

A Random Matrix Approach to the Finite Blocklength Regime of MIMO Fading Channels

Jakob Hoydis^{*§}, Romain Couillet[§], Pablo Piantanida^{*}, and M erouane Debbah[§]

^{*}Department of Telecommunications and [§]Alcatel-Lucent Chair on Flexible Radio, SUPELEC, France

{jakob.hoydis, romain.couillet, pablo.piantanida, merouane.debbah}@supelec.fr

Abstract—This paper provides a novel central limit theorem (CLT) for the information density of the MIMO Rayleigh fading channel under white Gaussian inputs, when the data blocklength n and the number of transmit and receive antennas K, N are large but of similar order of magnitude. This CLT is used to derive closed-form upper bounds on the error probability via Feinstein’s lemma and the second-order approximation of the coding rate. Numerical evaluations suggests that the normal approximation is tight for reasonably small values of n, K, N .

I. INTRODUCTION

The conventional notion of capacity focuses on the asymptotic limit of the tradeoff between accuracy and coding rate. When one considers the regime of finite-length codewords, only few results on this tradeoff are known and their evaluation is usually intractable by exact analysis. Thus, practical expressions of fundamental communication limits are mostly given by asymptotic approximations based on the large blocklength regime [1], [2]. Similarly, when multiple-input multiple-output (MIMO) systems are considered, one often relies on large system approximations where the number of transmit and receive antennas are assumed to grow without bounds [3]. For both scenarios, it is well known that these asymptotic approximations mimic closely the system performance in the non-asymptotic regimes. Motivated by this observation, we provide in this paper an asymptotic approximation of the error performance of MIMO channels in the finite blocklength regime, based on large random matrix theory.

One of the fundamental quantities of interest when exploring the tradeoff between achievable rate and block error probability is the information density or the information spectrum. This quantity was used by Feinstein in [4] to derive an upper bound on the block error probability for a given coding rate in the finite blocklength regime. Since this bound is in general not amenable to simple evaluation, asymptotic considerations were made, in particular by Strassen [1] who derived a general expression for the discrete memoryless channel with unconstrained inputs. In his work, the variance of the information density [5] appears as a fundamental quantity. Nevertheless, Strassen’s approach could not be generalized to channels with input constraints, such as the AWGN channel. To tackle this limitation, Hayashi [6] introduced the notion of second-order coding rate and provided an exact characterization of the so-called optimal average error probability when the channel inputs are coded within a vanishing set of rates around the critical rate. Similar considerations were

made in [2], specialized in [7] to the AWGN fading channel case. Further work on the asymptotic blocklength regime via information spectrum methods comprise the general capacity formula derived in [8] based on a lower bound on the error probability provided in [9]. Alternatively, in [10], Shannon derived bounds on the limit of the scaled logarithm of the error probability, known as the exponential rate of decrease. Simpler formulas for the latter were then provided by Gallager [11] which are still difficult to evaluate for practical channel models. To circumvent this issue, a Gaussian approximation of Gallager’s bound with higher-order correction terms was recently obtained in [12] for the Rayleigh fast-fading MIMO channel. In [13], an explicit expression of Gallager’s error exponent was derived for the block-fading MIMO channel. However, the computation of this result is quite involved.

The objective of this article is to investigate Feinstein’s bound on the error probability as well as Hayashi’s optimal average error probability for the Gaussian MIMO Rayleigh fading channel in the non-ergodic regime. Although exact expressions of the optimal error probability are extremely difficult to obtain in this setting, we derive a tight approximation of an upper bound on the error probability, which depends on the blocklength n , the number of transmit and receive antennas K and N , respectively, and the coding rate $r_{n,K}$. More precisely, using recent results from random matrix theory, we show that, given a probability of error $0 \leq \epsilon < 1$, and for n, K , and N sufficiently large, rates of the following form

$$r_{n,K} = \bar{C}_c(\sigma^2) - \frac{\theta_{c,\beta}}{\sqrt{nK}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{nK}}\right) \quad (1)$$

are achievable¹, where $\beta = n/K$, $c = N/K$, and both $\bar{C}_c(\sigma^2)$ and $\theta_{c,\beta}$ are given by simple *closed-form expressions*. Alternatively, for some desired rate $r_{n,K}$ within $O((nK)^{-\frac{1}{2}})$ of the ergodic channel capacity, the optimal error probability is upper-bounded as

$$\mathbb{P}_{e,N,K}^{(n)}(r_{n,K}) \leq Q\left(\frac{\sqrt{nK}}{\theta_{c,\beta}} (\bar{C}_c(\sigma^2) - r_{n,K})\right) + o(1). \quad (2)$$

This bound is useful to assess the backoff from the ergodic channel capacity in the finite blocklength regime and it is characterized by only a few important system parameters. Applications arise for example in the context of MIMO ARQ

¹We denote $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$.

block-fading channels where one is generally interested in minimizing the average delivery delay, rather than maximizing the transmission rate.

II. DEFINITION AND PROBLEM STATEMENT

A. Channel model and its information density

Consider the following MIMO memoryless fading channel:

$$\mathbf{y}_t = \mathbf{H}\mathbf{x}_t + \sigma\mathbf{w}_t, \quad t = \{1, \dots, n\} \quad (3)$$

where $\mathbf{y}_t \in \mathbb{C}^N$ is the channel output at time t , $\mathbf{H} \in \mathbb{C}^{N \times K}$ with independent $\mathcal{CN}(0, 1/K)$ entries is the channel transfer matrix, $\mathbf{x}_t \in \mathbb{C}^{K \times 1}$ is the channel input at time t assumed to be independent of \mathbf{H} , and $\sigma\mathbf{w}_t \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I}_N)$ is an additive noise at the receiver at time t . For later use, we define the following matrices: $\mathbf{X} = [\mathbf{x}_1 \dots \mathbf{x}_n] \in \mathcal{X}_K^n$, $\mathbf{W} = [\mathbf{w}_1 \dots \mathbf{w}_n] \in \mathbb{C}^{N \times n}$, and $\mathbf{Y} = [\mathbf{y}_1 \dots \mathbf{y}_n] \in \mathbb{C}^{N \times n}$. For $\alpha > 0$, the channel inputs \mathbf{X} must belong to the set of admissible inputs \mathcal{X}_K^n which satisfy the energy constraint

$$\mathcal{X}_K^n \triangleq \left\{ \mathbf{X} \in \mathbb{C}^{K \times n} \mid \frac{1}{nK} \text{tr} \mathbf{X}\mathbf{X}^H \leq 1 + \alpha \right\}. \quad (4)$$

Remark 2.1: For the case of independent inputs $\mathbf{x}_t \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, $\Pr\{\mathbf{X} \in \mathcal{X}_K^n\} = \chi_{2nK}^2(2nK(1 + \alpha))$ tends to one, where χ_k^2 denotes the distribution function of a chi-square random variable with k degrees of freedom.

The *information density* $i(\mathbf{X}; \mathbf{Y}\mathbf{H})$ of the channel $\{dP_{\mathbf{Y}\mathbf{H}|\mathbf{X}}\}$ (the joint probability density function (pdf) of (\mathbf{Y}, \mathbf{H}) conditioned on \mathbf{X}), is defined by [5]

$$i(\mathbf{X}; \mathbf{Y}\mathbf{H}) = \frac{1}{nK} \log \left(\frac{dP_{\mathbf{Y}\mathbf{H}|\mathbf{X}}(\mathbf{Y}, \mathbf{H}|\mathbf{X})}{dP_{\mathbf{Y}\mathbf{H}}(\mathbf{Y}, \mathbf{H})} \right) \quad (5)$$

where $dP_{\mathbf{Y}\mathbf{H}}$ denotes the pdf of (\mathbf{Y}, \mathbf{H}) . For the case of independent inputs $\mathbf{x}_t \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, this reads

$$\begin{aligned} I_{N,K}^{(n)}(\sigma^2) &\triangleq \frac{1}{nK} \sum_{t=1}^n \log \left(\frac{dP_{\mathbf{y}_t|\mathbf{H}, \mathbf{x}_t}(\mathbf{y}_t)}{dP_{\mathbf{y}_t|\mathbf{H}}(\mathbf{y}_t)} \right) \\ &= C_{N,K}(\sigma^2) + R_{N,K}^{(n)}(\sigma^2) \end{aligned} \quad (6)$$

where

$$\begin{aligned} C_{N,K}(\sigma^2) &\triangleq \frac{1}{K} \log \det \left(\mathbf{I}_N + \frac{1}{\sigma^2} \mathbf{H}\mathbf{H}^H \right) \\ R_{N,K}^{(n)}(\sigma^2) &\triangleq \frac{1}{nK} \text{tr} \left[(\mathbf{H}\mathbf{H}^H + \sigma^2\mathbf{I}_N)^{-1} \mathbf{Y}\mathbf{Y}^H - \mathbf{W}\mathbf{W}^H \right]. \end{aligned}$$

The information density will be exploited in this work to obtain bounds on two different definitions of error probability.

Definition 1 (Code and average error probability): An $(n, K, M_{n,K}, \varphi, \phi)$ -code for the channel model (3) consists of the following mappings:

- An encoder mapping:

$$\varphi : \mathcal{M}_{(n,K)} \mapsto \mathbb{C}^{K \times n}$$

for each (nK) -blocklength where n, K denote the number of channel uses and transmit antennas, respectively. The transmitted symbols are $\mathbf{X} = \varphi(m)$ for every

message m uniformly distributed over the set $\mathcal{M}_{(n,K)} = \{1, \dots, M_{n,K}\}$.

- A decoder mapping:

$$\phi : \mathbb{C}^{N \times n} \times \mathbb{C}^{N \times K} \mapsto \mathcal{M}_{(n,K)} \cup \{e\},$$

which produces the decoder's guess $\hat{m} = \phi(\mathbf{Y}, \mathbf{H})$ on the sent message m , or the error event e .

Given a code $\mathcal{C}_K^n \triangleq (n, K, M_{n,K}, \varphi, \phi)$, the average error probability is defined as

$$P_{e,N,K}^{(n)}(\mathcal{C}_K^n) \triangleq \frac{1}{M_{n,K}} \sum_{m=1}^{M_{n,K}} \mathbb{E}_{\mathbf{H}} [\Pr(\hat{m} \neq m | \mathbf{X} = \varphi(m), \mathbf{H})]. \quad (7)$$

Let $\text{supp}(\mathcal{C}_K^n)$ denote the codebook $\{\varphi(1), \dots, \varphi(M_{n,K})\}$. The optimal error probability $\mathbb{P}_{e,K,N}^{(n)}(r)$ is the infimum of all error probabilities over \mathcal{C}_K^n defined as²

$$\mathbb{P}_{e,N,K}^{(n)}(r) \triangleq \inf_{\substack{\mathcal{C}_K^n \\ \text{supp}(\mathcal{C}_K^n) \subset \mathcal{X}_K^n}} \left\{ P_{e,N,K}^{(n)}(\mathcal{C}_K^n) \mid \frac{1}{nK} \log M_{n,K} \geq r \right\}. \quad (8)$$

The exact characterization of the optimal error probability $\mathbb{P}_{e,N,K}^{(n)}(r)$ for fixed n, K, N and non-trivial channel models is generally intractable. An upper-bound for the exact optimal error probability was provided by Feinstein [4] as follows.

Theorem 1 (Feinstein [4] (see also [7, Thm. 24])): Let \mathbf{X} be an arbitrary input to the channel $\{dP_{\mathbf{Y}\mathbf{H}|\mathbf{X}}\}$ with output \mathbf{Y} and channel matrix \mathbf{H} . Given an arbitrary positive integer $M_{n,K}$, there exists a $\mathcal{C}_K^n = (n, K, M_{n,K}, \varphi, \phi)$ -code with codewords in the set \mathcal{X}_K^n satisfying

$$\begin{aligned} &P_{e,N,K}^{(n)}(\mathcal{C}_K^n) \Pr\{\mathbf{X} \in \mathcal{X}_K^n\} \\ &\leq \Pr \left\{ i(\mathbf{X}; \mathbf{Y}\mathbf{H}) \leq \frac{1}{nK} \log M_{n,K} + \delta_{n,K} \right\} + e^{-nK\delta_{n,K}} \end{aligned}$$

for all tuples (K, n, N) and $\delta_{n,K} > 0$.

There have been recent efforts [6], [2] to establish error probability approximations when the coding rate is within $O((nK)^{-\frac{1}{2}})$ of the ergodic capacity. In this scenario, a ‘‘second-order’’ expression is defined as follows.

Definition 2 (Second-order approximation): We define the *optimal average error probability* for the second-order coding rate r as [6], [2]

$$\begin{aligned} \mathbb{P}_e(r|\beta, c) &\triangleq \inf_{\{\mathcal{C}_K^n : \text{supp}(\mathcal{C}_K^n) \subset \mathcal{X}_K^n\}_{n=1}^{\infty}} \left\{ \limsup_{N \xrightarrow{(\beta, c)} \infty} P_{e,N,K}^{(n)}(\mathcal{C}_K^n) \mid \right. \\ &\left. \liminf_{N \xrightarrow{(\beta, c)} \infty} \sqrt{nK} \left(\frac{1}{nK} \log M_{n,K} - \mathbb{E}[C_{N,K}(\sigma^2)] \right) \geq r \right\} \end{aligned} \quad (9)$$

where $N \xrightarrow{(\beta, c)} \infty$ denotes $N, K, n \rightarrow \infty$, $\frac{n}{K} \rightarrow \beta$, $\frac{N}{K} \rightarrow c$.

We now provide closed-form approximations for the error probability given in the above definitions, using new asymptotic statistics on the information density.

²Although the focus is on the smallest average error probability at a given rate, by fixing the error probability and looking at the maximum achievable rate, similar results can be derived with essentially the same methods.

III. MAIN RESULTS

The first result is a central limit theorem (CLT) for the information density $I_{N,K}^{(n)}(\sigma^2)$ with Gaussian i.i.d. inputs \mathbf{x}_t .

Theorem 2 (Fluctuations of the information density): Let $n, K, N \rightarrow \infty$, such that $\frac{N}{K} \rightarrow c > 0$, $\frac{n}{K} \rightarrow \beta > 0$. Then,

$$(i) \quad \mathbb{E} \left[I_{N,K}^{(n)}(\sigma^2) \right] = \bar{C}_c(\sigma^2) + \mathcal{O}\left(\frac{1}{N^2}\right)$$

where

$$\bar{C}_c(\sigma^2) = \log(1 + cm) - \frac{cm}{1 + cm} + c \log\left(1 + \frac{1}{\sigma^2} \frac{1}{1 + cm}\right)$$

and

$$m = \frac{c-1}{2c\sigma^2} - \frac{1}{2c} + \frac{\sqrt{(1-c+\sigma^2)^2 + 4c\sigma^2}}{2c\sigma^2}.$$

$$(ii) \quad \frac{\sqrt{nK}}{\theta_{c,\beta}} \left(I_{N,K}^{(n)}(\sigma^2) - \bar{C}_c(\sigma^2) \right) \Rightarrow \mathcal{N}(0, 1)$$

$$\frac{\sqrt{nK}}{\theta_{c,\beta}} \left(I_{N,K}^{(n)}(\sigma^2) - \mathbb{E}[C_{N,K}(\sigma^2)] \right) \Rightarrow \mathcal{N}(0, 1)$$

where the asymptotic variance $\theta_{c,\beta}^2$ is given as

$$\theta_{c,\beta}^2 = -\beta \log\left(1 - \frac{cm^2}{(1+cm)^2}\right) + 2c(1 - \sigma^2 m).$$

Proof: A sketch of proof is provided in the appendix. ■

We now apply the CLT to provide a tight approximation of the upper bound in Theorem 1.

Corollary 1 (Upper bound on Feinstein's error probability): Let $\mathbf{x}_t \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$, independent across t . Then, for $\alpha > 0$ and any coding rate $r_{n,K}$,

$$\mathbb{P}_{e,N,K}^{(n)}(r_{n,K}) \chi_{2nK}^2(2nK(1+\alpha)) \leq \bar{\mathbb{P}}_{e,N,K}^{(n)}(r_{n,K}) + o(1)$$

where

$$\bar{\mathbb{P}}_{e,N,K}^{(n)}(r_{n,K}) = Q\left(\frac{\bar{C}_c(\sigma^2) - r_{n,K} - \delta_{n,K}^*}{(nK)^{-\frac{1}{2}} \theta_{c,\beta}}\right) + e^{-nK\delta_{n,K}^*}$$

with $\delta_{n,K}^* = u - \sqrt{u^2 - v}$,

$$u = \bar{C}_c(\sigma^2) - r_{n,K} + \theta_{c,\beta}^2$$

$$v = (\bar{C}_c(\sigma^2) - r_{n,K})^2 + \frac{\theta_{c,\beta}^2}{nK} \log(2\pi nK \theta_{c,\beta}^2).$$

Proof: A sketch of proof is provided in the appendix. ■

From Theorem 2, we can also obtain in a straight-forward fashion the following upper bound for (9).

Corollary 2 (Upper bound on the optimal average error): The optimal average error probability (9) with second-order coding rate r is upper bounded by

$$\mathbb{P}_e(r|\beta, c) \leq Q\left(-\frac{r}{\theta_{c,\beta}}\right) \quad (10)$$

where $\theta_{c,\beta}$ is given in Theorem 2.

Proof: A sketch of proof is provided in the appendix. ■

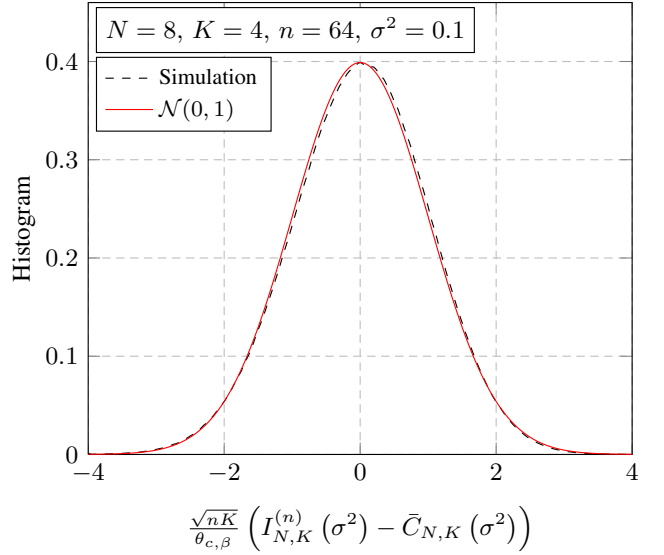


Fig. 1. Histogram of the fluctuations of the information density, for $N = 8$, $K = 4$, $n = 64$, and $\sigma^2 = 0.1$.

Remark 3.1: It is interesting to observe the transition from Corollary 1 to the second-order approximation when $r_{n,K}$ is close to the ergodic capacity, i.e., $r_{n,K} = \mathbb{E}[C_{N,K}(\sigma^2)] + \frac{r}{\sqrt{nK}}$. In this case, one can show that $\sqrt{nK}\delta_{n,K}^* \rightarrow 0$ while $nK\delta_{n,K}^* \rightarrow \infty$. Moreover, as $n, K \rightarrow \infty$, $\chi_{2nK}^2(2nK(1+\alpha)) \rightarrow 1$. Hence, the upper-bound on $\mathbb{P}_{e,N,K}^{(n)}(r_{n,K})$ can be approximated by (2). Letting $\mathbb{P}_{e,N,K}^{(n)}(r_{n,K}) = \epsilon$ and applying the inverse Q -function to both sides of (2) yields the approximated achievable rate (1).

IV. NUMERICAL RESULTS

In order to validate the accuracy of Theorem 2 (ii) for finite n, N , and K , we compare in Fig. 1 the empirical histogram of $\sqrt{nK}/\theta_{c,\beta}(I_{N,K}^{(n)}(\sigma^2) - \bar{C}_c(\sigma^2))$ against the standard normal distribution for $N = 8$, $K = 4$, $n = 64$, and $\sigma^2 = 0.1$. Even for these small system dimensions, we observe an almost perfect match between both results.

In Fig. 2, we then compare the error bound $\bar{\mathbb{P}}_{e,N,K}^{(n)}(r_{n,K})$ of Corollary 1 against a numerical evaluation of (25), both seen as functions of n for the same parameters as above. We suppose a coding rate of $r_{n,K} = 0.85 \times \mathbb{E}[C_{N,K}(\sigma^2)] = 3.41$ b/s/Hz. Under this assumption, the best possible error probability is the outage probability $P_{\text{out}} = \Pr\{C_{N,K}(\sigma^2) < r_{n,K}\} = 1.4\%$. Surprisingly, the approximation of (25) by $\bar{\mathbb{P}}_{e,N,K}^{(n)}(r_{n,K})$ is extremely accurate, even for very small values of n . We additionally provide the upper-bound of (2) in the same plot (the term $o(1)$ being discarded). For the chosen set of parameters, the error approximation (2) is not tight and leads to an overly optimistic error bound. Further simulations, not provided here for lack of space, confirm that this approximation becomes accurate as N, K, n , and $r_{n,K}$ increase.

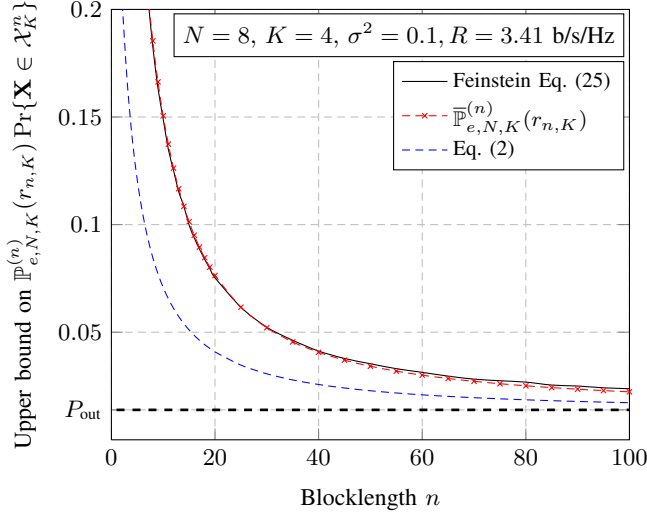


Fig. 2. Upper bounds on the (discounted) error probability $\mathbb{P}_{e,N,K}^{(n)}(r_{n,K})$ for $N = 8, K = 4, \sigma^2 = 0.1, r_{n,K} = 0.85 \times \mathbb{E}[C_{N,K}(\sigma^2)] = 3.41$ b/s/Hz, as a function of n , where $P_{\text{out}} = \Pr\{C_{N,K}(\sigma^2) < r_{n,K}\} = 1.4\%$ denotes the outage probability.

V. SUMMARY AND DISCUSSION

We have studied the error probability of quasi-static MIMO Rayleigh fading channels in the finite blocklength regime. Under a large system assumption, we have derived a CLT of the information density. This result was used to compute a tight closed-form approximation of Feinstein's upper bound on the optimal error probability and an achievable upper bound of the optimal average error probability in the second-order coding rate. Numerical results demonstrated that the Gaussian approximation is valid for very small blocklengths and realistic numbers of antennas. Some comments on relevant issues and on-going work are in order:

- *Relation to channel dispersion:* One can show that the asymptotic variance of the information density $\theta_{c,\beta}^2$ coincides asymptotically with the fundamental notion of channel dispersion [6], [7], defined as

$$V_{N,K} = \mathbb{E} \left[\left(I_{N,K}^{(n)}(\sigma^2) - \mathbb{E}[I_{N,K}^{(n)}(\sigma^2) | \mathbf{X}] \right)^2 \right].$$

- *Converse to Corollary 2:* Proving a converse to the optimal average error probability would require the derivation of a CLT of the information density for general input distributions. The proof of such a result is also related to the conjecture of Telatar on the outage-minimizing input distribution for multi-antenna fading channels, recently confirmed for the MISO channel in [14].
- *Extensions to other scenarios of interest:* The block-fading regime as well as tradeoffs between channel training and data transmission can also be addressed within the framework proposed in this article. Moreover, CLTs of the information density with linear receive filters can be obtained by similar techniques.

APPENDIX

Proof sketch of Theorem 2: Part (i) is [15, Theorem 1]. For notational convenience, we drop dependencies on σ^2 . To prove part (ii), we start by defining the following quantities: $\tilde{I}_{N,K}^{(n)} = I_{N,K}^{(n)} - \mathbb{E}[I_{N,K}^{(n)}]$, $\tilde{C}_{N,K} = C_{N,K} - \mathbb{E}[C_{N,K}]$, and $\tilde{R}_{N,K}^{(n)} = R_{N,K}^{(n)} - \mathbb{E}[R_{N,K}^{(n)}]$.

1) *Asymptotic variance:* With the above definitions, the variance of $I_{N,K}^{(n)}$ can be expressed as

$$\mathbb{E} \left[\left(\tilde{I}_{N,K}^{(n)} \right)^2 \right] = \mathbb{E} \left[\tilde{C}_{N,K}^2 \right] + \mathbb{E} \left[\left(R_{N,K}^{(n)} \right)^2 \right] - \left(\mathbb{E} \left[R_{N,K}^{(n)} \right] \right)^2 + 2\mathbb{E} \left[\tilde{C}_{N,K} \tilde{R}_{N,K}^{(n)} \right]. \quad (11)$$

After straightforward calculations, one can show that

$$\mathbb{E} \left[R_{N,K}^{(n)} \right] = 0 \quad \text{and} \quad \mathbb{E} \left[\tilde{C}_{N,K} \tilde{R}_{N,K}^{(n)} \right] = 0. \quad (12)$$

In a similar manner, one arrives after some calculus at

$$\mathbb{E} \left[\left(R_{N,K}^{(n)} \right)^2 \right] = \frac{2c}{\beta K^2} \left(1 - \mathbb{E} \left[\frac{\sigma^2}{N} \text{tr} \left(\mathbf{H}\mathbf{H}^H + \sigma^2 \mathbf{I}_N \right)^{-1} \right] \right). \quad (13)$$

From [15, Theorem 3], it follows that

$$\mathbb{E} \left[\frac{1}{N} \text{tr} \left(\mathbf{H}\mathbf{H}^H + \sigma^2 \mathbf{I}_N \right)^{-1} \right] = m + \mathcal{O} \left(\frac{1}{N^2} \right). \quad (14)$$

By [15, Theorem 2], we have

$$\mathbb{E} \left[\left(\sqrt{nK} \tilde{C}_{N,K} \right)^2 \right] \rightarrow -\beta \log \left(1 - \frac{cm^2}{(1+cm)^2} \right). \quad (15)$$

Equations (11)–(15) taken together finally prove that

$$\mathbb{E} \left[\left(\sqrt{nK} \tilde{I}_{N,K}^{(n)} \right)^2 \right] \rightarrow \theta_{c,\beta}^2. \quad (16)$$

2) *CLT:* Let us rewrite $R_{N,K}^{(n)}$ in the following way:

$$R_{N,K}^{(n)} = \frac{1}{nK} \sum_{t=1}^n z_t^n \quad (17)$$

where $z_t^n = \mathbf{y}_t^H (\mathbf{H}\mathbf{H}^H + \sigma^2 \mathbf{I}_N)^{-1} \mathbf{y}_t - \mathbf{w}_t^H \mathbf{w}_t$. Conditionally on \mathbf{H} , z_1^n, \dots, z_n^n are i.i.d. with zero mean and variance

$$\vartheta_n^2 = \frac{2nc}{\beta} \left(1 - \sigma^2 \frac{1}{N} \text{tr} \left(\mathbf{H}\mathbf{H}^H + \sigma^2 \mathbf{I}_N \right)^{-1} \right). \quad (18)$$

By Cauchy-Schwartz and Markov inequalities, for any $\varepsilon > 0$,

$$\begin{aligned} & \sum_{t=1}^n \frac{1}{n\vartheta_n^2} \mathbb{E} \left[|z_t^n|^2 \mathbb{1}_{|z_t^n| \geq \varepsilon \sqrt{n}\vartheta_n} \right] \\ & \leq \frac{1}{\vartheta_n^2} \sqrt{\mathbb{E}[|z_1^n|^2]} \sqrt{\mathbb{E} \left[\mathbb{1}_{|z_1^n| \geq \varepsilon \sqrt{n}\vartheta_n} \right]} \\ & = \frac{1}{\vartheta_n} \sqrt{\Pr \{ |z_1^n| \geq \varepsilon \sqrt{n}\vartheta_n \}} \\ & \leq \frac{1}{\vartheta_n} \sqrt{\frac{\mathbb{E}[|z_1^n|^2]}{\varepsilon^2 n \vartheta_n^2}} = \frac{1}{\varepsilon \vartheta_n \sqrt{n}}. \end{aligned} \quad (19)$$

Now, taking \mathbf{H} in a well-chosen space of probability one, we know from (14) that $\frac{1}{N}\text{tr}(\mathbf{H}\mathbf{H}^H + \sigma^2\mathbf{I}_N)^{-1} \rightarrow m > 0$ and, therefore, $\liminf_n \vartheta_n > 0$. This implies that $(\varepsilon\vartheta_n\sqrt{n})^{-1} \rightarrow 0$, and, as a consequence

$$\limsup_n \sum_{i=1}^n \frac{1}{n\vartheta_n^2} \mathbb{E} \left[|z_i^n|^2 \mathbb{1}_{|z_i^n| \geq \varepsilon\sqrt{n}\vartheta_n} \right] = 0 \quad (20)$$

which is the Lindeberg condition. By [16, Theorem 27.2], we therefore conclude that, almost surely,

$$\frac{1}{\sqrt{n}\vartheta_n} \sum_{t=1}^n z_t^n = \sqrt{\frac{K}{\vartheta_n^2}} \sqrt{nK} R_{N,K}^{(n)} \Rightarrow \mathcal{N}(0, 1).$$

Thus, by the continuity of the complex exponential, (14), and the dominated convergence theorem, we arrive at

$$\mathbb{E}_{\mathbf{H}} \left[\left| \mathbb{E}_{\mathbf{X}, \mathbf{W}} \left[e^{iu\sqrt{nK}R_{N,K}^{(n)}} \right] - e^{-u^2c(1-\sigma^2m)} \right| \right] \rightarrow 0. \quad (21)$$

We also know from [15, Theorem 2] that

$$\mathbb{E}_{\mathbf{H}} \left[e^{iu\sqrt{nK}\check{C}_{N,K}} - e^{\frac{1}{2}u^2\beta \log\left(1 + \frac{cm^2}{(1+cm)^2}\right)} \right] \rightarrow 0 \quad (22)$$

where $\check{C}_{N,K} = C_{N,K} - \bar{C}_c$. Define $\tilde{n} = \sqrt{nK}$ and write

$$\mathbb{E} \left[e^{iu\tilde{n}(I_{N,K}^{(n)} - \bar{C}_c)} \right] = \mathbb{E}_{\mathbf{H}} \left[e^{iu\tilde{n}\check{C}_{N,K}} \mathbb{E}_{\mathbf{X}, \mathbf{W}} \left[e^{iu\tilde{n}R_{N,K}^{(n)}} \right] \right]. \quad (23)$$

Thus,

$$\begin{aligned} & \left| \mathbb{E}_{\mathbf{H}} \left[e^{iu\tilde{n}\check{C}_{N,K}} \mathbb{E}_{\mathbf{X}, \mathbf{W}} \left[e^{iu\tilde{n}R_{N,K}^{(n)}} \right] \right] - e^{-\frac{1}{2}u^2\theta_{c,\beta}^2} \right| \\ & \leq \left| \mathbb{E}_{\mathbf{H}} \left[e^{iu\tilde{n}\check{C}_{N,K}} \left(\mathbb{E}_{\mathbf{X}, \mathbf{W}} \left[e^{iu\tilde{n}R_{N,K}^{(n)}} \right] - e^{-u^2c(1-\sigma^2m)} \right) \right] \right| \\ & + \left| \left(\mathbb{E}_{\mathbf{H}} \left[e^{iu\tilde{n}\check{C}_{N,K}} \right] - e^{\frac{1}{2}u^2\beta \log\left(1 + \frac{cm^2}{(1+cm)^2}\right)} \right) e^{-u^2c(1-\sigma^2m)} \right| \\ & \leq \mathbb{E}_{\mathbf{H}} \left[\left| \mathbb{E}_{\mathbf{X}, \mathbf{W}} \left[e^{iu\tilde{n}R_{N,K}^{(n)}} \right] - e^{-u^2c(1-\sigma^2m)} \right| \right] \\ & + \left| \mathbb{E}_{\mathbf{H}} \left[e^{iu\tilde{n}\check{C}_{N,K}} \right] - e^{\frac{1}{2}u^2\beta \log\left(1 + \frac{cm^2}{(1+cm)^2}\right)} \right|. \quad (24) \end{aligned}$$

By (21) and (22), the right-hand side of (24) tends to zero as $N, K, n \rightarrow \infty$. Thus, $\mathbb{E} \left[e^{iu\tilde{n}(I_{N,K}^{(n)} - \bar{C}_c)} \right] \rightarrow e^{-\frac{1}{2}u^2\theta_{c,\beta}^2}$ which, by Lévy's continuity theorem, terminates the proof. ■

Proof sketch of Corollary 1: From Theorem 1, Theorem 2 (ii), and [17, Lemma 2.11], we immediately obtain

$$\begin{aligned} & \mathbb{P}_{e,N,K}^{(n)}(r_{n,K}) \chi_{2nK}^2(2nK(1+\alpha)) \\ & \leq \inf_{\delta_{n,K}} \Pr \left\{ I_{N,K}^{(n)}(\sigma^2) \leq r_{n,K} + \delta_{n,K} \right\} + e^{-nK\delta_{n,K}} \quad (25) \\ & = \inf_{\delta_{n,K}} Q \left(\frac{\bar{C}_c(\sigma^2) - r_{n,K} - \delta_{n,K}}{(nK)^{-\frac{1}{2}}\theta_{c,\beta}} \right) + e^{-nK\delta_{n,K}} + o(1). \end{aligned}$$

Ignoring the negligible term, one can easily see that the last equation is minimized by $\delta_{n,K}^*$ as given in the theorem. ■

Proof sketch of Corollary 2: By restricting us to Gaussian inputs and codes of rate $\frac{1}{nK} \log M_{n,K} = \mathbb{E} [C_{N,K}(\sigma^2)] +$

r/\sqrt{nK} , we obtain by Theorem 1 the following upper bound on the optimal average error probability

$$\begin{aligned} & \mathbb{P}_e(r|\beta, c) \\ & \leq \limsup_{N \xrightarrow{(\beta,c)} \infty} \Pr \left\{ I_{N,K}^{(n)}(\sigma^2) \leq \mathbb{E} [C_{N,K}(\sigma^2)] + \frac{r}{\sqrt{nK}} \right\} \\ & \quad \text{supp}(C_K^n) \subset \mathcal{X}_K^n \}. \quad (26) \end{aligned}$$

Since $\frac{1}{nK} \text{tr} \mathbf{X}\mathbf{X}^H \rightarrow 1$ with probability one, the event $\text{supp}(C_K^n) \subset \mathcal{X}_K^n$ is satisfied with probability converging to one. Thus, by Theorem 2-(ii),

$$\begin{aligned} & \mathbb{P}_e(r|\beta, c) \\ & \leq \limsup_{N \xrightarrow{(\beta,c)} \infty} \Pr \left\{ \frac{\sqrt{nK}}{\theta_{c,\beta}} \left(I_{N,K}^{(n)} - \mathbb{E} [C_{N,K}] \right) \leq \frac{r}{\theta_{c,\beta}} \right\} \\ & = Q \left(\frac{-r}{\theta_{c,\beta}} \right). \quad (27) \end{aligned}$$

REFERENCES

- [1] V. Strassen, "Asymptotische Abschätzungen in Shannon's Informations-theorie," in *Proc. 3rd Prague Conf. Inf. Theory*, Czechoslovak Academy of Sciences, Prague, Czech Republic, 1962, pp. 689–723.
- [2] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [3] P. Katakopoulos, P. Mertikopoulos, A. L. Moustakas, and G. Caire, "Living at the edge: A large deviations approach to the outage MIMO capacity," *IEEE Trans. Inf. Theory*, vol. 57, no. 4, pp. 1984–2007, Apr. 2011.
- [4] A. Feinstein, "A new basic theorem of information theory," *IRE Transactions on Information Theory*, pp. 2–20, 1954.
- [5] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer-Verlag, 2003.
- [6] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.
- [7] Y. Polyanskiy and S. Verdú, "Scalar coherent fading channel: Dispersion analysis," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Aug. 2011, pp. 2959–2963.
- [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] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.
- [10] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, 1959.
- [11] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.
- [12] Y. Chen and M. R. McKay, "Coulumb fluid, Painlevé transcendents and the information theory of MIMO systems," *IEEE Trans. Inf. Theory*, submitted.
- [13] H. Shin and M. Z. Win, "Gallager's exponent for MIMO channels: A reliability-rate tradeoff," vol. 57, no. 4, pp. 972–985, Apr. 2009.
- [14] E. Abbe, S.-L. Huang, and E. Telatar, "Proof of the outage probability conjecture for MISO channels," *IEEE Trans. Inf. Theory*, March 2011, submitted. [Online]. Available: <http://arxiv.org/abs/1103.5478>
- [15] 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. 3987–4004, Sep. 2008.
- [16] P. Billingsley, *Probability and Measure*, 3rd ed. John Wiley & Sons, Inc., 1995.
- [17] A. W. van der Vaart, *Asymptotic Statistics*. Cambridge University Press, New York, 2000.