Sharpened Capacity Lower Bounds of Fading MIMO Channels with Imperfect CSI

Adriano Pastore
Dep. of Signal Theory & Communications
Universitat Politècnica de Catalunya
Barcelona, Spain
adriano.pastore@upc.edu

Jakob Hoydis
Bell Laboratories
Alcatel-Lucent
Stuttgart, Germany
jakob.hoydis@alcatel-lucent.com

Javier R. Fonollosa
Dep. of Signal Theory & Communications
Universitat Politècnica de Catalunya
Barcelona, Spain
javier.fonollosa@upc.edu

***Abstract*—A well-established capacity lower bound of multiple-input multiple-output (MIMO) single-user fading channels operating with imperfect receiver-side channel-state information (CSI) is improved using a simple rate-splitting and successive-decoding scheme. The potential improvement is shown to increase with the number of allowed decoding steps (layers) to such extent that the best layering strategy is approached in the limit as the number of layers tends to infinity. We give a general analytic expression of this limit, which constitutes a new capacity lower bound that is sharper than the conventional bound. Using large random matrix theory, we derive an asymptotic approximation of this novel bound, which is shown via numerical simulation to be highly accurate over the whole range of signal-to-noise ratios.**

I. Introduction

The capacity of fading channels with imperfect channel-state information (CSI) at the receiver side remains unknown and seems difficult to characterize. Therefore, a large body of work has concentrated on capacity and mutual information bounds. What is probably the best known capacity lower bound was derived by Médard for the single-antenna case [1] and extended to the MIMO case by Baltersee, Fock and Meyr [2], as well as Hassibi and Hochwald [3], who coined the term *worst-case noise* bound. Specifically, this bound corresponds to a lower bound on the mutual information for Gaussian codebooks. Thanks to its simplicity, it has become widely popular as a means to derive achievable rate expressions for the noncoherent setting (by inclusion of a training scheme as in [3]), or for the partially coherent setting, in which a channel estimate is assumed to be available to the receiver (e.g., [4]). As such, this bound has been used in a large body of work.

In [5], [6], it was shown that said bound can be improved by a combination of rate splitting at the transmitter and successive decoding at the receiver side, for any number of decoding steps (layers) larger than one. Within the family of such multiple-layer bounds, the conventional worst-case noise bound [3, Theorem 1] can be viewed as the single-layer bound. In the MIMO setting—similarly to the single-antenna setting [6]—if the number of layers can be arbitrarily large, the optimal layering strategy consists in driving the number of layers to infinity, in such a way that the powers and rates associated to each layer become infinitesimally small. An analytic expression for this limit is proposed which gives rise to an improved capacity lower bound of MIMO channels with imperfect receiver CSI. Its expression is more complex than that of the conventional bound, but fortunately, it lends itself well to large random-matrix approximations. Such an approximation is derived for the case of independent and identically distributed (i.i.d.) fading gains, and it is demonstrated by simulation that these approximations are very close to the exact value of the bound.

A. Pastore and J. R. Fonollosa have been supported by the Ministerio de Economía y Competitividad of Spain (TEC2010-19171 and CONSOLIDER-INGENIO 2010 CSD2008-00010 COMONSENS) and Generalitat de Catalunya (2009SGR-1236).

II. System Model

A discrete-time multiple-antenna channel with imperfect CSI at the receiver side is considered, whose time-k channel output $\mathbf{y}_k \in \mathbb{C}^r$ (where $\mathbb{C}$ denotes the set of complex numbers) corresponding to the time-k channel input $\mathbf{x}_k \in \mathbb{C}^t$ having a distribution with covariance $\mathbf{Q} = \mathsf{E}\big[\mathbf{x}_k\mathbf{x}_k^\dagger\big]$ and normalized power $\mathrm{tr}(\mathbf{Q}) = 1$, is given by

$$\mathbf{y}_k = \sqrt{\rho}(\hat{\mathbf{H}}_k + \tilde{\mathbf{H}}_k)\mathbf{x}_k + \mathbf{n}_k. \tag{1}$$

Here, $\mathbf{n}_k \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_r)$ is independent additive Gaussian (circularly symmetric) noise while ρ represents the signal-to-noise ratio (SNR), and $(\hat{\mathbf{H}}_k, \tilde{\mathbf{H}}_k) \in \mathbb{C}^{r\times t} \times \mathbb{C}^{r\times t}$ is a pair of complex-valued random matrices normalized such that the channel coefficients have an average second moment of one, that is,

$$\frac{1}{rt}\sum_{i=1}^{t}\sum_{j=1}^{r}\mathsf{E}\Big[\big|[\hat{\mathbf{H}}_k]_{i,j} + [\tilde{\mathbf{H}}_k]_{i,j}\big|^2\Big] = 1 \tag{2}$$

where $[\hat{\mathbf{H}}_k]_{i,j}$ and $[\tilde{\mathbf{H}}_k]_{i,j}$ stand for the (i,j)-th entry of $\hat{\mathbf{H}}_k$ and $\tilde{\mathbf{H}}_k$, respectively. For simplicity, we assume that all variables are stationary ergodic and memoryless, therefore we drop the time index k from now on.[1]

The receiver is cognizant of the joint distribution of $(\hat{\mathbf{H}}, \tilde{\mathbf{H}})$ and of the realization $\hat{\mathbf{H}}$, but ignores $\tilde{\mathbf{H}}$. Thus, $\hat{\mathbf{H}}$ can be viewed as a channel estimate (known channel component),

[1]Note, however, that the bounds presented in this article are versatile enough to be extended to other fading models, such as block fading or time-selective fading.

while $\tilde{\mathbf{H}}$ can be viewed as the channel estimation error (unknown channel component). Additionally, we assume that, conditioned on $\hat{\mathbf{H}}$, the channel estimate is unbiased, i.e.,[2]

$$\mathsf{E}\big[\tilde{\mathbf{H}} \bigm| \hat{\mathbf{H}}\big] = \mathbf{0}. \tag{3}$$

Also, we assume that the input $\mathbf{x}$, the noise $\mathbf{n}$, and the pair of channel components $(\hat{\mathbf{H}}, \tilde{\mathbf{H}})$ are mutually independent (though $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$ may be mutually dependent).

According to the channel coding theorem, given a fixed input distribution, the conditional mutual information

$$\begin{aligned} I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}) &= h(\mathbf{x}) - h(\mathbf{x}|\mathbf{y},\hat{\mathbf{H}}) \\ &= h(\mathbf{y}|\hat{\mathbf{H}}) - h(\mathbf{y}|\mathbf{x},\hat{\mathbf{H}}) \end{aligned} \tag{4}$$

represents the supremum of rates at which random channel coding can achieve an arbitrarily small block error probability as the blocklength tends to infinity.

The capacity of the channel (1) is unknown in general. For a fixed transmit covariance $\mathbf{Q}$, it is given by

$$C = \sup_{\mathbf{x}} I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}) \tag{5}$$

where the supremum is over all distributions of $\mathbf{x}$ having covariance $\mathbf{Q}$. Only in the *coherent* setting, where the channel-state is exactly known to the receiver (i.e., $\tilde{\mathbf{H}} = \mathbf{0}$ almost surely), the capacity is known to be [7]

$$C_{\text{coh}} \triangleq \sup_{\mathbf{x}} I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}) = \mathsf{E}\log\det\big(\mathbf{I}_t + \rho\hat{\mathbf{H}}^\dagger\hat{\mathbf{H}}\mathbf{Q}\big) \tag{6}$$

and is achieved by Gaussian inputs $\mathbf{x} \sim \mathcal{CN}(\mathbf{0},\mathbf{Q})$. In this situation, the receiver has *perfect* channel-state information. In contrast, for *imperfect* channel-state information as treated in this article (i.e., $\tilde{\mathbf{H}} \neq \mathbf{0}$ with non-zero probability), the capacity C is unknown. However, the lower bound [1], [3]

$$C \geq \mathsf{E}\log\det\big(\mathbf{I}_t + \hat{\mathbf{H}}^\dagger\boldsymbol{\Gamma}^{-1}\hat{\mathbf{H}}\mathbf{Q}\big) \triangleq R_{\text{M}} \tag{7}$$

is frequently used, where $\rho\boldsymbol{\Gamma}$ stands for the covariance of the effective noise $\sqrt{\rho}\tilde{\mathbf{H}}\mathbf{x} + \mathbf{n}$ conditioned on $\hat{\mathbf{H}}$, i.e.,

$$\boldsymbol{\Gamma} = \mathsf{E}\big[\tilde{\mathbf{H}}\mathbf{Q}\tilde{\mathbf{H}}^\dagger \bigm| \hat{\mathbf{H}}\big] + \rho^{-1}\mathbf{I}_r. \tag{8}$$

To be precise, the right-hand side (RHS) of (7) is a lower bound on the Gaussian-input mutual information $I(\mathbf{x}_{\text{G}};\mathbf{y}|\hat{\mathbf{H}})$, where $\mathbf{x}_{\text{G}} \sim \mathcal{CN}(\mathbf{0},\mathbf{Q})$. In the following, we show how the lower bound (7) can be sharpened using successive decoding.

III. Successive decoding

Consider the transmit covariance $\mathbf{Q}$ to be fixed throughout. In a system that performs rate splitting and successive decoding with L layers, we write the Gaussian transmitted signal $\mathbf{x}_{\text{G}}$ as a sum of $\mathbf{x}_\ell \sim \mathcal{CN}(\mathbf{0},\mathbf{Q}_\ell)$, which are mutually independent Gaussian random variables satisfying

$$\mathbf{x}_{\text{G}} = \sum_{\ell=1}^{L}\mathbf{x}_\ell \qquad\qquad \mathbf{Q} = \sum_{\ell=1}^{L}\mathbf{Q}_\ell \tag{9}$$

[2] In systems where $\hat{\mathbf{H}}$ is a function of some channel side information Ω independent of the input $\mathbf{x}$ (e.g., a training observation in training-based channel estimation), (3) means that $\hat{\mathbf{H}} = \mathsf{E}[\mathbf{H}|\Omega]$, i.e., $\hat{\mathbf{H}}$ is the minimum mean-square error estimator of $\mathbf{H}$ from Ω.

Let us denote a sequence $(A_1,\ldots,A_n)$ as A^n. The sequence of transmit covariances $(\mathbf{Q}_\ell)_{\ell=1,\ldots,L}$ is thus denoted as $\mathbf{Q}^L$. Since $\mathbf{y}$ depends on $\mathbf{x}^L$ only via the sum $\mathbf{x}_{\text{G}}$ of its elements, $\mathbf{y}$ has the same distribution whether it is conditioned on $\mathbf{x}_{\text{G}}$ or on $\mathbf{x}^L$, so

$$I\big(\mathbf{x}_{\text{G}};\mathbf{y}\big|\hat{\mathbf{H}}\big) = I\big(\mathbf{x}^L;\mathbf{y}\big|\hat{\mathbf{H}}\big) = \sum_{\ell=1}^{L} I\big(\mathbf{x}_\ell;\mathbf{y}\big|\mathbf{x}^{\ell-1},\hat{\mathbf{H}}\big). \tag{10}$$

The second equality follows from the chain rule for mutual information. By a direct application of the worst-case noise lower bound [3, Theorem 1], we can lower-bound each summand on the RHS of (10) as

$$I\big(\mathbf{x}_\ell;\mathbf{y}\big|\mathbf{x}^{\ell-1},\hat{\mathbf{H}}\big) \geq \mathsf{E}\log\det\big(\mathbf{I}_t + \hat{\mathbf{H}}^\dagger\boldsymbol{\Gamma}_\ell^{-1}\hat{\mathbf{H}}\mathbf{Q}_\ell\big) \tag{11}$$

wherein the random matrix $\boldsymbol{\Gamma}_\ell$ is defined as

$$\begin{aligned} \boldsymbol{\Gamma}_\ell \triangleq \mathsf{E}\big[\tilde{\mathbf{H}}\boldsymbol{\xi}_\ell\boldsymbol{\xi}_\ell^\dagger\tilde{\mathbf{H}}^\dagger \bigm| \boldsymbol{\xi}_\ell,\hat{\mathbf{H}}\big] &+ \mathsf{E}\big[\tilde{\mathbf{H}}\mathbf{Q}_\ell\tilde{\mathbf{H}}^\dagger \bigm| \hat{\mathbf{H}}\big] \\ &+ \hat{\mathbf{H}}\bar{\mathbf{Q}}_\ell\hat{\mathbf{H}}^\dagger + \mathsf{E}\big[\tilde{\mathbf{H}}\bar{\mathbf{Q}}_\ell\tilde{\mathbf{H}}^\dagger \bigm| \hat{\mathbf{H}}\big] + \rho^{-1}\mathbf{I}_r \end{aligned} \tag{12}$$

with $\boldsymbol{\xi}_\ell = \sum_{i<\ell}\mathbf{x}_i$ and $\bar{\mathbf{Q}}_\ell = \sum_{i=\ell+1}^{L}\mathbf{Q}_i$. By summing (11) over ℓ, with the chain rule (10) we obtain the so-called *successive-decoding lower bound* $R(\mathbf{Q}^L)$ induced by $\mathbf{Q}^L$:

$$\begin{aligned} I\big(\mathbf{x}_{\text{G}};\mathbf{y}\big|\hat{\mathbf{H}}\big) &\geq \sum_{\ell=1}^{L}\mathsf{E}\log\det\big(\mathbf{I}_t + \hat{\mathbf{H}}^\dagger\boldsymbol{\Gamma}_\ell^{-1}\hat{\mathbf{H}}\mathbf{Q}_\ell\big) \\ &\triangleq R(\mathbf{Q}^L). \end{aligned} \tag{13}$$

The following Theorem establishes that this successive-decoding technique allows one to sharpen the conventional lower bound R_{M}.

Theorem 1. *Any successive-decoding bound $R(\mathbf{Q}^L)$ is not smaller than the bound (7), i.e.,*

$$R_{\text{M}} \leq R(\mathbf{Q}^L) \leq I(\mathbf{x}_{\text{G}};\mathbf{y}|\hat{\mathbf{H}}). \tag{14}$$

Proof: Averaging $\boldsymbol{\Gamma}_\ell$ over $\boldsymbol{\xi}_\ell$ yields [cf. (8)]

$$\mathsf{E}[\boldsymbol{\Gamma}_\ell|\hat{\mathbf{H}}] = \boldsymbol{\Gamma} + \hat{\mathbf{H}}\bar{\mathbf{Q}}_\ell\hat{\mathbf{H}}^\dagger \tag{15}$$

and upon noting that the function $\mathbf{X} \mapsto \log\det(\mathbf{I}+\mathbf{A}\mathbf{X}^{-1})$ for positive semidefinite $\mathbf{A} \succeq \mathbf{0}$ is (strictly) convex over the set of positive definite $\mathbf{X} \succ \mathbf{0}$, we can apply Jensen's inequality to get

$$\begin{aligned} R(\mathbf{Q}^L) &\geq \sum_{\ell=1}^{L}\mathsf{E}\log\det\Big(\mathbf{I}_t + \hat{\mathbf{H}}^\dagger\,\mathsf{E}[\boldsymbol{\Gamma}_\ell|\hat{\mathbf{H}}]^{-1}\hat{\mathbf{H}}\mathbf{Q}_\ell\Big) \\ &= R_{\text{M}} \end{aligned} \tag{16}$$

with equality in the single-layer case ($L = 1$), or if the distribution of $(\hat{\mathbf{H}},\tilde{\mathbf{H}})$ is such that, almost surely, one at least among $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$ is zero (sufficient condition). ■

We would be interested in determining the best bound $R^\star$ (optimized over all layering strategies) for a prescribed covariance matrix $\mathbf{Q}$:

$$R^\star \triangleq \sup_{\substack{\mathbf{Q}^L,\,L\in\mathbb{N} \\ \sum_\ell \mathbf{Q}_\ell = \mathbf{Q}}} R(\mathbf{Q}^L) \tag{17}$$

As this problem is difficult, we will decompose it into a two-stage optimization, one stage of which we can solve analytically. For this purpose, let $\mathbb{C}_+^{t\times t}$ denote the set of complex positive semidefinite $t\times t$ matrices, and consider the following Definitions.

Definition 1. *A function $\mathbf{K}\colon [0;1] \to \mathbb{C}_+^{t\times t}, \iota \mapsto \mathbf{K}(\iota)$ with $\mathbf{K}(0)=\mathbf{0}$ and $\mathbf{K}(1)=\mathbf{Q}$ that satisfies the two conditions*

$$\iota_1 < \iota_2 \Rightarrow \mathbf{K}(\iota_1) \preceq \mathbf{K}(\iota_2) \quad \text{and} \quad \operatorname{tr}(\mathbf{K}(\iota)) = \iota$$

is called a $\mathbf{Q}$-layering function.

Definition 2. *An finite set of distinct real numbers $\mathcal{I} = \{\iota_0, \iota_1, \ldots, \iota_L\} \in [0;1]^{L+1}$ labelled in ascending order $0 = \iota_0 < \iota_1 < \ldots < \iota_L = 1$ is referred to as an L-indexing.*

Using the above definitions, we can cast any pair $(\mathbf{K},\mathcal{I})$ into a layering $\mathbf{Q}^L = (\mathbf{Q}_\ell)_{\ell=1,\ldots,L}$ by means of

$$\mathbf{Q}_\ell = \mathbf{K}(\iota_\ell) - \mathbf{K}(\iota_{\ell-1}), \quad \ell = 1,\ldots,L. \tag{18}$$

Conversely, for any layering $\mathbf{Q}^L$ one can find a pair $(\mathbf{K},\mathcal{I})$ from which the $\mathbf{Q}_\ell$ can be constructed via (18). Due to this one-to-one correspondence, we write $R(\mathbf{Q}^L) = R(\mathbf{K},\mathcal{I})$ without distinction of meaning.

With these definitions in hand, we can rewrite (17) as

$$R^\star = \sup_{\mathbf{K}} \sup_{\mathcal{I}} R(\mathbf{K},\mathcal{I}). \tag{19}$$

The outer supremum over all $\mathbf{Q}$-layering functions $\mathbf{K}$ seems too difficult to solve in the general case, but can be shown to disappear for a single transmit antenna. As to the inner supremum, we can compute it by exploiting the fact that

$$\mathcal{I} \subset \mathcal{I}' \Rightarrow R(\mathbf{K},\mathcal{I}) \le R(\mathbf{K},\mathcal{I}'). \tag{20}$$

This is a straightforward generalization of Theorem 1 and is proved in a similar way (using Jensen's inequality). As a direct consequence, the inner supremum in (19) is approached as the number of layers (or the cardinality of $\mathcal{I}$) tends to infinity. The following Theorem provides an explicit expression for this limit, which we shall denote as $R^\star(\mathbf{K}) = \sup_{\mathcal{I}} R(\mathbf{K},\mathcal{I})$.

Theorem 2. *For a prescribed layering function $\mathbf{K}$, the best successive-decoding bound is given by the Riemann-Stieltjes integral*[3]

$$R^\star(\mathbf{K}) = \mathsf{E}\int_0^1 \operatorname{tr}\Big[\hat{\mathbf{H}}^\dagger \boldsymbol{\Gamma}(\iota)^{-1}\hat{\mathbf{H}}\,\mathrm{d}\mathbf{K}(\iota)\Big] \tag{21}$$

which is always well-defined as a consequence of the properties of layering functions set forth by Definition 1, and where

$$\boldsymbol{\Gamma}(\iota) = \mathsf{E}\big[\tilde{\mathbf{H}}\mathbf{K}(\iota)^{\frac12}\boldsymbol{\xi}\boldsymbol{\xi}^\dagger\mathbf{K}(\iota)^{\frac12}\tilde{\mathbf{H}}^\dagger \,\big|\, \boldsymbol{\xi},\hat{\mathbf{H}}\big] + \hat{\mathbf{H}}(\mathbf{Q}-\mathbf{K}(\iota))\hat{\mathbf{H}}^\dagger + \mathsf{E}\big[\tilde{\mathbf{H}}(\mathbf{Q}-\mathbf{K}(\iota))\tilde{\mathbf{H}}^\dagger \,\big|\, \hat{\mathbf{H}}\big] + \rho^{-1}\mathbf{I}_r \tag{22}$$

with $\boldsymbol{\xi} \sim \mathcal{CN}(\mathbf{0},\mathbf{I}_t)$ independent of $\hat{\mathbf{H}}$.

Proof: To be published in [8]. The special case of a single-antenna channel $t=r=1$ is treated in depth in [6]. ∎

[3]The notation $\int_0^1 \operatorname{tr}\big[\mathbf{A}(\iota)\,\mathrm{d}\mathbf{B}(\iota)\big]$ with matrix-valued functions $\mathbf{A}(\iota) = [A_{i,j}(\iota)]_{i,j}$ and $\mathbf{B}(\iota) = [B_{i,j}(\iota)]_{i,j}$ stands for the sum of Riemann-Stieltjes integrals $\sum_{i,j}\int_0^1 A_{i,j}(\iota)\,\mathrm{d}B_{j,i}(\iota)$.

IV. Isotropic fading

We shall consider the special case of mutually independent $\operatorname{vec}(\hat{\mathbf{H}}) \sim \mathcal{CN}(\mathbf{0},\hat{V}\mathbf{I}_{rt})$ and $\operatorname{vec}(\tilde{\mathbf{H}}) \sim \mathcal{CN}(\mathbf{0},\tilde{V}\mathbf{I}_{rt})$ (i.i.d. Rayleigh fading) with $\hat{V}+\tilde{V}=1$. Setting $\mathbf{Q} = \frac1t\mathbf{I}_t$ and choosing a spatially uniform layering function $\mathbf{U}(\iota) = \frac{\iota}{t}\mathbf{I}_t$, the L-layer successive-decoding bound induced by the layering $\mathbf{U}$ and by an L-indexing $\mathcal{I} = (0,\iota_1,\ldots,\iota_{L-1},1)$ reads as

$$R(\mathbf{U},\mathcal{I}) = \sum_{\ell=1}^{L}\Bigg(\mathsf{E}\log\det\left(\mathbf{I}_t + \frac{1-\iota_{\ell-1}}{A_\ell t}\hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger\right) - \mathsf{E}\log\det\left(\mathbf{I}_t + \frac{1-\iota_\ell}{A_\ell t}\hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger\right)\Bigg) \tag{23}$$

where

$$A_\ell = \tilde{V}\iota_{\ell-1}\tfrac{\Xi_t}{t} + \tilde{V}(1-\iota_{\ell-1}) + \rho^{-1} \tag{24}$$

with Ξ_t being a Gamma distributed random variable with shape t and scale 1 whose probability density function is given by $f(x)$ in (35b).

When considering an infinite number of layers, Theorem 2 specializes to the Riemann integral

$$R^\star(\mathbf{U}) = \frac1t\int_0^1 \mathsf{E}\operatorname{tr}\Big[\hat{\mathbf{H}}^\dagger\boldsymbol{\Gamma}(\iota)^{-1}\hat{\mathbf{H}}\Big]\,\mathrm{d}\iota \tag{25}$$

where

$$\boldsymbol{\Gamma}(\iota) = \iota\tilde{V}\tfrac{\Xi_t}{t}\mathbf{I}_r + (1-\iota)\big(\tfrac1t\hat{\mathbf{H}}\hat{\mathbf{H}}^\dagger + \tilde{V}\mathbf{I}_r\big) + \rho^{-1}\mathbf{I}_r. \tag{26}$$

V. Large random-matrix approximations

As the general expressions in (23) and (25) are rather involved and do not allow for a simple evaluation, we will consider the large system limit $t,r\to\infty$ such that $\limsup(r/t)<\infty$, which will be denoted as $t\to\infty$ in the following. This allows us to derive tight and easily computable approximations of the previously developed bounds. We first recall the following result which provides an asymptotically exact approximation of the coherent capacity.

Theorem 3 ([9, Theorem 8])**.** *Let $\mathbf{h} = \operatorname{vec}(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0},\mathbf{I}_{rt})$, where $\mathbf{H}\in\mathbb{C}^{r\times t}$. Assuming $\mathbf{Q} = \frac1t\mathbf{I}_t$ and $\rho>0$, the coherent capacity C_{coh} as defined in (6) satisfies*

$$C_{\text{coh}} = \bar{I} + \mathcal{O}\left(t^{-1}\right) \tag{27}$$

as $t\to\infty$, where

$$\bar{I} = \int_{\frac1\rho}^{\infty}\left(\frac{r}{u} - t\delta(u)\right)\mathrm{d}u \tag{28}$$

$$= t\log(1+\delta) + r\log\left(1+\frac{1}{\rho(1+\delta)}\right) - \frac{t\delta}{1+\delta} \tag{29}$$

and $\delta = \delta(1/\rho)$ is given in Theorem 5 in Appendix A.

We would like to remark that Theorem 3 is a much stronger result than the well-known convergence of the *per-antenna* mutual information to its asymptotic limit (see, e.g., [10]) which holds for channels composed of arbitrary i.i.d. entries with finite second-order moment. As a direct consequence of

Theorem 3, we can obtain the following approximation of the lower bound in (7):

$$R_{\mathrm{M}} = \bar{R}_{\mathrm{M}} + \mathcal{O}\left(t^{-1}\right) \tag{30}$$

where $\bar{R}_{\mathrm{M}} = \bar{I}\big(\hat{V}/(\tilde{V}+\rho^{-1})\big)$.

Since one can show (cf. [9]) that the $\mathcal{O}(t^{-1})$-term in Theorem 3 is integrable over any compact $[0,\rho]$, we can approximate the L-layer successive-decoding bound in (23) in a similar fashion by

$$R(\mathbf{U},\mathcal{I}) = \bar{R}(\mathbf{U},\mathcal{I}) + \mathcal{O}\left(t^{-1}\right) \tag{31}$$

where

$$\bar{R}(\mathbf{U},\mathcal{I}) = \sum_{\ell=1}^{L}\int_{0}^{\infty}\left[\bar{I}\left(\frac{1-\iota_{\ell-1}}{\sigma_\ell^2(x)}\right) - \bar{I}\left(\frac{1-\iota_\ell}{\sigma_\ell^2(x)}\right)\right] f(x)\,\mathrm{d}x \tag{32}$$

with $f(x)$ as given in (35b) and

$$\sigma_\ell^2(x) = \frac{\tilde{V}\iota_{\ell-1}\frac{x}{t} + \tilde{V}(1-\iota_{\ell-1}) + \rho^{-1}}{\hat{V}}, \quad l = 1,\ldots,L. \tag{33}$$

This result is especially useful as it can be used to find an approximation of the optimal indexing $\mathcal{I}$. An example will be shown later in Section VI.

Relying on Theorem 5 in Appendix A, we can also derive an approximation of the bound with an infinite number of layers $R^\star(\mathbf{U})$. This result is summarized in next Theorem.

Theorem 4. *For $\rho > 0$, let*

$$\bar{R}^\star(\mathbf{U}) = \int_0^\infty\int_{\sigma^2(x)}^{\infty}\frac{r - tg(x,u)\delta(g(x,u))}{u} f(x)\,\mathrm{d}u\,\mathrm{d}x \tag{34}$$

where

$$\sigma^2(x) = \frac{\frac{x}{t}\tilde{V}+\rho^{-1}}{\hat{V}} \tag{35a}$$

$$f(x) = \frac{x^{t-1}}{(t-1)!}e^{-x} \tag{35b}$$

$$g(x,u) = \left(1-\frac{x}{t}\right)\frac{\tilde{V}}{\hat{V}} + u \tag{35c}$$

and $\delta(x)$ is defined in (39)*. Then, as $t\to\infty$,*

$$R^\star(\mathbf{U}) = \bar{R}^\star(\mathbf{U}) + \mathcal{O}\left(t^{-1}\right). \tag{36}$$

Proof: The proof is provided in Appendix B. ■

Remark 1. *By comparison with* (28)*, the inner integral in* (34) *evaluated for $x = t$ corresponds exactly to $\bar{I}$ for $\rho = \sigma^{-2}(t) = \hat{V}/(\tilde{V}+\rho^{-1})$, and hence the lower bound $\bar{R}_{\mathrm{M}}$.*

Remark 2. *Although the computation of $\bar{R}^\star(\mathbf{U})$ requires the numerical evaluation of two integrals, it can be computed very efficiently and, most importantly, much faster than $R^\star(\mathbf{U})$.*

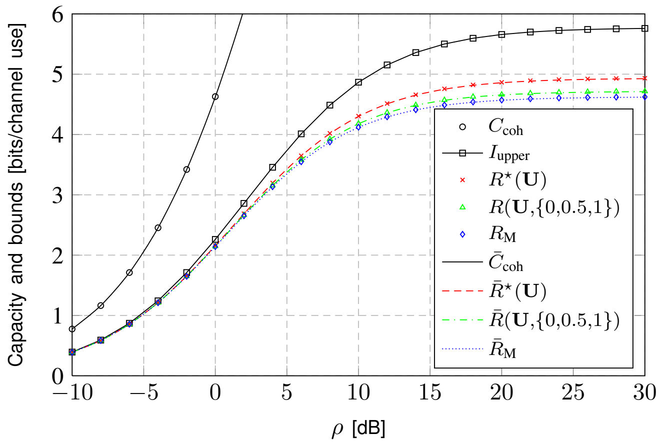

Fig. 1. Coherent capacity, bounds, and asymptotic approximations versus the signal-to-noise ratio ρ for $r = 6$, $t = 4$, $\hat{V} = \tilde{V} = 0.5$. The 2-layer successive-decoding bound $R(\mathbf{U},\{0,\iota_1,1\})$ is evaluated for $\iota_1 = 0.5$.

VI. Numerical Results

We would now like to compare the different bounds developed in the previous sections and assess the quality of their asymptotic approximations. Figure 1 depicts the coherent capacity, the bound R_{M} from (7), the successive-decoding bounds for $L = 2$ and $\iota_1 = 0.5$, and $R^\star(\mathbf{U})$, together with their respective asymptotic approximations from Section V, versus the signal-to-noise ratio ρ. To assess how much closer the bound $R^\star(\mathbf{U})$ is to the exact mutual information $I(\mathbf{x}_{\mathrm{G}};\mathbf{y}|\hat{\mathbf{H}})$ as compared to R_{M}, an upper bound $I_{\mathrm{upper}} \geq I(\mathbf{x}_{\mathrm{G}};\mathbf{y}|\hat{\mathbf{H}})$ is also shown, which reads as (e.g., [2], [4], [6])

$$I_{\mathrm{upper}} = R_{\mathrm{M}} + r\,\mathsf{E}\left[\log\frac{\tilde{V}+\rho^{-1}}{\tilde{V}\frac{\Xi_t}{t}+\rho^{-1}}\right]. \tag{37}$$

In this figure, we have used the parameters $(r,t) = (6,4)$, and $\hat{V} = \tilde{V} = 0.5$.[4] We can observe that the asymptotic approximations are almost indistinguishable from the simulation results over the entire range of SNRs. The improvement of the successive-decoding bound $R^\star(\mathbf{U})$ over the conventional bound R_{M} becomes especially pronounced at high SNR. We can see that the 2-layer successive-decoding bound is well below the best successive-decoding bound $R^\star(\mathbf{U})$. This demonstrates the impact of a large number of layers.

In order to visualize the impact of the power allocation across layers, in Figure 2 we show the 2-layer successive-decoding bound $R(\mathbf{U},\mathcal{I})$ with its asymptotic approximation $\bar{R}(\mathbf{U},\mathcal{I})$ as a function of $\iota_1 \in [0,1]$. We have chosen the same parameters as in Figure 1, and $\rho = 10\,\mathrm{dB}$. The accuracy of asymptotic approximations suggests that it can be directly used (instead of the exact value) to optimize the value of ι_1. For $\iota_1 = 0$ or $\iota_1 = 1$, the 2-layer bound coincides with R_{M}.

[4] In many practical scenarios such as training-based channels, we would rather have $(\hat{V},\tilde{V}) \to (1,0)$ as $\rho\to\infty$. As a consequence of choosing $(\hat{V},\tilde{V})$ and the distribution of $\mathbf{x}_{\mathrm{G}}$ (Gaussian) to not depend on the SNR, the curves in Figure 1 are bounded, except for C_{coh}. It is known that the capacity C (not plotted) would exhibit a double-logarithmic growth at high SNR [11].

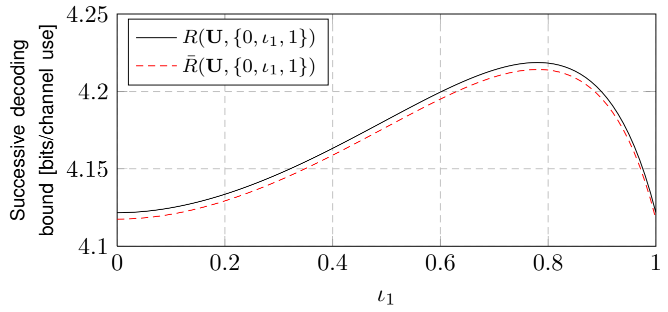

Fig. 2. 2-layer successive decoding bound $R(\mathbf{U},\{0,\iota_1,1\})$ and asymptotic approximation $\bar{R}(\mathbf{U},\{0,\iota_1,1\})$ versus ι_1 for $\rho = 10\,\text{dB}$, $r = 6$, $t = 4$, and $\hat{V} = \tilde{V} = 0.5$.

VII. Conclusion

We have demonstrated that a rate-splitting/successive-decoding approach on a fading MIMO channel with imperfect CSI yields capacity lower bounds which outperform the conventional lower bound from [1], [3]. We have provided an analytic expression of the optimal successive-decoding bound (obtained when optimizing over all possible layerings), which generalizes the single-antenna expression from [6] to the MIMO case. With the help of large random-matrix theory, we have computed an approximation of this new bound, which allows for an easier manipulation. Interesting extensions of this work include, for example, a general treatment of spatially correlated channels with random-matrix approximations.

Appendix A
Related Results

Theorem 5. *Let* $\mathbf{h} = \text{vec}(\mathbf{H}) \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{rt})$, *where* $\mathbf{H} \in \mathbb{C}^{r\times t}$. *Then, for any* $x > 0$,

$$\left|\mathsf{E}\left[\operatorname{tr}\left(\tfrac{1}{t}\mathbf{H}\mathbf{H}^\dagger + x\mathbf{I}_r\right)^{-1}\right] - t\delta(x)\right| \le \frac{2r}{x^4 t^2} \tag{38}$$

where

$$\delta(x) = \frac{\frac{r}{t}-1}{2x} - \frac{1}{2} + \frac{\sqrt{\left(1-\frac{r}{t}+x\right)^2 + 4\frac{r}{t}x}}{2x}. \tag{39}$$

Sketch: The result can be easily proved along the lines of the proofs of [9, Theorem 3, Proposition 3 and 5]. ■

Appendix B
Proof of Theorem 4

Denote $\Xi_t = \boldsymbol{\xi}^\dagger\boldsymbol{\xi}$ which is a Gamma distributed random variable with shape t and scale 1 whose probability density function is given by $f(x)$ for $x \ge 0$ [cf. (35b)]. Upon performing the variable change $u = \frac{\sigma^2(x)}{1-\iota}$, applying the identity $\mathbf{A}(\mathbf{A}+z\mathbf{I})^{-1} = \mathbf{I} - z(\mathbf{A}+z\mathbf{I})^{-1}$ for some matrix $\mathbf{A}$ and scalar z, and using the Fubini theorem, one can express the successive-decoding bound as

$$R^\star(\mathbf{U}) = \int_0^\infty \int_{\sigma^2(x)}^\infty \left(\frac{r}{u} - \frac{g(x,u)}{u} \times \right.$$
$$\left. \times\, \mathsf{E}\left[\operatorname{tr}\left(\frac{\mathbf{H}\mathbf{H}^\dagger}{t} + g(u,x)\mathbf{I}_r\right)^{-1}\right]\right) f(x)\,\mathrm{d}u\,\mathrm{d}x.$$

Applying Theorem 5 to the trace term on the RHS of the last equation yields

$$R^\star(\mathbf{U}) = \int_0^\infty \int_{\sigma^2(x)}^\infty \left(\frac{r - tg(x,u)\delta(g(x,u))}{u} + \varepsilon\right) f(x)\,\mathrm{d}u\,\mathrm{d}x \tag{40}$$

for some ε, satisfying $|\varepsilon| \le \frac{2r}{ug(x,u)^3t^2}$. One can verify that $tg(x,u)\delta(g(x,u)) \le r$ and continue by bounding the error term:

$$\begin{aligned}
&\left|\int_0^\infty \int_{\sigma^2(x)}^\infty \varepsilon f(x)\,\mathrm{d}u\,\mathrm{d}x\right| \\
&\quad\le \int_0^\infty \int_{\sigma^2(x)}^\infty \frac{2r}{ug(x,u)^3t^2} f(x)\,\mathrm{d}u\,\mathrm{d}x \\
&\quad\le \int_0^\infty \int_{\sigma^2(x)}^\infty \frac{2r}{\sigma^2(x)g(x,u)^3t^2} f(x)\,\mathrm{d}u\,\mathrm{d}x \\
&\quad= \int_0^\infty \frac{2r}{\sigma^2(x)(g(x,\sigma^2(x)))^2t^2} f(x)\,\mathrm{d}x \\
&\quad\le \frac{2r\rho\hat{V}^3}{\left(\tilde{V}+\rho^{-1}\right)^2 t^2} = \mathcal{O}(t^{-1}).
\end{aligned} \tag{41}$$

Since this integral is finite, the integral over the first two integrands in (40) must also exist. This concludes the proof.

References

[1] M. Médard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.

[2] J. Baltersee, G. Fock, and H. Meyr, "Achievable rate of MIMO channels with data-aided channel estimation and perfect interleaving," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 12, pp. 2358–2368, Dec. 2001.

[3] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[4] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Trans. Inf. Theory*, vol. 52, no. 5, pp. 2203–2214, May 2006.

[5] A. Pastore, T. Koch, and J. Fonollosa, "Improved capacity lower bounds for fading channels with imperfect CSI using rate splitting," in *IEEE 27th Convention of Electrical Electronics Engineers in Israel (IEEEI)*, Nov. 2012, pp. 1–5.

[6] ——, "A rate-splitting approach to fading channels with imperfect channel-state information," *IEEE Trans. Inf. Theory*, 2013, submitted. [Online]. Available: *arXiv preprint*.

[7] E. Telatar, "Capacity of multi-antenna Gaussian channels," *European Trans. on Telecommunications*, vol. 10, pp. 585–595, 1999.

[8] A. Pastore and J. Fonollosa, "Improved mutual information lower bounds for fading MIMO channels with imperfect CSI," in preparation.

[9] 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.

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

[11] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.