Asymptotic Performance of Linear Receivers in MIMO Fading Channels

K. Raj Kumar, *Student Member, IEEE*, Giuseppe Caire, *Fellow, IEEE*, and Aris L. Moustakas, *Senior Member, IEEE*

***Abstract*—Linear receivers are an attractive low-complexity alternative to optimal processing for multiple-antenna multiple-input multiple-output (MIMO) communications. In this paper, we characterize the information-theoretic performance of MIMO linear receivers in two different asymptotic regimes. For fixed number of antennas, we investigate the limit of error probability in the high-signal-to noise-ratio (SNR) regime in terms of the diversity–multiplexing tradeoff (DMT). Following this, we characterize the error probability for fixed SNR in the regime of large (but finite) number of antennas.**

As far as the DMT is concerned, we report a negative result: we show that both linear zero-forcing (ZF) and linear minimum mean-square error (MMSE) receivers achieve the same DMT, which is largely suboptimal even in the case where outer coding and decoding is performed across the antennas. We also provide an approximate quantitative analysis of the markedly different behavior of the MMSE and ZF receivers at finite rate and nonasymptotic SNR, and show that while the ZF receiver achieves poor diversity at any finite rate, the MMSE receiver error curve slope flattens out progressively, as the coding rate increases.

When SNR is fixed and the number of antennas becomes large, we show that the mutual information at the output of an MMSE or ZF linear receiver has fluctuations that converge in distribution to a Gaussian random variable, whose mean and variance can be characterized in closed form. This analysis extends to the linear receiver case a well-known result previously obtained for the optimal receiver. Simulations reveal that the asymptotic analysis captures accurately the outage behavior of systems even with a moderate number of antennas.

***Index Terms*—Diversity–multiplexing tradeoff (DMT), large-system limit, linear receivers, multiple-input multiple-output (MIMO) channels, spatial multiplexing.**

I. Introduction

THE next generation of wireless communication systems is expected to capitalize on the large gains in spectral efficiency and reliability promised by multiple-input multiple-output (MIMO) multiple-antenna communications [4], [3], [13], [14] and include MIMO technology as a fundamental component of their physical layer [1]. The information-theoretic analysis and the efficient design of space–time (ST) codes for transmission over these MIMO systems have been active areas of research over the past decade. Also, suboptimal receiver schemes have been widely proposed and investigated as a low-complexity alternative to the optimal maximum-likelihood (ML) or ML-like receivers [10], [11]. These schemes range from the iterative interference (soft) cancellation (e.g., [7]), to successive interference (hard) cancellation (e.g., [5], [6]), to the even lower complexity "separated" architecture, based on linear spatial equalization followed by standard single-input single-output (SISO) decoding.[1]

In this paper, we present two types of asymptotic performance analysis of this low-complexity MIMO architecture. First, we consider the diversity–multiplexing tradeoff (DMT) [3], which captures the performance tradeoff between rate and block-error probability in the high-signal-to-noise ratio (SNR), high spectral efficiency regime. We determine the DMT achieved by MIMO architectures that use zero-forcing (ZF) or minimum mean-square error (MMSE) linear receivers and apply conventional SISO outer coding *before* the MIMO transmitter and conventional SISO decoding to the output of the linear receiver. The DMT analysis reveals that both ZF and MMSE linear receivers are very suboptimal in terms of their achievable diversity. Furthermore, we observe that while the DMT analysis accurately predicts the behavior of the ZF receiver at all finite rates, the performance of the MMSE receiver is in stark contrast to that predicted by the DMT analysis at low rates. In fact, we observe that for sufficiently low rates, the MMSE receiver exhibits an ML-like performance. On the contrary, when working at higher rates (and correspondingly higher SNR), the MMSE receiver approaches the ZF performance. We provide an approximate analysis that explains this behavior both qualitatively and quantitatively.

In the second part of this paper, we take a closer look at the performance of the linear MMSE and ZF receivers at finite SNR. Since this is very difficult to capture in closed form, we explore a second type of asymptotic regime, where we fix SNR and let the number of antennas become large. Using random matrix theory, we show that in this case the limiting distribution of the mutual information of the parallel channels induced by the linear receiver is Gaussian, with mean and variance that can be computed in closed form. The analysis provides accurate results even for a moderate number of antennas and allows to quantify how the performance loss in terms of diversity suffered by linear receivers may be recovered by increasing the number

Manuscript received October 03, 2008; revised May 22, 2009. Current version published September 23, 2009. This work was supported in part by the European Commission through the project "Newcom++" No. EU-IST-NoE-FP6-2007-216715, by the NSF-CCF Grant 0635326, and by an Oakley fellowship from the University of Southern California. The material in this paper was presented in part at the IEEE Information Theory Workshop (ITW-07), Lake Tahoe, CA, September 2007.

K. R. Kumar and G. Caire are with the Department of Electrical Engineering–Systems, University of Southern California, Los Angeles, CA 90089 USA (e-mail: rkkrishn@usc.edu; caire@usc.edu).

A. L. Moustakas is with the Department of Physics, National and Capodistrian University of Athens, 15784 Athens, Greece (e-mail: arislm@phys.uoa.gr).

Communicated by G. Taricco, Associate Editor for Communications.

Color versions of Figures 2–9 in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIT.2009.2027493

[1]It should be noticed that the current MIMO WLAN standard [1] is based on MIMO-OFDM, therefore, linear equalization is performed in the space and in the frequency domains. For simplicity, in this work we restrict ourselves to the standard frequency-flat case where equalization is purely spatial.

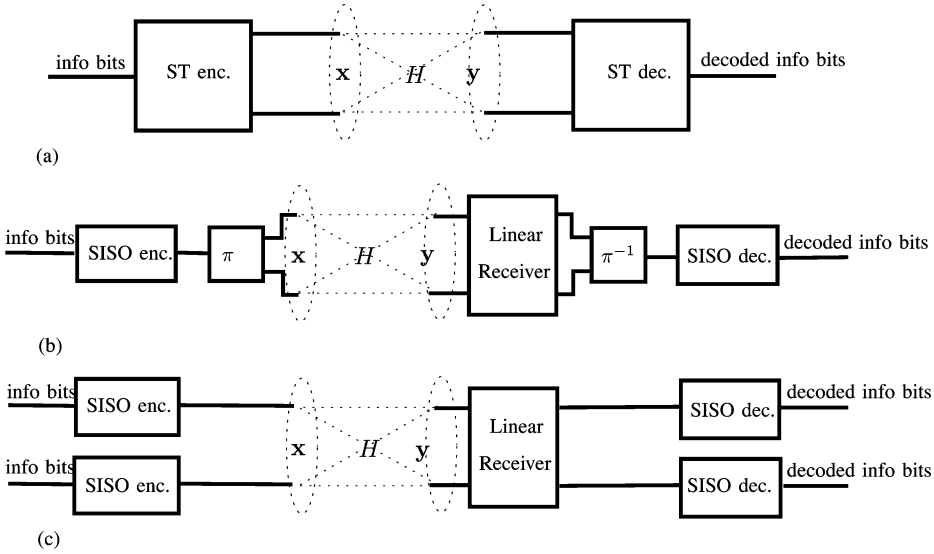

Fig. 1. Three possible ST architectures: (a) unrestricted ST coding scheme; (b) coding across the antennas, with linear spatial equalization; (c) pure spatial multiplexing with linear spatial equalization. π and π^{-1} in (b) denote interleaving and de-interleaving.

of antennas. This prompts to the conclusion that in order to achieve a desired target spectral efficiency and block-error rate, at given SNR and receiver complexity, increasing the number of antennas and using simple linear receiver processing may be, in fact, a good design option.

The paper is organized as follows. In the remainder of this section, we briefly comment on concurrent and related literature. In Section II, we define the system model and recall the main facts of the ZF and MMSE linear receivers considered in this work. Section III presents the DMT analysis and some illustrative numerical examples. Section IV is devoted to the fixed-rate analysis of the MMSE receiver performance with coding across the antennas and provides an approximate quantitative analysis of the slope of the error probability versus SNR. Section V deals with the limiting distribution of the mutual information for the MMSE and ZF receivers for a large number of antennas and provides some illustrative numerical examples on the validity and limitations of this analysis. Conclusions are pointed out in Section VI and some technical details of the proofs are deferred to the Appendix.

A. Related Literature

Since its introduction in the seminal work [3], the DMT has become a *standard tool* in the characterization of the performance of slowly varying fading channels in the high-SNR, large spectral efficiency regime. ST coding schemes have been characterized in terms of their achievable DMT in a series of works, including lattice coding and decoding [16] and ZF or MMSE *decision feedback* receivers (see, for example, [17], [18]). The multipath diversity achievable by *linear equalizers* in frequency-selective SISO channels has also attracted some attention and was recently solved in [19]. The spatial diversity achievable by MIMO linear receivers and separated detection and decoding was investigated in [21], [20], and in [23].[2] In this respect, it is worthwhile to stress the differences between the present work and [20]: 1) we investigate the full DMT curve, while [20] considers only on the fixed-rate case (corresponding to zero multiplexing gain); 2) [20] develops only lower bounds to the diversity order, while we have both lower and upper bounds and show that they are tight; 3) the analysis on the diversity order of the ZF receiver in [20] is based on the conjecture that "the diversity arising from coding across channels with independent power levels is no better than coding across channels whose power levels are correlated." Our analysis, instead, is based on the proper correlation structure of the signal-to-interference plus noise ratios (SINRs) and does not rely on their assumed independence. We notice, in passing, that if the SINRs were independent, then the diversity order would be different, as explained in a comment at the end of Section III-A.

With respect to the large-system analysis of linear receivers presented in Section V, we notice that asymptotic Gaussianity was shown for the MIMO channel mutual information given by the "log-det" formula, whose cumulative distribution function (cdf) yields the block-error rate achievable under *optimal* decoding. This was shown in various works, such as [24]–[27]. At the same time, the *marginal* asymptotic Gaussianity of SINR of a single MMSE and ZF receiver channel was derived in [28], [29], without looking at the *joint* Gaussianity of all SINRs for all these channels. While the marginal Gaussianity is useful in the case of pure spatial multiplexing, where each antenna (or "spatial stream") is independently encoded and decoded, we would like to remark here that the joint Gaussianity is crucial in the analysis of the most relevant case where outer coding is applied across antennas. In Section V, we characterize the limiting joint Gaussian distribution of the SINRs and obtain the statistics of the mutual information of linear MMSE and ZF receivers for the case of coding across the antennas. Our approach is novel and does not follow as a simple extension of the analysis of the marginal statistics as done previously.

II. System Model, DMT and Linear Receivers

Fig. 1 shows three types of MIMO architectures, employing M transmit and N receive antennas. Since the focus of

[2]The present paper provides the detailed proofs of the DMT results presented in [23] and presents the novel large-system finite-SNR analysis of the MMSE receiver, which is not given in [23].

this paper is on linear receivers, we shall assume $N \geq M$ throughout this paper. Scheme of Fig. 1(a) puts no restriction on the choice of the ST coding and decoding scheme: the M-channel inputs are jointly encoded, and the N-channel outputs are jointly and possibly optimally decoded. Scheme of Fig. 1(b) is based on interleaving and demultiplexing over the M inputs the codewords of a SISO code. A linear spatial equalizer (referred briefly as "linear receiver" in the following) processes each N-dimensional channel output vector (purely spatial processing) and creates M virtual *approximately parallel* channels (details are given later on). The output of these virtual channels are then demultiplexed and de-interleaved, and eventually fed to a SISO decoder that treats them as scalar observations, thus disregarding the possible dependencies introduced by the underlying MIMO channel. Notice that in scheme of Fig. 1(b) coding is applied *across the antennas*. Finally, the scheme in Fig. 1(c) is based solely on "spatial multiplexing," that is, M independently encoded streams drive the M transmit antennas and are approximately separated by the linear receiver, the outputs of which are fed to M independent decoders.

The output of the underlying frequency-flat slowly varying MIMO channel is given by

$$\boldsymbol{y}_t = \boldsymbol{H}\boldsymbol{x}_t + \boldsymbol{w}_t, \quad t = 1, \ldots, T \tag{1}$$

where $\boldsymbol{x}_t \in \mathbb{C}^M$ denotes the channel input vector at channel use t, $\boldsymbol{w}_t \sim \mathcal{CN}(\boldsymbol{0}, N_0\boldsymbol{I})$ is the additive spatially and temporally white Gaussian noise and $\boldsymbol{H} \in \mathbb{C}^{N\times M}$ is the channel matrix. In this work, we make the standard assumption that the entries of $\boldsymbol{H}$ are independent and identically distributed (i.i.d.) $\sim \mathcal{CN}(0,1)$, and that $\boldsymbol{H}$ is random but constant over the duration T of a codeword (quasi-static Rayleigh i.i.d. fading [4], [3]). The input is subject to the total power constraint

$$\frac{1}{MT}\mathbb{E}\left[\|\boldsymbol{X}\|_F^2\right] \leq E_s \tag{2}$$

where $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T]$ denotes an ST codeword, uniformly distributed over the ST codebook $\mathcal{X}$, and $\|\cdot\|_F$ denotes the Frobenius norm. Furthermore, following the standard literature of MIMO channels and ST coding, we define the transmit SNR ρ as the total transmit energy per time slot over the noise power spectral density, i.e., $\rho = ME_s/N_0$.

We assume no channel state information (CSI) at the transmitter. In this work, we consider the case of very large block length (and, consequently, of very slowly varying fading). Under the quasi-static assumption, it is well known that the capacity and the outage capacity (or ϵ-capacity) are independent of the assumption on CSI at the receiver [9]. Hence, assuming perfect CSI at the receiver incurs no loss of generality.

We focus on the MIMO detector/decoder blocks in Fig. 1. Under the fully unconstrained ST architecture of Fig. 1(a), the optimum receiver for the MIMO channel in (1) is the ML decoder, with minimum distance decision rule given by

$$\hat{\boldsymbol{X}} = \arg\min_{\boldsymbol{X}\in\mathcal{X}} \|\boldsymbol{Y} - \boldsymbol{H}\boldsymbol{X}\|_F^2.$$

This entails joint processing of the symbols across all antennas at the receiver, over the whole block length T, and is typically implemented using algorithms like Sphere Decoding (see [10] and reference therein) and their tree search sequential decoding generalization [11], possibly coupled with ML Viterbi algorithm if the underlying code has a trellis structure (e.g., [31], [32]). The performance of this decoder is characterized by the information outage probability given by

$$P_{\text{out}}(R,\rho) = \inf_{\substack{\boldsymbol{S}:\boldsymbol{S}\succeq 0\\ \text{tr}(\boldsymbol{S})\leq 1}} P(\log\det(I + \rho\boldsymbol{H}\boldsymbol{S}\boldsymbol{H}^{\mathsf{H}}) \leq R) \tag{3}$$

where the optimization is over the Hermitian symmetric nonnegative definite matrix $\boldsymbol{S}$ subject to a trace constraint, reflecting the channel input power constraint (2). Several lower complexity suboptimal decoders have been proposed in the literature. In particular, architectures of parts (b) and (c) in Fig. 1 involve a linear *memoryless* receiver defined by the matrix $\boldsymbol{G}$, such that the output of the linear receiver is $\boldsymbol{y}'_t = \boldsymbol{G}\boldsymbol{y}_t$. Classical choices for $\boldsymbol{G}$ are the ZF or the MMSE spatial filters, or any diagonal scaling thereof. Under the assumption of Gaussian inputs, very large block length T, and ideal interleaving, the linear receiver creates M "virtual" parallel channels that, without loss of generality, can be described by

$$y'_{k,t} = \sqrt{\gamma_k}x_{k,t} + w'_{k,t}, \quad k = 1, \ldots, M \tag{4}$$

where we normalize the input and output such that $\mathbb{E}[|x_{k,t}|^2] = \mathbb{E}[|w'_{k,t}|^2] = 1$, and where γ_k denotes the SINR at the kth linear receiver output.[3]

Under the above assumptions, the performance of such schemes is characterized by the following two outage probabilities. With coding across antennas (scheme (b)), the outage probability of interest is given by

$$P_{\text{out}}^{\text{lin}}(R,\rho) \triangleq P\left(\sum_{k=1}^{M}\log(1+\gamma_k) \leq R\right). \tag{5}$$

Under pure spatial multiplexing (scheme (c)), the relevant outage probability is given by

$$P_{\text{out}}^{\text{sp mult}}(R,\rho) \triangleq P\left(\bigcup_{k=1}^{M}\left\{\log(1+\gamma_k) \leq \frac{R}{M}\right\}\right) \tag{6}$$

where we used the fact that, by symmetry, without CSI at the transmitter the optimal performance of spatial multiplexing with linear receivers is achieved by allocating the same rate R/M to each stream.

For completeness and for later use, we recall here the expressions of the SINRs for the ZF and the MMSE linear receivers. Here we follow closely the approach presented in [20].

a) ZF Receiver: In this case, the matrix $\boldsymbol{G}$ is chosen as $\boldsymbol{G} = \boldsymbol{D}\boldsymbol{H}^{+}$, where $\boldsymbol{D}$ is a suitable diagonal scaling matrix and $\boldsymbol{H}^{+}$ is the Moore–Penrose pseudoinverse of $\boldsymbol{H}$ [12]. Since $\boldsymbol{H}$ has rank M with probability 1, this takes on the form

$$\boldsymbol{H}^{+} = (\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathsf{H}}.$$

In the absence of transmitter CSI, the signal power is allocated uniformly across the transmitter antennas. It is immediate to

[3]In order to avoid any misunderstanding, it should be noticed here that "interference" is uniquely caused by the generally nonperfect separation of the transmitted symbols in $\boldsymbol{x}_t$ by the linear receiver $\boldsymbol{G}$. We consider a strictly single-user setting, with no multiuser interference.

show that the SINRs on the resulting M parallel channels are given by

$$\gamma_k = \frac{\rho/M}{[(\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H})^{-1}]_{kk}} \tag{7}$$

where the notation $[\boldsymbol{A}]_{kk}$ indicates the kth diagonal entry of a matrix $\boldsymbol{A}$.

b) MMSE Receiver: In this case, the matrix $\boldsymbol{G}$ is chosen in order to maximize the SINR γ_k for each k, over all linear receivers. It is well known that this is achieved by choosing $\boldsymbol{G} = \boldsymbol{D}\boldsymbol{H}_{\mathrm{mmse}}$, where $\boldsymbol{D}$ is a suitable diagonal scaling matrix and $\boldsymbol{H}_{\mathrm{mmse}}$ is the linear MMSE filter [12] that minimizes the mean-square error (MSE) $\mathbb{E}[||\boldsymbol{x}_t - \boldsymbol{H}_{\mathrm{mmse}}\boldsymbol{y}_t||^2]$. Using the orthogonality principle, we find

$$\begin{aligned}\boldsymbol{H}_{\mathrm{mmse}} &= \frac{\rho}{M}\boldsymbol{H}^{\mathsf{H}}\left[\boldsymbol{I} + \frac{\rho}{M}\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}\right]^{-1} \\ &= \left[\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H} + \frac{M}{\rho}\boldsymbol{I}\right]^{-1}\boldsymbol{H}^{\mathsf{H}}.\end{aligned} \tag{8}$$

A standard calculation [12] yields the SINRs γ_k of the resulting set of virtual parallel channels in the form

$$\begin{aligned}\gamma_k &= \frac{\rho}{M}\boldsymbol{h}_k^{\mathsf{H}}\left[\boldsymbol{I} + \frac{\rho}{M}\boldsymbol{H}_k\boldsymbol{H}_k^{\mathsf{H}}\right]^{-1}\boldsymbol{h}_k \\ &= \frac{1}{\left[\left(\boldsymbol{I} + \frac{\rho}{M}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}\right)^{-1}\right]_{kk}} - 1\end{aligned} \tag{9}$$

where $\boldsymbol{H}_k$ denotes the $N\times(M-1)$ matrix obtained by removing the kth column, $\boldsymbol{h}_k$, from $\boldsymbol{H}$.

III. Diversity–Multiplexing Tradeoff (DMT)

A compact and convenient characterization of the tradeoff between rate and block-error probability of MIMO quasi-static fading channels in the high-SNR regime is provided by the DMT introduced by [3]. Consider a family of ST coding systems, each of which operates at SNR ρ with rate $R(\rho)$ and error probability $P_e(\rho)$. We say that this family achieves multiplexing gain r and the diversity gain d (i.e., the point (r,d) on the DMT plane) if

$$\lim_{\rho\to\infty}\frac{R(\rho)}{\log\rho} = r, \qquad \lim_{\rho\to\infty}\frac{\log P_e(\rho)}{\log\rho} = -d.$$

The latter relation is written briefly as $P_e(\rho) \doteq \mathrm{SNR}^{-d}$ in the exponential equality notation of [3].

The optimal DMT is the best possible error probability exponent $d^*(r)$ achievable by any ST scheme at multiplexing gain r. The standard theory of ϵ-capacity [8] readily yields that $d^*(r)$ is equal to the negative ρ-exponent of the information outage probability (3). For the ST channel in (1), $d^*(r)$ is given by the piecewise linear function interpolating the points (r,d) with coordinates

$$r = k, \quad d = (M-k)(N-k)$$

for $k = 0, 1, \ldots, \min\{M,N\}$, and is zero for $r > \min\{M,N\}$ [3].

While $d^*(r)$ is achievable under the optimal receiver (a) in Fig. 1, the following result characterizes the DMT of the MIMO channel in (1) under schemes (b) and (c), when the linear receiver is either the ZF or the MMSE receiver defined above:

Theorem 1: The DMT of the M-transmit, N-receive i.i.d. Rayleigh MIMO channel with $N \geq M$, constrained to use Gaussian codes under either MMSE or ZF linear receivers is given by[4]

$$d^*_{\mathrm{lin}}(r) = (N-M+1)\left(1 - \frac{r}{M}\right)^+ \tag{10}$$

for both the cases of coding across antennas or pure spatial multiplexing.

Proof: The theorem is proved by developing upper and lower bounds on $P^{\mathrm{lin}}_{\mathrm{out}}(R,\rho)$ for the MMSE receiver in the configuration (b) of the block diagram of Fig. 1. A simple upper bound on the outage probability for the ZF receiver extends immediately the result to this case. For configuration (c), the result follows as an immediate corollary.

Lower bound on the outage exponent. Let $\lambda_{\min}(\boldsymbol{A})$ and $\lambda_{\max}(\boldsymbol{A})$ denote the minimum and maximum eigenvalues of a Hermitian symmetric matrix $\boldsymbol{A}$, and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_M$ denote the ordered eigenvalues of the $M \times M$ Wishart matrix $\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}$, with joint probability density function (pdf) given by [4]

$$p(\boldsymbol{\lambda}) = K_{M,N}\prod_{i=1}^{M}\lambda_i^{N-M}\cdot\prod_{i<j}(\lambda_i - \lambda_j)^2\exp\left(-\sum_{i=1}^{M}\lambda_i\right) \tag{11}$$

where $K_{M,N}$ is a normalization constant and we have assumed $M \leq N$.

Using (9), we can write the mutual information with Gaussian coding across the antennas and the MMSE receiver as

$$I_{\mathrm{mmse}}(\boldsymbol{H}) = -\sum_{k=1}^{M}\log\left(\left[\left(\boldsymbol{I} + \frac{\rho}{M}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}\right)^{-1}\right]_{kk}\right). \tag{12}$$

Since the function $-\log(\cdot)$ is convex, using Jensen's inequality we have

$$\begin{aligned}I_{\mathrm{mmse}}(\boldsymbol{H}) &\geq -M\log\left(\frac{1}{M}\sum_{k=1}^{M}\left[\left(\boldsymbol{I} + \frac{\rho}{M}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}\right)^{-1}\right]_{kk}\right) \\ &= -M\log\left(\frac{1}{M}\mathrm{Tr}\left[\left(\boldsymbol{I} + \frac{\rho}{M}\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}\right)^{-1}\right]\right) \\ &= -M\log\left(\frac{1}{M}\sum_{k=1}^{M}\frac{1}{1+\frac{\rho}{M}\lambda_k}\right).\end{aligned}$$

Using this bound in (5) we obtain

$$\begin{aligned}P^{\mathrm{mmse}}_{\mathrm{out}}(R,\rho) &\leq P\left(\log\left(\frac{1}{M}\sum_{k=1}^{M}\frac{1}{1+\frac{\rho}{M}\lambda_k}\right) \geq -\frac{R}{M}\right) \\ &= P\left(\frac{1}{M}\sum_{k=1}^{M}\frac{1}{1+\frac{\rho}{M}\lambda_k} \geq \rho^{-\frac{r}{M}}\right)\end{aligned} \tag{13}$$

where in the last line we let $R = r\log\rho$. Finally, we can use the trivial asymptotic upper bound

$$P\left(\frac{1}{M}\sum_{k=1}^{M}\frac{1}{1+\frac{\rho}{M}\lambda_k} \geq \rho^{-\frac{r}{M}}\right) \dot{\leq} P\left(\frac{1}{\rho\lambda_1} \geq \rho^{-\frac{r}{M}}\right) \tag{14}$$

First, we notice that the asymptotic outage probability upper bound in the right-hand side (RHS) of (14) vanishes only if

[4]Note: $(x)^+ \triangleq \max\{x, 0\}$.

$r/M < 1$. Hence, the outage exponent lower bound is zero for $r/M \geq 1$. When $r/M < 1$, we can write

$$\begin{aligned} P\left(\lambda_1 \leq \rho^{\frac{r}{M}-1}\right) &= \int_0^{\rho^{\frac{r}{M}-1}} d\lambda_1 \prod_{i=2}^{M}\left[\int_{\lambda_1}^{\infty} d\lambda_i\right] p(\boldsymbol{\lambda}) \\ &\leq \int_0^{\rho^{\frac{r}{M}-1}} d\lambda_1 \prod_{i=2}^{M}\left[\int_{0}^{\infty} d\lambda_i\right] p(\boldsymbol{\lambda}) \\ &= \int_0^{\rho^{\frac{r}{M}-1}} p_1(\lambda_1) d\lambda_1 \\ &= \kappa_1 \rho^{(N-M+1)(r/M-1)} \end{aligned} \tag{15}$$

where κ_1 is a constant and where we have used the well-known fact [4] that the marginal pdf of $\lambda_1 = \lambda_{\min}(\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H})$, denoted by $p_1(\lambda)$ in (15), satisfies $p_1(\lambda) \propto \lambda^{N-M}$ for small argument $\lambda \ll 1$. The resulting outage exponent lower bound is

$$d^*_{\mathrm{mmse}}(r) \geq (N-M+1)\left(1-\frac{r}{M}\right)^{+}. \tag{16}$$

The same result can be obtained by following the by-now standard technique of [3] based on the change of variable $\lambda_i = \rho^{-\alpha_i}$, integrating the resulting pdf of $\alpha_1, \ldots, \alpha_M$ over the outage region and applying Varadhan's lemma [3].

Upper bound on the outage exponent. Using the concavity and the monotonicity of the $\log(\cdot)$ function, we obtain from (12) and Jensen's inequality that

$$I_{\mathrm{mmse}}(\boldsymbol{H}) \leq M \log\left(\frac{1}{M}\sum_{k=1}^{M}\frac{1}{[(\boldsymbol{I}+\rho\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H})^{-1}]_{kk}}\right). \tag{17}$$

Consider the decomposition $\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H} = \boldsymbol{U}^{\mathsf{H}}\boldsymbol{\Lambda}\boldsymbol{U}$, where $\boldsymbol{U}$ is unitary and $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues of $\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}$ on the diagonal. Defining $\boldsymbol{u}_k$ to be the kth column of $\boldsymbol{U}$ and $\boldsymbol{e}_k$ to be the column vector that has a one in the kth component and zeros elsewhere, we have that

$$\begin{aligned} [(\boldsymbol{I}+\rho\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H})^{-1}]_{kk} &= \boldsymbol{e}_k^{\mathsf{H}}\boldsymbol{U}^{\mathsf{H}}(\boldsymbol{I}+\rho\boldsymbol{\Lambda})^{-1}\boldsymbol{U}\boldsymbol{e}_k \\ &= \boldsymbol{u}_k^{\mathsf{H}}(\boldsymbol{I}+\rho\boldsymbol{\Lambda})^{-1}\boldsymbol{u}_k \\ &= \sum_{\ell=1}^{M}\frac{|u_{\ell k}|^2}{1+\rho\lambda_\ell}. \end{aligned}$$

Hence, the term inside the logarithm in (17) can be upper-bounded as

$$\begin{aligned} \frac{1}{M}\sum_{k=1}^{M}\frac{1}{[(\boldsymbol{I}+\rho\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H})^{-1}]_{kk}} &= \frac{1}{M}\sum_{k=1}^{M}\frac{1}{\sum_{\ell=1}^{M}\frac{|u_{\ell k}|^2}{1+\rho\lambda_\ell}} \\ &= \frac{1}{M}\sum_{k=1}^{M}\frac{1}{\frac{|u_{1k}|^2}{1+\rho\lambda_1}} \\ &\quad \cdot \frac{1}{\left[1+\sum_{\ell=2}^{M}\frac{|u_{\ell k}|^2}{|u_{1k}|^2}\frac{1+\rho\lambda_1}{1+\rho\lambda_\ell}\right]} \\ &\leq \frac{1+\rho\lambda_1}{M}\sum_{k=1}^{M}\frac{1}{|u_{1k}|^2}. \end{aligned} \tag{18}$$

Let $\mathcal{A}$ denote the event $\left\{\frac{1}{M}\sum_{k=1}^{M}\frac{1}{|u_{1k}|^2} \leq c\right\}$, where c is some constant (independent of ρ). We have that

$$\begin{aligned} P_{\mathrm{out}}^{\mathrm{mmse}}(R,\rho) &\geq P(\mathcal{A}) P\left(\log\left((1+\rho\lambda_1)\frac{1}{M}\sum_{k=1}^{M}\frac{1}{|u_{1k}|^2}\right) \leq \frac{R}{M}\middle|\mathcal{A}\right) \\ &\geq P(\mathcal{A}) P\left(\log((1+\rho\lambda_1)c) \leq \frac{R}{M}\right) \\ &\doteq P\left(\log(1+\rho\lambda_1) \leq \frac{r}{M}\log\rho\right) \end{aligned} \tag{19}$$

where the last exponential equality holds if $P(\mathcal{A})$ is a $O(1)$ nonzero term, i.e., it is a constant with respect to ρ bounded away from zero. This is indeed the case, as shown rigorously in Appendix A.

It is immediate to check that the last line of (19) is asymptotically equivalent to (14). Therefore, applying the same argument as in (15), we find that the upper bound on the outage probability exponent coincides with the previously found lower bound.

The proof of Theorem 1 is completed by observing that in the case of the ZF receiver a lower bound on the SINR γ_k is readily obtained from the inequality

$$[(\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H})^{-1}]_{kk} \leq \lambda_{\max}[(\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H})^{-1}] = \frac{1}{\lambda_{\min}(\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H})} = \frac{1}{\lambda_1}$$

that holds for all $k = 1, \ldots, M$. Using this in the mutual information expression for the ZF receiver with coding across the antennas we obtain

$$P_{\mathrm{out}}^{\mathrm{zf}}(R,\rho) \leq P\left(\log(1+\rho\lambda_1) \leq \frac{r}{M}\log\rho\right). \tag{20}$$

Noticing that (20) coincides with the asymptotic lower bound (19) for the MMSE receiver, and that the MMSE receiver maximizes the mutual information over all linear receivers, under Gaussian inputs and the system assumptions made here, we immediately obtain that the ZF also achieves the outage exponent $d^*_{\mathrm{lin}}(r)$ given in (10).

Finally, as far as spatial multiplexing is concerned (no coding across the antennas), it is clear from (5) and (6) that, for any linear receiver $\boldsymbol{G}$, $P_{\mathrm{out}}^{\mathrm{lin}}(R,\rho) \leq P_{\mathrm{out}}^{\mathrm{sp\ mult}}(R,\rho)$. On the other hand, it is immediate to show that spatial multiplexing achieves the same DMT (10). Details are trivial, and are omitted. □

A. Discussion and Numerical Results

Theorem 1 shows that, in terms of DMT, there is no advantage in using interleaving and coding across the antennas when a linear receiver is used in order to spatially separate the transmitted symbols. In order words, the linear receiver front-end kills the *transmit* diversity gain offered by the MIMO channel. In fact, the DMT $(N-M+1)(1-\frac{r}{M})^{+}$ of Theorem 1 has the following intuitive interpretation: this coincides with the DMT of a SIMO (single-input, multiple-output) channel (receiver diversity only) with $N-M+1$ receive antennas, used at a rate R/M.

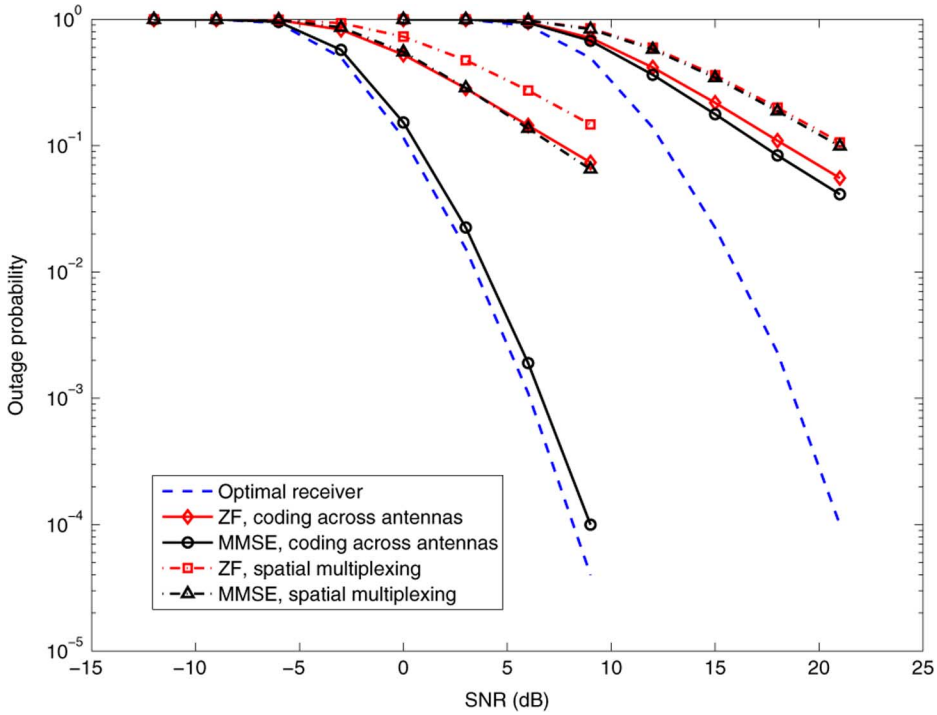

Fig. 2. Outage probabilities of ZF and MMSE receivers, 2×2 i.i.d. Rayleigh channel, $R = 1$ and 5 bpcu.

This fact shows also that the channel gains of the virtual parallel channels are *strongly* statistically dependent. For example, it is well known that the ZF receiver applied to an $M \times N$ channel with $N \geq M$ and i.i.d. Rayleigh fading yields channel gains $\gamma_k = \frac{1}{[(\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H})^{-1}]_{kk}}$ that are marginally distributed as central Chi-squared random variables with $2(N - M + 1)$ degrees of freedom [22]. If the gains $\gamma_1, \ldots, \gamma_M$ were statistically independent, by coding across the antennas we would obtain the DMT of the parallel independent channels, given by [15]

$$d^*_{\text{parallel, i.i.d.}}(r) \geq (N - M + 1)(M - r)^+$$

which is much larger than the DMT given by Theorem 1. In contrast, the channel gains in the regime of high SNR are essentially dominated by the minimum eigenvalue of the matrix $\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}$ and therefore are strongly correlated: if one subchannel is in deep fade, they are all in deep fade with high probability. This is the reason why coding across the transmit antennas does not buy any improvement in terms of DMT with respect to simple spatial multiplexing.

Having said so, we should also remark that the picture about linear receivers is not totally grim as it may appear from the high-SNR DMT analysis. Indeed, coding across antennas yields a very significant performance advantage with the linear MMSE receiver at fixed and not too large rate (notice that fix rate R corresponds to the case of zero multiplexing gain, $r = 0$.) In order to illustrate these claims, we provide simulations results for the following outage probabilities under i.i.d. Rayleigh fading:

- MIMO outage probability (3) with input covariance $(\rho/M)\boldsymbol{I}$ (scheme (a) in Fig. 1);
- outage probability (5) with ZF and MMSE receivers with coding across antennas (scheme (b) in Fig. 1);
- outage probability (6) with ZF and MMSE receivers under pure spatial multiplexing, i.e., without coding across antennas (scheme (c) in Fig. 1).

Fig. 2 shows the corresponding plots at rates $R = 1$ and 5 bits per channel use (bpcu).

Several interesting observations can be drawn from this figure. We observe that while at high rates the MMSE with coding across antennas behaves as predicted by the DMT analysis, the behavior at low rates is in stark contrast to the asymptotic result (this fact was also noticed in [20]). In fact, the MMSE exhibits an apparent "full diversity" behavior at small rate (e.g., $R = 1$ bpcu in Fig. 2). In contrast, the behavior of the ZF receiver is accurately predicted by the asymptotic analysis at all rates. This remarkable behavior of the MMSE receiver is explained through an approximate analysis in Section IV.

From Fig. 2 we observe also that coding across antennas does achieve an advantage over spatial multiplexing. For the MMSE receiver operating at small rates the advantage is very significant, and corresponds to the diversity advantage discussed above. At high rates the advantage is moderate and consists only of a horizontal shift (decibel gain) of the error curve, not in a steeper slope.

IV. MMSE Receiver With Coding Across Antennas

The difference between the performances of the ZF and MMSE receivers is best explained by comparing their corresponding upper bounds on outage probability in (20) and (13). While only the minimum eigenvalue appears in the ZF case in (20), all eigenvalues play a role in the case of the MMSE receiver in (13). Although at asymptotically high SNR and high coding rates the minimum eigenvalue dominates (and therefore determines the corresponding DMT), the other eigenvalues

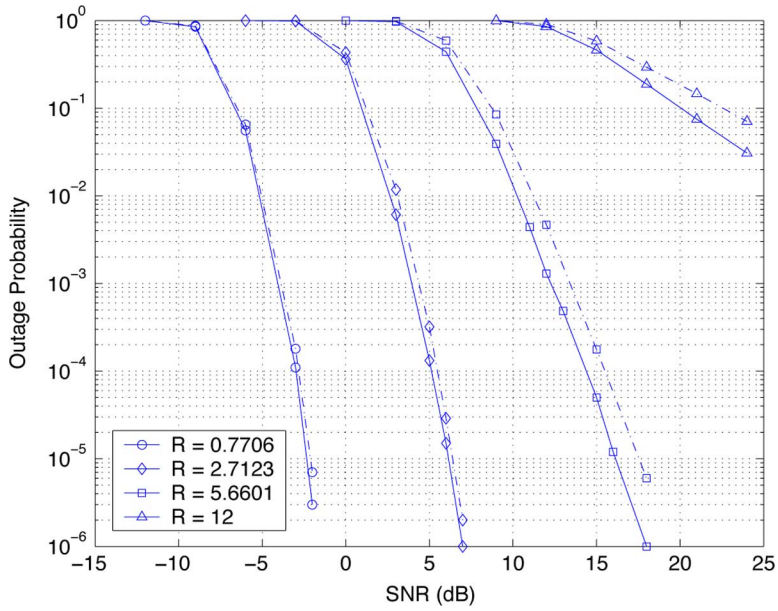

Fig. 3. Diversity of the MMSE receiver with joint spatial encoding: solid lines represent the outage probability in (5) and the dash-dot lines represent the corresponding upper bounds (13). $M = N = 4$, rates R are in bpcu.

appear to be relevant at lower rates and provide higher effective diversity for the MMSE receiver. In order to substantiate this intuition, we compare in Fig. 3 the outage probability of the MMSE receiver with coding across antennas for the case $M = N = 4$ with the corresponding upper bound in (13). The upper bound is found to be very accurate across a wide range of rates and SNRs. The particular choice of rates for this plot will be made clear in the sequel, where we analyze the high SNR behavior of the outage probability upper bound (13).

Define $\mathfrak{T}_k \triangleq \frac{1}{1+\frac{\rho}{M}\lambda_k}$ and $\mathfrak{T} \triangleq M2^{-\frac{R}{M}}$. We use a change of variables $\lambda_k = \rho^{-\alpha_k}$, where α_k denotes the *level of singularity* of the corresponding eigenvalue [3]. For ease of analysis, we make the assumption that the channel eigenvalues fall into one of the following two categories:

- $\alpha_k < 1$, i.e., λ_k is "much larger" than the inverse SNR $1/\rho$: in this case, $\mathfrak{T}_k \to 0$ as $\rho \to \infty$;
- $\alpha_k > 1$, i.e., λ_k is "much smaller" than $1/\rho$: in this case, $\mathfrak{T}_k \to 1$ as $\rho \to \infty$.

Recall that the $\{\alpha_i\}$ are ordered according to $\alpha_1 \geq \cdots \geq \alpha_M$. Suppose that the rate R is such that $m-1 < \mathfrak{T} \leq m$, for some integer $m = 1, 2, \ldots, M$, i.e.,

$$M \log \frac{M}{m} \leq R < M \log \frac{M}{m-1}. \tag{21}$$

For all $i = 1, \ldots, M$ define the event

$$\mathcal{E}_i = \{\alpha_1, \ldots, \alpha_i > 1\} \cap \{\alpha_{i+1}, \ldots, \alpha_M < 1\}. \tag{22}$$

Then, for large ρ, the following approximation holds:

$$\begin{aligned}\left\{\sum_{k=1}^{M} \mathfrak{T}_k \geq \mathfrak{T}\right\} &\approx \bigcup_{i=m}^{M} \{\alpha_1, \ldots, \alpha_i > 1\} \\ &\quad \cap \{\alpha_{i+1}, \ldots, \alpha_M < 1\} \\ &= \mathcal{E}_m \cup \mathcal{E}_{m+1} \cup \cdots \cup \mathcal{E}_M. \end{aligned} \tag{23}$$

In the preceding approximation we are neglecting the cases where the eigenvalues take on values that are comparable with $1/\rho$, and therefore contribute to the sum $\sum_{k=1}^{M} \mathfrak{T}_k$ in (13) by a quantity between 0 and 1. It can be expected that as $\rho \to \infty$, the probability of such intermediate values decreases, and our approximation becomes tight.

Using the union bound, we find an approximate upper bound on (13) given by

$$P\left(\sum_{k=1}^{M} \mathfrak{T}_k \geq \mathfrak{T}\right) \lesssim \sum_{i=m}^{M} P(\mathcal{E}_i). \tag{24}$$

Defining $P(\mathcal{E}_i) \doteq \rho^{-\tilde{d}_i(R)}$, $i = 1, \ldots, M$, using the joint pdf of the α_k's, given by [3]

$$\begin{aligned} p(\boldsymbol{\alpha}) &= K_{M,N}[\log(\rho)]^M \prod_{i=1}^{M} \rho^{-(N-M+1)\alpha_i} \\ &\quad \cdot \prod_{i<j} (\rho^{-\alpha_i} - \rho^{-\alpha_j})^2 \exp\left(-\sum_{i=1}^{M} \rho^{-\alpha_i}\right) \\ &\doteq \left[\prod_{i=1}^{M} \rho^{-(2i-1+N-M)\alpha_i}\right] \exp\left(-\sum_{i=1}^{M} \rho^{-\alpha_i}\right) \end{aligned}$$

and applying Varadhan's lemma as in [3], we obtain

$$\begin{aligned} \tilde{d}_i(R) &= \inf_{\substack{\alpha_j > 1 \ \forall j \leq i \\ \alpha_j < 1 \ \forall j > i \\ \alpha_j \geq 0 \ \forall \ j}} \sum_{j=1}^{M} (2j - 1 + M - N)\alpha_j \\ &= \sum_{j=1}^{i} (2j - 1 + M - N) \times 1 \\ &\quad + \sum_{j=i+1}^{M} (2j - 1 + M - N) \times 0 \\ &= i(i + N - M). \end{aligned} \tag{25}$$

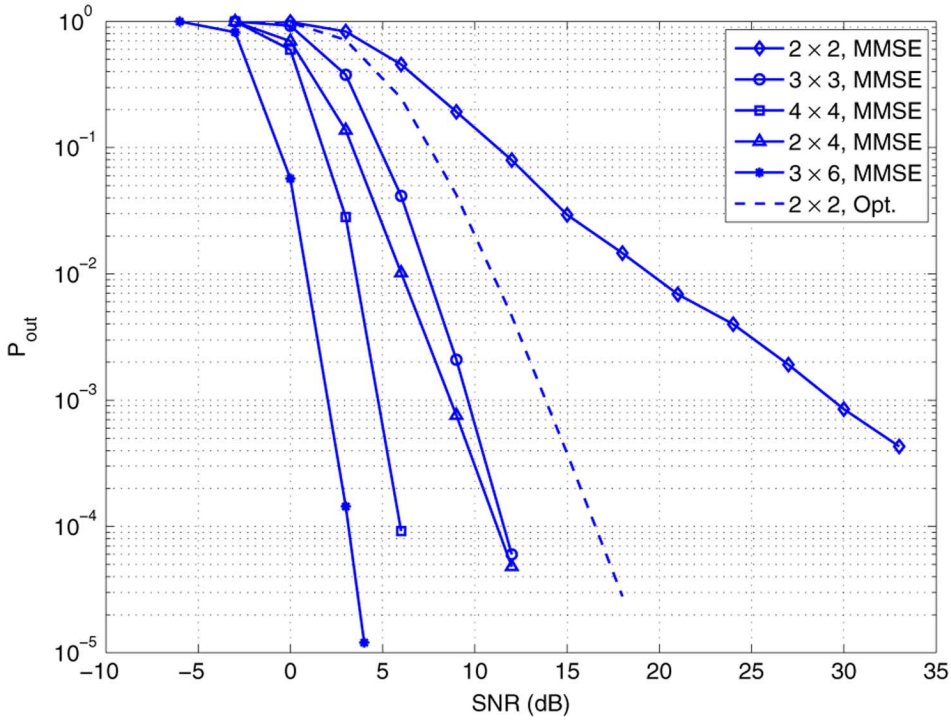

Fig. 4. Comparing the outage probability of optimal and MMSE receivers, $R = 3$ bpcu.

From (24) and (25), we eventually conclude that

$$P\left(\sum_{k=1}^{M} \mathfrak{T}_k \geq \mathfrak{T}\right) \dot{\leq} P(\mathcal{E}_m).$$

This yields the diversity of the MMSE receiver with spatial encoding at a finite rate R as

$$d_{\text{mmse}}(R) \approx m(m+N-M). \tag{26}$$

In particular, when $M = N$, $d_{\text{mmse}}(R) \approx m^2$ where m and R are related by (21).

To illustrate the effectiveness of the above approximation, consider the plots in Fig. 3 for the case $M = N = 4$. The coding rates are $R = 0.7706$, 2.7123, 5.6601, and 12 bpcu, corresponding to $\mathfrak{T} = 3.5, 2.5, 1.5$, and 0.5, respectively. The diversities $16, 9, 4$, and 1 predicted by the analysis in (26) well approximate the measured slopes (for high SNR) of the outage curves, that are $15.15, 10.69, 5.55$, and 1.3 in the $\log P_{\text{out}}^{\text{mmse}}(R,\rho)$ versus $\log \rho$ chart observed in Fig. 3.

V. Outage Probability of Linear Receivers in the Large Antenna Regime

In order to motivate this section, consider the following system design issue: for a given target spectral efficiency, block-error rate, operating SNR, and receiver computational complexity (including power consumption, VLSI chip area, etc.) how many antennas do we need at the transmitter and receiver? Consider the outage probability curves of Fig. 4 and suppose that we wish to achieve a rate of $R = 3$ bpcu with block-error rate of 10^{-3} at SNR not larger than 15 dB. With $M = N = 2$ antennas, this target performance is achieved by an optimal receiver, but is not achieved by the MMSE receiver. However, with $M = 2, N = 4$, or $M = N = 3$, the target performance is achieved also by the MMSE receiver. It turns out that, in some cases, adding antennas may be more convenient than insisting on high-complexity receiver processing.

It is therefore interesting to analyze the outage probability of a linear receiver with coding across the antennas in the regime of fixed SNR ρ and rate R. This analysis is difficult due to the fact that, for finite M, N, the joint distribution of the channel SINRs $\{\gamma_k\}$ in (4) escapes a closed-form expression. This problem can be overcome by considering the system in the limit of a large number of antennas. Specifically, we will show that the mutual information for the linear MMSE and ZF receivers becomes asymptotically Gaussian. Therefore, the outage probability for large but finite dimensions and fixed SNR can be accurately approximated by a Gaussian cdf with appropriate mean and variance, that we shall give in closed form.

In the next subsection, we will discuss the methodology used to show the asymptotic Gaussianity of the mutual information. The method is general and applies to both MMSE and ZF linear receivers. Subsequently, in Section V-B we will calculate the first and second cumulant moments of the SINR for the MMSE and ZF receivers, which suffice to characterize the mutual information limiting distribution.

A. Asymptotic Gaussianity of the Mutual Information

The mutual information at the output of a linear receiver with M transmit and $N \geq M$ receive antennas and coding across the antennas is given by

$$I_N \triangleq \sum_{k=1}^{M} \log(1+\gamma_k) \tag{27}$$

with γ_k given by (9) for the MMSE case and by (7) for the ZF case. In the following, we fix the ratio $\beta = M/N \leq 1$ and consider the limit for large N and the "fluctuations" around this limit. In order to prove the asymptotic Gaussianity of these fluctuations, we will analyze the characteristic function of the mutual information, given by

$$\Phi_N(\omega) \triangleq \mathbb{E}[e^{j\omega I_N}]. \tag{28}$$

We start by considering the *cumulant generating function* [33], defined as

$$\phi_N(\omega) \triangleq \log(\Phi_N(\omega)) = \sum_{n=1}^{\infty} \frac{(j\omega)^n}{n!}\mathcal{C}_n \tag{29}$$

where the coefficient $\mathcal{C}_n$ is the nth cumulant moment of the mutual information. In general, the *joint cumulant* of m random variables $X_1, \ldots, X_m$ is defined as

$$\mathbb{E}_c(X_1; \ldots; X_m) \triangleq \sum_{\pi} (|\pi| - 1)!(-1)^{|\pi|-1} \prod_{B \in \pi} \mathbb{E}\left[\prod_{i \in B} X_i\right]$$

where π runs through all partitions of $\{1, \ldots, m\}$, $|\pi|$ denotes the number of blocks in π, and B runs through the list of all blocks of π. We will call the above moment irreducible, with respect to the random variables $X_1, \ldots, X_m$, when in each argument of the cumulant moment only one random variable X_i appears. By contrast, a *reducible* cumulant moment with respect to the same random variables has arguments containing mixed products of these random variables. In general, an n-order reducible cumulant moment can be written in terms of a sum of products of irreducible cumulant moments, with each term in the sum having moments with order summing up to n.

The nth cumulant moment $\mathcal{C}_n$ of a random variable X is defined to be

$$\mathcal{C}_n \triangleq \mathbb{E}_c(\underbrace{X; \ldots; X}_{n \text{ times}}).$$

For example, the first few cumulant moments of X are

$$\begin{aligned} \mathcal{C}_1 &= \mathbb{E}[X] && \text{mean} \\ \mathcal{C}_2 &= \mathrm{Var}[X] && \text{variance} \\ \mathcal{C}_3 &= \mathrm{Sk}[X] && \text{skewness.} \end{aligned} \tag{30}$$

The probability density of I_N can be expressed in terms of (28) and (29) as follows:

$$\begin{aligned} p(y) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-j\omega y}\Phi_N(\omega)d\omega \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-j\omega(y - \mathcal{C}_1) - \frac{\omega^2}{2}\mathcal{C}_2 \right. \\ &\qquad \left. + \sum_{n>2} \frac{(j\omega)^n}{n!}\mathcal{C}_n\right) d\omega. \end{aligned} \tag{31}$$

In Section V-B, we will show that in the limit of large N and $M = \beta N$ with $\beta \leq 1$

$$\mathcal{C}_1 = m_1 + o(1) \tag{32}$$

$$\mathcal{C}_2 = \sigma^2 + o(1) \tag{33}$$

where $m_1 = Mc_{10} + c_{11}$, and where c_{10}, c_{11}, and σ^2 are constants independent of N for which we give closed-form expressions for both MMSE and ZF cases. In Appendix D, we will also show that all higher order cumulants of the mutual information asymptotically vanish for large N. Therefore, $\phi_N(\omega)$ is a quadratic function of ω with corrections that vanish as $N \to \infty$. As a result, the mutual information is asymptotically Gaussian, i.e.,

$$\frac{I_N - m_1}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1). \tag{34}$$

This follows directly from (31) by setting $y = z + m_1$ and taking the large N limit

$$\begin{aligned} p(z) &= \lim_{N\to\infty} \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left(-j\omega z - \frac{\omega^2}{2}\sigma^2 + o(1)\right) d\omega \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{z^2}{2\sigma^2}}. \end{aligned}$$

Before moving on to the proofs, we would like to comment on the nature of this result. This states that the probability $P(|I_N - m_1| > z)$ approaches a Gaussian probability for sufficiently large N and *fixed* distance z of the mutual information from its mean. This is quite different from stating that for fixed N the mutual information distribution falls off like a Gaussian random variable for any z and ρ. As a matter of fact, for *fixed* N and large enough SNR this Gaussian approximation is no longer valid, since the higher order cumulants will no longer be small.

It is also worth pointing out here that the variance of I_N is $O(1)$ (a *finite* constant) for large N. This is another manifestation of the fact that the SINRs of the parallel channels $\{\gamma_k\}$ are strongly correlated, in agreement with the outage analysis of previous sections. In contrast, if they were independent, or nearly independent, the variance would be roughly linear in N, as the central limit theorem would suggest. This fact is in line with the well-known behavior of the mutual information $\log\det(\boldsymbol{I} + \frac{\rho}{M}\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}})$ under the optimal receiver [24]–[27], where again the variance is $O(1)$ for large N, indicating the strong correlation among the eigenvalues of $\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H}$.

B. Joint Cumulant Moments of the SINRs of Order 1 and 2

Our goal is to calculate the cumulant moments of I_N. Since I_N consists of a sum of mutual informations of the virtual channels (see (27)), the nth cumulant moment of I_N can be written as

$$\mathcal{C}_n = \sum_{k_1,\ldots,k_n=1}^{M} \mathbb{E}_c\left[\log(1 + \gamma_{k_1}); \ldots; \log(1 + \gamma_{k_n})\right]. \tag{35}$$

The building blocks of the above cumulant moments are the joint cumulant moments of the SINRs $\{\gamma_k\}$, i.e.,

$$\mathbb{E}_c\left[\gamma_{k_1}; \gamma_{k_2}; \ldots; \gamma_{k_n}\right]. \tag{36}$$

In fact, by expanding the logarithms in (35) in Taylor series, we can express (35) in terms of (36). Even calculating these

joint cumulants amounts generally to a formidable task. However, with the help of Theorem 2 (Novikov's theorem) given in Appendix B and due to simplifications that occur in the large N limit, we will show that this computation is possible. To obtain a feel for the computation, we will first calculate the first two joint cumulants of $\{\gamma_k\}$ and defer the proof that the higher order cumulants vanish sufficiently fast with N to Appendix D.

1) Cumulant Moments for the MMSE Receiver: Starting with the case of the MMSE receiver, we recall from (9) that the SINR of the kth virtual channel induced by the MMSE receiver can be written as

$$\gamma_k = \alpha \boldsymbol{h}_k^{\mathsf{H}} \left[\boldsymbol{I} + \alpha \boldsymbol{H}_k \boldsymbol{H}_k^{\mathsf{H}} \right]^{-1} \boldsymbol{h}_k \tag{37}$$

where $\boldsymbol{H}_k$ is the $N \times (M-1)$ matrix obtained by eliminating the kth column $\boldsymbol{h}_k$ from the channel matrix $\boldsymbol{H}$, and contains i.i.d. Gaussian elements $\sim \mathcal{CN}(0, 1/N)$ and we have defined for convenience $\alpha = \rho N/M = \rho/\beta$.

The asymptotic mean of γ_k in the limit of large N and $M = \beta N$ has been calculated in [35] in the context of large-system analysis of code-division multiple access (CDMA) with random spreading, and successively rederived in various ways (e.g., [34], [28], [30]). Due to symmetry, the result does not depend on the index k. Hence, without loss of generality we can choose $k = 1$. We have

$$\mathbb{E}[\gamma_1] = \alpha \mathbb{E}\left[\frac{1}{N} \mathrm{Tr}\left(\left[\boldsymbol{I} + \alpha \boldsymbol{H}_1 \boldsymbol{H}_1^{\mathsf{H}} \right]^{-1} \right) \right]. \tag{38}$$

The leading order in N of the above trace can be evaluated as

$$\begin{aligned} g_1^{\mathsf{mmse}}(\alpha, \beta) &= \lim_{N \to \infty} \alpha \mathbb{E}\left[\frac{1}{N} \mathrm{Tr}\left(\left[\boldsymbol{I} + \alpha \boldsymbol{H}_1 \boldsymbol{H}_1^{\mathsf{H}} \right]^{-1} \right) \right] \\ &= \frac{\alpha}{1 + \frac{\alpha \beta}{1 + g_1^{\mathsf{mmse}}(\alpha, \beta)}}. \end{aligned} \tag{39}$$

Solving for $g_1^{\mathsf{mmse}}(\alpha, \beta)$ in (39), we obtain

$$g_1^{\mathsf{mmse}}(\alpha, \beta) = \frac{1}{2}\left[\alpha(1-\beta) - 1 + \sqrt{(\alpha(1-\beta)-1)^2 + 4\alpha} \right]. \tag{40}$$

To be able to calculate the $O(1)$ correction to the mean mutual information, we need to evaluate the next to leading $(O(1/N))$ correction to $\mathbb{E}[\gamma_1]$. The correction follows by noticing that the term β in (39) should be replaced by the aspect ratio of the matrix $\boldsymbol{H}_1$. For large but finite N, this is equal to $(M-1)/N = \beta - 1/N$. Therefore, the correction can be evaluated by replacing β by $\beta - 1/N$ in (40). Using the Taylor series expansion, this amounts to computing

$$\begin{aligned} \mathbb{E}[\gamma_1] &= g_1^{\mathsf{mmse}}\left(\alpha, \beta - \frac{1}{N} \right) \\ &= g_1^{\mathsf{mmse}}(\alpha, \beta) - \frac{1}{N} \frac{\partial}{\partial \beta} g_1^{\mathsf{mmse}}(\alpha, \beta) + O(N^{-2}) \end{aligned} \tag{41}$$

where

$$\frac{\partial}{\partial \beta} g_1^{\mathsf{mmse}}(\alpha, \beta) = -\frac{\alpha}{2}\left[1 + \frac{\alpha(1-\beta)-1}{\sqrt{(\alpha(1-\beta)-1)^2 + 4\alpha}} \right].$$

For later use, we also define the following asymptotic moments:

$$g_m^{\mathsf{mmse}}(\alpha, \beta) \triangleq \lim_{N \to \infty} \alpha^m \mathbb{E}\left[\frac{1}{N} \mathrm{Tr}\left(\left[\boldsymbol{I} + \alpha \boldsymbol{H}_1 \boldsymbol{H}_1^{\mathsf{H}} \right]^{-m} \right) \right]$$

which can be obtained by repeatedly differentiating $g_1^{\mathsf{mmse}}(\alpha, \beta)$ with respect to α using the recursive relation

$$g_{m+1}^{\mathsf{mmse}}(\alpha, \beta) = \frac{\alpha^2}{m} \frac{\partial}{\partial \alpha} g_m^{\mathsf{mmse}}(\alpha, \beta), \quad m \geq 1. \tag{42}$$

Thus, we have

$$\begin{aligned} g_2^{\mathsf{mmse}}(\alpha, \beta) &= \alpha^2 \frac{\partial}{\partial \alpha} g_1^{\mathsf{mmse}}(\alpha, \beta) \\ g_3^{\mathsf{mmse}}(\alpha, \beta) &= \frac{\alpha^3}{2}\left(\alpha \frac{\partial^2}{\partial \alpha^2} g_1^{\mathsf{mmse}}(\alpha, \beta) + 2 \frac{\partial}{\partial \alpha} g_1^{\mathsf{mmse}}(\alpha, \beta) \right). \end{aligned}$$

For large SNR, i.e., $\alpha = \rho/\beta \gg 1$ and $\beta < 1$, $g_1^{\mathsf{mmse}}(\alpha, \beta)$ is approximately $\alpha(1-\beta)$. This result indicates that only the $\approx N(1-\beta)$ zero eigenvalues of the matrix $\boldsymbol{H}_1 \boldsymbol{H}_1^{\mathsf{H}}$ contribute to the SINR for large α. Similarly, $g_m^{\mathsf{mmse}}(\alpha, \beta) \approx (1-\beta)\alpha^m$ for large ρ and $\beta < 1$, while for $\beta = 1$, $g_m^{\mathsf{mmse}}(\alpha, \beta) \approx k_m \rho^{m-1/2}$, where the constant k_m satisfies $k_{m+1} = \prod_{j=1}^{m}(1 - 1/2j)$.

Next we calculate the matrix of the joint cumulants of order 2 with elements

$$\Sigma_{i,j}^{\mathsf{mmse}} = \mathbb{E}_c[\gamma_i; \gamma_j] \equiv \mathbb{E}[\gamma_i \gamma_j] - \mathbb{E}[\gamma_i]\mathbb{E}[\gamma_j].$$

Given the symmetry, all diagonal elements $(i = j)$ are equal, and so are all off-diagonal ones $(i \neq j)$. Therefore, it is sufficient to compute $\Sigma_{1,1}^{\mathsf{mmse}}$ and $\Sigma_{1,2}^{\mathsf{mmse}}$.

We start with $\mathbb{E}_c[\gamma_1; \gamma_1]$. For convenience, we define $\boldsymbol{B}_1 \triangleq (\boldsymbol{I} + \alpha \boldsymbol{H}_1 \boldsymbol{H}_1^{\mathsf{H}})^{-1}$, and let $(\boldsymbol{B}_1)_{ij}$ denote the (i,j)th element of $\boldsymbol{B}_1$ and h_{1i} denote the ith element of $\boldsymbol{h}_1$. Then, a direct application of (37) and (77) yields

$$\begin{aligned} \Sigma_{1,1}^{\mathsf{mmse}} &= \mathbb{E}_c[\gamma_1; \gamma_1] \\ &= \alpha^2 \sum_{a,b,c,d} \mathbb{E}\left[(\boldsymbol{B}_1)_{ab} (\boldsymbol{B}_1)_{cd} \left(h_{1a}^* h_{1b} h_{1c}^* h_{1d} \right.\right. \\ &\qquad \left.\left. - \frac{\delta_{a,b}}{N} \frac{\delta_{c,d}}{N} \right) \right] \\ &= \alpha^2 \sum_{a,b,c,d} \mathbb{E}[(\boldsymbol{B}_1)_{ab} (\boldsymbol{B}_1)_{cd}] \mathbb{E}_c[h_{1a}^* h_{1b}; h_{1c}^* h_{1d}] \\ &= \alpha^2 \mathbb{E}\left[\frac{1}{N^2} \mathrm{Tr}(\boldsymbol{B}_1^2) \right] \\ &\to \frac{v_d^{\mathsf{mmse}}}{M} + O(1/N^2) \end{aligned} \tag{43}$$

where $v_d^{\mathsf{mmse}} = \beta g_2^{\mathsf{mmse}}(\alpha, \beta - \frac{1}{N})$. We see that the leading correction in the autocorrelation is nonvanishing only due to the random character of the vector $\boldsymbol{h}_1$ [28].

We now turn to the more complicated computation of $\Sigma_{1,2}^{\mathsf{mmse}}$ to leading order in N. To simplify notation, we define the matrices $\boldsymbol{B}_i = (\boldsymbol{I} + \alpha \boldsymbol{H}_i \boldsymbol{H}_i^{\mathsf{H}})^{-1}$, for $i = 1, 2$, as before, and $\boldsymbol{B}_{12} = (\boldsymbol{I} + \alpha \boldsymbol{H}_{12} \boldsymbol{H}_{12}^{\mathsf{H}})^{-1}$ where $\boldsymbol{H}_{12}$ is obtained by striking out from $\boldsymbol{H}$ both columns $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$. Therefore

$$\begin{aligned} \gamma_1 &= \alpha \boldsymbol{h}_1^{\mathsf{H}} \boldsymbol{B}_1 \boldsymbol{h}_1 \\ \gamma_2 &= \alpha \boldsymbol{h}_2^{\mathsf{H}} \boldsymbol{B}_2 \boldsymbol{h}_2. \end{aligned} \tag{44}$$

Using the same notation as before, we rewrite the cumulant moment of γ_1, γ_2 as

$$\mathbb{E}_c[\gamma_1; \gamma_2] = \alpha^2 \sum_{abcd} \mathbb{E}_c[h_{1a}^* (\boldsymbol{B}_1)_{ab} h_{1b}; h_{2c}^* (\boldsymbol{B}_2)_{cd} h_{2d}]. \tag{45}$$

In the following, we will make extensive use of the following matrix identities, obtained by applying the Sherman–Morrison matrix inversion lemma

$$\begin{aligned}\boldsymbol{B}_2 &= \boldsymbol{B}_{12} - \boldsymbol{B}_{12}\boldsymbol{h}_1\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{B}_{12}\frac{\alpha}{1+\alpha\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{B}_{12}\boldsymbol{h}_1}\\ \boldsymbol{B}_1 &= \boldsymbol{B}_{12} - \boldsymbol{B}_{12}\boldsymbol{h}_2\boldsymbol{h}_2^{\mathsf{H}}\boldsymbol{B}_{12}\frac{\alpha}{1+\alpha\boldsymbol{h}_2^{\mathsf{H}}\boldsymbol{B}_{12}\boldsymbol{h}_2}.\end{aligned} \tag{46}$$

We will now use Novikov's theorem (Theorem 2 in Appendix B) to successively average over the variables $\boldsymbol{h}_1$ and $\boldsymbol{h}_2$. For example, considering the general term for indices (a,b,c,d) in (45) we write

$$\begin{aligned}&\mathbb{E}_c[h_{1a}^*(\boldsymbol{B}_1)_{ab}h_{1b};h_{2c}^*(\boldsymbol{B}_2)_{cd}h_{2d}]\\ &\quad=\mathbb{E}[h_{1a}^*(\boldsymbol{B}_1)_{ab}h_{1b}h_{2c}^*(\boldsymbol{B}_2)_{cd}h_{2d}]\\ &\qquad-\mathbb{E}[h_{1a}^*(\boldsymbol{B}_1)_{ab}h_{1b}]\mathbb{E}[h_{2c}^*(\boldsymbol{B}_2)_{cd}h_{2d}]\\ &\quad=\frac{1}{N}\mathbb{E}\left[\frac{\partial}{\partial h_{1a}}\left((\boldsymbol{B}_1)_{ab}(\boldsymbol{B}_2)_{cd}h_{1b}h_{2c}^*h_{2d}\right)\right]\\ &\qquad-\frac{1}{N}\mathbb{E}\left[\frac{\partial}{\partial h_{1a}}\left((\boldsymbol{B}_1)_{ab}h_{1b}\right)\right]\mathbb{E}\left[h_{2c}^*(\boldsymbol{B}_2)_{cd}h_{2d}\right] \qquad (47)\\ &\quad=\frac{1}{N}\mathbb{E}\left[\frac{\partial h_{1b}}{\partial h_{1a}}(\boldsymbol{B}_1)_{ab}(\boldsymbol{B}_2)_{cd}h_{2c}^*h_{2d}\right]\\ &\qquad+\frac{1}{N}\mathbb{E}\left[\frac{\partial(\boldsymbol{B}_2)_{cd}}{\partial h_{1a}}(\boldsymbol{B}_1)_{ab}h_{1b}h_{2c}^*h_{2d}\right]\\ &\qquad-\frac{1}{N}\mathbb{E}\left[\frac{\partial h_{1b}}{\partial h_{1a}}(\boldsymbol{B}_1)_{ab}\right]\mathbb{E}[h_{2c}^*(\boldsymbol{B}_2)_{cd}h_{2d}] \qquad (48)\end{aligned}$$

where in (47) we have applied Novikov's theorem formally replacing h_{1a}^* with $\frac{1}{N}\frac{\partial}{\partial h_{1a}}$ inside the expectations. We remind the reader that, as explained in Section B, in the above manipulations we treat the complex variables h_{ka} and h_{nb}^* as distinct and independent for all k,n,a,b, such that partial derivatives are performed individually with respect to these variables. In order to compute $\frac{\partial(\boldsymbol{B}_2)_{cd}}{\partial h_{1a}}$, we use the matrix inversion lemma (46) for $\boldsymbol{B}_2$. After some algebra, we obtain

$$\mathbb{E}_c[\gamma_1;\gamma_2] = \frac{\alpha^2}{N}\mathbb{E}_c\left[\mathrm{Tr}(\boldsymbol{B}_1);\boldsymbol{h}_2^{\mathsf{H}}\boldsymbol{B}_2\boldsymbol{h}_2\right] \tag{49}$$

$$-\frac{\alpha^3}{N}\mathbb{E}\left[\frac{\boldsymbol{h}_2^{\mathsf{H}}\boldsymbol{B}_{12}\boldsymbol{B}_1\boldsymbol{h}_1\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{B}_{12}\boldsymbol{h}_2}{1+\alpha\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{B}_{12}\boldsymbol{h}_1}\right] \tag{50}$$

$$+\frac{\alpha^4}{N}\mathbb{E}\left[\frac{\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{B}_{12}\boldsymbol{B}_1\boldsymbol{h}_1\boldsymbol{h}_2^{\mathsf{H}}\boldsymbol{B}_{12}\boldsymbol{h}_1\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{B}_{12}\boldsymbol{h}_2}{\left(1+\alpha\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{B}_{12}\boldsymbol{h}_1\right)^2}\right]. \tag{51}$$

The term in (49) results by summing over all indices the first two terms in (48), and the terms in (50) and (51) result by summing the last term in (48) after applying the partial derivative with respect to the elements of $\boldsymbol{h}_1$ appearing in the numerator and the denominator of the matrix inversion lemma expansion of $\boldsymbol{B}_2$. It is important to notice that the order of magnitude of the first term is $O(1)$, while the last two terms are $O(1/N)$. The reason is that the last two terms are the result of applying the partial derivative in h_{1a} to $\boldsymbol{B}_2$, where the term that depends on $\boldsymbol{h}_1$ is scaled by a factor $O(1/N)$ compared to the remaining matrix.

We proceed now by applying Novikov's theorem to the random variables h_{2a}^* appearing in the numerator of (50) and (51) and exchanging the corresponding expectation with a derivative $\partial/\partial h_{2a}$. However, with some hindsight we only apply the derivative to $\boldsymbol{h}_2$ and not to $\boldsymbol{B}_1$, which would give a subleading term in $1/N$. Therefore, to leading order in $1/N$, we have

$$\begin{aligned}(50)\ &\approx -\frac{\alpha^3}{N^2}\mathbb{E}\left[\frac{\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{B}_{12}^2\boldsymbol{B}_1\boldsymbol{h}_1}{1+\alpha\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{B}_{12}\boldsymbol{h}_1}\right]\\ &\approx -\frac{\alpha^3}{N^2}\frac{\frac{\mathrm{Tr}(\boldsymbol{B}_{12}^3)}{N}}{1+\frac{\alpha}{N}\mathrm{Tr}(\boldsymbol{B}_{12})}\\ &\approx -\frac{1}{N^2}\frac{g_3^{\mathsf{mmse}}\left(\alpha,\beta-\frac{2}{N}\right)}{1+g_1^{\mathsf{mmse}}\left(\alpha,\beta-\frac{2}{N}\right)} \qquad (52)\\ (51)\ &\approx \frac{\alpha^4}{N^2}\mathbb{E}\left[\frac{\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{B}_{12}\boldsymbol{B}_1\boldsymbol{h}_1\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{B}_{12}^2\boldsymbol{h}_1}{\left(1+\alpha\boldsymbol{h}_1^{\mathsf{H}}\boldsymbol{B}_{12}\boldsymbol{h}_1\right)^2}\right]\\ &\approx \frac{\alpha^4}{N^2}\frac{\left(\frac{\mathrm{Tr}(\boldsymbol{B}_{12}^2)}{N}\right)^2}{\left(1+\frac{\alpha}{N}\mathrm{Tr}(\boldsymbol{B}_{12})\right)^2}\\ &\approx \frac{1}{N^2}\frac{g_2^{\mathsf{mmse}}\left(\alpha,\beta-\frac{2}{N}\right)^2}{\left(1+g_1^{\mathsf{mmse}}\left(\alpha,\beta-\frac{2}{N}\right)\right)^2} \qquad (53)\end{aligned}$$

where the approximation sign $\approx$ means to leading order in $1/N$. The second expression in each line occurred by averaging over $\boldsymbol{h}_1$ to leading order, i.e., only on the numerator. In the last equation in each line we used the fact that $\frac{1}{N}\mathrm{Tr}(\boldsymbol{B}_{12}) \approx g_1^{\mathsf{mmse}}(\alpha,\beta-\frac{2}{N})$.

Next we may go back to (49) and expand $\boldsymbol{B}_2$ using (46). After applying exactly the same methods as above we arrive at the following expression:

$$\begin{aligned}(49)\ &\approx \frac{\alpha^2}{N^2}\mathbb{E}_c[\mathrm{Tr}(\boldsymbol{B}_{12});\mathrm{Tr}(\boldsymbol{B}_{12})]\\ &\quad+\frac{2}{N^2}\frac{g_2^{\mathsf{mmse}}\left(\alpha,\beta-\frac{2}{N}\right)^2}{\left(1+g_1^{\mathsf{mmse}}\left(\alpha,\beta-\frac{2}{N}\right)\right)^2}\\ &\quad-\frac{1}{N^2}\frac{g_3^{\mathsf{mmse}}\left(\alpha,\beta-\frac{2}{N}\right)}{1+g_1^{\mathsf{mmse}}\left(\alpha,\beta-\frac{2}{N}\right)}.\end{aligned} \tag{54}$$

We collect all terms and use (82) to reach the final result

$$\begin{aligned}\Sigma_{1,2}^{\mathsf{mmse}} &= \mathbb{E}_c[\gamma_1;\gamma_2]\\ &\approx \frac{1}{N^2}\frac{\left(\beta-\frac{2}{N}\right)\alpha^4}{\left(1+2\alpha\left(1+\beta-\frac{2}{N}\right)+\alpha^2\left(1-\beta+\frac{2}{N}\right)^2\right)^2}\\ &\quad+\frac{1}{N^2}\left(\frac{3g_2^{\mathsf{mmse}}\left(\alpha,\beta-\frac{2}{N}\right)^2}{\left(1+g_1^{\mathsf{mmse}}\left(\alpha,\beta-\frac{2}{N}\right)\right)^2}\right.\\ &\qquad\left.-\frac{2g_3^{\mathsf{mmse}}\left(\alpha,\beta-\frac{2}{N}\right)}{1+g_1^{\mathsf{mmse}}\left(\alpha,\beta-\frac{2}{N}\right)}\right)\\ &= \frac{v_{od}^{\mathsf{mmse}}}{M^2}+O(1/N^3)\end{aligned} \tag{55}$$

where we let

$$v_{od}^{\mathsf{mmse}} = \beta^2 \left(\frac{3g_2^{\mathsf{mmse}}\left(\alpha, \beta - \frac{2}{N}\right)^2}{\left(1 + g_1^{\mathsf{mmse}}\left(\alpha, \beta - \frac{2}{N}\right)\right)^2} + \frac{\left(\beta - \frac{2}{N}\right)\alpha^4}{\left(1 + 2\alpha\left(1 + \beta - \frac{2}{N}\right) + \alpha^2\left(1 - \beta + \frac{2}{N}\right)^2\right)^2} - \frac{2g_3^{\mathsf{mmse}}\left(\alpha, \beta - \frac{2}{N}\right)}{1 + g_1^{\mathsf{mmse}}\left(\alpha, \beta - \frac{2}{N}\right)} \right). \quad (56)$$

For large α, $v_{od}^{\mathsf{mmse}} \approx \rho^2$ when $\beta < 1$, and $v_{od}^{\mathsf{mmse}} \approx \rho^2/16$ when $\beta = 1$.

We collect the results of (55) and (43) by writing the correlation matrix for the SINRs $\{\gamma_k\}$ to leading order as

$$\Sigma_{i,j}^{\mathsf{mmse}} = \delta_{i,j}\frac{v_d^{\mathsf{mmse}}}{M} + (1 - \delta_{i,j})\frac{v_{od}^{\mathsf{mmse}}}{M^2}. \quad (57)$$

It is worth pointing out that despite the fact that the off-diagonal elements are much smaller compared to the diagonal ones, they all contribute to the eigenvalues of $\mathbf{\Sigma}$. In fact, these can be computed in closed form and are given by

$$\lambda_1(\mathbf{\Sigma}^{\mathsf{mmse}}) = \frac{v_d^{\mathsf{mmse}}}{M} + (M-1)\frac{v_{od}^{\mathsf{mmse}}}{M^2} \approx \frac{v_d^{\mathsf{mmse}} + v_{od}^{\mathsf{mmse}}}{M}$$

and

$$\lambda_k(\mathbf{\Sigma}^{\mathsf{mmse}}) = \frac{v_d^{\mathsf{mmse}}}{M} - \frac{v_{od}^{\mathsf{mmse}}}{M^2} \approx \frac{v_d^{\mathsf{mmse}}}{M}$$

for all $k = 2, \ldots, M$.

2) Cumulant Moments for the ZF Receiver: The corresponding results for the ZF receiver can be derived directly from the previous section by observing that the SINR for the kth channel of the ZF receiver, given by (7), can be deduced from the corresponding expression (9) for the MMSE receiver in the limit of infinite α, i.e.,

$$\gamma_k^{\mathsf{zf}} = \frac{\alpha}{[(\boldsymbol{H}^{\mathsf{H}}\boldsymbol{H})^{-1}]_{kk}} = \alpha \lim_{\alpha_0 \to \infty} \frac{\gamma_k^{\mathsf{mmse}}(\alpha_0)}{\alpha_0} = \alpha \lim_{\alpha_0 \to \infty} \{\alpha_0 [(\boldsymbol{I} + \alpha_0 \boldsymbol{H}^{\mathsf{H}}\boldsymbol{H})^{-1}]_{kk}\}^{-1}.$$

A subtle point needs to be stressed here: the results for the ZF receiver cannot be obtained simply as the "limit for high SNR" of the results for the MMSE receiver. Rather, we have to distinguish between the channel SNR (contained in the parameter α) and the SNR parameter in the linear receiver matrix expression (indicated by α_0 above) that we let to infinity in order to obtain the ZF results. It can be shown that for $\beta < 1$ the analysis of the previous section involving the matrices $\bar{\boldsymbol{B}}_1, \bar{\boldsymbol{B}}_2$ and $\boldsymbol{B}_{12}$ can be carried out in this limiting case. In addition, as seen in Appendix C, the condition for the validity of the manipulation of the first term of (55) is that $f(x) = (1 + \alpha x)^{-1}$ is a smooth function of x in the region of support of the eigenvalue spectrum. This is not true in the vicinity of $x = 0$ for arbitrarily large α, specifically when $\alpha = O(N)$. Thus, when $\beta = 1$, in which case the asymptotic eigenvalue spectrum includes $x = 0$, the above approximation is not valid. As a result, this method breaks down at $\beta = 1$.

From (41) we get the mean SINR for the ZF receiver[5]

$$\mathbb{E}[\gamma_1^{\mathsf{zf}}] = \begin{cases} \alpha(1 - \beta + 1/N), & \beta < 1 \\ 0, & \beta = 1 \end{cases} \quad (58)$$

in agreement with [28]. Similarly, the second-order moments can be obtained from

$$\mathbb{E}_c\left[\gamma_i^{\mathsf{zf}}; \gamma_j^{\mathsf{zf}}\right] = \alpha^2 \lim_{\alpha_0 \to \infty} \frac{\mathbb{E}_c\left[\gamma_i^{\mathsf{mmse}}(\alpha_0); \gamma_j^{\mathsf{mmse}}(\alpha_0)\right]}{\alpha_0^2}. \quad (59)$$

Thus, we get

$$\Sigma_{11}^{\mathsf{zf}} = \frac{v_d^{\mathsf{zf}}}{M} = \begin{cases} \frac{\alpha^2\beta(1-\beta+1/N)}{M}, & \beta < 1 \\ 0, & \beta = 1 \end{cases} \quad (60)$$

and

$$\Sigma_{12}^{\mathsf{zf}} = \frac{v_{od}^{\mathsf{zf}}}{M^2} = \begin{cases} \frac{\alpha^2\beta^2}{M^2}, & \beta < 1 \\ \frac{\rho^2}{16M^2}, & \beta = 1. \end{cases} \quad (61)$$

While in the case of $\beta < 1$ the covariance matrix $\mathbf{\Sigma}$ is well-defined and positive-definite with eigenvalues

$$\lambda_1(\mathbf{\Sigma}^{\mathsf{zf}}) = \frac{v_d^{\mathsf{zf}}}{M} + (M-1)\frac{v_{od}^{\mathsf{zf}}}{M^2} \approx \frac{\alpha^2\beta^2}{\beta M}$$

and

$$\lambda_k(\mathbf{\Sigma}^{\mathsf{zf}}) = \frac{v_d^{\mathsf{zf}}}{M} - \frac{v_{od}^{\mathsf{zf}}}{M^2} \approx \frac{\alpha^2\beta^2(1-\beta)}{\beta M}$$

for all $k = 2, \ldots, M$, the case $\beta = 1$ is problematic. Specifically, it results in (narrowly) negative eigenvalues for $\mathbf{\Sigma}$, thereby invalidating the Gaussian approximation for the γ_i's and, as a result, the further treatment of the mutual information as a Gaussian variable. The case $\beta = 1$ for the ZF receiver is therefore excluded in the subsequent Gaussian approximation of the mutual information.

C. Gaussian Approximation and Outage Probability

In this subsection, we use the previous results together with the asymptotic Gaussianity that follows from the fact that higher order moments are vanishing (see Appendix D) to give an explicit Gaussian approximation for the outage probability of linear MMSE and ZF receivers with coding across the antennas in the regime of fixed SNR and large number of antennas.

We start with the mean $\mathbb{E}[I_N]$. Due to the symmetry with respect to the terms γ_k, we have

$$\begin{aligned} \mathcal{C}_1 = \mathbb{E}[I_N] &= M\mathbb{E}[\log(1 + \gamma_1)] \\ &= M\mathbb{E}[\log(1 + \mathbb{E}[\gamma_1] + \gamma_1 - \mathbb{E}[\gamma_1])] \\ &= M\log(1 + \mathbb{E}[\gamma_1]) - M\sum_{n=1}^{\infty} \frac{(-1)^n}{n} \frac{\mathbb{E}[(\gamma_1 - \mathbb{E}[\gamma_1])^n]}{(1 + \mathbb{E}[\gamma_1])^n}. \end{aligned}$$

[5]For simplicity we neglect the subleading terms in the following equalities, i.e., we omit $O(1/N^2)$ in (58) and $O(1/N^3)$ in (60), (61), respectively.

In the preceding expansion, all terms $n > 2$ involve cumulants of γ_1 higher than 2, thus can be neglected. For the MMSE receiver, this yields

$$\begin{aligned}\mathcal{C}_1^{\mathsf{mmse}} &= \mathbb{E}[I_N] \\ &= M\log\left(1+g_1^{\mathsf{mmse}}\left(\alpha,\beta-\frac{1}{N}\right)\right) \\ &\quad - \frac{v_d^{\mathsf{mmse}}}{2\left(1+g_1^{\mathsf{mmse}}\left(\alpha,\beta-\frac{1}{N}\right)\right)^2}+o(1) \\ &\approx M\log\left(1+g_1^{\mathsf{mmse}}(\alpha,\beta)-\frac{1}{N}\frac{\partial}{\partial\beta}g_1^{\mathsf{mmse}}(\alpha,\beta)\right) \\ &\quad - \frac{v_d^{\mathsf{mmse}}}{2(1+g_1^{\mathsf{mmse}}(\alpha,\beta))^2} \\ &\approx M\log\left(1+g_1^{\mathsf{mmse}}(\alpha,\beta)\right)-\beta\frac{\frac{\partial}{\partial\beta}g_1^{\mathsf{mmse}}(\alpha,\beta)}{1+g_1^{\mathsf{mmse}}(\alpha,\beta)} \\ &\quad - \frac{v_d^{\mathsf{mmse}}}{2(1+g_1^{\mathsf{mmse}}(\alpha,\beta))^2} \end{aligned} \tag{62}$$

where $g_1^{\mathsf{mmse}}(\alpha,\beta)$ and v_d^{mmse} are given by (40) and (43), respectively.

For the ZF receiver the mean is given by

$$\mathcal{C}_1^{\mathsf{zf}} = M\log(1+\alpha(1-\beta)) + \frac{\alpha\beta\left(1+\frac{\alpha(1-\beta)}{2}\right)}{(1+\alpha(1-\beta))^2}+o(1). \tag{63}$$

Similarly, we can calculate the variance of the mutual information as follows:

$$\begin{aligned}\mathcal{C}_2 &= M\mathbb{E}_c[\log(1+\gamma_1);\log(1+\gamma_1)] \\ &\quad + M(M-1)\mathbb{E}_c[\log(1+\gamma_1);\log(1+\gamma_2)] \\ &= M\sum_{m,n=1}^{\infty}\frac{(-1)^{m+n}}{mn}\frac{\mathbb{E}_c[(\gamma_1-\mathbb{E}[\gamma_1])^m;(\gamma_1-\mathbb{E}[\gamma_1])^n]}{(1+\mathbb{E}[\gamma_1])^{n+m}} \\ &\quad + M(M-1)\sum_{m,n=1}^{\infty}\frac{(-1)^{m+n}}{mn} \\ &\quad \cdot\frac{\mathbb{E}_c[(\gamma_1-\mathbb{E}[\gamma_1])^m;(\gamma_2-\mathbb{E}[\gamma_2])^n]}{(1+\mathbb{E}[\gamma_1])^{n+m}}. \end{aligned} \tag{64}$$

As we see above, the first terms in both the above summations give the leading order of the variance of the mutual information, which we denoted in (33) by σ^2. The variance of the MMSE mutual information is given by

$$\mathcal{C}_2^{\mathsf{mmse}} = \frac{v_d^{\mathsf{mmse}}+v_{od}^{\mathsf{mmse}}}{\left(1+g_1^{\mathsf{mmse}}\left(\alpha,\beta-\frac{1}{N}\right)\right)^2}+o(1) \tag{65}$$

where v_{od}^{mmse} is given by (56). The corresponding variance for the ZF receiver is

$$\mathcal{C}_2^{\mathsf{zf}} = \frac{\beta\alpha^2(1+1/N)}{(1+\alpha(1-\beta+1/N))^2}+o(1) \tag{66}$$

for $\beta < 1$. As mentioned above, the case $\beta = 1$ does not result in a well-behaved jointly Gaussian behavior of the γ_k's, and therefore the Gaussian approximation of the mutual information cannot be derived with this approach.

As anticipated at the beginning of this section, from (62) and (63) we see that the mean mutual information is expressed in the form $m_1 = Mc_{10} + c_{11}$, where the coefficient c_{10} was found in previous works (e.g., [35]), considering the limit $N \to \infty$ of the normalized mutual information per transmit antenna, and the term c_{11} is a correction term that captures the correlation between the SINRs γ_k.

Under this Gaussian approximation, we can easily evaluate the outage probability for fixed SNR, β and number of antennas M as follows:

$$P_{\text{out}}^{\text{lin}}(R,\rho) \approx Q\left(\frac{R-\mathcal{C}_1}{\sqrt{\mathcal{C}_2}}\right) \tag{67}$$

where $Q(x)=\int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt$ is the Gaussian tail function.

We conclude this subsection with a discussion on the range of validity of the Gaussian approximation. For the MMSE receiver, in the large ρ limit we have that $\mathcal{C}_2^{\mathsf{mmse}} = O(\rho)$ for $\beta = 1$, while $\mathcal{C}_2^{\mathsf{mmse}} = O(1)$ for $\beta < 1$. This fast increase of the variance of the distribution for $\beta = 1$ and $\rho \gg 1$ is a spurious result in this approximation, due to the neglected terms which are negligible for fixed ρ and increasingly large N, but become important for fixed N and large ρ.

The behavior of the ZF receiver for $\beta = 1$, when the jointly Gaussian behavior of the γ_k's breaks down, exacerbates the above large ρ behavior of the MMSE receiver. In fact, as was discussed before, the ZF case is in some sense the "infinite ρ limit" of the MMSE case. Therefore, the problematic situation appearing in the ZF receiver for $\beta = 1$ has the same roots as the problems faced in the large (but finite) ρ limit when N is also finite but not large enough.

D. Simulations and Comparisons

In this subsection, we first validate the asymptotic analysis by comparing the asymptotic approximation for $\mathcal{C}_1$ and $\mathcal{C}_2$ with the exact moments obtained by finite-dimensional Monte Carlo simulation. We then compare the Gaussian approximation to the outage probability with finite-dimensional Monte Carlo simulation.

For the sake of comparison, we also consider the outage probability of the optimal receiver, given by the log-det cdf $P(I_N^{\mathsf{opt}} \le R)$, where $I_N^{\mathsf{opt}} = \log\det(\boldsymbol{I}+\frac{\rho}{M}\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}})$. As said in the Introduction, the asymptotic Gaussianity of the log-det mutual information is well known and holds under very general models of channel correlation across the antennas (not considered in this work). For completeness, we recall its expressions under the i.i.d. channel coefficient assumptions and in the notation of this paper. We have

$$\frac{I_N^{\text{opt}}-M\mu}{\nu} \xrightarrow{d} \mathcal{N}(0,1) \tag{68}$$

where

$$\begin{aligned}\mu &= \log(1+g_1^{\mathsf{mmse}}(\alpha,\beta)) + \frac{g_1^{\mathsf{mmse}}(\alpha,\beta)}{\alpha\beta} \\ &\quad + \frac{1}{\beta}\log(1-\alpha(1-\beta)+g_1^{\mathsf{mmse}}(\alpha,\beta)) - \frac{1}{\beta}\end{aligned} \tag{69}$$

and

$$\nu^2 = -\log\left(1-\frac{1}{\beta}\left(1-\frac{g_1^{\mathsf{mmse}}(\alpha,\beta)}{\alpha}\right)^2\right). \tag{70}$$

The first term in (69) coincides with the coefficient of M in the first term of (62), i.e., it is the asymptotic capacity per antenna

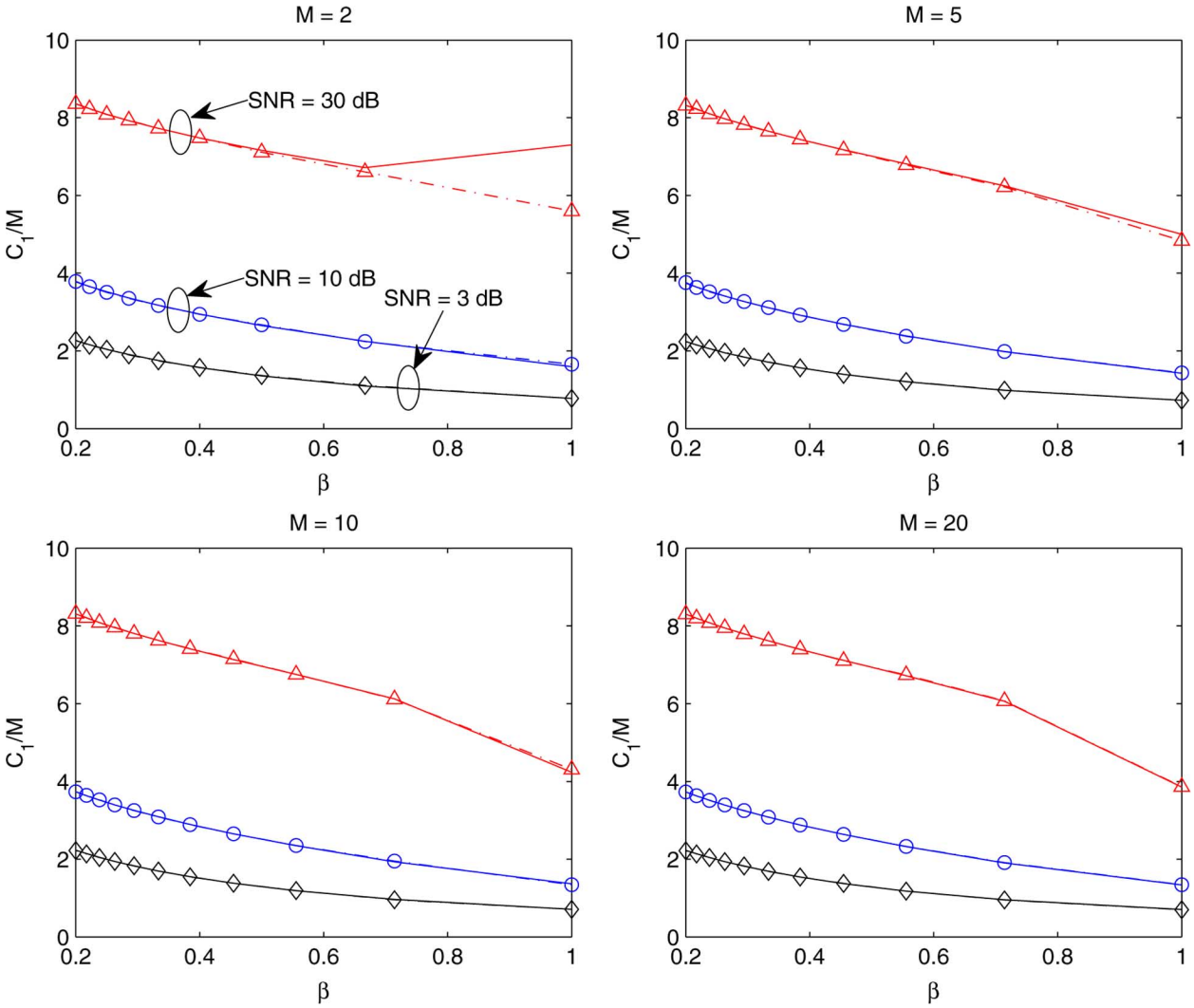

Fig. 5. Mean of the MMSE mutual information per antenna ($\mathcal{C}_1^{\mathsf{mmse}}/M$) as a function of β, for $M = 2, 5, 10$ and 20. The solid lines are analytical results, and the corresponding dash-dot lines are empirical results obtained from Monte Carlo simulation. Diamonds denote 3 dB, circles 10 dB, and triangles 30 dB.

of the linear MMSE receiver. The additional terms in (69) represent the so-called "nonlinear" gain of the optimal versus linear MMSE receiver, as discovered in [36].

It is worth pointing out that for large SNR, the asymptotic mean capacity is given by

$$\mathbb{E}[I_N^{\text{opt}}] \approx M \log \rho \tag{71}$$

while the variance has the following behavior:

$$\nu^2 \approx \begin{cases} -\log(1-\beta), & \beta < 1 \\ \frac{1}{2}\log\rho, & \beta = 1. \end{cases} \tag{72}$$

Using the Gaussian approximation

$$P_{\text{out}}(R,\rho) \approx Q\left(\frac{R - M\mu}{\nu}\right) \approx \exp\left(-\frac{(R-M\mu)^2}{2\nu^2}\right)$$

with $R = r \log \rho$, in the large ρ limit we find

$$\begin{cases} -\frac{\log P_{\text{out}}(R,\rho)}{\log\rho} \approx (M-r)^2 \frac{\log\rho}{2|\log(1-\beta)|}, & \beta < 1 \\ -\frac{\log P_{\text{out}}(R,\rho)}{\log\rho} \approx (M-r)^2, & \beta = 1. \end{cases} \tag{73}$$

For $\beta = 1$, the Gaussian approximation yields an outage probability exponent equal to $(r-M)^2$ for $r \in [0, M]$, that closely approximates the exact exponent $d^*(r)$ [3]. However, for $\beta < 1$, the Gaussian approximation yields a completely inaccurate behavior. In fact, in this case the exponent obtained through (73) would be infinite, while we know from [3] that $d^*(r) \leq MN$. The reason for this spectacular failure of the Gaussian approximation is that, in the large-N approximation, it is implicitly assumed that for $\beta \neq 1$ the eigenvalue distribution at very small eigenvalues is zero, as described by the Marcenko–Pastur law [38]. Instead, for large ρ, the eigenvalues that dominate the outage probability are exactly the very small ones, of the order of $1/\rho$, i.e., exactly the ones that the Gaussian approximation neglects.

We conclude this section by presenting some numerical results. Fig. 5 compares the analytical mean of the MMSE mutual information per antenna $\mathcal{C}_1^{\mathsf{mmse}}/M$ with the corresponding empirical mean obtained from Monte Carlo simulation. The corresponding comparison between the analytical variance $\mathcal{C}_2^{\mathsf{mmse}}$ and the empirical variance is presented in Fig. 6. Using the results for

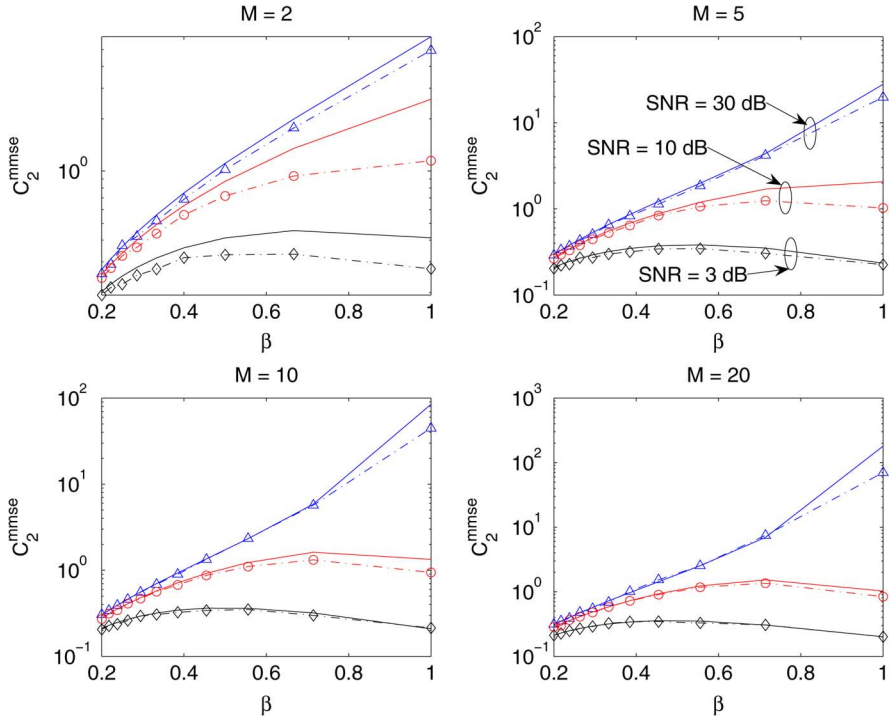

Fig. 6. Variance of the MMSE mutual information ($\mathcal{C}_2^{\mathsf{mmse}}$) as a function of β, for $M = 2, 5, 10$ and 20. The solid lines are analytical results, and the corresponding dash-dot lines are empirical results obtained from Monte Carlo simulation. Diamonds denote 3 dB, circles 10 dB, and triangles 30 dB.

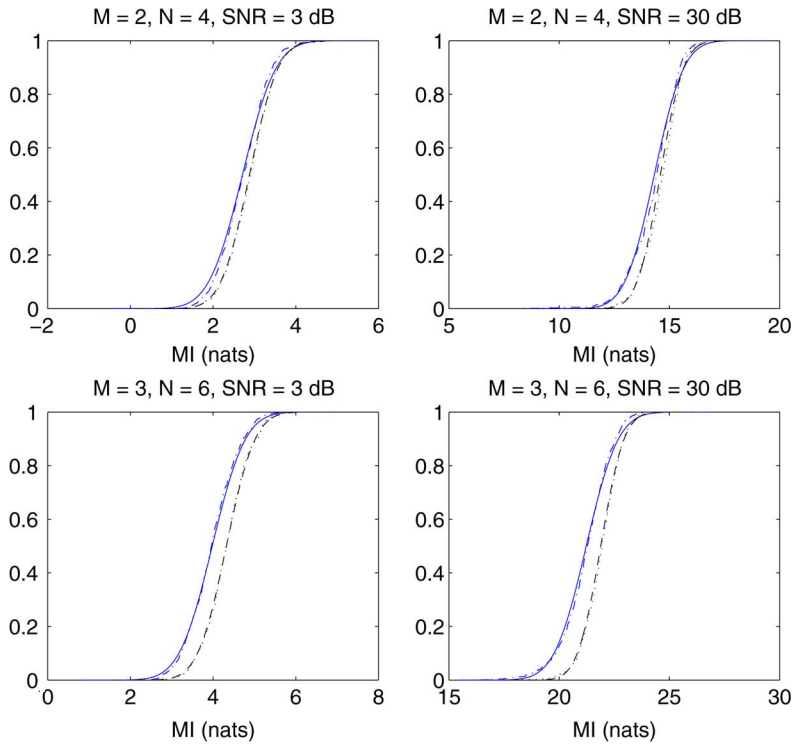

Fig. 7. CDF of the mutual information (MI) for the MMSE and optimal receivers, for $M = 2, 3$, $\beta = 0.5$, and $\rho = 3, 30$ dB. The solid blue line is the analytical result for the MMSE, the dot-dash blue is MMSE empirical, the dashed black is optimal receiver analytical, and the dotted black is the optimal receiver empirical.

the mean and the variance, we plot the cdf of the (Gaussian) mutual information for the MMSE and optimal receiver in Figs. 7, 8. Both analytical and empirical results are plotted, for a wide range of M, N and SNRs. For brevity, we have only reported plots of the cdf for the mutual information for the ZF case, see Fig. 9. The plots are for $M = 3, 10$, $\beta = 0.5$, and $\rho = 3, 30$ dB. The results are similar in flavor to the MMSE case.

We notice that the analytical and empirical results match closely, for even moderate number of antennas and not too large SNRs, in line with the comments made earlier regarding

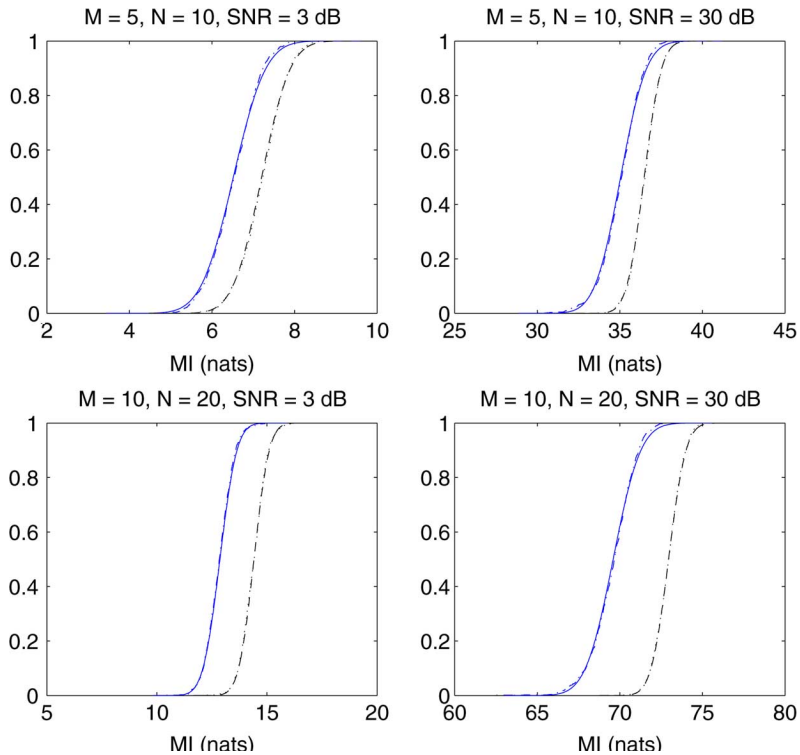

Fig. 8. CDF of the mutual information (MI) for the MMSE and optimal receivers, for $M = 5, 10$, $\beta = 0.5$, and $\rho = 3, 30$ dB. The solid blue line is the analytical result for the MMSE, the dot-dash blue is MMSE empirical, the dashed black is optimal receiver analytical, and the dotted black is the optimal receiver empirical.

the validity of the analysis. It is also worthwhile noticing that the accuracy of the Gaussian approximation for linear receivers appears to be slightly inferior to that of the Gaussian approximation for the optimal receiver case, especially for very small N and large SNR.

VI. Conclusion

Novel wireless communication systems are targeting very large spectral efficiencies and will operate at high SNR thanks to hot spots and pico-cell arrangements. For example, a system with bandwidth of 20 MHz and operating at 100 Mb/s requires a spectral efficiency of 5 bit/s/Hz, corresponding to coding rate $R = 5$ bpcu in notation adopted here, if one neglects nonideal effects such as pilot symbols, guard band and guard intervals, cyclic prefix redundancy for orthogonal frequency-division multiplexing (OFDM), etc. For such systems, the use of low-complexity linear receivers in a *separated detection and decoding* architecture as those examined in this paper may be mandatory because of complexity and power consumption.

In this paper, we investigated the asymptotic performance of such separated linear detection and decoding architectures in two relevant asymptotic regimes. In the regime of fixed number of antennas and increasing SNR and coding rate, we showed that linear detection may be very suboptimal. Furthermore, due to the strong correlation between the SINRs of the parallel channels induced by the linear receiver, coding across the antennas does not help in terms of the achievable DMT. We also illuminated the very peculiar behavior of the linear MMSE receiver with coding across the antennas, that exhibits a diversity order (slope of the outage probability curve) that changes depending on the rate. Then, we analyzed the asymptotic behavior of the linear MMSE and the ZF receivers with coding across the antennas in the regime of fixed SNR and large (but finite) number of antennas. We showed that the corresponding mutual information has statistical fluctuations that converge in distribution to a Gaussian random variable, and we computed its mean and variance in closed form. This yields a simple Gaussian approximation of the outage probability in this asymptotic regime, within the limitations that have been thoroughly discussed.

Based on the analysis carried out in this work, we may summarize some considerations on system design. In order to achieve a required target spectral efficiency at given block-error rate and SNR operating point, an attractive design option may consists of increasing the number of antennas (especially at the receiver) and using a low-complexity linear receiver. However, pure spatial multiplexing (independent coded streams directly fed into the transmit antennas) and/or linear ZF receivers should be avoided. In contrast, coding across antennas and a linear MMSE receiver can achieve a very good tradeoff between performance and complexity in a wide range of system operating points.

Appendix A
Proof of $P(\mathcal{A}) = O(1)$

Our aim is to provide a lower bound to the quantity

$$P(\mathcal{A}) = P\left(\frac{1}{M}\sum_{k=1}^{M}\frac{1}{|u_{1k}|^2} \le c\right).$$

It is well known that $\boldsymbol{U}$ and $\boldsymbol{\Lambda}$ are independent random matrices and that $\boldsymbol{U}$ is Haar distributed (i.e., is an isotropically random unitary matrix distributed uniformly over the Stiefel

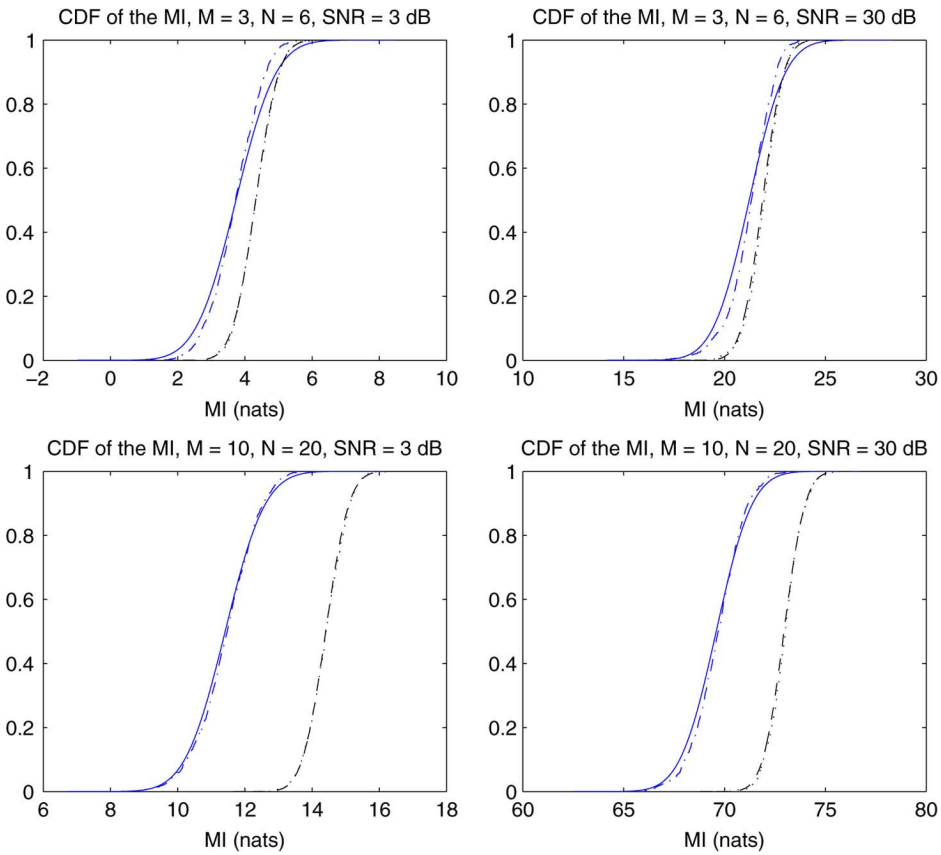

Fig. 9. CDF of the mutual information (MI) for the ZF and optimal receivers, for $M = 3, 10$, $\beta = 0.5$, and $\rho = 3, 30$ dB. The solid blue line is the analytical result for the ZF receiver, the dot-dash blue is ZF empirical, the dashed black is optimal receiver analytical, and the dotted black is the optimal receiver empirical.

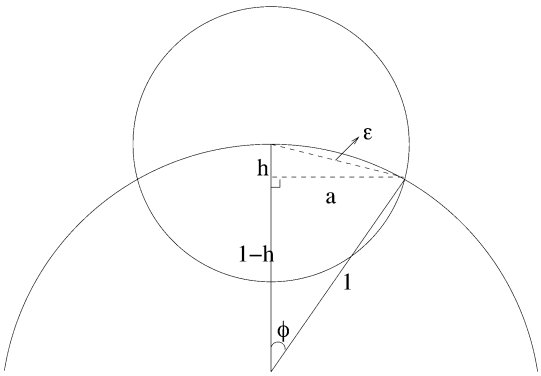

Fig. 10. The unit hemisphere and a spherical cap.

manifold). Therefore, the vector $\boldsymbol{u}_1 \triangleq (u_{1,1}, u_{1,2}, \ldots, u_{1,M})$ corresponding to the first row of $\boldsymbol{U}$ is uniformly distributed on the unit M-dimensional hypersphere (or M-sphere) $\mathcal{O}_M(\mathbf{0}, 1)$,[6] and satisfies $|\boldsymbol{u}_1|^2 = \sum_{i=1}^{M} |u_{1i}|^2 = 1$. The point $\boldsymbol{p} = (\frac{1}{\sqrt{M}}, \frac{1}{\sqrt{M}}, \ldots, \frac{1}{\sqrt{M}})$ is a point on the unit M-sphere. Consider the spherical cap $\mathcal{C}$ of the unit sphere that is cut out by the sphere $\mathcal{O}_M(\boldsymbol{p}, \epsilon)$, where ϵ is a small positive number (see Fig. 10).

The coordinates of any point $\boldsymbol{u}$ in this spherical cap $\mathcal{C}$ are lower-bounded by

$$u_j \geq \frac{1}{\sqrt{M}} - \epsilon, \quad \forall j = 1, \ldots, M.$$

[6]We will use the notation $\mathcal{O}_M(\boldsymbol{c}, \delta)$ to denote an M-sphere centred at $\boldsymbol{c}$ with radius δ.

Therefore

$$\frac{1}{M} \sum_{j=1}^{M} \frac{1}{|u_j|^2} \leq \frac{1}{\left(\frac{1}{\sqrt{M}} - \epsilon\right)^2}.$$

Defining the constant $c = \left(\frac{1}{\sqrt{M}} - \epsilon\right)^{-2}$, we have that

$$\begin{aligned} P(\mathcal{A}) &\geq P(\boldsymbol{u}_1 \in \mathcal{C}) \\ &= \frac{\text{Surface area of } \mathcal{C}}{\text{Surface area of } \mathcal{O}_M(\mathbf{0}, 1)} \end{aligned}$$

where the latter equality holds since $\boldsymbol{u}_1$ is isotropic. In order to compute the above surface areas, consider Fig. 10. The surface area of $\mathcal{C}$, denoted by $\Omega(\phi)$ is given by [2]

$$\Omega(\phi) = \frac{(M-1)\pi^{(M-1)/2}}{\Gamma\left(\frac{M+1}{2}\right)} \int_0^{\phi} (\sin\theta)^{M-2} d\theta,$$

and the surface area of an unit M-sphere is [2] $S_M(1) = M\pi^{M/2}/\Gamma(M/2+1)$. All that remains is to compute the angle ϕ, which is accomplished by solving the following equations obtained from the two right-angled triangles in Fig. 10:

$$\begin{aligned} h^2 + a^2 &= \epsilon^2 \\ a^2 + (1-h)^2 &= 1. \end{aligned}$$

Solving for a, h, we obtain

$$h = \frac{\epsilon^2}{2}, \quad a = \epsilon\sqrt{1 - \frac{\epsilon^2}{4}}.$$

Therefore

$$\phi = \tan^{-1}\left(\frac{\epsilon\sqrt{1-\frac{\epsilon^2}{4}}}{1-\frac{\epsilon^2}{2}}\right)$$

and

$$P(\mathcal{A}) \geq \frac{\Omega(\phi)}{S_M(1)} > 0$$

as desired.

Appendix B
Novikov's Theorem

We introduce here a useful trick which makes the connection between Gaussian integration and differentiation over Gaussian random variables.

Theorem 2: [37] (Novikov) Let $\boldsymbol{H}$ be an $N \times M$ matrix with i.i.d. elements drawn from $\mathcal{CN}(0, 1/N)$. Let $f(\boldsymbol{H}, \boldsymbol{H}^{\mathsf{H}})$ be a scalar function of the matrix elements of $\boldsymbol{H}$ and their complex conjugates, which does not grow faster than the inverse of the probability density of $\boldsymbol{H}$, i.e., such that

$$\lim_{|h_{ij}|\to\infty} p(\boldsymbol{H}) f(\boldsymbol{H}, \boldsymbol{H}^H) = 0 \tag{74}$$

for all matrix elements h_{ij}. Then, for any set of indices i, j, the following relation holds:

$$\mathbb{E}[h_{ij} f(\boldsymbol{H}, \boldsymbol{H}^{\mathsf{H}})] = \frac{1}{N}\mathbb{E}\left[\frac{\partial f(\boldsymbol{H}, \boldsymbol{H}^{\mathsf{H}})}{\partial h_{ij}^*}\right] \tag{75}$$

where h_{ij} and h_{ij}^* are to be treated as individual variables in the differentiation.[7]

Sketch of the Proof: Even though the general proof is involved, we present here a simple proof for a single real Gaussian variable $\mathcal{N}(0, 1)$ by integrating by parts

$$\begin{aligned}\mathbb{E}[zf(z)] &= \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} z f(z) e^{-\frac{z^2}{2}} \\ &= -\frac{e^{-\frac{z^2}{2}} f(z)}{\sqrt{2\pi}}\bigg|_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{dz}{\sqrt{2\pi}} f'(z) e^{-\frac{z^2}{2}} \\ &= \mathbb{E}[f'(z)]. \end{aligned} \tag{76}$$

□

Example: A useful and illustrative example of the application of Novikov theorem is to evaluate the fourth moment of an i.i.d. complex Gaussian vector with elements drawn from $\mathcal{CN}(0, 1/N)$. We have

$$\begin{aligned}\mathbb{E}[h_i^* h_j h_k^* h_l] &= \frac{1}{N}\mathbb{E}\left[\frac{\partial}{\partial h_i}(h_j h_k^* h_l)\right] \\ &= \frac{\delta_{ij}}{N}\mathbb{E}[h_k^* h_l] + \frac{\delta_{il}}{N}\mathbb{E}[h_k^* h_j] \\ &= \frac{\delta_{ij}\delta_{kl}}{N^2} + \frac{\delta_{il}\delta_{kj}}{N^2}. \end{aligned} \tag{77}$$

[7]This means that $\frac{\partial h_{i,j}^*}{\partial h_{i,j}} = 0$ when computing the partial derivative in (75).

Appendix C
Fluctuations of Eigenvalues

Let

$$\boldsymbol{B} = (\boldsymbol{I} + \alpha \boldsymbol{H}\boldsymbol{H}^{\mathsf{H}})^{-1} \tag{78}$$

where $\boldsymbol{H} \in \mathbb{C}^{N\times M}$ has i.i.d. elements $\mathcal{CN}(0, 1/N)$, and let $\beta = M/N$. The normalized trace of $\boldsymbol{B}$ is given by

$$\frac{1}{N}\mathrm{Tr}(\boldsymbol{B}) = \frac{1}{N}\sum_{k=1}^{N} \frac{1}{1+\alpha\lambda_k(\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}})} = \eta^{(N)}(\alpha)$$

where $\eta^{(N)}(\alpha)$ is the η-transform [38] of the empirical eigenvalue distribution of $\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}$. For large N, the variance of $\mathrm{Tr}(\boldsymbol{B}) = N\eta^{(N)}(\alpha)$ is of order unity and can be calculated in closed form [39], [40]

$$\mathbb{E}_c[\mathrm{Tr}(\boldsymbol{B}); \mathrm{Tr}(\boldsymbol{B})] = \mathcal{P}\int_{\lambda_{\min}}^{\lambda_{\max}} \int_{\lambda_{\min}}^{\lambda_{\max}} \frac{\mathcal{K}_2(\lambda,\mu) d\mu d\lambda}{(1+\alpha\lambda)(1+\alpha\mu)} \tag{79}$$

where $\mathcal{P}$ denotes the principal part of the integral, $\lambda_{\min,\max}$ denote the extremal values of the support of the Marcenko–Pastur law [38], given by

$$\lambda_{\min,\,\max} = (1 \pm \sqrt{\beta})^2, \tag{80}$$

and $\mathcal{K}_2(\lambda, \mu)$ is an integral kernel representing the deviation of the joint eigenvalue distribution of the eigenvalues λ, μ from the product of their marginal distributions and is given asymptotically by

$$\begin{aligned}\mathcal{K}_2(\lambda,\mu) = &\frac{1}{2\pi^2}\frac{1}{\sqrt{(\lambda-\lambda_{\min})(\lambda_{\max}-\lambda)}} \\ &\cdot\frac{\partial}{\partial\mu}\left[\frac{\sqrt{(\mu-\lambda_{\min})(\lambda_{\max}-\mu)}}{\lambda-\mu}\right]. \end{aligned} \tag{81}$$

It can be verified that the above function is symmetric in λ and μ. After integration by parts, we get

$$\begin{aligned}\mathbb{E}_c[\mathrm{Tr}(\boldsymbol{B}); \mathrm{Tr}(\boldsymbol{B})] &= \frac{1}{2\pi^2}\mathcal{P}\int_{\lambda_{\min}}^{\lambda_{\max}} d\lambda \int_{\lambda_{\min}}^{\lambda_{\max}} d\mu \\ &\quad\cdot\left[\frac{(\mu-\lambda_{\min})(\lambda_{\max}-\mu)}{(\lambda-\lambda_{\min})(\lambda_{\max}-\lambda)}\right]^{1/2} \\ &\quad\cdot\frac{\alpha}{(\mu-\lambda)(1+\alpha\lambda)(1+\alpha\mu)^2} \\ &= \frac{\alpha^2\beta}{(1+2\alpha(1+\beta)+\alpha^2(1-\beta)^2)^2} \\ &= O(1). \end{aligned} \tag{82}$$

The result (82) is used in the calculation of the correlations of γ_1, γ_2 in (55).

Furthermore, in [39], [40] two important and more general results are shown. In particular, for any functions $f_1(x)$, $f_2(x)$ the following result is true in the limit of large N:

$$\begin{aligned}\mathbb{E}_c[\mathrm{Tr}(f_1(\boldsymbol{B})); \mathrm{Tr}(f_2(\boldsymbol{B}))] &= \mathcal{P}\int_{\lambda_{\min}}^{\lambda_{\max}} d\lambda \int_{\lambda_{\min}}^{\lambda_{\max}} d\mu \\ &\quad\cdot\mathcal{K}_2(\lambda,\mu) f_1(\lambda) f_2(\mu) \\ &= O(1) \end{aligned} \tag{83}$$

as long as these functions are bounded and smooth enough within the support of the asymptotic eigenvalue spectrum (for example, $f(x) = [x]_+$ is not smooth, while $f(x) = \alpha/(1+\alpha x)^2$ is smooth).

Also, in [39], [40] it is shown that all higher order cumulants of such smooth functions vanish in the large N limit, i.e., for $n > 2$

$$\mathbb{E}_c[\mathrm{Tr}(f_1(\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}})); \ldots; \mathrm{Tr}(f_n(\boldsymbol{H}\boldsymbol{H}^{\mathsf{H}}))] = \mathcal{Q}_n(N) = o(1) \tag{84}$$

where we have denoted the (arbitrary for our purposes) scaling of the above cumulant moment with N as $\mathcal{Q}_n(N)$ for future reference. We will use this result in Appendix D to prove that all cumulant moments of I_N of order higher than two vanish for large N.

Appendix D
Higher Order Cumulants Are Vanishing

After the calculation of the mean (62) and the variance (64) of the mutual information in Section V we now need to show that the higher order cumulants of the mutual information vanish in the limit of large N. This will conclude the proof of the asymptotic Gaussianity of the mutual information, as discussed in the beginning of Section V.

We need to show that the cumulant moments defined in (35) as

$$\mathcal{C}_n = \sum_{k_1,\ldots k_n=1}^{M} \mathbb{E}_c\left[\log\left(1+\gamma_{k_1}\right); \ldots; \log\left(1+\gamma_{k_n}\right)\right]$$

vanish for $n > 2$ when $N \to \infty$. Despite the fact that $\mathcal{C}_n$ is a sum of $O(N^n)$ terms, we shall show that it is in fact of order $o(1)$.

A. MMSE Receiver Higher Order Cumulants

We discuss in some detail the case of the MMSE receiver. At the end of this appendix, a short argument is given in order to reach the same conclusions for the ZF case.

While a formal proof would be lengthy and tedious and would not add much value to the paper, we shall provide a sketch the basic steps of the proof leaving out several technicalities. We start by recalling that each γ_{k_i} in the above sum is defined as $\gamma_{k_i} = \alpha \boldsymbol{h}_{k_i}^{\mathsf{H}} \boldsymbol{B}_{k_i} \boldsymbol{h}_{k_i}$, where $\boldsymbol{B}_{k_i} = (\boldsymbol{I} + \alpha \boldsymbol{H}\boldsymbol{H}^{\mathsf{H}} - \alpha \boldsymbol{h}_{k_i}\boldsymbol{h}_{k_i}^{\mathsf{H}})^{-1}$. Using the matrix inversion lemma, we see that the denominator in the expression of $\boldsymbol{B}_{k_i}$ includes all columns of $\boldsymbol{H}$ other than $\boldsymbol{h}_{k_i}$. For every n-tuple $\{k_1, k_2, \ldots, k_n\}$, we define the matrix $\boldsymbol{B}_{\{k_i\}}$, such that

$$\boldsymbol{B}_{\{k_i\}} = \left(\boldsymbol{I} + \alpha \boldsymbol{H}_{\{k_i\}} \boldsymbol{H}_{\{k_i\}}^{\mathsf{H}}\right)^{-1}$$

where $\boldsymbol{H}_{\{k_i\}}$ is obtained by removing columns $\boldsymbol{h}_{k_1}, \boldsymbol{h}_{k_2}, \ldots, \boldsymbol{h}_{k_n}$ from $\boldsymbol{H}$. If for some $i \neq j$ we have $k_i = k_j$, then we only remove column k_i once. For any finite n, the following approximation holds:

$$\boldsymbol{B}_{k_i} = \boldsymbol{B}_{\{k_i\}} + O\left(\frac{1}{N}\right), \tag{85}$$

in the sense that the elements of the matrix $N(\boldsymbol{B}_{k_i} - \boldsymbol{B}_{\{k_i\}})$ are almost surely finite in the limit of large N. Roughly speaking, since the elements of $\boldsymbol{H}$ are zero-mean Gaussian with variance $1/N$, the difference $\boldsymbol{B}_{k_i}^{-1} - \boldsymbol{B}_{\{k_i\}}^{-1} = \alpha \sum_{k_j \neq k_i} \boldsymbol{h}_{k_j}\boldsymbol{h}_{k_j}^{\mathsf{H}}$ adds a term of order $O(1/N)$ to each element of $\boldsymbol{B}$, which can be neglected for large N. For example, if the above approximation can be proved in an iterative fashion by showing that $\boldsymbol{B}_{k_1} - \boldsymbol{B}_{k_1,k_2}$ is $O(1/N)$, then adding to writing $\boldsymbol{B}_{k_1} = \boldsymbol{B}_{k_1} - \boldsymbol{B}_{k_1,k_2} + \boldsymbol{B}_{k_1,k_2} - \boldsymbol{B}_{k_1,k_2,k_3}$ and then showing that $\boldsymbol{B}_{k_1,k_2} - \boldsymbol{B}_{k_1,k_2,k_3}$ is also $O(1/N)$, etc. Each such difference can be shown to be almost surely $O(1/N)$ by applying the matrix inversion lemma and observing that the elements of $\boldsymbol{h}_{k_i}$ are $\mathcal{CN}(0, 1/N)$.

As a result, to leading order in $1/N$, we have

$$\mathcal{C}_n \approx \sum_{\{k_i\}} \mathbb{E}_c\left[\log\left(1+\alpha \boldsymbol{h}_{k_1}^{\mathsf{H}} \boldsymbol{B}_{\{k_i\}} \boldsymbol{h}_{k_1}\right); \ldots; \log\left(1+\alpha \boldsymbol{h}_{k_n}^{\mathsf{H}} \boldsymbol{B}_{\{k_i\}} \boldsymbol{h}_{k_n}\right)\right]$$

where $\sum_{\{k_i\}}$ is a sum over all possible $1 \leq k_1, \ldots, k_n \leq M$.

Since now the random vectors $\boldsymbol{h}_{k_1}, \ldots, \boldsymbol{h}_{k_n}$ do not appear in $\boldsymbol{B}_{\{k_i\}}$, we can explicitly average over them in the above expression. At this point, it is convenient to expand the logarithms in Taylor series, such that each term in the sum above becomes

$$\sum_{l_1,\ldots,l_n=1}^{\infty} \frac{(-\alpha)^{l_1+\ldots+l_n}}{l_1 l_2 \ldots l_n} \sum_{\{k_i\}} \mathbb{E}_c\left[\left(\boldsymbol{h}_{k_1}^{\mathsf{H}} \boldsymbol{B}_{\{k_i\}} \boldsymbol{h}_{k_1}\right)^{l_1}; \ldots; \left(\boldsymbol{h}_{k_n}^{\mathsf{H}} \boldsymbol{B}_{\{k_i\}} \boldsymbol{h}_{k_n}\right)^{l_n}\right]. \tag{86}$$

We may now apply Novikov theorem to average over the $\boldsymbol{h}_{k_i}$'s. This is in general a formidable exercise in combinatorics. Instead, we only need to find how the leading terms scale with N. Specifically, in the following, we will fix the n-tuple $l_1, l_2 \ldots, l_n$ and show that the corresponding term of the type $\sum_{\{k_i\}} \mathbb{E}_c[\cdot]$ is $o(1)$ as $N \to \infty$.

We first notice that since there are $L_{\{l_i\}} = l_1 + \cdots + l_n$ pairs of $\boldsymbol{h}^{\mathsf{H}}$'s and $\boldsymbol{h}$'s, we will have an overall factor of $N^{-L_{\{l_i\}}}$ after averaging over all $\boldsymbol{h}$'s. Also, we can decompose the sum over $\{k_i\}$ into subsets or "shells," where each shell has the same number of distinct indices k_i's. For example, there are $\binom{M}{n} = O(N^n)$ terms containing all distinct indices, and $nM!/((M-n+1)!(n-1)!) = O(N^{n-1})$ terms having $n-1$ distinct indices and one repeated index. In general, the shell with q distinct indices contains $O(N^q)$ terms.

If the k_i's are all distinct, then the resulting cumulant moment will be of order n. Possible terms that may appear include, for example,

$$\begin{aligned} &\mathbb{E}_c\left[\mathrm{Tr}\left[\boldsymbol{B}_{\{k_i\}}\right]^{l_1}; \ldots; \mathrm{Tr}\left[\boldsymbol{B}_{\{k_i\}}\right]^{l_n}\right] \\ &\mathbb{E}_c\left[\mathrm{Tr}\left[\boldsymbol{B}_{\{k_i\}}\right]^{l_1-2} \mathrm{Tr}\left[\boldsymbol{B}_{\{k_i\}}^2\right]; \ldots; \mathrm{Tr}\left[\boldsymbol{B}_{\{k_i\}}\right]^{l_n}\right] \end{aligned} \tag{87}$$

where a term as in the second line can only appear for $l_1 \geq 2$. Therefore, after averaging over the $\boldsymbol{h}_{k_i}$'s we are left with order-n moments, having as arguments products of the random variables $\mathrm{Tr}\boldsymbol{B}_{\{k_i\}}^{m_j}$, for $1 \geq m_j \geq l_j$, with $j = 1, \ldots, n$. Let x_j be the total number of traces appearing in the jth argument of a particular term, such that $x_j \leq l_j$. For example, in the first line of (87), we have $x_1 = 1$, while in the second

line we have $x_1 = 2$. Since generally $x_i \geq 1$, such cumulant moments can be reduced to sums of products of irreducible moments with respect to these random variables. To estimate the leading scaling in N of these reducible moments, we recall that the first moment of the trace is $\mathbb{E}_c[\mathrm{Tr}\boldsymbol{B}_{\{k_i\}}^m] = O(N)$, the second cumulant moment is $O(1)$ (83), while all higher cumulant moments are $o(1)$ (84). Let the leading term in the expansion of the reducible moments into irreducible ones have d_1 cumulants of order 1, d_2 cumulants of order 2, d_3 cumulants of order 3, etc. The only constraint we need to impose is that

$$d_1 \leq d_{1\max} = \sum_{i=1}^{n} [x_i - 1]_+ \leq L_{\{l_i\}} - n$$

which is valid due to the shift-invariance of irreducible cumulant moments of order higher than one. In the case that $d_1 = d_{1\max}$, we need to have $d_n = 1$ and $d_k = 0$ for all $k \neq 1, n$. Collecting all powers of N this term will be of order $N^{n-L_{\{l_i\}}+d_{1\max}}\mathcal{Q}_n(N) \leq \mathcal{Q}_n(N) = o(1)$, where we recall that the factor N^n is due to the $O(N^n)$ possible combinations of the distinct k_i's that appear in the corresponding sum in (86), while the factor $N^{-L_{\{l_i\}}}$ comes from the averages over the $\boldsymbol{h}_{k_i}$. Otherwise, when $d_1 < d_{1\max}$, the leading term will have, if that is at all possible, $d_2 = (\sum_i x_i - d_1)/2$ and $d_k = 0$ for $k > 2$. In this case, the scaling of this term with N is $N^{n-L_{\{l_i\}}+d_1} \leq N^{-1} = o(1)$.

The above argument can be extended to the case when there are $q < n$ distinct k_i's. The difference is that now additional terms may appear, since Novikov's formula may give derivatives across different arguments of the cumulant moment, thus reducing the order of the cumulant, e.g., if $k_1 = k_2$ we will get the term

$$\mathbb{E}_c\left[\mathrm{Tr}\left[\boldsymbol{B}_{\{k_i\}}\right]^{l_1-1}\mathrm{Tr}\left[\boldsymbol{B}_{\{k_i\}}\right]^{l_2-1}\mathrm{Tr}\left[\boldsymbol{B}_{\{k_i\}}^2\right];\right.$$
$$\left.\mathrm{Tr}\left[\boldsymbol{B}_{\{k_i\}}\right]^{l_3}\ldots;\mathrm{Tr}\left[\boldsymbol{B}_{\{k_i\}}\right]^{l_n}\right]. \quad (88)$$

In general, for the shell with $q < n$ distinct indices, the resulting terms will include cumulant moments with orders m, such that $q \leq m \leq n$. When $m = n$, we can directly apply the argument used when $q = n$, only replacing the number of possible combinations of the distinct combinations from N^n to N^q.

In the case that $m < n$, when one expresses the reducible cumulant moments in terms of sums over products of irreducible ones, the maximum number of order one cumulants that may appear is now bounded by

$$d_1 \leq d'_{1\max} = \sum_{i=1}^{m} [x_i - 1]_+ < L_{\{l_i\}} - q.$$

The crucial difference is that $d'_{1\max} < L_{\{l_i\}} - q$, which is due to the fact that, as seen in (88), in order to reduce the order of the cumulant moment to $m < n$, one needs to produce traces that span different arguments of the original cumulant moments. In this case, the leading term will be of order $N^{q-L_{\{l_i\}}+d_1} \leq N^{q-L_{\{l_i\}}+d'_{1\max}} = o(1)$.

Following the above argument, we can show that all cumulant moments of the mutual information with order $n > 2$ are negligible in the limit $N \to \infty$.

As far as the ZF receiver is concerned, recall that (see Section V-B2) we can obtain the SINR of the virtual channels of the ZF receiver by setting the parameter α inside the corresponding SINRs expression of the MMSE receiver to ∞. Specifically, expressing the relation in terms of the matrices $\boldsymbol{B}_k$, where we have explicitly denoted the dependence on α, we have

$$\gamma_k^{\mathrm{zf}} = \alpha \lim_{\alpha_0 \to \infty} \boldsymbol{h}_k^{\mathsf{H}} \boldsymbol{B}_k(\alpha_0)\boldsymbol{h}_k$$
$$= \alpha \lim_{\alpha_0 \to \infty} \boldsymbol{h}_k^{\mathsf{H}} \left[\boldsymbol{I} + \alpha_0 \boldsymbol{H}_k \boldsymbol{H}_k^{\mathsf{H}}\right]^{-1} \boldsymbol{h}_k. \quad (89)$$

As mentioned in Section V-B2, the above analysis involving the matrices $\boldsymbol{B}_{\{k_i\}}$, etc., can be carried out in the case of ZF if $\beta < 1$. In addition, as seen in Appendix C, for $\beta < 1$ all nth-order cumulant moments of traces of products of $\boldsymbol{B}_{\{k_i\}}$ are given by (83) and (84). As a result, all finite $n > 2$ order cumulant moments of the mutual information are $o(1)$ for the ZF receiver too.

REFERENCES

[1] *Draft Standardization Document*, IEEE P802.11n/D2.00, Feb. 2007.

[2] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, May 1959.

[3] 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.

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

[5] P. Wolniansky, J. Foschini, G. Golden, and R. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. 1998 URSI Int. Symp. Signals, Systems, and Electronics*, Pisa, Italy, Sep. 29–Oct. 2, 1998, pp. 295–300.

[6] G. Caire and G. Colavolpe, "On low-complexity space-time coding for quasi-static channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 6, pp. 1400–1416, Jun. 2003.

[7] H. El Gamal and R. Hammons, "A new approach to layered space-time coding and signal processing," *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2321–2334, Sep. 2001.

[8] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

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

[10] M. O. Damen, H. El Gamal, and G. Caire, "On maximum-likelihood detection and the search for the closest lattice point," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2389–2402, Oct. 2003.

[11] A. D. Murugan, H. El Gamal, M. O. Damen, and G. Caire, "A unified framework for tree search decoding: Rediscovering the sequential decoder," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 933–953, Mar. 2006.

[12] S. Verdú, *Multiuser Detection*. Cambridge, U.K.: Cambridge Univ. Press, 1998.

[13] E. Biglieri and G. Taricco, "Transmission and reception with multiple antennas: Theoretical foundations," in *Foundations and Trends in Communications and Information Theory*. Hanover, MA: Now Publishers, 2004.

[14] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2003.

[15] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[16] H. El Gamal, G. Caire, and M. O. Damen, "Lattice coding and decoding achieve the optimal diversity–multilpexing tradeoff of MIMO channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 968–985, Jun. 2004.

[17] Y. Jiang and M. K. Varanasi, "Spatial multiplexing architectures with jointly designed rate-tailoring and ordered BLAST decoding—Part I: Diversity-multiplexing tradeoff analysis," *IEEE Trans. Wireless Commun.*, vol. 7, no. 8, pp. 3252–3261, Aug. 2008.

[18] Y. Jiang, M. K. Varanasi, and J. Li, "Performance analysis of ZF and MMSE equalizers for MIMO systems: A closer study in high SNR regime," *IEEE Trans. Inf. Theory*, accepted for publication.

[19] A. Tajer, A. Nosratinia, and N. Al-Dhahir, "MMSE infinite length symbol-by-symbol linear equalization achieves full diversity," in *Proc. IEEE Int. Symp. Information Theory (ISIT 2007)*, Nice, France, Jun. 2007, pp. 1706–1710.
[20] A. Hedayat and A. Nosratinia, "Outage and diversity of linear receivers in flat-fading MIMO channels," *IEEE Trans. Signal Process.*, vol. 55, no. 12, pp. 5868–5873, Dec. 2007.
[21] A. Hedayat, A. Nosratinia, and N. Al-Dhahir, "Linear equalizers for flat fading MIMO channels," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, PA, Mar. 2005, vol. 3, pp. iii/445–iii/448.
[22] J. H. Winters, J. Salz, and R. D. Gitlin, "The impact of antenna diversity on the capacity of wireless communication systems," *IEEE Trans. Commun.*, vol. 42, no. 2/3/4, pp. 1740–1751, Feb./Mar./Apr. 1994.
[23] K. R. Kumar, G. Caire, and A. L. Moustakas, "The diversity-multiplexing tradeoff of linear MIMO receivers," in *Proc. IEEE Information Theory Workshop, ITW' 07*, Porto, Portugal, Sep. 2007, pp. 487–492.
[24] 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.
[25] A. L. Moustakas, S. H. Simon, and A. M. Sengupta, "MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2545–2561, Oct. 2003.
[26] P. J. Smith and M. Shafi, "On the Gaussian approximation to the capacity of wireless MIMO systems," in *Proc. IEEE Int. Conf. Communications*, New York, Apr. 28–May 2, 2002, pp. 406–410.
[27] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, "A new approach for mutual information analysis of large dimensional multi-antenna channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 9, pp. 3987–2561, Sep. 2008.
[28] D. N. C. Tse and O. Zeitouni, "Linear multiuser receivers in random environments," *IEEE Trans. Inf. Theory*, vol. 46, no. 1, p. 171, Jan. 2000.
[29] Y. C. Liang, G. Pan, and Z. D. Bai, "Asymptotic performance of MMSE receivers for large systems using random matrix theory," *IEEE Trans. Inf. Theory*, vol. 53, no. 11, p. 4173, Nov. 2007.
[30] M. Debbah *et al.*, "MMSE analysis of certain large isometric random precoded systems," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, p. 1293, May 2003.
[31] K. R. Kumar and G. Caire, "Space-time codes from structured lattices," *IEEE Trans. Inform. Theory*, vol. 55, no. 2, pp. 547–556, Feb. 2009.
[32] Y. Hong, E. Viterbo, and J. C. Belfiore, "Golden space–time trellis coded modulation," *IEEE Trans. Inform. Theory*, vol. 53, no. 5, pp. 1689–1705, May 2007.
[33] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and systems theory: Theoretical results and some applications," *Proc. IEEE*, vol. 79, no. 3, pp. 278–305, Mar. 1991.
[34] D. N. C Tse and S. Hanly, "Linear multiuser receivers: Effective interference, effective bandwidth and user capacity," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 641–657, Mar. 1999.
[35] S. Verdú and S. Shamai (Shitz), "Spectral efficiency of CDMA with random spreading," *IEEE Trans. Inf. Theory*, vol. 45, no. 2, pp. 622–640, Mar. 1999.
[36] S. Verdú and S. Shamai (Shitz), "The impact of frequency-flat fading on the spectral efficiency of CDMA," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1302–1327, May 2001.
[37] Novikov, *Sov. Phys.—JETP*, vol. 20, p. 1290, 1965.
[38] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Foundations and Trends in Communications and Information Theory*, vol. 1, no. 1, pp. 1–182, 2004.
[39] C. W. J. Beenakker, "Random-matrix theory of quantum transport," *Rev. Mod. Phys.*, vol. 69, pp. 731–808, 1997.
[40] D. H. Politzer, "Random matrix description of the distribution of mesoscopic conductance," *Phys. Rev. B*, vol. 40, no. 17, p. 11917, 1989.

K. Raj Kumar (S'02) received the B.E. degree from the University of Madras, Madras, India, in 2003 and the M.Sc. (Eng.) degree from the Indian Institute of Science, Bangalore, in 2005.

He is currently working toward the Ph.D. degree in the Department of Electrical Engineering-Systems, University of Southern California (USC), Los Angeles. His current research interests include MIMO systems, cooperative communications, cognitive radios, coded modulation, and multiuser information theory.

Mr. Kumar is a recipient of the 2006 Best Student Paper Award from the Department of Electrical Engineering–Systems, USC, and an Oakley Fellowship from USC for the 2007–2008 academic year.

Giuseppe Caire (S'92–M'94–SM'03–F'05) was born in Torino, Italy, in 1965. He received the B.Sc. degree in electrical engineering from Politecnico di Torino, Torino, Italy, in 1990, the M.Sc. degree in electrical engineering from Princeton University, Princeton, NJ, in 1992, and the Ph.D. degree from Politecnico di Torino in 1994.

He has been visiting Princeton University in summer 1997 and Sydney University, Sydney, Australia, in summer 2000. He was with the European Space Agency (ESTEC, Noordwijk, The Netherlands) from May 1994 to February 1995, has been Assistant Professor in Telecommunications at the Politecnico di Torino, Associate Professor at the University of Parma, Parma, Italy, Professor with the Department of Mobile Communications at the Eurecom Institute, Sophia-Antipolis, France, and is now Professor with the Electrical engineering Department of the Viterbi School of Engineering, University of Southern California, Los Angeles. His current interests are in the field of communications theory, information theory, and coding theory with particular focus on wireless applications.

Prof. Caire served as Associate Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS during 1998–2001 and as Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY from 2001 to 2003. He was a recipient of the AEI G. Someda Scholarship in 1991, was a recipient of the COTRAO Scholarship in 1996, and of a CNR Scholarship in 1997. He received the Jack Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, and the Joint IT/Comsoc Best Paper Award in 2004. Since November 2004 he has been member of the Board of Governors of the IEEE Information Theory Society.

Aris L. Moustakas (SM'04) received the B.S. degree in physics from the California Institute of Technology (Caltech), Pasadena, in 1990 and the M.S. and Ph.D. degrees in theoretical condensed matter physics from Harvard University, Cambridge, MA, in 1992 and 1996, respectively.

In 1998, he joined Bell Labs, Lucent Technologies, NJ, first in the Physical Sciences Division and then also in the Wireless Advanced Technology Laboratory. Since 2005, he has been an Assistant Professor at the Physics Department of the National Kapodistrian University of Athens, Athens, Greece. His current research interests lie in the areas of multiple antenna systems and the applications of game theory and statistical physics to the theory of communications and networks.