beta-1)^6P^5\Big]\frac{1}{2\beta^2\left(\beta^2+2(\beta+1)\beta P+(\beta-1)^2P^2\right)^{5/2}} \tag{19}$$

the first-order correction term for the ℓth cumulant [i.e., characterized by $C_{\ell,1}, \ell = 1, 2, \ldots$ in (9)] by the following integral:

$$C_{\ell,1} = \ell! \int_{\infty}^{\beta/P} \frac{z_\ell(x)}{x} dx, \qquad \ell = 1, 2, \ldots. \tag{21}$$

Substituting (20) into (8) (with $\lambda \to n\lambda$) and setting the second leading-order terms (in n) to be zero, we obtain the equation involving $Z(x)$:

$$\begin{aligned} &\Big\{2\beta\lambda - \left[\lambda^2 + 2(1+\beta-x)\lambda + x^2 + 2(1+\beta)x + (\beta-1)^2\right] Y' \\ &+ (x - \lambda + \beta + 1)Y - 4Y'Y + 6x(Y')^2\Big\} Z' \\ &= \left(2(Y')^2 - (x + \beta + 1 - \lambda) Y' - Y\right) Z - \frac{x^2}{2}(Y'')^2. \end{aligned} \tag{22}$$

This equation captures the *exact* first-order corrections to all leading-order cumulant approximations. Substituting $Y(x) = \lambda y_1(x) + \lambda^2 y_2(x) + \cdots$ and $Z(x) = \lambda z_1(x) + \lambda^2 z_2(x) + \cdots$ into (22) and matching the coefficients of λ^ℓ, we are able compute the $z_\ell(x)'$s systematically. Full details can be found in Appendix I-B, giving the recursive formula (23), at the bottom of the page, for $\ell > 2$, with the initial values

$$z_1(x) = \frac{-\beta x^2}{\left(x^2 + 2(\beta+1)x + (\beta-1)^2\right)^{5/2}} \tag{24}$$

and $z_2(x)$ given by (25), at the bottom of the page.

With this result, we can compute $z_\ell(x)$ in closed form for any value of ℓ. For example, $z_3(x)$ is expressed as (26), at the bottom of the page. Integrating $z_\ell(x)$ via (21), we obtain the first-order correction terms in closed form. For example, $C_{1,1}$ and $C_{2,1}$ are given as (27) and (28), at the bottom of the next page. respectively. The first-order correction terms to the higher cumulants $\kappa_3, \kappa_4, \ldots$ (i.e., $C_{3,1}, C_{4,1}, \ldots$) are omitted here in order to keep the presentation concise, though they follow trivially.

Note that if we aim to compute further high-order correction terms (i.e., $C_{i,j}, i = 1, 2, \ldots, j \geq 2$) for each cumulant, we can invoke the same procedure as what has been used for computing the first two leading-order terms, which only takes more algebraic effort.

$$z_\ell = \frac{\sum_{i=1}^{\ell-1} z_i \left[2\sum_{j=1}^{\ell-i} y_j' y_{\ell-i-j+1}' + y_{\ell-i}' - (x+\beta+1) y_{\ell-i+1}' + y_{\ell-i+1}\right] - \frac{x^2}{2}\sum_{i=1}^{\ell} y_i'' y_{\ell-i+1}''}{(x+\beta+1) y_1' - y_1}, \quad \ell > 2 \tag{23}$$

$$\begin{aligned} z_2(x) = &\frac{\left[(x+\beta+1)y_1' - y_1\right] z_1 + \left[(x+\beta+1)y_2' - 2(y_1')^2 - y_2 - y_1'\right]}{(x+\beta+1)y_1' - y_1} \\ &+ \frac{x^2}{2} \frac{(\beta+1)x^3 + \left(3\beta^2 - 14\beta + 3\right)x^2 + (3\beta-1)(\beta-3)(\beta+1)x + \left(\beta^2 + 10\beta + 1\right)(\beta-1)^2}{\left(x^2 + 2(\beta+1)x + (\beta-1)^2\right)^4} \end{aligned} \tag{25}$$

$$\begin{aligned} z_3(x) = &-\frac{x^2}{2\left(x^2+2(\beta+1)x+(\beta-1)^2\right)^6}\Big[-x^7 + 8(\beta+1)x^6 + \left(49\beta^2 - 38\beta + 49\right)x^5 + 6(\beta+1)\left(15\beta^2 - 38\beta + 15\right)x^4 \\ &+ \left(65\beta^4 - 76\beta^3 - 42\beta^2 - 76\beta + 65\right)x^3 + 4(\beta+1)\left(\beta^2 + 26\beta + 1\right)(\beta-1)^2 x^2 - \left(17\beta^2 - 46\beta + 17\right)(\beta-1)^4 x - 6(\beta+1)(\beta-1)^6\Big] \\ &- \frac{x^2}{2\left(x^2 + 2(\beta+1)x + (\beta-1)^2\right)^{11/2}}\Big[x^6 - 9(\beta+1)x^5 - 8\left(5\beta^2 - 11\beta + 5\right)x^4 - 10(\beta+1)\left(5\beta^2 - 13\beta + 5\right)x^3 \\ &- \left(15\beta^4 + 134\beta^3 - 330\beta^2 + 134\beta + 15\right)x^2 + (\beta+1)\left(11\beta^2 - 100\beta + 11\right)(\beta-1)^2 x + 6\left(\beta^2 + 5\beta + 1\right)(\beta-1)^4\Big] \end{aligned} \tag{26}$$

C. Large-n Series Computation for a Given Cumulant

In addition to the machinery used in the previous section for systematically computing the high-order cumulants one after the other for a given order in n, we can also derive recursions for systematically computing the large-n series for a *given* cumulant. This procedure was also considered in [30] for the specific case $\beta = 1$. The generalization to arbitrary β, however, is quite straightforward, and thus, here we simply present the result using the mean as an example. Specifically, we evaluate $C_{1,\ell},\ \ell = 0, 1, \ldots$ via the following integral:

$$C_{1,\ell} = \int_{\infty}^{\beta/P} \frac{Q_\ell(x)}{x} dx, \qquad \ell = 1, 2, \ldots \tag{29}$$

where Q_ℓ admits the general recursion (30) at the bottom of the page with $\ell \geq 2$ and $Q_0(x), Q_1(x)$ equal to (14) and (24), respectively. Note that this recursion is a generalization of [30, eq. (132)], for arbitrary β. We write $C_{1,2}$ in (31), at the bottom of the page, as an example, and omit the presentation of the higher order correction formulas for the sake of conciseness. In the same way, systematic expansions for the variance and higher order cumulants can also be computed to arbitrary degree of accuracy.

IV. Large SNR Analysis of the Gaussian Deviation

The systematic high-order cumulant expansions derived in the previous section allow us to closely investigate the behavior of the mutual information distribution under various conditions. For example, our analysis in Section III-A recovered the fact that as the number of antennas n grows large, with the SNR P kept fixed, the distribution approaches a Gaussian. However, in practice, the number of antennas is finite and is not typically huge (though some recent trends have considered such systems [40]–[44]), while the SNR may vary significantly depending on the application. Thus, a natural question is, for asymptotically high SNR, if and by how much the mutual information distribution deviates from Gaussian.

In this section, we focus on the high-SNR regime of the mutual information. In this case, as shown in Fig. 1(a) and (b), the

$$C_{1,1} = -\frac{\beta+1}{24\beta} + \frac{\beta^3(\beta+1) + 3\beta^2(\beta+1)^2 P + 3\beta(\beta+1)\left(\beta^2+1\right)P^2 + \left(\beta^2+1\right)(\beta-1)^2 P^3}{24\beta\left(\beta^2 + 2\beta(\beta+1)P + (\beta-1)^2P^2\right)^{3/2}} \tag{27}$$

$$\begin{aligned} C_{2,1} = {} & \frac{\beta^2+1}{24\beta^2} - \frac{4(\beta+1)\beta^3P^3 + \left(9\beta^2-42\beta+9\right)\beta^2P^4 + 6\beta(\beta+1)(\beta-1)^2P^5 + (\beta-1)^4P^6}{12\left(\beta^2 + 2\beta(\beta+1)P + (\beta-1)^2P^2\right)^3} \\ & - \Bigg[\left(\beta^2+1\right)\beta^7 + 7\beta^6(\beta+1)\left(\beta^2+1\right)P + 7\beta^5\left(\beta^2+1\right)\left(3\beta^2+4\beta+3\right)P^2 + (\beta+1)\left(35\beta^4+62\beta^2+35\right)\beta^4P^3 \\ & + \left(35\beta^6+9\beta^4+8\beta^3+9\beta^2+35\right)\beta^3P^4 + (\beta+1)(\beta-1)^2\left(21\beta^4-14\beta^3-2\beta^2-14\beta+21\right)\beta^2P^5 \\ & + 7(\beta+1)^2(\beta-1)^6\beta P^6 + (\beta+1)(\beta-1)^8P^7\Bigg] \frac{1}{24\beta^2\left(\beta^2 + 2\beta(\beta+1)P + (\beta-1)^2P^2\right)^{7/2}} \end{aligned} \tag{28}$$

$$Q_\ell(x) = \frac{\sum_{i=1}^{\ell-1}\left[Q_{\ell-i}Q_i + \left(x^2 + 2(\beta+1)x + (\beta-1)^2\right)Q'_{\ell-i}Q'_i - 2(x+\beta+1)Q_{\ell-i}Q'_i\right] - x^2\sum_{i=0}^{\ell-1}Q''_{\ell-i-1}Q''_i}{2\left[(x+\beta+1)Q'_0 - Q_0\right]}, \quad \ell \geq 2 \tag{30}$$

$$\begin{aligned} C_{1,2} = {} & \frac{1+\beta^3}{240\beta^3} - \frac{1}{240\beta^3\left((\beta-1)^2P^2 + 2(\beta+1)P + \beta^2\right)^{9/2}}\Bigg[9(\beta+1)\left(\beta^6-3\beta^5+3\beta^4+3\beta^2-3\beta+1\right)(\beta-1)^4\beta P^8 \\ & + 18(\beta+1)\left(2\beta^2+3\beta+2\right)\left(\beta^2-\beta+1\right)\beta^7P^2 + 42\left(2\beta^2+\beta+2\right)\left(\beta^2-\beta+1\right)(\beta+1)^2\beta^6P^3 \\ & + 42\left(2\beta^2+\beta+2\right)\left(\beta^2-\beta+1\right)(\beta+1)^2\beta^5P^4 + \left(126+126\beta^8-384\beta^4+126\beta^5+126\beta^3\right)\beta^4P^5 \\ & + 6\left(6\beta^8-9\beta^7-3\beta^6+9\beta^5+56\beta^4+9\beta^3-3\beta^2-9\beta+6\right)(\beta-1)^2\beta^2P^7 + \left(1+\beta^3\right)\beta^9 \\ & + 6(\beta+1)\left(14\beta^8-35\beta^7+35\beta^6-21\beta^5-16\beta^4-21\beta^3+35\beta^2-35\beta+14\right)\beta^3P^6 \\ & + 9\left(\beta^2-\beta+1\right)(\beta+1)^2\beta^8P + \left(\beta^6-3\beta^5+3\beta^4+3\beta^2-3\beta+1\right)(\beta-1)^6P^9\Bigg] \end{aligned} \tag{31}$$

distribution appears to deviate from Gaussian as the SNR becomes large, while the deviation appears to be much *stronger* for the case of equal numbers of transmit and receive antennas (i.e., $n_t = n_r$) compared with the case $n_t \neq n_r$. This is an interesting observation which thus far has resisted theoretical explanation. Here, we draw insight into this phenomenon by employing our new closed-form high-order cumulant expansions given in the previous section.

Interestingly, by taking P large in (1), (2), (18), (27), and (28), it turns out that we obtain very different limiting results depending on whether $\beta = 1$ (i.e., $n_t = n_r$) or not. For $\beta = 1$, as P grows large, we obtain[3]:

$$\mu \sim n \ln P + \frac{1}{16}\frac{\sqrt{P}}{n} + \frac{3}{1024}\left(\frac{\sqrt{P}}{n}\right)^3 + \frac{45}{32768}\left(\frac{\sqrt{P}}{n}\right)^5 + \cdots \tag{32}$$

$$\sigma^2 \sim \frac{1}{2}\ln\left(\frac{P}{2}\right) + \frac{1}{32}\left(\frac{\sqrt{P}}{n}\right)^2 + \cdots \tag{33}$$

$$\kappa_3 \sim \frac{1}{4n} + \frac{1}{32}\left(\frac{\sqrt{P}}{n}\right)^3 + \cdots. \tag{34}$$

These large-n large-P series expansions were also documented in [30], which focused primarily on the case $n_t = n_r$. From these expressions, we see that for each cumulant, the correction terms grow with P, and indeed grow *faster* than the leading-order terms. As such, approximating each cumulant via their leading-order term will be inaccurate when P is large.

Now, considering $\beta \neq 1$, we obtain

$$\mu \sim n\left[\ln P - (\beta-1)\ln\left(\frac{\beta-1}{\beta}\right) - 1\right] + \frac{1}{n}\left[\frac{1}{12(\beta-1)} - \frac{1}{12\beta}\right] + \frac{1}{n^3}\left[-\frac{1}{120(\beta-1)^3} + \frac{1}{120\beta^3}\right] + \frac{1}{n^5}\left[-\frac{1}{252(\beta-1)^5} + \frac{1}{252\beta^5}\right] + \cdots \tag{35}$$

$$\sigma^2 \sim \ln\left(\frac{\beta}{\beta-1}\right) + \frac{1}{n^2}\left[-\frac{1}{12(\beta-1)^2} + \frac{1}{12\beta^2}\right] + \cdots \tag{36}$$

$$\kappa_3 \sim \frac{1}{n}\left[-\frac{1}{(\beta-1)} + \frac{1}{\beta}\right] + \frac{1}{n^3}\left[\frac{1}{6(\beta-1)^3} - \frac{1}{6\beta^3}\right] + \cdots. \tag{37}$$

From these, we can make some key observations, which we summarize in the following remarks:

Remark 1: In contrast to the results in (32)–(34), for $n_t \neq n_r$, all terms other than the leading term of μ are strictly bounded as P increases, converging to constants depending on β. Moreover, as n or β increases, the correction terms have less effect, eventually becoming negligible, even when P is very large.

Remark 2: For $n_t \neq n_r$, the n-asymptotic power series representations for the cumulants remain valid for arbitrary SNR P, while in contrast, for $n_t = n_r$, they break down for sufficiently large P. This clearly indicates that the commonly assumed Gaussian approximation (based on the leading-order terms of μ and σ only) is quite robust at high SNRs for the case $n_t \neq n_r$, but not for the case $n_t = n_r$.

These remarks are illustrated in Fig. 3(a)–(c), showing σ^2 and κ_3. Fig. 3(a) shows that when $n_t = n_r$, the leading-order term of σ^2 (i.e., representing the variance of the Gaussian approximation) and the leading-order term of κ_3 both deviate strongly from simulations when P is sufficiently large. In contrast, when $n_t \neq n_r$ [i.e., Fig. 3(b) and (c)], the leading-order cumulants capture the simulations accurately for arbitrary P, even with smaller n. We also see that increasing β enhances the accuracy of the leading-order results, while including the first-order correction term provides improved accuracy as well. These observations explain technically why the Gaussian approximation, based purely on the leading-order terms of μ and σ^2, is relatively more robust to increasing SNRs in Fig. 1, compared with the results in Fig. 1(b). Nevertheless, even in the former case, the higher cumulants [e.g., of order $O(1/n)$] still contribute to some deviations from Gaussian. These deviations become particularly significant in the tail region, yielding the Gaussian approximation unsuitable for capturing low outage probabilities of practical interest. This will be considered further in the subsequent sections, where we will make use of our closed-form cumulant expansions to "correct" for these deviations, both in the tail and around the mean, and thereby refine the Gaussian approximation.

Before presenting these refinements, we would like to briefly connect our large-n-large-P cumulant expansions with existing results, which were obtained via different means by considering n fixed and taking P large at the beginning. This approach was used to obtain the leading-order terms of μ and σ^2 in [28], and these results coincide exactly with the leading-order terms in (32) and (33). In addition, the first few terms of the mean and variance expansions (i.e., (35) and (36), respectively) match exactly with the results in [45, eq. (134)], and [46, Lemma A.2]. For example, [45] has the large-P mean representation:

$$\mu \sim n\ln\left(\frac{P}{m}\right) - n + m\psi(m) - (m-n)\psi(m-n), \quad P \to \infty \tag{38}$$

which gives (35) by expanding the digamma function $\psi(x)$ at large x as

$$\psi(x) = \ln x - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} - \frac{1}{252x^6} + O\left(\frac{1}{x^8}\right). \tag{39}$$

Our results can also be shown to be consistent with the large-P MGF representation in [24].

V. Characterization With the Edgeworth Expansion

Armed with closed-form expressions for $C_{i,j}$ in (9), in this section, we draw upon the Edgeworth expansion technique. This approach allows us to start with a Gaussian distribution and to systematically correct this by including higher cumulant effects (i.e., other than the mean and variance), giving an explicit expression for the corrected PDF. Moreover, the CDF, which directly defines the outage probability, can be obtained explicitly through a straightforward integration.

[3]Note that the correction term for the third cumulant κ_3 [i.e., $C_{3,1}$ in (9)] was obtained in the derivation described in Section III-B, but was not presented explicitly in this paper for the sake of conciseness.

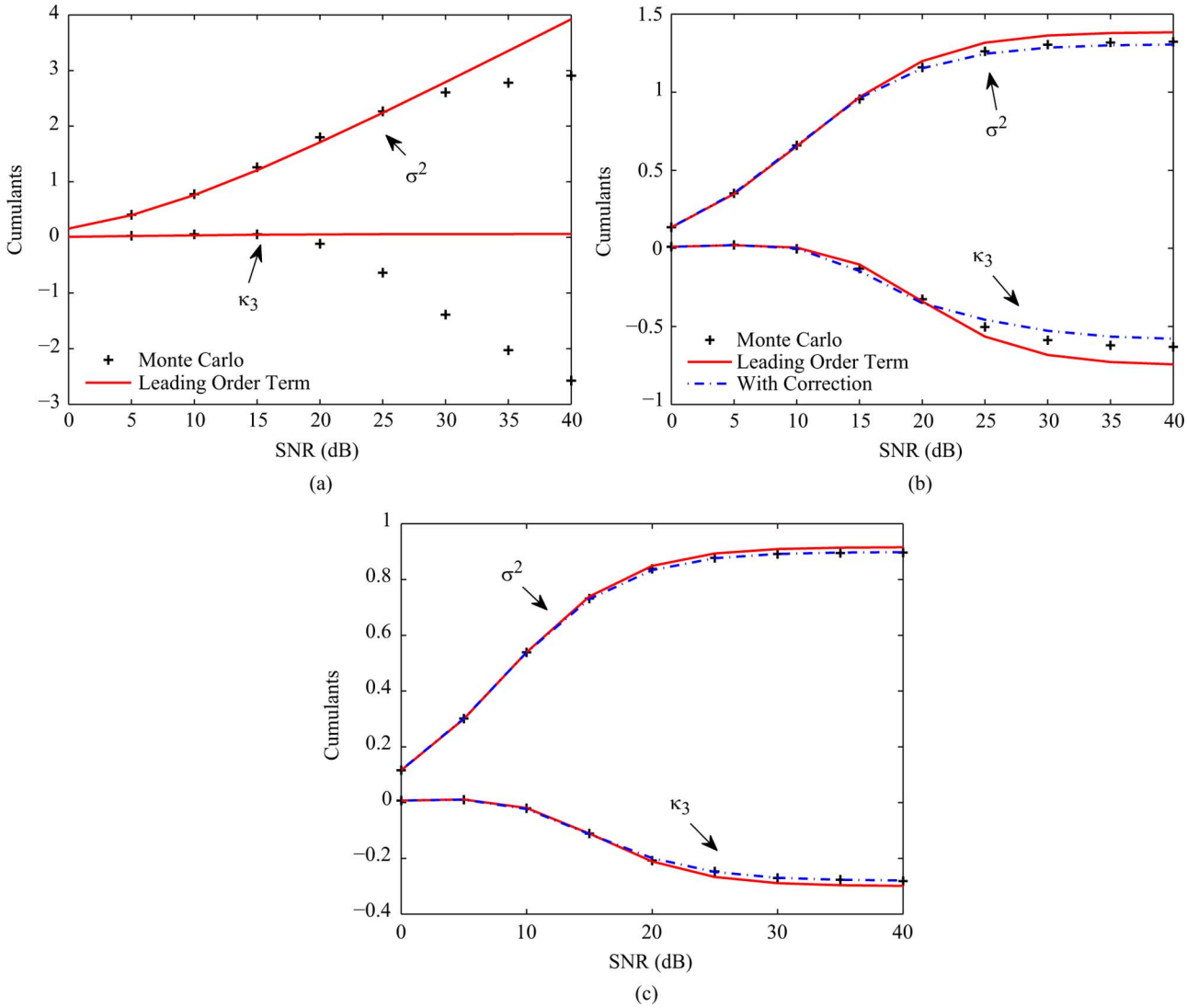

Fig. 3. Comparison of σ^2 to leading order (in n) and with first-order correction, κ_3 to leading order and with first-order correction, and Monte Carlo simulations. Results are shown for different antenna configurations. (a) $n_t = 4, n_r = 4$; (b) $n_t = 4, n_r = 3$; (c) $n_t = 5, n_r = 3$.

The PDF of the mutual information in Edgeworth expansion takes the form [47, eq. (43)]:

$$p_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(t) \approx \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-z^2/2} \left[1 + \mathcal{D}(z)\right] \tag{40}$$

where $z := (t - \mu)/\sigma$ and

$$\mathcal{D}(z) = \sum_{s=1}^{L} \sum_{\{k\}} \frac{\mathrm{He}_{s+2r}(z)}{\sigma^{s+2r}} \prod_{\ell=1}^{s} \frac{1}{k_\ell!} \left(\frac{\kappa_{\ell+2}}{(\ell+2)!} \right)^{k_\ell} \tag{41}$$

is the quantity which determines any deviation from Gaussian. Here, L is a positive integer characterizing how many cumulants are included in the corrected PDF (i.e., $\kappa_\ell, \ell = 3, 4, \ldots, L+2$ are involved). Note that the second summation enumerates all sets $k = \{k_1, k_2, \ldots, k_s\}$ containing the nonnegative integer solutions of the Diophantine equation $k_1 + 2k_2 + \cdots + sk_s = s$. For each of these solutions, a corresponding constant r is defined as $r = k_1 + k_2 + \cdots + k_s$. In [47], a practical algorithm for computing the $\{k\}$ solutions is proposed in general. $\mathrm{He}_\ell(z)$ is the ℓth Chebyshev–Hermite polynomial, with the explicit form [47, eq. (13)]

$$\mathrm{He}_\ell(z) = \ell! \sum_{k=0}^{\lfloor \ell/2 \rfloor} \frac{(-1)^k z^{\ell-2k}}{k!(\ell-2k)!2^k} \tag{42}$$

where $\lfloor \cdot \rfloor$ denotes the floor function. These are generated by differentiating the standard normal distribution:

$$\mathrm{He}_\ell(z) = (-1)^\ell g_0^{(\ell)}(z)/g_0(z) \tag{43}$$

where $g_0(z) = \exp\left(-z^2/2\right)/\sqrt{2\pi}$ and the superscript (ℓ) denotes the ℓth derivative w.r.t. z. In theory, we can approximate the mutual information distribution with arbitrary accuracy, by taking L sufficiently large in $\mathcal{D}(z)$. We first set $L = 2$ (i.e., including κ_3 and κ_4) and compare the Edgeworth expansion with the Monte Carlo simulations and the Gaussian approximation in Fig. 4. Note that for the Edgeworth expansion curves, we only use the leading-order terms of the cumulants [i.e., $C_{\ell,0}/n^{2-\ell}$ in (9)], since as we have shown, the leading-order terms of the

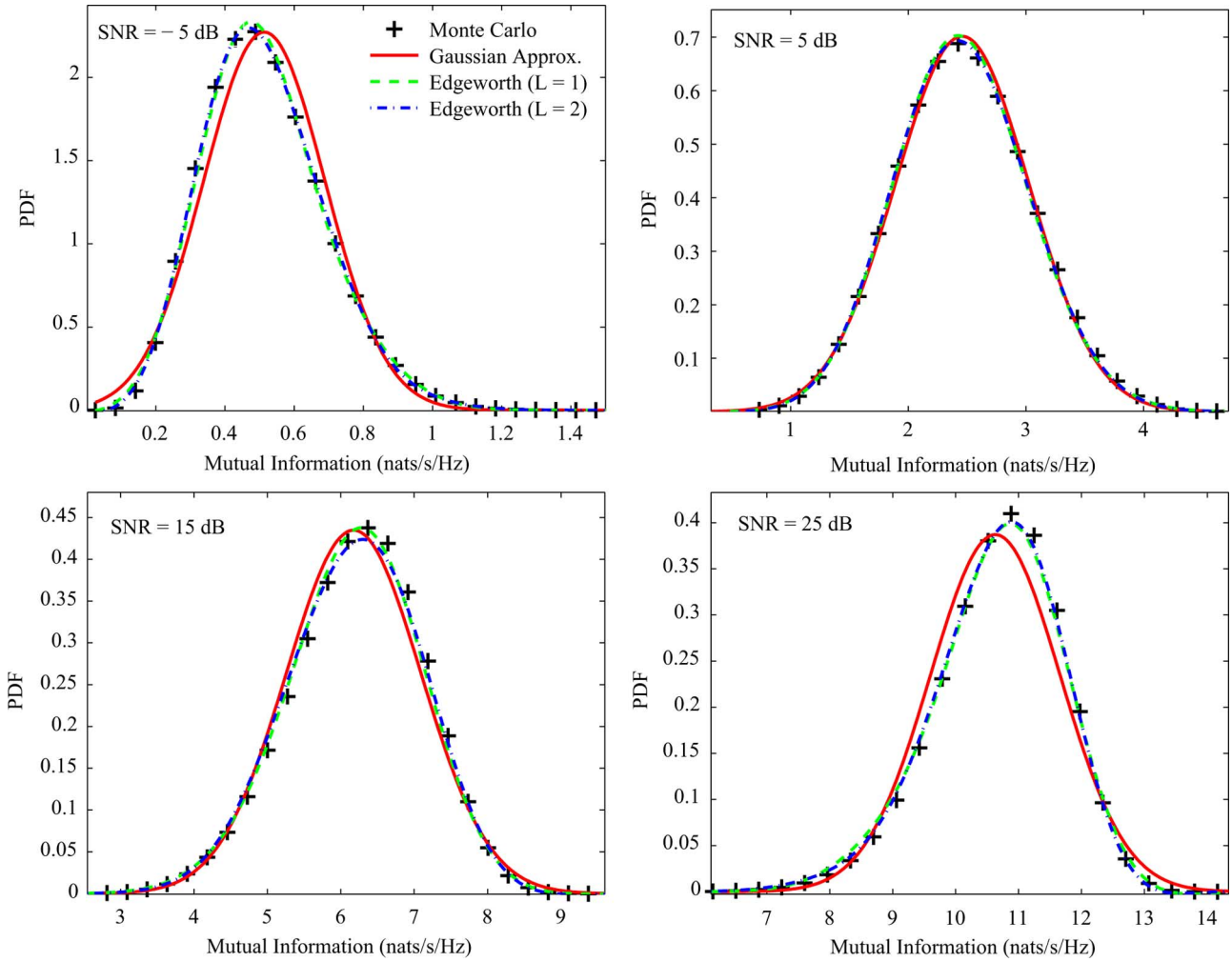

Fig. 4. PDF of mutual information, comparing the Gaussian approximation, Edgeworth expansions (with $L = 1$ and $L = 2$), and Monte Carlo simulations. Results are shown for $n_t = 3, n_r = 2$ and for different SNRs.

first few cumulants give valid approximations for arbitrary P if $\beta \neq 1$ (cf., *Remark 2*).

To examine the outage probability, we need to derive the CDF of the mutual information. Recalling (43), we have

$$\int_{-\mu/\sigma}^{z} \mathrm{He}_\ell(u)\, g_0(u)\, du \approx \int_{-\infty}^{z} \mathrm{He}_\ell(u)\, g_0(u)\, du = \mathrm{He}_{\ell-1}(z) g_0(z). \tag{44}$$

Therefore, the Edgeworth PDF formula (40) can be immediately integrated to give the CDF:

$$F_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(t) \approx 1 - Q(z) - g_0(z) \sum_{s=1}^{L} \sum_{\{k\}} \frac{\mathrm{He}_{s+2r-1}(z)}{\sigma^{s+2r}} \prod_{\ell=1}^{s} \frac{1}{k_\ell!} \left(\frac{\kappa_{\ell+2}}{(\ell+2)!} \right)^{k_\ell} \tag{45}$$

where $Q(z) := \frac{1}{\sqrt{2\pi}} \int_z^\infty e^{-u^2/2} du$. Comparisons between the CDF curves computed by the Edgeworth expansion and the Gaussian approximation are made in Fig. 5. We see that in the tail region, the Edgeworth expansions with higher cumulants nicely approach the simulations (in this case, $L = 6$ was required to achieve good accuracy), while the Gaussian curve strongly deviates. This confirms that, by the virtue of our new cumulant expressions and the Edgeworth expansion, we can compute the outage probability of MIMO systems with high accuracy in closed form for arbitrary SNR.

Moreover, from these numerical tests, for small SNR (i.e., $\mathrm{SNR} = -10\,\mathrm{dB}, 0\,\mathrm{dB}$), we see that the Gaussian approximation overestimates the outage probability in the tail, while for large SNR, the outage probability is underestimated. These tail deviations can be analyzed by examining the points far away from the mean (i.e., $z \to \infty$). Consequently, Chen and McKay [30] proposed to take $z \to \infty$ in (41) and collect the dominant terms (letting $L \to \infty$); the PDF of the mutual information became [30, eq. (217)]:

$$p_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(t) \approx \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(t-\mu)^2}{2\sigma^2} + \frac{\kappa_3}{6\sigma^6}(t-\mu)^3} \tag{46}$$

which gives a more accurate distribution than the Gaussian in the tails for finite n, since the effect of κ_3 is considered. However, Chen and McKay [30] only investigated (46) for the case where $n_t = n_r$. Here, we provide more general insights with

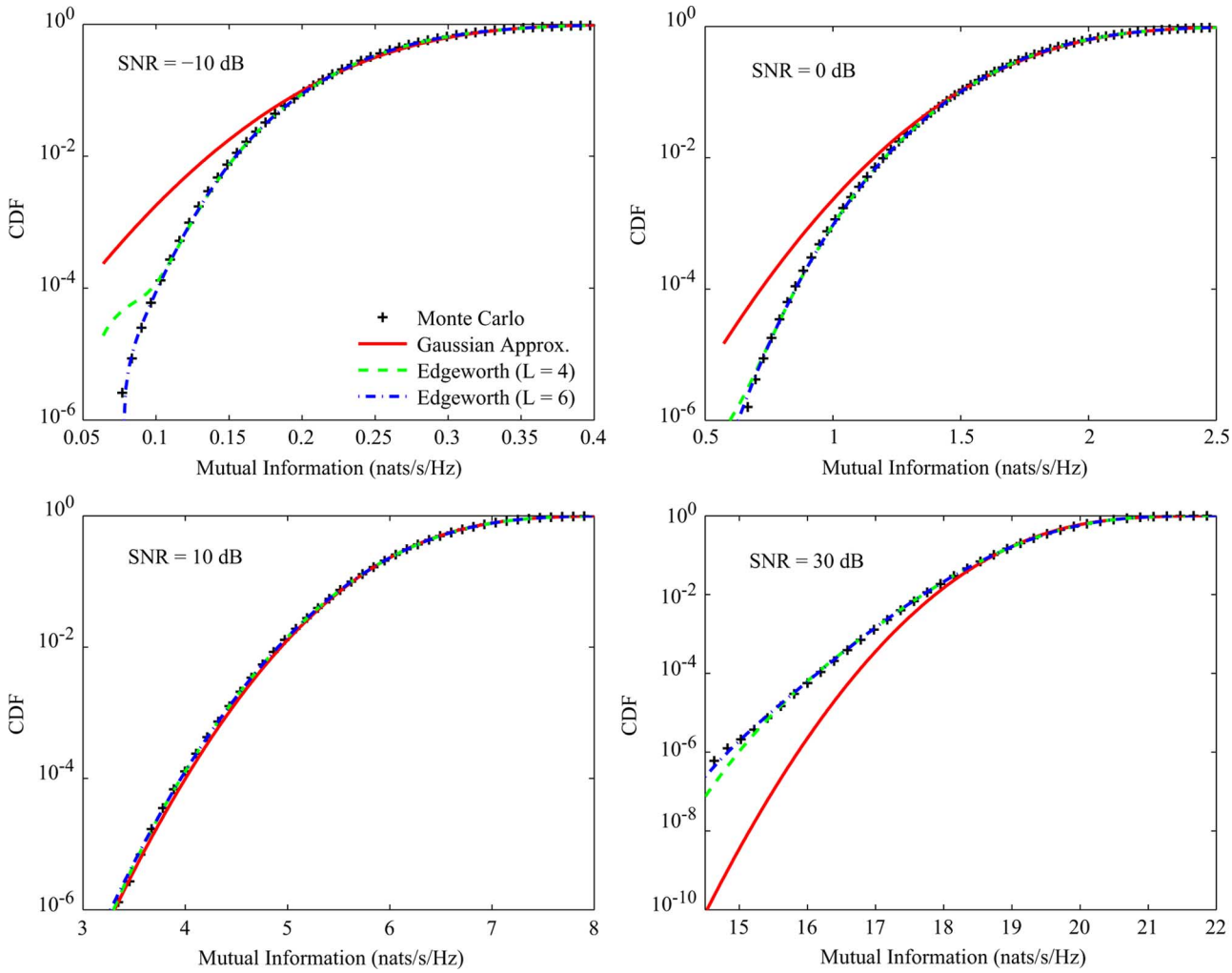

Fig. 5. CDF of mutual information, comparing the Gaussian approximation, Edgeworth expansions (with $L = 4$ and $L = 6$), and Monte Carlo simulations. Results are shown for $n_t = 6, n_r = 3$ and for different SNRs.

our new cumulant expressions which apply for both $n_t = n_r$ and $n_t \neq n_r$.

First, (46) indicates that the Gaussian approximation is accurate when n grows large under the condition:

$$|t-\mu| \ll \sqrt[3]{\left|\frac{6\sigma^6}{\kappa_3}\right|} \approx \sqrt[3]{\left|\frac{6C_{2,0}^3}{C_{3,0}}\right|}\, n^{1/3}. \tag{47}$$

Thus, the large-n Gaussian approximation is valid for deviations of $|t-\mu| \sim O(n^{1/3})$ or less from the mean. Outside this regime, the second term of the exponent in (46) becomes nonnegligible. In this case, by examining the factor κ_3/σ^6 in (46) to leading order in n (i.e., $\frac{\kappa_3}{\sigma^6} \approx \frac{C_{3,0}}{nC_{2,0}^3}$), we can further understand how the distribution behaves compared with the Gaussian approximation. To this end, considering $C_{2,0} = \sigma_0^2$ given in (2), and $C_{3,0}$ given in (18), for small P,

$$\frac{C_{3,0}}{nC_{2,0}^3} \approx \frac{2\beta}{n\,P^3}, \qquad P \to 0, \tag{48}$$

while for large P,

$$\frac{C_{3,0}}{nC_{2,0}^3} \approx \begin{cases} \frac{2}{n\,(\ln P)^3}, & \beta = 1 \\ \frac{1}{n\,\beta(\beta-1)(\ln(1-\beta^{-1}))^3}, & \beta > 1 \end{cases}, \qquad P \to \infty. \tag{49}$$

These results are in perfect agreement with [33, eq. (70)] obtained via Coulomb fluid arguments, and also [30, eq. (218)] obtained for the special case $\beta = 1$. Therefore, seen from (46), for small SNR, (48) indicates that $\kappa_3/\sigma^6 > 0$, and the left tail of the PDF should be always above the Gaussian approximation (similarly for the CDF in the left tail). For large SNR and $\beta > 1$ on the other hand, (49) indicates that $\kappa_3/\sigma^6 < 0$, and the situation in the tail is the opposite.

Note that the interpretation of the large SNR results above for the particular case $\beta = 1$ should be taken with caution since, as discussed in Section IV and also in [30], the leading-order expressions for σ^2 and κ_3 (upon which the arguments are based) become inaccurate in that scenario, unless n is also very large. For the case $\beta \neq 1$, however, there is no such problem.

VI. Refining the Tail Distribution via the Saddle Point Method

While the Edgeworth expansion technique provides an accurate closed-form characterization of the mutual information distribution, it becomes unwieldy when too many cumulants are needed for obtaining the desired accuracy. Particularly, in the case when one is interested in the tail region of $O(n)$ away from the mean (i.e., the "large deviation" region discussed in [33]), we need to consider the effect of *all* cumulants to obtain high accuracy. Therefore, to supplement the Edgeworth expansion results, in this section, we will draw upon the saddle point method and the cumulant results from Section III to further investigate the large deviation scenario.

When discussing the saddle point method, it will be important to keep in mind the large-n structure (especially the leading-order behavior) for the cumulants, shown in (9).

A. Preliminaries of the Saddle Point Method and its Implications

Given the MGF (equivalently, the CGF) of the mutual information, its PDF can be derived through the inversion formula:

$$p_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(t) = \frac{1}{2\pi i}\int_{\tau-i\infty}^{\tau+i\infty} e^{\mathcal{K}(\lambda)-\lambda t}d\lambda, \tag{50}$$

where $\mathcal{K}(\lambda)$ is the CGF and τ is a real number determining the integration path (i.e., a line in the complex plane which is parallel to the imaginary axis and crosses the real axis at τ). In most cases, it is infeasible to evaluate (50) in closed form. However, if n is large, the saddle point method in [48] can be used to provide an accurate approximation for this integral. Here, we briefly demonstrate how this asymptotic method applies for our case.

Choose the path of integration to pass through a saddle point $\tau^\star(t)$ which, for a given t, is computed as the real-valued solution to

$$\mathcal{K}'(\tau^\star(t)) - t = 0 \tag{51}$$

where $'$ denotes derivative w.r.t. λ. On this path, the integration variable in (50) is then written as $\lambda = \tau^\star(t) + iy$. Knowing that the integrand attains its maximum modulus at the saddle point, we expand the exponent of the integrand in (50) as a series about this point:

$$\begin{aligned}\mathcal{K}(\lambda) - \lambda t = \mathcal{K}(\tau^\star(t)) - \tau^\star(t)t - \frac{\mathcal{K}''(\tau^\star(t))}{2}y^2 \\ - i\frac{\mathcal{K}'''(\tau^\star(t))}{6}y^3 + \frac{\mathcal{K}''''(\tau^\star(t))}{24}y^4 + \cdots.\end{aligned} \tag{52}$$

Setting $y = \nu/\sqrt{\mathcal{K}''(\tau^\star(t))}$ in (52) and substituting into the inversion formula(50), we have

$$\begin{aligned}p_{\mathcal{I}(x;y|H)}(t) = \frac{e^{\mathcal{K}(\tau^\star(t))-\tau^\star(t)t}}{2\pi\sqrt{\mathcal{K}''(\tau^\star(t))}} \\ \times \int_{-\infty}^{\infty} e^{-\frac{\nu^2}{2} - i\frac{\mathcal{K}'''(\tau^\star(t))}{6(\sqrt{\mathcal{K}''(\tau^\star(t))})^3}\nu^3 + \frac{\mathcal{K}''''(\tau^\star(t))}{24(\sqrt{\mathcal{K}''(\tau^\star(t))})^4}\nu^4 + \cdots} d\nu.\end{aligned} \tag{53}$$

To establish the asymptotic behavior of (53) as $n \to \infty$, we first need to characterize the asymptotic behavior of the saddle point $\tau^\star(t)$. To this end, using the power series expansion of the CGF (6), we get

$$\mathcal{K}'(\tau^\star(t)) = \mu + \sigma^2\tau^\star(t) + \frac{\kappa_3}{2}(\tau^\star(t))^2 + \frac{\kappa_4}{3!}(\tau^\star(t))^3 + \cdots \tag{54}$$

and therefore, (51) implies that

$$t - \mu = \sigma^2\tau^\star(t) + \frac{\kappa_3}{2}(\tau^\star(t))^2 + \frac{\kappa_4}{3!}(\tau^\star(t))^3 + \cdots. \tag{55}$$

The asymptotic behavior of the saddle point $\tau^\star(t)$ is thus characterized by the asymptotic behavior of $t-\mu$, i.e., how far one deviates from the mean. To this end, we will consider

$$t = \mu + an^\epsilon, \tag{56}$$

for $0 \le \epsilon \le 1$ and a a real constant independent of n. Recalling that as $n \to \infty$, $\sigma^2 = O(1)$, while $\kappa_\ell = O(n^{2-\ell})$ for $\ell \ge 3$, (55) immediately leads to

$$\tau^\star(\mu + an^\epsilon) = O(n^\epsilon). \tag{57}$$

(Note that since $\mu = O(n)$, it does not make sense to consider $\epsilon > 1$.) Furthermore, since the saddle point scales at a maximum rate of $O(n)$ (i.e., for $\epsilon = 1$), we may now characterize the asymptotic behavior of the higher order derivative terms in (53) as

$$\begin{aligned}&\mathcal{K}''(\tau^\star(t)) = \sigma^2 + \kappa_3\tau^\star(t) + \frac{\kappa_4}{2}(\tau^\star(t))^2 + \cdots = O(1),\\ &\mathcal{K}'''(\tau^\star(t)) = \kappa_3 + \kappa_4\tau^\star(t) + \frac{\kappa_5}{2}(\tau^\star(t))^2 + \cdots = O\left(\frac{1}{n}\right),\\ &\quad\vdots\\ &\mathcal{K}^{(\ell)}(\tau^\star(t)) = O\left(\frac{1}{n^{\ell-2}}\right),\\ &\quad\vdots\end{aligned}$$

Thus, as $n \to \infty$, all the terms of the exponent in (53) vanish except for the leading term $-\frac{\nu^2}{2}$. This term is trivially integrated, giving the following PDF expression to leading order in n:

$$p_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(t) \approx \frac{e^{-I(t)}}{\sqrt{2\pi\mathcal{K}''(\tau^\star(t))}} \tag{58}$$

where

$$I(t) := t\tau^\star(t) - \mathcal{K}(\tau^\star(t)) \tag{59}$$

is the so-called *rate function*. Equations (58) and (59) together define the saddle point approximation that we wish to compute in the sequel.

Before going further, it is instructive to investigate the behavior of the asymptotic PDF obtained via the saddle point method, as one considers varying deviations about the mean. More specifically, we will examine the asymptotic PDF for deviations $|t-\mu| = O(n^\epsilon)$, with ϵ varying in the range $[0,1]$.

To this end, we start by substituting the CGF series expansion (6) into (59) giving

$$I(t) = (t-\mu)\tau^\star(t) - \frac{\sigma^2}{2}(\tau^\star(t))^2 - \frac{\kappa_3}{3!}(\tau^\star(t))^3 - \frac{\kappa_4}{4!}(\tau^\star(t))^4 - \cdots, \quad (60)$$

which along with (57) and (55) allows us to draw the following remarks:

1) *Constant deviations*, $|t-\mu| = O(1)$: We know from (57) that $\tau^\star(t) = O(1)$, or more specifically from (55) that $\tau^\star(t) \approx (t-\mu)/\sigma^2$. Substituting this solution into (60) yields

$$I(t) = (t-\mu)^2/(2\sigma^2) - (t-\mu)^3\kappa_3/(3!\sigma^6) - (t-\mu)^4\kappa_4/(4!\sigma^8) - \cdots. \quad (61)$$

Since $|t-\mu| = O(1)$, all terms in (61) involving higher cumulants $(\kappa_\ell, \ell \geq 3)$ vanish as $n \to \infty$, resulting in the Gaussian exponent. This is the region where, as argued in [33], the central limit theorem is valid, and thus, the Gaussian approximation is asymptotically accurate.

2) *Sublinear deviations*, $|t-\mu| = O(n^\epsilon), 0 < \epsilon < 1$: Here, we know from (57) that $\tau^\star(t) = O(n^\epsilon)$. Once again, as $n \to \infty$, upon substituting $t = \mu + an^\epsilon$ on the left-hand side of (55), we obtain the saddle point $\tau^\star(t) \approx (t-\mu)/\sigma^2$, which gives the same $I(t)$ as (61). However, since $|t-\mu| = O(n^\epsilon)$, in this case, some of the terms in (61) involving the higher cumulants $(\kappa_\ell, \ell \geq 3)$ do not vanish as n grows large. More specifically, this includes all cumulants κ_ℓ for which $(\ell-2)/\ell \leq \epsilon$. That is, for large n, κ_3 becomes effective (provides a nonnegligible contribution) for deviations of $O(n^{1/3})$ or more, κ_4 is effective for deviations of $O(n^{1/2})$ or more, κ_5 is effective for deviations of $O(n^{3/5})$ or more, and so on. This behavior is consistent with the discussion in the previous section and in [30]. For example, considering $|t-\mu| = O(n^\epsilon)$, $\epsilon \in [1/3, 1/2)$ and taking $n \to \infty$, (61) behaves asymptotically as

$$I(t) \approx (t-\mu)^2/(2\sigma^2) - (t-\mu)^3\kappa_3/(3!\sigma^6), \quad (62)$$

which is the same exponent as in (46). In these scenarios, our Edgeworth expansion results presented in the previous section provide an accurate closed-form approximation, by accounting for a fixed number of higher cumulant effects. However, when the deviations become stronger (e.g., as ϵ increases toward 1), more and more terms are required in the Edgeworth expansion in order to account for the increasing number of nonnegligible high-order cumulants, thereby significantly increasing the computational complexity.

3) *Linear deviations*, $|t-\mu| = O(n)$: Here, (57) indicates that $\tau^\star(t) = O(n)$. In this case, all cumulants to leading order in n (i.e., an infinite number) contribute in (55) and (60). Therefore, no simple asymptotic expression of $\tau^\star(t)$ or $I(t)$ can be derived as in (61)–(62). (Nonetheless, (55) and (60) tell us that $I(t) = O(n^2)$ in this case, which is crucial information for our further study.) In this "large deviation" region which is sufficiently deep in the tails of the distribution, the Gaussian approximation strongly misses the correct behavior, while the Edgeworth expansion in (45) also becomes intractable.

Importantly, the above discussions provide a unified picture of the mutual information distribution for large-n. To be specific, they show that the well-known Gaussian approximation, our Edgeworth expansion approximation, and the large deviation results (also considered in [33]) have their own region of validity, depending on how far one looks into the tail of the distribution as n increases. While the Gaussian approximation and the Edgeworth expansion have been well characterized in the previous sections, further work is required in relation to the large deviation region, i.e., $|t-\mu| = O(n)$ (meanwhile, $I(t) = O(n^2)$ from the second bullet point above), which is considered in the sequel.

B. Computation of the Saddle Point Approximation

We note that a key problem encountered with (58) is how to compute $I(t)$. In addition, while in principle the CDF of the mutual information can be obtained by integrating the PDF expression (58) (i.e., $F_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(t) = \int_0^t p_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(u)du$), this is difficult and generally does not yield a closed-form expression. We will first address this integration problem, then deal with the problem of computing the rate function $I(t)$.

As discussed previously, for the large deviation region of interest, $|t-\mu| \sim O(n)$, the rate function $I(t)$ in (58) is of order $O(n^2)$, and thus can be explicitly written as $n^2 I_0(x)$. Hence, the Laplace integration method is admissible for computing the n-asymptotic CDF (see e.g., [49, Ch. 2]). In particular, since $-I(u)$ is an increasing function for $u \leq \mu$ (decreasing function for $u > \mu$), to leading order in n, the integration of (58) is dominated by the region in the neighborhood of the upper limit t for $u \leq \mu$ (lower limit t for $u > \mu$). Thus, we expand $I(u)$ about t as

$$I(u) = I(t) + (u-t)I'(t) + \frac{(u-t)^2}{2}I''(t) + O((u-t)^3). \quad (63)$$

Additionally, in the neighborhood of t, the function $\mathcal{K}''(\tau^\star(u))$ is nearly constant and can be approximated by its value at t. Based on these arguments (see [49, Ch. 2] for more details), we obtain the approximated integral (64), shown at the bottom of the next page. By neglecting the terms in the exponent of order $(u-t)^2$ or higher, we have

$$F_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(t) \approx \begin{cases} -\frac{e^{-I(t)}}{I'(t)\sqrt{2\pi\mathcal{K}''(\tau^\star(t))}}, & t \leq \mu \\ 1 + \frac{e^{-I(t)}}{I'(t)\sqrt{2\pi\mathcal{K}''(\tau^\star(t))}}, & t > \mu. \end{cases} \quad (65)$$

If one were to instead neglect the terms of order $(u-t)^3$ or higher in (64), then the (less) asymptotic formula proposed in [33, eqs. (63) and (64)] results. However, it is important to point out that (65) and [33, eqs. (63) and (64)] each involve the derivatives of $I(t)$ [e.g., $I'(t)$ and $I''(t)$], which are difficult to handle, both analytically and computationally. Thus, it is of interest to derive a more manageable representation, which is now pursued.

Recall that the Gaussian approximation is accurate around the bulk of the distribution (which captures the vast majority of the area under the PDF curve), while the absolute difference

between the true PDF and the Gaussian approximation is typically small outside of the bulk (since the PDF naturally takes extremely small values in this region). Thus, the Gaussian exponent (i.e., $(t-\mu)^2/(2\sigma^2)$) should capture the leading effect in $I(t)$. As such, we expand the integrand around the mean (equivalently, around $\tau^\star(t)=0$), giving:

$$F_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(t) \approx \int_0^t \frac{1}{\sqrt{2\pi\mathcal{K}''(0)}} e^{-\frac{(u-\mu)^2}{2\sigma^2}-O((u-\mu)^3)}du = \frac{1}{\sqrt{2\pi}\sigma}\int_0^t e^{-\widetilde{I}(u)}e^{-\frac{(u-\mu)^2}{2\sigma^2}}du \tag{66}$$

where

$$\widetilde{I}(u) := I(u) - \frac{(u-\mu)^2}{2\sigma^2} \quad . \tag{67}$$

Noting that $|\widetilde{I}(u)| \ll (u-\mu)^2/(2\sigma^2)$, we adopt Laplace's expansion method again as follows. Since $(u-\mu)^2/\sigma^2$ is a decreasing function for $u \le \mu$ (increasing function for $u > \mu$), to leading order in n, the integral (66) is dominated by the region in the neighborhood of the upper limit t for $u \le \mu$ (lower limit t for $u > \mu$). In this neighborhood, the function $\widetilde{I}(u)$ can be regarded as effectively a constant and can be approximated by its value at t, which leads to the following:

$$\frac{1}{\sqrt{2\pi}\sigma}\int_0^t e^{-\widetilde{I}(u)}e^{-\frac{(u-\mu)^2}{2\sigma^2}}du \approx \begin{cases} e^{-\widetilde{I}(t)}\left[\frac{1}{\sqrt{2\pi}\sigma}\int_0^t e^{-\frac{(u-\mu)^2}{2\sigma^2}}du\right], & t \le \mu \\ 1-e^{-\widetilde{I}(t)}\left[\frac{1}{\sqrt{2\pi}\sigma}\int_t^\infty e^{-\frac{(u-\mu)^2}{2\sigma^2}}du\right], & t > \mu. \end{cases} \tag{68}$$

Thus, the CDF of the mutual information is represented by the concise formula

$$F_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(t) \approx \begin{cases} Q\left(-\frac{t-\mu}{\sigma}\right)e^{-\widetilde{I}(t)}, & t \le \mu \\ 1-Q\left(\frac{t-\mu}{\sigma}\right)e^{-\widetilde{I}(t)}, & t > \mu. \end{cases} \tag{69}$$

In the next section, we will show that this asymptotic CDF captures the distributional behavior very accurately.

Now we address the remaining challenge: Computing a manageable expression for $I(t)$, and therefore $\widetilde{I}(t)$ via (67). Recalling the large-n expansion structure in (9), summing up the leading terms of the cumulants (i.e., $C_{\ell,0}/n^{2-\ell}, \ell = 1,2,\ldots$) gives us the n-asymptotic CGF. However, in general, the complexity of the expressions for the $C_{\ell,0}$'s makes it difficult to derive a generic closed-form formula for this asymptotic CGF. Thus, to make solid analytical progress, we focus on the scenarios of small P and large P.

C. Asymptotic CGF at Low SNR (Small P Scenario)

We first consider the n-asymptotic cumulants for small P. In this case, the expansion of the leading order (in n) cumulants (i.e., μ_0 given in (1), σ_0^2 in (2), $C_{3,0}$ in (18) and $C_{4,0}$ in (19), etc.) for small P reveals the generic expression as $P \to 0$:

$$\kappa_\ell \sim (\ell-1)!\, mn\left(\frac{P}{m}\right)^\ell, \quad \ell = 1,2,\ldots. \tag{70}$$

Here, $\kappa_1 = \mu$ and $\kappa_2 = \sigma^2$. Thus, the CGF is obtained as

$$\mathcal{K}(\lambda) \sim \sum_{\ell=1}^{\infty}\frac{mn}{\ell}\left(\frac{P}{m}\lambda\right)^\ell = -mn\ln\left(1-\frac{P}{m}\lambda\right). \tag{71}$$

Interestingly, even without giving such generic expression as (70), we can still recover (71) by scaling and solving the exact equation for the n-asymptotic CGF in (12), where the derivatives are taken w.r.t. x. To this end, given that $Y(x) = \lambda y_1(x) + \lambda^2 y_2(x) + \cdots$, and $y_\ell(x) \sim O(x^{-\ell})$ (indicated by the expressions of $y_\ell(x)$'s computed in Section III), in order to keep the terms to leading order in x of $Y(x)$, we introduce the following variable substitution:

$$\frac{n\lambda}{x} := y. \tag{72}$$

With this, (12) becomes

$$\left[-y\left(1+\frac{\beta+1}{x}+y\right)Y' - Y\right]^2 = 4y\left(yY'+Y-\beta\right)\left(-\frac{y}{x^2}Y'+y\right)Y' \tag{73}$$

where $'$ is the derivative w.r.t. y. Note that here we consider the large x (equivalently, small P) but finite y scenario. Letting $x \to \infty$, we have

$$\left[Y'\left(y^2+y\right)+Y\right]^2 = 4y^2\left(yY'+Y-\beta\right)Y', \tag{74}$$

with the solution

$$Y(y) = \frac{\beta y}{y-1}. \tag{75}$$

Integrating $Y(y)$, we obtain the large-n-small-P CGF:

$$\mathcal{K}(\lambda) \sim n^2\int_\infty^{\beta/P}\frac{Y(\frac{\lambda}{nx})}{x}dx = -mn\ln\left(1-\frac{P}{m}\lambda\right), \quad \lambda < \frac{m}{P} \tag{76}$$

$$\int_0^t \frac{e^{-I(u)}}{\sqrt{2\pi\mathcal{K}''(\tau^\star(u))}}du \approx \begin{cases} \frac{e^{-I(t)}}{\sqrt{2\pi\mathcal{K}''(\tau^\star(t))}}\int_0^t e^{-I'(t)(u-t)-I''(t)(u-t)^2/2+O((u-t)^3)}\,du, & t \le \mu \\ 1-\frac{e^{-I(t)}}{\sqrt{2\pi\mathcal{K}''(\tau^\star(t))}}\int_t^\infty e^{-I'(t)(u-t)-I''(t)(u-t)^2/2+O((u-t)^3)}\,du, & t > \mu. \end{cases} \tag{64}$$

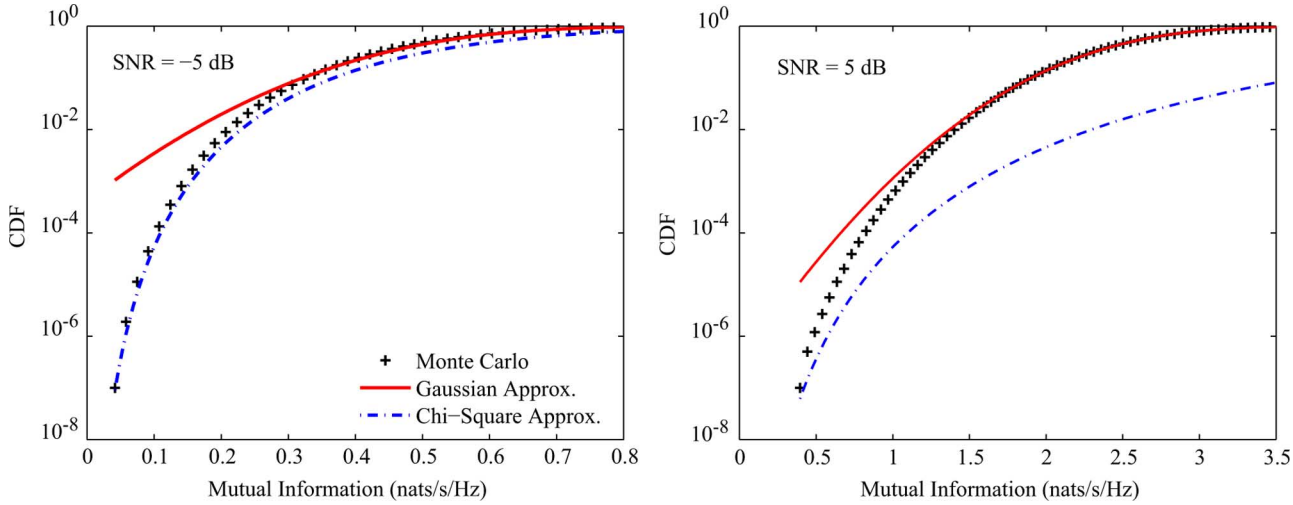

Fig. 6. CDF of mutual information for small SNRs, comparing the Gaussian approximation, chi-square approximation, and Monte Carlo simulations. Results are shown for $n_t = 4$, $n_r = 2$.

in agreement with (71), which was obtained by summing the cumulants. Note that the CGF in (76) corresponds to that of a chi-square distribution, indicating that in this low-P scenario:

$$\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) \overset{d}{\sim} \frac{P}{2m}\chi^2(2mn). \tag{77}$$

This approximation is in fact quite well known, and it can be readily established by noting that $\ln\det(\mathbf{I}_n + \frac{P}{m}\mathbf{H}\mathbf{H}^\dagger) \approx \frac{P}{m}\mathrm{tr}(\mathbf{H}\mathbf{H}^\dagger)$ for small P (i.e., obtained by expanding $\ln\det(\mathbf{I}_n + \frac{P}{m}\mathbf{H}\mathbf{H}^\dagger)$ around $P = 0$, and keeping the first term), and the obvious fact that $\mathrm{tr}(\mathbf{H}\mathbf{H}^\dagger) \overset{d}{\sim} \chi^2(2mn)/2$.

As depicted in Fig. 6, if P is very small (e.g., $P = -5$ dB), then the chi-square approximation lines up quite well with the simulations; however, beyond this very small regime (e.g., for $P = 5$ dB), it is inaccurate. Thus, for greater validity, further refinement beyond the leading-order chi-square approximation is necessary.

This requires computing the higher order correction terms (in P), a task which appears difficult via direct expansion of the $\ln\det(\mathbf{I}_n + \frac{P}{m}\mathbf{H}\mathbf{H}^\dagger)$ formula, as indicated above for the leading-order chi-square approximation. To our knowledge, such refinement has not been computed thus far.

To develop a systematic method for solving this refinement problem, we may once again make use of our Painlevé representation. Noting that x is large while the new variable y is finite, we assume the following large-x expansion:

$$Y(y) = Y_0(y) + \frac{Y_1(y)}{x} + \frac{Y_2(y)}{x^2} + \cdots. \tag{78}$$

Substituting (78) into (12) and matching the coefficient of x^{-k}, we can compute $Y_k(y)$ systematically in closed form. For example,

$$Y_1(y) = -\frac{\beta(\beta+1)y}{(y-1)^3}. \tag{79}$$

Integrating $Y_1(y)$ through (76), we obtain the CGF with first-order correction term (in P):

$$\mathcal{K}(\lambda) \approx -mn\ln\left(1 - \frac{P}{m}\lambda\right) - \frac{n}{2}\frac{(1+1/\beta)\lambda P^2}{(\lambda P/m - 1)^2}. \tag{80}$$

With this, the saddle point (51) is obtained as

$$t = \frac{1}{2(\tau^\star P - m)^3}\Big[mnP\big(-2(\tau^\star)^2P^2 + mnP + m^2P + (nP^2 + mP^2 + 4mP)\tau^\star - 2m^2\big)\Big]. \tag{81}$$

This expression can be solved in closed form, with the resulting expression involving a cubic equation. Alternatively, one may trivially compute the solution numerically for any given value of t. With $\tau^\star$ solved for a given t, the value of the rate function $I(t)$ and thus $\widetilde{I}(t)$ follow immediately according to the definitions (59) and (67) respectively (with μ and σ^2 in (67) approximated via μ_0 and σ_0^2). By invoking the CDF formula (69), we can then compute the saddle point approximation for the mutual information distribution. This approximation is illustrated in Fig. 7. Compared with the chi-square approximation, it is shown that this refined CDF is remarkably accurate, even for moderate P values (e.g., $P = 10$ dB). As P further increases (beyond, for example, $P = 15$ dB), we have found that the saddle point approximation starts to miss the correct behavior, and higher order correction terms (equivalently, $Y_2(y), Y_3(y), \ldots$) are needed. These can be systematically computed using the same procedure as before.

D. Asymptotic CGF at Large SNR (Large P Scenario)

Now we consider the large-n-large-P scenario. Based on the expressions that have been computed for $C_{\ell,0}, \ell = 1, 2, \ldots$ for

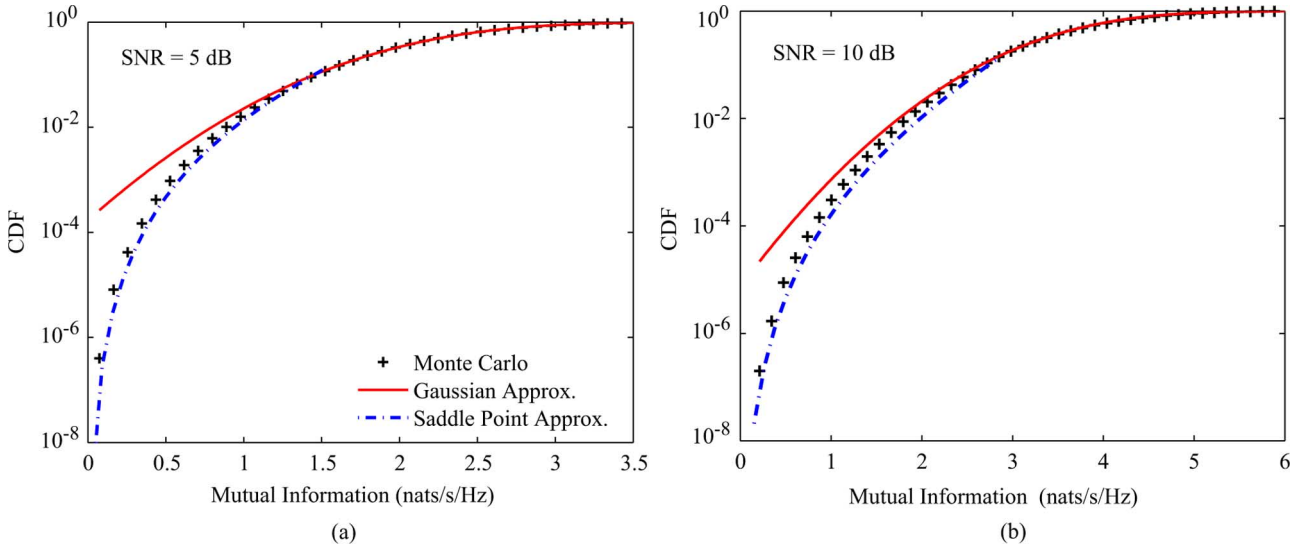

Fig. 7. CDF of mutual information for small SNRs, comparing the Gaussian approximation, saddle point approximation, and Monte Carlo simulations. Results are shown for $n_t = 2$, $n_r = 2$ and for different SNRs.

$P \to \infty$ [e.g., (35)–(37)], we obtain the generic expression for the leading term of the ℓth cumulant ($\ell \geq 3$):

$$\kappa_\ell \sim (-1)^\ell (\ell-3)! \left[\frac{1}{(m-n)^{\ell-2}} - \frac{1}{m^{\ell-2}} \right], \quad \ell = 1, 2, \ldots . \tag{82}$$

With this, upon summing the CGF series, the asymptotic CGF is computed in closed form:

$$\begin{aligned} \mathcal{K}(\lambda) &\approx \mu_0 \lambda + \frac{\sigma_0^2}{2}\lambda^2 \\ &\quad + \sum_{\ell=3}^{\infty} \frac{(\ell-3)!}{\ell!} (-\lambda)^\ell \left[\frac{1}{(m-n)^{\ell-2}} - \frac{1}{m^{\ell-2}} \right] \qquad (83) \\ &= \left(\mu_0 - \frac{n}{2}\right)\lambda + \frac{\sigma_0^2}{2}\lambda^2 + \frac{(\lambda+m)^2}{2} \ln\left(1+\frac{\lambda}{m}\right) \\ &\quad - \frac{(\lambda+m-n)^2}{2} \ln\left(1+\frac{\lambda}{m-n}\right). \qquad (84) \end{aligned}$$

Now, from (1) and (2), we have

$$\mu_0 \approx n \ln P - (m-n) \ln\left(\frac{m-n}{m}\right) - n, \quad P \to \infty, \tag{85}$$

$$\sigma_0^2 \approx \ln\left(\frac{m}{m-n}\right), \qquad P \to \infty, \tag{86}$$

giving the following large-n large-P CGF:

$$\begin{aligned} \mathcal{K}(\lambda) \approx &\left[n \ln\left(\frac{P}{m}\right) - \frac{3}{2}n \right]\lambda + \frac{(m+\lambda)^2}{2}\ln(m+\lambda) - \frac{m^2}{2}\ln m \\ &- \frac{(m-n+\lambda)^2}{2}\ln(m-n+\lambda) + \frac{(m-n)^2}{2}\ln(m-n), \end{aligned} \tag{87}$$

valid for $\lambda \in (n-m, \infty)$.

Remark 3: Interestingly, although (87) was obtained based on the cumulant expressions which are valid only for $\beta \neq 1$ [i.e., since a singularity exists for $\beta = 1$ in the n-asymptotic cumulants, as seen in (35)–(37)], by setting $\beta = 1$, we have

$$\begin{aligned} \mathcal{K}(\lambda) \approx &\, n\lambda \ln\left(\frac{P}{n}\right) - \frac{3}{2}n\lambda + \frac{(n+\lambda)^2}{2}\ln(n+\lambda) \\ &- \frac{\lambda^2}{2}\ln\lambda - \frac{n^2}{2}\ln n. \end{aligned} \tag{88}$$

This will be shown to describe the correct behavior of the mutual information.

The saddle point (51) can be computed by

$$\begin{aligned} t = &\, n\ln P - n\ln m - n + (m+\tau^\star)\ln(m+\tau^\star) \\ &- (m-n+\tau^\star)\ln(m-n+\tau^\star). \end{aligned} \tag{89}$$

While a closed-form solution for (89) is intractable, it can be trivially computed numerically for any given value of t.

By invoking (58) and (69), we can compute the distribution (both the PDF and CDF) of the mutual information. We should point out that t in (89) is a monotonically increasing function of λ (i.e., $dt(\lambda)/d\lambda > 0$); thus, for any $t \geq n\ln(P/\beta) - n$, there exists a real root $\tau^\star$. This, in turn, implies that the distribution cannot be captured explicitly by (89) if t is sufficiently small such that $\mu - t > m\ln(\beta) - (m-n)\ln(\beta-1)$ (i.e., when looking sufficiently far into the left tail region). Nevertheless, the right-hand side of this inequality is increasing with m (fixed n); thus, as m grows, the valid region of (89) extends further into the left tail, allowing smaller outage probabilities to be calculated. In fact, scenarios for which m is reasonably large compared with n is quite realistic in many applications; for example, in cellular systems, for which the base station may be equipped with a reasonably large number of antennas, while the number of antennas on the mobile device is more restricted due to limited space constraints.

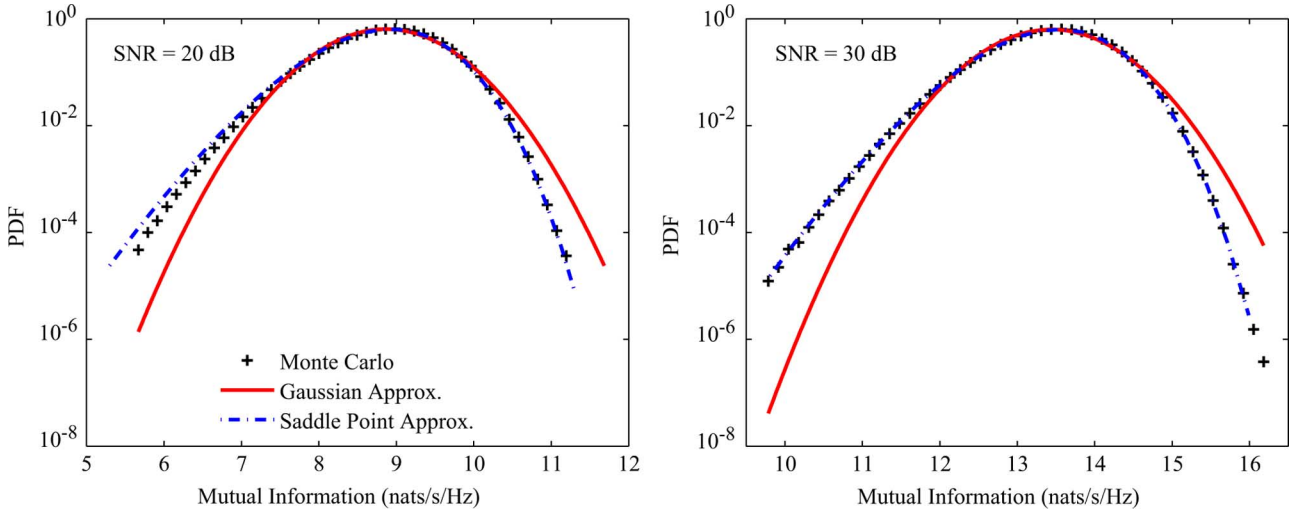

Fig. 8. PDF of mutual information, comparing the Gaussian approximation, saddle point approximation, and Monte Carlo simulations. Results are shown for $n_t = 6,\ n_r = 2$ and for different SNRs.

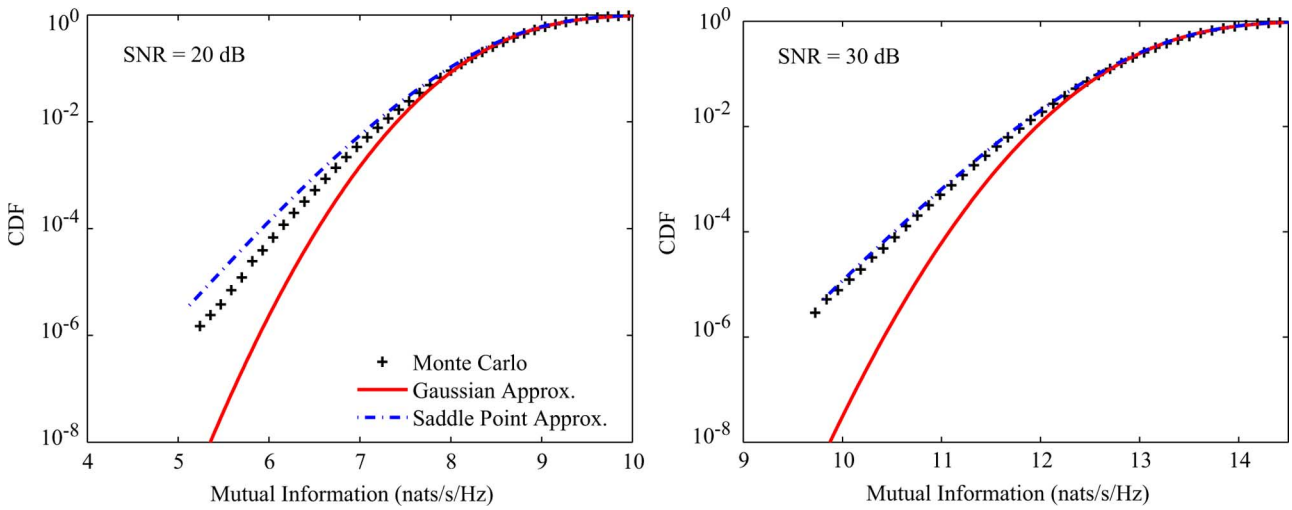

Fig. 9. CDF of mutual information, comparing the Gaussian approximation, saddle point approximation, and Monte Carlo simulations. Results are shown for $n_t = 6,\ n_r = 2$ and for different SNRs.

Fig. 8 depicts the saddle point approximation of the PDF (58) with $I(t)$ computed from (87) and (89), comparing with the Gaussian approximation and Monte Carlo simulations. The saddle point result is clearly much more accurate than the Gaussian, and lines up almost perfectly with the simulations when the SNR is sufficiently large (i.e., at 30 dB). Similar observations are made in Fig. 9, which shows the corresponding CDF curves based on (69) and the same $I(t)$. Note that for these results, we have chosen $m = 6,\ n = 2$, where m is comparatively large enough such that the validity of (89) extends deep enough into the left tail region to capture outage probabilities of interest.

Fig. 10 depicts the complementary CDF (CCDF) of the mutual information, comparing the saddle point approximation based on $I(t)$ computed from (87) and (89), the Gaussian approximation, and Monte Carlo simulations. As for the left tail, we see that the saddle point approximation becomes extremely accurate when the SNR is sufficiently high, and significantly outperforms the Gaussian approximation. In fact, quite surprisingly, even for moderate SNRs of 20 dB, the saddle point approximation in the right-hand tail traces the simulated curve very closely.

As done for the small-SNR scenario, for large SNR, we can also evaluate the higher order correction terms (in P) in order to draw insight into the accuracy of the leading-order results. To this end, it is convenient to first introduce the change of variables $x \to \beta/x$ in the Painlevé representation of the CGF (7) and the large-n(12). The CGF to leading order in n can then be written as

$$\mathcal{K}(n\lambda) \sim -n^2 \int_0^P \frac{Y(\beta/x)}{x} dx, \qquad n \to \infty, \tag{90}$$

with $Y(\beta/x)$ satisfying

$$\left[xY' + \left(1 + \frac{\lambda+1}{\beta}\right) x^2 Y' + Y\right]^2 = 4\left(-xY' - Y + \beta\right)\left(\frac{x^2}{\beta}Y' - \lambda\right)\left(\frac{x^2}{\beta}Y'\right) \tag{91}$$

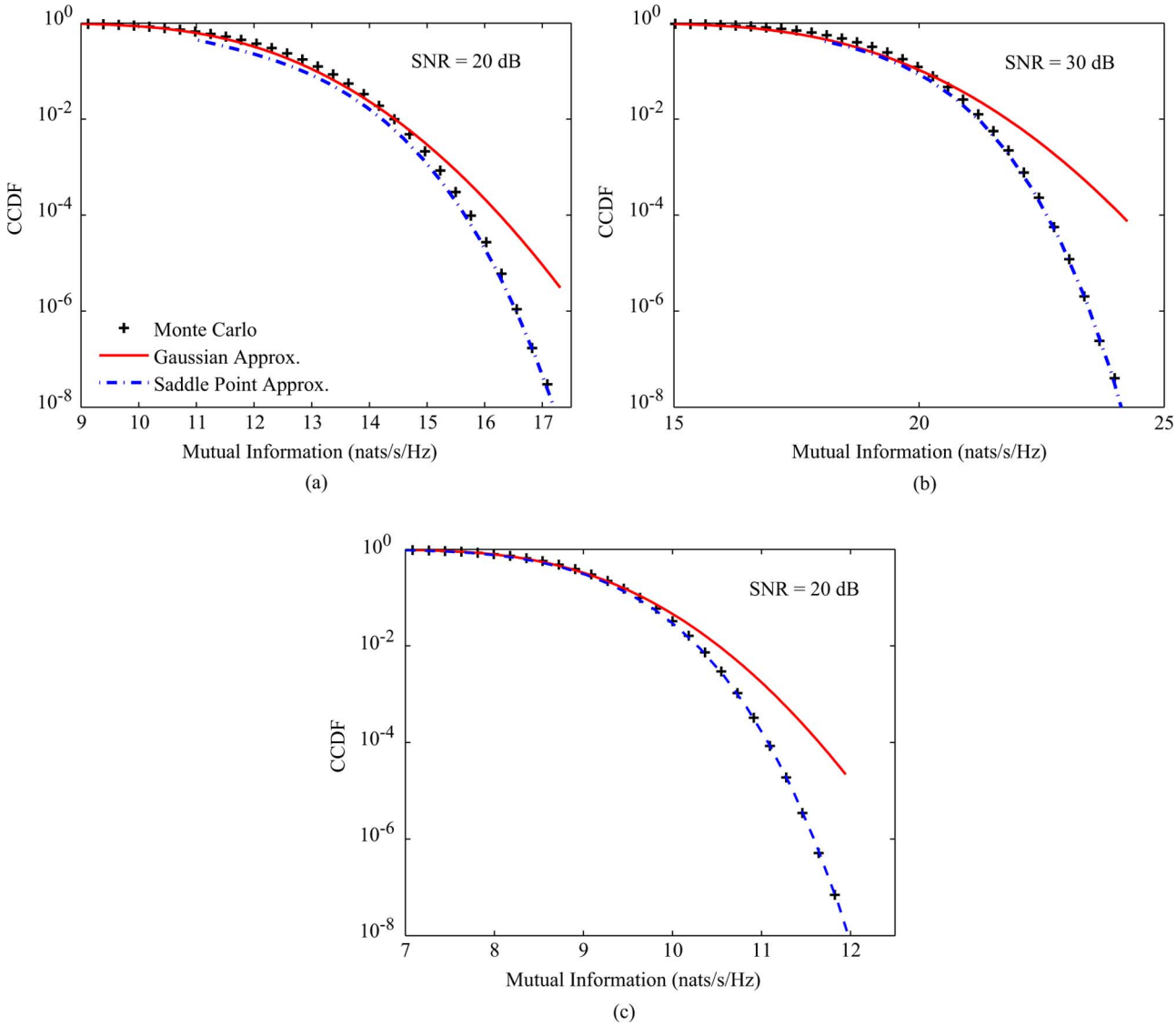

Fig. 10. CCDF of mutual information, comparing the Gaussian approximation, saddle point approximation, and Monte Carlo simulations. Results are shown for different antenna configurations and different SNRs. (a) $n_t = n_r = 3$; (b) $n_t = n_r = 3$; (c) $n_t = 4,\ n_r = 2$.

where $Y' := dY(\beta/x)/dx$. By noting that in (87), there exists first-order terms in P which are $O(\ln P)$ and second-order terms which are $O(1)$, we assume that the CGF admits the following generic large-P expansion:

$$\mathcal{K}(n\lambda) = n^2\left(\lambda \ln P + b_0 + \frac{b_1}{P} + \frac{b_2}{P^2} + \cdots\right) \tag{92}$$

where the coefficient b_0 denotes the constant term in (87) and $b_i, i = 1, 2, \ldots$ depend on β and λ. Taking the derivative of $\mathcal{K}(n\lambda)$ w.r.t. P, we obtain the following power series expansion:

$$\begin{aligned} Y\left(\frac{\beta}{P}\right) &= -P\frac{d}{dP}\left(\frac{\mathcal{K}(n\lambda)}{n^2}\right), \qquad n \to \infty \\ &= -\lambda + \frac{b_1}{P} + \frac{2b_2}{P^2} + \cdots. \end{aligned} \tag{93}$$

Substituting (93) into (91) and matching coefficients of the powers series of P on the left- and right-hand sides, we solve the b_i's:

$$\begin{aligned} b_1 &= \frac{\lambda\beta}{\beta+\lambda-1}, \\ b_2 &= -\frac{\beta^2\lambda\,(\beta-1)\,(\beta+\lambda)}{2\,(\beta+\lambda-1)^4}, \\ b_3 &= \frac{\beta^3\lambda(\beta-1)(\beta+\lambda)(\beta+\lambda+1)(\beta-1-\lambda)}{3\,(\beta+\lambda-1)^7}, \\ &\vdots \end{aligned} \tag{94}$$

Together with (92), we obtain the large-n–large-P CGF with higher correction terms in P:

$$\begin{aligned} \mathcal{K}(\lambda) \sim \mathcal{K}_0(\lambda) + \frac{mn\lambda}{m-n+\lambda}\frac{1}{P} - \frac{m^2(m-n)(m+\lambda)}{2n(m-n+\lambda)^4}\frac{1}{P^2} \\ + \cdots, \quad n \to \infty, \quad P \to \infty \end{aligned} \tag{95}$$

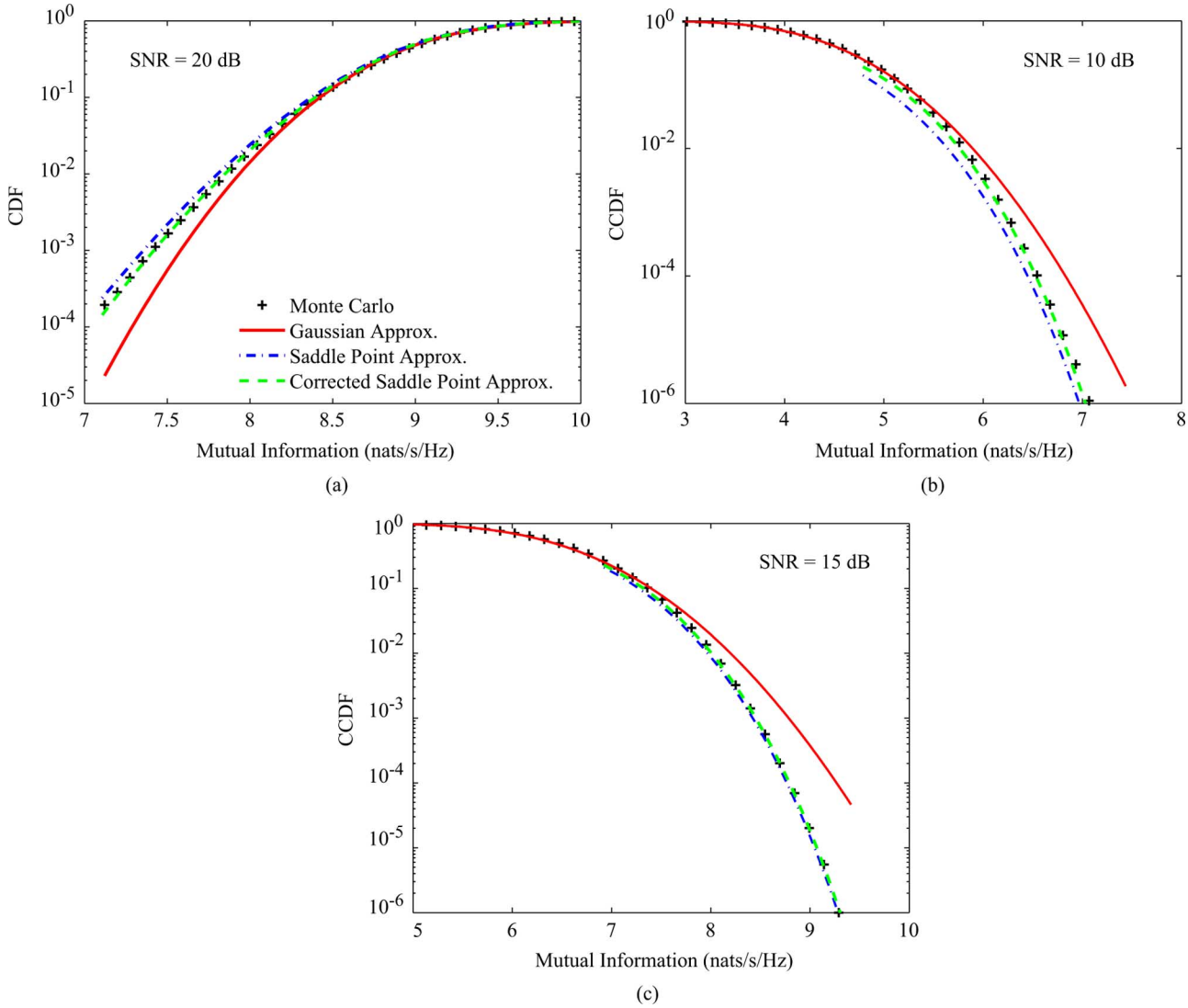

Fig. 11. CDF and CCDF of mutual information, comparing the Gaussian approximation, the saddle point approximation with and without the first-order corrections (in P), and Monte Carlo simulations. Results are shown for different antenna configurations and different SNRs. (a) $n_t = 10, n_r = 2$; (b) $n_t = 4, n_r = 2$; (c) $n_t = 4, n_r = 2$.

where $\mathcal{K}_0(\lambda)$ denotes the leading-order large-n–large-P CGF expression in (87). Based on this formula, we draw the following remarks:

1) First, we see that the higher correction terms vanish rapidly as $m - n + \lambda$ increases (equivalently, λ grows for fixed m and n). Meanwhile, as λ decreases and approaches $n - m$, the correction terms become large and eventually invalidate the expansion. This indicates that the leading-order large-n-large-P approximation for the right-hand tail (corresponding to positive λ) is more robust for finite values of P, compared with the approximation for the left-hand tail (corresponding to negative λ). This is consistent with the results shown in Figs. 9 and 10.
2) Interestingly, as we keep computing b_i's, it is found that $b_i, i > 1$ have the common factor of $m - n$. Assuming this is true for all higher correction terms, then for equal antenna arrays (i.e., $m = n$), we have

$$\mathcal{K}(\lambda) \sim \mathcal{K}_0(\lambda) + \frac{n^2}{P}, \qquad n \to \infty. \tag{96}$$

Note that the single correction term n^2/P is independent of λ, meaning that the saddle point (89) (after setting $m = n$) is unaffected by including correction terms for finite P. This agrees with the formula in [33], (53), and the corresponding argument made therein.

Based on (95), we plot the saddle point approximation with first-order correction term (in P) in Fig. 11. The refinement brought by including the correction term is clearly evident. Moreover, we find that the right-hand tail is captured accurately for moderate $P = 15$ dB even without the first-order correction, which is in line with the discussion in the first point above.

E. Important Large-SNR Behavior in Left and Right Tails (Including DMT)

In this section, we develop further our analytical results under the assumption that $t = qn \ln P,\ P \to \infty$, where $q > 0$ is a fixed constant. This represents the setting where the data rate is specified to grow as a nonvanishing fraction (i.e., q) of the

mean mutual information for large P (i.e., $n \ln P$). This assumption is important in various contexts; for example, in specifying the fundamental DMT [35], as well as capturing the scheduling gains in opportunistic multiuser downlink transmissions [28].

We begin with the scenario $q \geq 1$, in which case we are interested in the mutual information distribution at the right-hand side of the mean. This scenario is relevant for evaluating the performance of scheduling algorithms for which the multiantenna base-station transmits to the multiantenna user with the best channel (i.e., $\mathcal{I}_{\text{best}} = \max\{\mathcal{I}_1, \mathcal{I}_2, \ldots\}$, with $\mathcal{I}_k$ denoting the mutual information between the base-station and the kth user, all of which are assumed to be independent and to undertake the same distribution). In this case, the mutual information achieved by the system will typically be above the mean mutual information for each user μ, and the performance gains of such "best user" selection algorithms can be characterized by studying the distribution of the per-user mutual information in the regime $t \geq \mu$. See [28] and [33].

Substituting $t = qn \ln P, q \geq 1$ into (89), we have the following asymptotic solution:

$$\tau^\star \sim nP^{q-1}, \quad P \to \infty \tag{97}$$

which further yields

$$\begin{aligned} I(t) &\sim n^2 P^{q-1} \ln P^q \\ &= \frac{nt}{P} e^{t/n}, \qquad P \to \infty. \end{aligned} \tag{98}$$

Therefore, for high P, the CCDF admits

$$\Pr(\mathcal{I}(\mathbf{x}; \mathbf{y}|\mathbf{H}) > t) \sim e^{-\frac{nt}{P} e^{t/n}}, \quad t \geq \mu. \tag{99}$$

This agrees with a result derived recently in [33, eq. (79)], by asymptotically solving a set of coupled equations obtained via a Coulomb fluid formulation. From (99), we find the probability of the mutual information taking greater values than the mean drops very sharply (doubly exponentially with t), indicating that for large SNR, the best-user scheduling algorithm indicated above will not enhance the overall data rate significantly.

Now we consider the alternative regime, $q < 1$, corresponding to the DMT framework seminally proposed by [35]. In this case, however, we find that one cannot simply adopt the direct approach of substituting $t = qn \ln P$ with $q < 1$ into (89), since a solution for the asymptotic $\tau^\star$ does not exist. This can be explained by noting that the solution in fact lies in the range $\tau^\star < n - m$ which cannot be described by (87), because $t = qn \ln P < n \ln(P/\beta) - n$ [i.e., the smallest value of t that can be covered by (89)] for $q < 1$ and large P. Nevertheless, in the following, we are able to draw upon the exact characterization of the large-n–large-P CGF (12) to derive the DMT formula.

Since in the large deviation regime, $\tau^\star \sim O(n)$ such that all the cumulants to leading order in n remain effective for large n, we scale the CGF variable $\lambda \to n\lambda$ before taking $n \to \infty$ in (100). Further, by recalling the definition of the CGF, its leading-order representation in P should be $O(\ln P)$. Consequently, the asymptotic characterization of the CGF admits

$$\begin{aligned} \mathcal{K}(n\lambda) &\approx -n^2 \int_0^P \frac{Y(\beta/x)}{x} dx \\ &\approx A\, n^2 \ln P, \quad n \to \infty, \quad P \to \infty, \end{aligned} \tag{100}$$

where $Y(\beta/x)$ satisfies (91) and A denotes a certain function of β and λ. In light of the DMT formulation [35], we require the coefficient of the $O(n^2 \ln P)$ term of the CGF, i.e., the quantity

$$A = \lim_{P \to \infty} \left(-\frac{1}{\ln P} \int_0^P \frac{Y(\beta/x)}{x} dx \right). \tag{101}$$

Note here that we cannot employ the assumption $Y(\beta/x) = \sum_{k=0}^{\infty} b_k/x^k$, since integrating this power series diverges for each term. This motivates us to introduce suitable variable transformations to (90) and (91) as described below, which are aimed at scaling x to increase with P, while keeping the new variable of integration finite. To this end, observe that

$$\begin{aligned} &-\frac{1}{\ln P} \int_0^P \frac{Y(\beta/x)}{x} dx \\ &= -\frac{1}{\ln P} \int_\epsilon^P \frac{Y(\beta/x)}{x}\, dx - \frac{1}{\ln P} \int_0^\epsilon \frac{Y(\beta/x)}{x}\, dx \qquad (102) \\ &= -\int_{\frac{\ln \epsilon}{\ln P}}^1 Y(\beta/x)\, d\left(\frac{\ln x}{\ln P}\right) - \frac{1}{\ln P} \int_0^\epsilon \frac{Y(\beta/x)}{x}\, dx \qquad (103) \\ &\sim -\int_0^1 Y(\beta/P^s)\, ds, \quad P \to \infty \qquad (104) \end{aligned}$$

where the new variable $s := \ln x / \ln P$ and ϵ is an arbitrarily small positive constant. Here, ϵ is introduced to avoid the singularity at the lower limit when changing dx/x on the right-hand side of (102) to $d(\ln x)$ in (103). Nevertheless, for fixed ϵ, the second integral in (103) vanishes as $P \to \infty$ and the lower limit of the first integral becomes zero, thus we arrive to the asymptotic formula (104). With (104), we have

$$A = \lim_{P \to \infty} \left(-\int_0^1 Y(\beta/P^s)\, ds \right) \tag{105}$$

with $Y(\beta/P^s)$ satisfying

$$\begin{aligned} &\left[\frac{Y'}{\ln P} + \left(\frac{\beta + \lambda + 1}{\beta} \right) \frac{P^s}{\ln P} Y' + Y \right]^2 = \\ &\quad 4 \left(-\frac{Y'}{\ln P} - Y + \beta \right) \left(\frac{P^s}{\beta \ln P} Y' - \lambda \right) \frac{P^s}{\beta \ln P} Y' \end{aligned} \tag{106}$$

where $Y' := dY(\beta/P^s)/ds$. By comparing the coefficient of $P^s/\ln P$ (i.e., the terms corresponding to the leading order in P) in (106), we have the equation involving $Y(\beta/P^s)$:

$$\left(\frac{P^s}{\beta \ln P} Y' \right)^2 \left[4Y - 4\beta + (\beta + \lambda + 1)^2 \right] = 0, \tag{107}$$

which gives a constant solution

$$Y(\beta/P^s) = -\frac{(\beta+\lambda+1)^2}{4} + \beta. \tag{108}$$

Thus, we obtain the large-n-large-P CGF via (104) and (100):

$$\mathcal{K}(n\lambda) \sim n^2 \left[\frac{(\beta+\lambda+1)^2}{4} - \beta\right] \ln P, \quad n \to \infty,\ P \to \infty. \tag{109}$$

By invoking the saddle point (51): $qn \ln P = \frac{d}{d(n\tau^\star)}\mathcal{K}(n\tau^\star)$, we have $\tau^\star = 2q - \beta - 1$. This result reconfirms our statement that the DMT cannot be described by (89), which implicitly requires $n\tau^\star > n - m$. With this saddle point, the rate function $I(t)$ (with $t = qn \ln P,\ q < 1$) is evaluated as

$$\begin{aligned} I(qn \ln P) &= qn^2\tau^\star \ln P - \mathcal{K}(n\tau^\star) \\ &= n^2 \left[q^2 - q(\beta+1) + \beta\right] \ln P. \end{aligned} \tag{110}$$

Consequently, the CDF of the mutual information in the left tail for large P (and large n) becomes

$$\begin{aligned} &\ln \Pr\left(\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H}) < qn \ln P\right) \\ &\qquad \sim -(qn-m)(qn-n) \ln P, \quad q < 1. \end{aligned} \tag{111}$$

This agrees precisely with the well-known DMT result in [35].

VII. Conclusion

Capitalizing upon the exact Painlevé-based representation for the MGF of the MIMO mutual information in [30], we have systematically computed new expansions for the high-order cumulants of the mutual information distribution which apply for arbitrary SNRs and for asymmetric antenna arrays. In particular, closed-form expressions were given for the leading-order terms (in n), as well as the first-order correction terms which capture finite-antenna deviations. Based on these new expressions, we established key novel insights into the behavior of the distribution under different conditions; for example, explaining why the n-asymptotic Gaussian approximation is more robust to increasing SNRs for asymmetric systems compared with symmetric systems. This is an interesting phenomenon which appears difficult to capture with other methods. In addition, we called upon the Edgeworth expansion technique along with the high-order cumulant formulas to provide closed-form refinements to the Gaussian approximation for the tail region corresponding to $O(n^\epsilon)$ $(0 < \epsilon \leq 1)$ deviations from the mean. For deviations of $O(n)$, the so-called large deviation region, the Edgeworth expansion requires summing over *all* cumulants and becomes unwieldy; thus, in this region, we employed a saddle point approximation technique along with asymptotic integration tools to derive very simple and concise formulas for the CDF for the cases of low and high SNRs. Simulations showed that our results captured the tail distribution very accurately for outage probabilities of practical interest. Moreover, while formally derived based on a large-antenna framework, they were shown to be very accurate even when the antenna numbers are small. To emphasize the utility of our framework even further, in the end, we recovered well-known properties of the tail distribution of the MIMO mutual information, including the DMT.

We conclude by noting that the key analytical tool underpinning the analysis in this paper, the Painlevé-based MGF representation in Proposition 1, is extremely valuable since, as we have shown, it facilitates a "unified" investigation of the mutual information distribution under a wider range of scenarios than appear possible with previous existing tools. To the best of our knowledge, together with our recent work [30], this is the first time that such tools have been applied to problems in information theory. It turns out that these tools are also applicable to other problems in information theory and wireless communications, and such topics are currently being pursued. Some preliminary results in wireless relay communication can be found in [50].

Appendix I
Derivations of the Recursion for the Cumulants

A) Recursion for the Cumulants to Leading Order in n: In this section, we show how to establish the recursion equation for $y_\ell(x)$'s in (13) and its initial values in (14) and (15).

We substitute $Y(x) = \lambda y_1(x) + \lambda^2 y_2(x) + \cdots$ into (12) and match the coefficients of λ^ℓ on the left- and right-hand side, resulting in the equations for $y_\ell(x),\ \ell = 1, 2, \ldots$ in sequence, which is elaborated as follows.

The case $\ell = 1$ characterizes the mean of the mutual information. In this case, we obtain the nonlinear differential equation involving $y_1(x)$:

$$\begin{aligned} &\left[x^2 + 2(\beta+1)x + (\beta-1)^2\right](y_1')^2 \\ &\qquad - 2(x+\beta+1)y_1'y_1 - 4\beta y_1' + y_1^2 = 0, \end{aligned} \tag{112}$$

which can be explicitly solved to give the expression of $y_1(x)$ in (14).

$$\begin{aligned} &\sum_{i=1}^{\ell} y_i' \left\{\left[x^2 + 2(\beta+1)x + (\beta-1)^2\right] y_{\ell-i+1}' - 2(x+\beta+1)y_{\ell-i+1}\right\} + \sum_{i=1}^{\ell} y_i y_{\ell-i+1} - 4\beta y_\ell' \\ &= 2\sum_{i=1}^{\ell-1} y_i' \left[\sum_{j=1}^{\ell-i}\left(2xy_j'y_{\ell-i-j+1}' - 2y_j'y_{\ell-i-j+1}\right) + (x-\beta-1)y_{\ell-i}' - y_{\ell-i}\right] - \sum_{i=1}^{\ell-2} y_i'y_{\ell-i-1}' \end{aligned} \tag{115}$$

The case $\ell = 2$ characterizes the variance of the mutual information. In this case, we obtain the nonlinear differential equation involving $y_1(x)$ and $y_2(x)$:

$$\Big\{ -2\left[x^2 + 2(\beta+1)x + (\beta-1)^2\right] y_1' + 2(x+\beta+1)y_1 + 4\beta \Big\} y_2' + 2\left[(x+\beta+1)y_1' - y_1\right] y_2 + y_1' \left\{ 4x(y_1')^2 + 2(x-\beta-1)y_1' - 4y_1'y_1 - 2y_1 \right\} = 0. \tag{113}$$

While seemingly complicated, quite remarkably, once we substitute $y_1(x)$ with (14), the coefficient of $y_2'(x)$ vanishes, i.e.,

$$-2\left[x^2 + 2(\beta+1)x + (\beta-1)^2\right] y_1' + 2(x+\beta+1)y_1 + 4\beta = 0\,, \tag{114}$$

and the differential (113) collapses to a simple *algebraic equation* in $y_2(x)$. The solution given by (15) is then easily obtained.

For $\ell > 2$, $y_\ell(x)$ are found to satisfy the recursion (115), at the bottom of the previous page, with initial conditions $y_1(x)$ and $y_2(x)$ given in (14) and (15), respectively. In theory, the recursive (115) enables us to systematically compute the leading-order expressions (in n) in closed form for any desired number of higher order cumulants. In its current form, however, (115) appears very complicated. Fortunately, this expression can be simplified considerably by observing that the only term in (115) which involves $y_\ell(x)$ (or its derivative) is

$$\left\{ \left[x^2 + 2(\beta+1)x + (\beta-1)^2\right] y_1' - (x+\beta+1)y_1 - 2\beta \right\} y_\ell' - \left[(x+\beta+1)y_1' - y_1\right] y_\ell, \tag{116}$$

with all other terms involving the previously computed $y_i(x), i < \ell$. As for the variance, quite remarkably, the term (116) simplifies considerably upon noting that the coefficient of $y_\ell'(x)$ is precisely zero, by virtue of (114). This interesting observation indicates that the computation of the higher cumulants, for $\ell > 2$, only involves solving simple algebraic equations in $y_\ell(x)$, rather than nonlinear differential equations involving $y_\ell(x)$ and $y_\ell'(x)$, and thus, the recursive (115) collapses to the simpler formula (13).

A. Recursion for the Cumulants to Second Order in n

To derive the recursion for $z_\ell(x)$'s in (23) and its initial values, we substitute $Y(x) = \lambda y_1(x) + \lambda^2 y_2(x) + \cdots$ and $Z(x) = \lambda z_1(x) + \lambda^2 z_2(x) + \cdots$ into (22) and match the coefficients of λ^ℓ to obtain the equations for $z_\ell(x),\ \ell = 1, 2, \ldots$ in sequence.

The case $\ell = 1$ corresponds to the correction term for the mean. In this case, we obtain the equation involving $z_1(x)$:

$$\left\{2\beta - [x^2 + 2(\beta+1)x + (\beta-1)^2]y_1' + (x+\beta+1)y_1\right\} z_1' + \left[(x+\beta+1)y_1' - y_1\right] z_1 + \frac{x^2}{2} y_1''^2 = 0. \tag{117}$$

Again, once we substitute $y_1(x)$ with (14), the coefficient of the $z_1'(x)$ vanishes and (117) collapses to an algebraic equation whose solution is given by (24).

The case $\ell = 2$ corresponds to the correction term for the variance. In this case, we obtain the equation involving $z_2(x)$:

$$\begin{aligned} &\left\{2\beta - [x^2 + 2(\beta+1)x + (\beta-1)^2]y_1' + (x+\beta+1)y_1\right\} z_2' \\ &+ \left[(x+\beta+1)y_1' - y_1\right] z_2 + \Big\{2\beta - [x^2 + 2(\beta+1)x + (\beta-1)^2]y_2' \\ &+ (x+\beta+1)y_2 + 6x(y_1')^2 - 2(1+\beta-x)y_1' - 4y_1'y_1 - y_1\Big\} z_1' \\ &+ \left[(x+\beta+1)y_1' - y_1\right] z_1 + \left[(x+\beta+1)y_2' - 2(y_1')^2 - y_2 - y_1'\right] = 0. \end{aligned} \tag{118}$$

The coefficient of $z_2'(x)$ is the same as in (117) and vanishes, which means (118) is an algebraic equation for $z_2(x)$. More interestingly, after substituting $y_1(x)$ with (14) and $y_2(x)$ with (113), the coefficient of $z_1'(x)$ also becomes zero, which further simplifies the computation. Thus, the expression for $z_2(x)$ is obbtained as (25).

For the case $\ell > 2$, we can derive a recursive equation in (119), at the bottom of the page. The only term in (119) that involves $z_\ell(x)$ (or its derivative) is

$$\left\{-\left[x^2 + 2(\beta+1)x + (\beta-1)^2\right] y_1' + (x+\beta+1)y_1 + 2\beta\right\} z_\ell' + \left[(x+\beta+1)y_1' - y_1\right] z_\ell \tag{120}$$

where the coefficient of $z_\ell'(x)$ is exactly the same as the one of $y_\ell'(x)$ in (116), which has been shown to be zero. This observation indicates that we only need to solve an algebraic equation for $z_\ell(x)$. Interestingly, apart from $z_\ell'(x)$, it is found that the

$$\begin{aligned} &\sum_{i=1}^{\ell} z_i' \left\{[x^2 + 2(\beta+1)x + (\beta-1)^2]y_{\ell-i+1}' + (x+\beta+1)y_{\ell-i+1}\right\} + 2\beta z_\ell' \\ &= \sum_{i=1}^{\ell-1} z_i' \left\{ y_{\ell-i} + 2(1+\beta-x)y_{\ell-i}' + 4\sum_{j=1}^{\ell-i} y_{\ell-i-j+1}'y_j - 6x\sum_{j=1}^{\ell-j} y_{\ell-i-j+1}'y_j' \right\} + \sum_{i=1}^{\ell-2} z_i'y_{\ell-i-1}' \\ &\quad - \sum_{i=1}^{\ell} z_i \left[(x+\beta+1)y_{\ell-i+1}' - y_{\ell-i+1}\right] + \sum_{i=1}^{\ell-1} z_i \left[2\sum_{j=1}^{\ell-i} y_j'y_{\ell-i-j+1}' + y_{\ell-i}'\right] - \frac{x^2}{2}\sum_{i=1}^{\ell} y_i''y_{\ell-i+1}'' \end{aligned} \tag{119}$$

coefficients of $z_i'(x), i = 1, 2, \ldots, \ell - 1$ in (119) are also identically equal to zero. Given that this vanishing property holds for any ℓ, eliminating terms involving $z_i'(x), i = 1, 2, \ldots, \ell$ in (119) immediately leads to the general recursive solution (23).

Appendix II
Relation With Coulomb Fluid Method in [33]

Here, we draw the connections between our saddle point results and those derived based on a Coulomb fluid large deviation approximation in [33]. We start by recasting the formulation of [33] in terms of the MGF of the mutual information. To this end, consider the exact MGF (5) represented in multi-integral form:

$$\mathcal{M}(\lambda) = \frac{\int_{\mathbb{R}_+^n} \prod_{k=1}^n (1+Py_k)^\lambda y_k^\alpha e^{-ny_k} \prod_{i<j}(y_i - y_j)^2 d\mathbf{y}}{\int_{\mathbb{R}_+^n} \prod_{k=1}^n y_k^\alpha e^{-ny_k} \prod_{i<j}(y_i - y_j)^2 d\mathbf{y}} := \frac{Z(\lambda)}{Z(0)} \tag{121}$$

where $\mathbf{y} = (y_1, \ldots, y_n)$ denotes the eigenvalues of $\mathbf{HH}^\dagger$ and

$$Z(\lambda) = \int_{\mathbb{R}_+^n} \exp\left\{ \lambda \sum_{k=1}^n (1+Py_k) + \sum_{k=1}^n [(m-n)\ln y_k - ny_k] + 2\sum_{i<j} \ln|y_i - y_j| \right\} d\mathbf{y}. \tag{122}$$

Based on the Coulomb fluid interpretation (see [51]–[55] for details), as n grows large, the CGF associated with (121) is anticipated to be well approximated with the following:

$$\mathcal{K}(\lambda) := \ln \mathcal{M}(\lambda) \approx -\min_{\sigma(y)} F[\sigma(y), \lambda] + \min_{\sigma(y)} F[\sigma(y), 0] \tag{123}$$

where

$$F[\sigma, \lambda] = \int_a^b \sigma(y) \left\{ n^2 [y - (\beta - 1)\ln y] \right\} dy - \lambda n \int_a^b \ln(1+Py)\sigma(y)dy - n^2 \int_a^b \int_a^b \ln|y-z|\sigma(y)\sigma(z)dydz \tag{124}$$

is the so-called free energy, while $\sigma(y)$ is a PDF with support $[a, b]$. Meanwhile, we have the saddle point approximation of the density function:

$$p_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(t) \sim \exp\{-\max_\lambda [t\lambda - \mathcal{K}(\lambda)]\}, \tag{125}$$

which, combined with (123), is equivalent to

$$p_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(t) \sim \exp\left\{ -\max_\lambda \min_{\sigma(y)} \{F[\sigma, \lambda] + t\lambda\} + \min_{\sigma(y)} F[\sigma, 0] \right\}. \tag{126}$$

Recalling that in the "large deviation" regime of interest, $t \sim O(n), \lambda \sim O(n)$ (see the discussions in Section VI-A), we scale $t \to nt, \lambda \to n\lambda$, and (126) becomes

$$p_{\mathcal{I}(\mathbf{x};\mathbf{y}|\mathbf{H})}(nt) \sim \exp\left\{ n^2 \left\{ -\max_\lambda \min_{\sigma(y)} f[\sigma, \lambda, t] + \min_{\sigma(y)} f[\sigma, 0, 0] \right\} \right\}, \tag{127}$$

with

$$\begin{aligned} f[\sigma, \lambda, t] &= \frac{F(n\lambda)}{n^2} + t\lambda \\ &= \int_a^b \sigma(y) \left([y - (\beta - 1)\ln y] \right) dy \\ &\quad + \lambda \left[t - \int_a^b \ln(1+Py)\sigma(y)dy \right] \\ &\quad - \int_a^b \int_a^b \ln|y-z|\sigma(y)\sigma(z)dydz. \end{aligned} \tag{128}$$

We find that (127) coincides exactly with [33, eq. (23)], where the optimization problems $\max_\lambda \min_{\sigma(y)}(\cdot)$ are solved *jointly*, eventually resulting in three coupled nonlinear equations in general[4].

In contrast, our method first employs the Painlevé equation to obtain the n-asymptotic CGF as

$$\mathcal{K}(n\lambda) \sim n^2 \int_\infty^{\beta/P} \frac{G(x)}{x} dx, \qquad n \to \infty$$

with $G(x)$ exactly characterized by (8). This representation corresponds to evaluating (123) explicitly, without any intuitive Coulomb fluid analogy. Then, armed with this CGF result, we separately draw upon the saddle point equation to capture the tail distribution, which corresponds to solving (125). Thus, in essence, our saddle point approximation is solving the equivalent problem to that considered in [33] by *explicitly* finding the asymptotic CGF (while in [33] it is implicit). Quite remarkably, under the high- and low-SNR regimes considered, it also leads to simplified results.

References

[1] H. Bolcskei, R. U. Nabar, O. Oyman, and A. J. Paulraj, "Capacity scaling laws in MIMO relay networks," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1433–1444, Jun. 2006.

[2] Y. Fan and J. Thompson, "MIMO configurations for relay channels: Theory and practice," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1774–1786, May 2007.

[3] J. Wagner, B. Rankov, and A. Wittneben, "Large n analysis of amplify-and-forward MIMO relay channels with correlated Rayleigh fading," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5735–5746, Dec. 2008.

[4] P. Dharmawansa, M. R. McKay, and R. K. Mallik, "Analytical performance of amplify-and-forward MIMO relaying with orthogonal space-time block codes," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 2147–2158, Jul. 2010.

[5] S. Jin, M. R. McKay, C. Zhong, and K.-K. Wong, "Ergodic capacity analysis of amplify-and-forward MIMO dual-hop systems," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2204–2224, May 2010.

[4]For the special case, $n_t = n_r$, there is one nonlinear equation for the right-hand tail and two nonlinear equations for the left-hand tail.

[6] M. Yuksel and E. Erkip, "Multiple-antenna cooperative wireless systems: A diversity-multiplexing tradeoff perspective," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3371–3393, Oct. 2007.

[7] D. Gesbert, S. Hanly, H. Huang, S. Shamai, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Select. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[8] H. Dahrouj and W. Yu, "Coordinated beamforming for the multicell multi-antenna wireless system," *IEEE Trans. Wireless Commun.*, vol. 9, no. 5, pp. 1748–1759, May 2010.

[9] T. Liu and S. Shamai, "A note on the secrecy capacity of the multiple antenna wiretap channel," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2547–2553, Jun. 2009.

[10] A. Mukherjee and A. L. Swindlehurst, "Robust beamforming for security in MIMO wiretap channels with imperfect CSI," *IEEE Trans. Signal Process.*, vol. 59, no. 1, pp. 351–361, Jan. 2011.

[11] F. Oggier and B. Hassibi, "The secrecy capacity of the MIMO wiretap channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4961–4972, Aug. 2011.

[12] B. Chen and M. J. Gans, "MIMO communications in ad hoc networks," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2773–2783, Jul. 2006.

[13] A. M. Hunter, J. G. Andrews, and S. Weber, "Transmission capacity of ad hoc networks with spatial diversity," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5058–5071, Dec. 2008.

[14] R. H. Y. Louie, M. R. Mckay, and I. B. Collings, "Open-loop spatial multiplexing and diversity communications in ad hoc networks," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, pp. 317–344, Jan. 2011.

[15] Y. Wu, R. H. Y. Louie, M. R. McKay, and I. B. Collings, "Generalized framework for the analysis of linear MIMO transmission schemes in decentralized wireless ad hoc networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 8, pp. 2815–2827, Aug. 2012.

[16] A. Lozano and N. Jindal, "Are yesterday's information-theoretic fading models and performance metrics adequate for the analysis of today's wireless systems?," *IEEE Commun. Mag.*, vol. 50, no. 11, pp. 210–217, Nov. 2012.

[17] I.E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

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

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

[20] H. Shin, M. Z. Win, and J. H. Lee, "Saddlepoint approximation to the outage capacity of MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 10, pp. 2697–2684, Oct. 2006.

[21] H. Shin, M. Z. Win, J. H. Lee, and M. Chiani, "On the capacity of doubly correlated MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 8, pp. 2253–2265, Aug. 2006.

[22] H. Shin, "Capacity and error exponents for multiple-input multiple-output wireless channels," Ph.D. dissertation, Seoul National Univ., Seoul, Korea, Aug. 2004.

[23] H. Shin, M. Z. Win, and M. Chiani, "Asymptotic statistics of mutual information for doubly correlated MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 7, no. 2, pp. 562–573, Feb. 2008.

[24] O. Oyman, R.U. Nabar, H. Bolcskei, and A.J Paulraj, "Characterizing the statistical properties of mutual information in MIMO channels," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2784–2795, Nov. 2003.

[25] P. J. Smith, L. M. Garth, and S. Loyka, "Exact capacity distribution for MIMO systems with small numbers of antennas," *IEEE Commun. Lett.*, vol. 7, no. 10, pp. 481–483, Oct. 2003.

[26] P. J. Smith and L. M. Garth, "Exact capacity distribution for dual MIMO systems in Ricean fading," *IEEE Commun. Lett.*, vol. 8, no. 1, pp. 18–20, Jan. 2004.

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

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

[29] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of large dimensional multiantenna channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–4004, Sep. 2008.

[30] Y. Chen and M. R. McKay, "Coulomb fluid, Painlevé transcendents and the information theory of MIMO systems," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4594–4634, Jul. 2012.

[31] P. B. Rapajic and D. Popescu, "Information capacity of a random signature multiple-input multiple-output channel," *IEEE Trans. Commun.*, vol. 48, no. 8, pp. 1245–1248, Aug. 2000.

[32] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "Statistical mechanics of multi-antenna communications: Replicas and correlations," *Acta Physica Polonica Ser. B*, vol. 36, no. 9, pp. 2719–2732, Sep. 2005.

[33] P. Kazakopoulos, P. Mertikopoulos, A. L. Moustakas, and G. Caire, "Living at the edge: A large deviations approach to the outage MIMO capacity," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1984–2007, Apr. 2011.

[34] S. Li, Y. Chen, and M. R. McKay, "Mutual information distribution of interference-limited MIMO: A joint Coulomb fluid and Painlevé based approach," in *Proc. 45th Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 9, 2011, pp. 774–778.

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

[36] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.

[37] A. Magnus, "Painlevé-type differential equations for the recurrence coefficients of semi-classical orthogonal polynomials," *J. Comput. Appl. Math.*, vol. 57, no. 1-2, pp. 215–237, 1995.

[38] Y. Chen and A. R. Its, "Painlevé III and a singular linear statistics in Hermitian random matrix ensembles," *J. Approx. Theory*, vol. 162, no. 2, pp. 270–297, 2010.

[39] E. Basor and Y. Chen, "Perturbed Hankel determinants," *J. Phys. A: Math. Gen.*, vol. 38, no. 47, pp. 10101–10106, 2005.

[40] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[41] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Analysis of the pilot contamination effect in very large multicell multiuser MIMO systems for physical channel models," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Prague, Czech Republic, May 2011, pp. 3464–3467.

[42] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.

[43] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Process. Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[44] B. Gopalakrishnan and N. Jindal, "An analysis of pilot contamination on multi-user MIMO cellular systems with many antennas," in *Proc. IEEE 12th Int. Workshop Signal Process. Adv. Wireless Commun.*, San Francisco, CA, USA, Jun. 2011, pp. 381–385.

[45] A. Lozano, A. M. Tulino, and S. Verdú, "High-SNR power offset in multi-antenna communication," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4134–4151, Dec. 2005.

[46] A. Grant, "Rayleigh fading multiple-antenna channels," *EURASIP J. Appl. Signal Process., Spec. Issue Space-Time Coding (Part I)*, vol. 2002, no. 3, pp. 316–329, Mar. 2002.

[47] S. Blinnikov and R. Moessner, "Expansions for nearly Gaussian distributions," *Astron. Astrophys. Suppl. Ser.*, no. 130, pp. 193–205, May 1998.

[48] H. E. Daniels, "Saddlepoint approximations in statistics," *Ann. Math. Statist.*, vol. 25, no. 4, pp. 631–650, 1954.

[49] I. Avramidi, Lecture Notes on Asymptotic Expansion Oct. 2000 [Online]. Available: http://www.nmt.edu/~iavramid/notes/asexp.pdf

[50] Y. Chen, N. S. Haq, and M. R. McKay, "Random matrix models, double-time Painlevé equations, and wireless relaying," *J. Math. Phys.*, vol. 54, no. 6, p. 063506, 2013.

[51] F. J. Dyson, "Statistical theory of energy levels of complex systems I-III," *J. Math. Phys.*, vol. 3, no. 1, pp. 140–175, 1962.

[52] Y. Chen and S. M. Manning, "Asymptotic level spacing of the Laguerre ensemble: A Coulomb fluid approach," *J. Phys. A.:Math. Gen.*, vol. 27, no. 11, pp. 3615–3620, 1994.

[53] Y. Chen and N. D. Lawrence, "On the linear statistics of Hermitian random matrices," *J. Phys. A: Math. Gen.*, vol. 31, no. 4, pp. 1141–1152, 1998.

[54] Y. Chen and S. M. Manning, "Distribution of linear statistics in random matrix models (metallic conductance fluctuations)," *J. Phys.: Condens. Matter*, vol. 6, no. 16, pp. 3039–3044, 1994.

[55] Y. Chen and M. E. H. Ismail, "Thermodynamic relations of the Hermitian matrix ensembles," *J. Phys. A: Math. Gen.*, vol. 30, no. 19, pp. 6633–6654, 1997.

Shang Li received the B.Sc. degree in electronics from Peking University, China, in 2010 and the M.Phil. degree in electronic and computer engineering (ECE) from the Hong Kong University of Science and Technology (HKUST), Hong Kong, in 2012.

He is currently working towards the Ph.D. degree in electrical engineering at Columbia University, New York. His research interests include random matrix theory with application in MIMO communications, statistical signal processing, decentralized detection/estimation and sequential analysis.

Shang has been serving as a reviewer for the IEEE TRANSACTIONS ON INFORMATION THOERY and several conferences including International Conference on Communications (ICC), the IEEE International Conference of Acoustics, Speech and Signal Processing (ICASSP), and the IEEE Information Theory Workshop (ITW).

Matthew R. McKay (S'08–M'12) received the combined Bachelor's degree in electrical engineering and computer science from the Queensland University of Technology, Australia, in 2002, and the Ph.D. degree in electrical engineering from the University of Sydney, Australia, in 2007. He then worked as a Research Scientist at the Commonwealth Science and Industrial Research Organization (CSIRO), Sydney, Australia, prior to joining the faculty at the Hong Kong University of Science and Technology (HKUST) in 2007, where he is currently the Hari Harilela Associate Professor of Electronic and Computer Engineering. He is also a member of the Centre for Wireless Information Technology at HKUST, and an Affiliate faculty member of the Division of Biomedical Engineering. His research interests include communications and signal processing; in particular the analysis and design of MIMO systems, random matrix theory, information theory, and wireless ad-hoc and sensor networks.

Matthew was awarded the University Medal upon graduating from the Queensland University of Technology. He and his co-authors have been awarded a Best Student Paper Award at IEEE ICASSP 2006, Best Student Paper Award at IEEE VTC 2006-Spring, Best Paper Award at ACM IWCMC 2010, Best Paper Award at IEEE Globecom 2010, and Best Paper Award at IEEE ICC 2011. Recently, Matthew has also received a Young Author Best Paper Award by the IEEE Signal Processing Society (awarded at IEEE ICASSP 2011), the Stephen O. Rice Prize in the Field of Communication Theory by the IEEE Communication Society (awarded at IEEE ICC 2011), and the 2011 Young Investigator Research Excellence Award by the School of Engineering at HKUST. Matthew serves on the editorial boards of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the mathematics journal, *Random Matrices: Theory and Applications*, and he served as the 2011 Chair of the Hong Kong Chapter of the IEEE Information Theory Society.

Yang Chen received his Ph.D. at the University of Massachusetts Amherst in 1987. He was a Post-doctoral Research Fellow at Cavendish Laboratory (1987–1990) and Max Planck Institute, Stuttgart (1990–1992). Then he joined the Department of Mathematics at Imperial College London and had been a Lecturer, Reader and Professor there (1992–2012). Currently, he is Professor of Mathematics at the University of Macau, Taipa, China.

Over the past twenty years, Yang and his collaborators worked on problems arising from unitary random matrix ensembles, intimately related to orthogonal polynomials. His current research activities include: moduli space in Super-symmetric QCD, information theoretic problem in wireless communications. He is founding editor of *Random Matrices*: *Theory and Applications*.