Short Packets Over Block-Memoryless Fading Channels: Pilot-Assisted or Noncoherent Transmission?

Johan Östman, *Student Member, IEEE*, Giuseppe Durisi, *Senior Member, IEEE*, Erik G. Ström, *Senior Member, IEEE*, Mustafa C. Coşkun, *Student Member, IEEE*, and Gianluigi Liva, *Senior Member, IEEE*

***Abstract*— We present nonasymptotic upper and lower bounds on the maximum coding rate achievable when transmitting short packets over a Rician memoryless block-fading channel for a given requirement on the packet error probability. We focus on the practically relevant scenario in which there is no *a priori* channel state information available at the transmitter or at the receiver. An upper bound built upon the min–max converse is compared with two lower bounds: the first one relies on a noncoherent transmission strategy in which the fading channel is not estimated explicitly at the receiver and the second one employs pilot-assisted transmission (PAT) followed by maximum-likelihood channel estimation and scaled mismatched nearest-neighbor decoding at the receiver. Our bounds are tight enough to unveil the optimum number of diversity branches that a packet should span so that the energy per bit required to achieve a target packet error probability is minimized, for a given constraint on the code rate and the packet size. Furthermore, the bounds reveal that noncoherent transmission is more energy efficient than PAT, even when the number of pilot symbols and their power is optimized. For example, in Rayleigh fading, for the case when a coded packet of 168 symbols is transmitted using a channel code of rate 0.48-bits/channel use, over a block-fading channel with block size equal to eight symbols, PAT requires an additional 1.2 dB of energy per information bit to achieve a packet error probability of 10^{-3} compared with a suitably designed noncoherent transmission scheme. Finally, we devise a PAT scheme based on punctured tail-biting quasi-cyclic codes and ordered-statistics decoding, whose performance is close (1-dB gap at 10^{-3} packet error probability) to the ones predicted by our PAT lower bound. This shows that the PAT lower bound provides useful guidelines on the design of actual PAT schemes.**

***Index Terms*— Ultra-reliable low-latency communications, fading channels, noncoherent communications, pilot-assisted transmission, nearest-neighbor decoding, finite-blocklength information theory, error-exponent analysis, short-packet channel coding.**

Manuscript received December 18, 2017; revised May 13, 2018 and August 16, 2018; accepted September 29, 2018. Date of publication October 9, 2018; date of current version February 14, 2019. This work was partly supported by the Swedish Research Council under grants 2014-6066 and 2016-03293. This paper was presented in part at the IEEE International Workshop on Signal Processing Advances in Wireless Communications, Sapporo, Japan, July 2017 [1]. The associate editor coordinating the review of this paper and approving it for publication was L. Cottatellucci. *(Corresponding author: Johan Östman.)*

J. Östman, G. Durisi, and E. G. Ström are with the Department of Electrical Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden (e-mail: johanos@chalmers.se; durisi@chalmers.se; erik.strom@chalmers.se).

M. C. Coşkun and G. Liva are with the Institute of Communications and Navigation, German Aerospace Center (DLR), 82234 Weßling, Germany (e-mail: mustafa.coskun@tum.de; gianluigi.liva@dlr.de).

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

Digital Object Identifier 10.1109/TCOMM.2018.2874993

I. Introduction

SUPPORTING the transmission of short packets under stringent latency and reliability constraints is critically required for next-generation wireless communication networks to address the needs of future autonomous systems such as connected vehicles, automated factories, and smart grids [2], [3]. Classic information-theoretic performance metrics, i.e., the *ergodic* and the *outage capacity*, provide inaccurate benchmarks to the performance of short-packet communication systems, because of the assumption of asymptotically large blocklength [4]. In particular, these performance metrics are unable to capture the tension between the throughput gains in the transmission of short packets over wireless fading channels that are attainable by exploiting channel diversity, and the throughput losses caused by the insertion of pilot symbols, which are often used to estimate the wireless fading channel at the receiver [5].

A more useful performance metric for short-packet communication systems is the so-called *maximum coding rate* $R^*(n,\epsilon)$, which is the largest rate achievable for a fixed blocklength n, and a fixed packet error probability ϵ. No closed-form expressions for $R^*(n,\epsilon)$ are available for the channel models of interest in wireless communication systems. However, tight bounds on $R^*(n,\epsilon)$ as well as second-order expansions in the limit $n \to \infty$ have been recently reported for a variety of wireless channel models. These results rely on the nonasymptotic information-theoretic tools developed in [6].

In this paper, we study the maximum coding rate achievable over Rician memoryless block-fading channels, for the case in which no *a priori* channel state information (CSI) is available at the transmitter and at the receiver. Such a setup is of particular interest in sporadic short-packet transmissions subject to stringent latency constraints. Indeed, the CSI that may have been acquired at the receiver during previous packet transmissions is often outdated due to the sporadic nature of the transmissions, and delay constraints may prevent the use of a feedback link, which is necessary for the transmitter to obtain CSI. In practical wireless systems, the receiver typically obtains CSI through the use of pilot-assisted transmission (PAT) schemes [5], which involve multiplexing known pilot symbols among the data symbols within each packet. Our goal is to investigate the performance of such schemes when packets are short, using a nonasymptotic information-theoretic analysis.

A. Prior Art

The Nonfading AWGN Channel: Tight upper (converse) and lower (achievability) bounds on $R^*(n,\epsilon)$, based on cone packing, were obtained by Shannon [7]. Polyanskiy, Poor, and Verdú [6] showed recently that Shannon's converse bound is a special case of the so-called *min-max converse* [6, Th. 27], [8], a general converse bound that involves a binary hypothesis test between the channel law and a suitably chosen auxiliary distribution. Furthermore, they obtained an alternative achievability bound—the $\kappa\beta$-bound [9, Th. 25]—also based on binary hypothesis testing. This bound, although less tight than Shannon's achievability bound, is easier to evaluate numerically and to analyze asymptotically. Indeed, Shannon's achievability bound relies on the transmission of codewords that are uniformly distributed on the surface of an $(n-1)$-dimensional hypersphere in $\mathbb{R}^n$ (a.k.a., *spherical* or *shell codes*), which makes the induced output distribution unwieldy. Min-max and $\kappa\beta$ bounds solve this problem by replacing the above-mentioned output distribution by a product Gaussian distribution, which is easier to analyze analytically.

Characterizing the asymptotic behavior of the min-max converse and the $\kappa\beta$ bounds in the large-blocklength regime, Polyanskiy, Poor, and Verdú established the following asymptotic expansion for $R^*(n,\epsilon)$ (see [6] and also the refinement in [10]), which, for convenience, we state for the case of a *complex* AWGN channel:

$$R^*(n,\epsilon) = C - \sqrt{n^{-1}V}Q^{-1}(\epsilon) + \mathcal{O}\big(n^{-1}\log n\big)\,. \tag{1}$$

Here, $C=\log(1+\rho)$, where ρ denotes the SNR, is the channel capacity, $V=\rho(2+\rho)/(1+\rho)^2$ is the so-called channel *dispersion*, $Q(\cdot)$ is the Gaussian Q function, and $\mathcal{O}(n^{-1}\log n)$ comprises remainder terms of order $n^{-1}\log n$. The expansion (1), which is commonly referred to as *normal approximation* relies on a central-limit-theorem analysis and is accurate when R^* is close to capacity. Note, however, that when the blocklength is short and the target packet error probability is low, which implies that the maximum coding rate is far from capacity, large-deviation analyses resulting in the classical Gallager's random-coding error exponent (RCEE) [11] yield more accurate approximations than (1).

Fading Channels–no A-Priori CSI: Bounds on R^* for multiantenna fading channels offering multiple diversity branches in time and/or frequency were reported in [4]. Specifically, the authors of [4] considered a multiantenna Rayleigh memoryless block-fading channel and assumed that coding can be performed across a fixed number of independently fading blocks. The converse bound in [4] relies on the min-max converse, whereas the achievability bound is built upon the so-called dependence-testing (DT) bound [6, Th. 17]. The input distribution used in [4] to compute the DT bound is the one induced by unitary space-time modulation (USTM) [12], according to which the matrices describing the signal transmitted within each coherence block over the available transmit antennas are drawn independently from the uniform distribution on the set of unitary matrices. Then, they are scaled so as to satisfy the power constraint. This distribution, which achieves capacity at high SNR [13] (provided that the sum of transmit and receive antennas does not exceed the length of the coherence block), corresponds—in the single-input single-output (SISO) case—to the transmission of independent shell codes over each coherence block. Note that the resulting signaling scheme is noncoherent in that no pilot symbols are transmitted to learn the channel. Rather, information is conveyed through the choice of the subspace spanned by the row of each matrix, a quantity that is not affected by the fading. It is also worth remarking that the resulting bound assumes the adoption of an optimal receiver, able to compute the log-likelihood ratio of each codeword, which may be impractical. The auxiliary distribution used in [4] to compute the min-max converse is the one induced by USTM.

No asymptotic expansions of the form (1) are available for fading channels with no *a priori* CSI. Indeed, not even capacity is known in closed form in the ergodic setting. An attempt to analyze the scenario of imperfect CSI at the receiver for the case of multiple-input multiple-output (MIMO) Rayleigh block-fading channels was undertaken in [14]. The analysis, however, contains several inaccuracies.

For the multiple-antenna Rayleigh memoryless block-fading case, the input distribution achieving the RCEE was studied by Abou-Faycal and Hochwald [15]. They showed that it has the same structure as the ergodic-capacity-achieving input distribution [16]. Namely, the optimum input matrix is the product of a real, nonnegative, diagonal matrix and an isotropically distributed unitary matrix. Furthermore, for the SISO case, they proved that for large SNR, the real-valued component becomes deterministic, and the input vector becomes a shell code. The results in [15] were partly extended to single-antenna Rician memoryelss fading channels (coherence block of size one) in [17] where it is shown that the optimal scalar input has uniform phase and its amplitude is supported on a finite number of mass-points.

An upper bound on the packet error probability based on the RCEE was derived in [18] for the MIMO case using USTM as input distribution. Through numerical simulations, the authors showed that this bound is close to the DT bound obtained in [4] already at moderate error probabilities ($\epsilon \approx 10^{-4}$) in some scenarios.

Pilot-Assisted Transmission and Mismatched Decoding: Analyses of PAT schemes in which the channel estimate is treated as perfect by a decoder that operates according to the scaled nearest-neighbor (SNN) rule, fall into the general framework of mismatched decoding [19]–[23]. A study of the performance of SNN decoders over fading channels under different assumptions on the availability of CSI was presented in [23]. The analysis relies on using a Gaussian codebook and on the generalized mutual information (GMI)—an asymptotic quantity introduced in [19] that provides a lower bound on the maximum coding rate achievable for a fixed (possibly mismatched) decoding rule.[1] The rate gains obtainable with more sophisticated decoders that process jointly pilot and data symbols are studied in [24].

Nonasymptotic lower bounds on the maximum coding rate achievable with mismatch decoding are presented in [25] for

[1]The authors of [19] analyze also the performance achievable over quasi-static Rician and Nakagami fading channels for the case of perfect CSI and no CSI with both matched and mismatched decoders, using the cut-off rate as asymptotic performance metric.

the case of identical and independently distributed (i.i.d.), constant-composition, and cost-constrained codes. The analysis is based on the random-coding union bound with parameter s (RCUs) [26], an adaptation and relaxation of the random-coding union bound (RCU) in [6] for the case of mismatch decoder that recovers the generalized RCEE introduced in [19].

An analysis of the performance of PAT schemes using mutual information as asymptotic performance metric (and without imposing any restriction on the receiver structure) was carried out in [27] for the case of MIMO Rayleigh block-fading channels. There, it is shown that when one is allowed to optimize the power allocation between pilot and data symbols, it is optimal to use as many pilot symbols per coherence block as the number of transmit antennas. If instead pilot and data symbols need to be transmitted at the same power, the optimum number of pilot symbols becomes SNR dependent, and a number of pilot symbols much larger than the number of transmit antennas is needed in the low-SNR regime. This investigation has been generalized to MIMO Rician-fading channels in [28]. Finally, a comprehensive asymptotic analysis of the performance of SNN decoders (and generalizations thereof) over MIMO fading channels using GMI as performance metric can be found in [29].

Channel Codes for Short Packets: Recent developments in the design of efficient codes in the short blocklength regime lead to a variety of solutions enabling different trade-offs between decoding complexity and performance (i.e., coding gain). We refer the reader to recent surveys on the topic (see [30], [31]) for an extensive review of some of the most effective coding schemes. In this paper, we consider short quasi-cyclic (QC) binary block codes with good distance spectra obtained from tail-biting trellises. Furthermore, we focus on ordered-statistics decoding (OSD) [32], which is a general soft-decision decoding algorithm that can be applied to any binary linear block code. OSD is capable of achieving near-maximum likelihood (ML) decoding performance for codes of dimension up to several tens of bits with manageable decoding complexity. A few variants of the OSD algorithms have been proposed during the past two decades (see [33], [34]), enabling further remarkable performance savings at moderate blocklength with respect to the original algorithm presented in [32]. In this paper, we stick to the original algorithm due to the very short blocklengths under consideration.

B. Contributions

We study the maximum coding rate achievable over a SISO Rician memoryless block-fading channel under the assumption of no *a priori* CSI. The purpose of this paper is to adapt tools from finite-blocklength information theory to wireless channels of practical interest in order to provide design guidelines for low-complexity short-packet transmission.

Specifically, we present converse and achievability bounds on the maximum coding rate that generalize and tighten the bounds previously reported in [1] and [4]. As in [1] and [4], our converse bound relies on the min-max converse. Our two achievability bounds, which are built upon the RCUs bound, allow us to compare noncoherent and PAT schemes in terms of performance. Specifically, the first bound relies on the transmission of i.i.d. shell codes per coherence block and does not require explicit channel estimation at the receiver (while imposing no complexity constraint on the receiver architecture). The second one, which has a more practical flavor and has not been analyzed before in the literature (including in our previous contribution [1]), assumes PAT combined with shell codes for the transmission of the data symbols; furthermore, the receiver is constrained to perform ML channel estimation based on the pilot symbols followed by SNN detection. From a technical perspective, two critical steps allow us to obtain easy-to-evaluate expressions for our bounds. i) We obtain a compact expression for $\mathbb{E}\left[P_{\boldsymbol{Y}|\boldsymbol{X}}(\mathbf{y}|\boldsymbol{X})^{s}\right]$ where $P_{\boldsymbol{Y}|\boldsymbol{X}}$ denotes the Rician memoryless block-fading channel law, s is a positive real number, and $\boldsymbol{X}$ is shell-distributed, i.e., it is isotropically distributed and has constant modulus. ii) We account for the availability of imperfect CSI by transforming the Rician memoryless block-fading channel into an equivalent Rician channel whose parameters depend on the channel estimate and the estimation error.

Through a numerical investigation, we show that our converse and achievability bounds delimit tightly the maximum coding rate, for a large range of SNR and Rician κ-factor values, and allow one to identify—for given coding rate and packet size—the optimum number of coherence blocks to code over in order to minimize the energy per bit required to attain a target packet error probability.

Furthermore, our achievability bounds reveal that noncoherent transmission is more energy efficient than PAT, even when the number of pilot symbols and their power is optimized. For example, for the case when a coded packet of 168 symbols is transmitted using a channel code of rate 0.48 bits/channel use over a Rayleigh block-fading channel with block size equal to 8 symbols, the gap between the noncoherent and the PAT bound is about $1.2\,\mathrm{dB}$ at a packet error probability of 10^{-3}. This gap increases by a further $0.5\,\mathrm{dB}$ if pilot and data symbols are transmitted at the same power. When the power of the pilot symbols is optimized, one pilot symbol per coherence block turns out to suffice—a nonasymptotic counterpart of the result obtained in [27].

We finally design an actual PAT scheme based on punctured tail-biting QC codes and a decoder that, using OSD, performs SNN detection based on ML channel estimates. The performance of this coding scheme is remarkably close to what predicted by our PAT–SNN achievability bound: $1\,\mathrm{dB}$ gap at 10^{-3} packet error probability for a packet of 168 symbols, a code rate of 0.48 bit/channel use, and transmission over a Rayleigh-fading channel with coherence block of 24 symbols. This shows that our bound provides useful guidelines on the design of actual PAT schemes. We also discuss how the performance of the decoder can be further improved (without hampering its relatively low computational complexity) by accounting for the inaccuracy of the channel estimates.

Notation: Uppercase letters such as X and $\boldsymbol{X}$ are used to denote scalar random variables and vectors, respectively; their realizations are written in lowercase, e.g., x and $\mathbf{x}$. The identity matrix of size $a \times a$ is written as I_a. The distribution of a

circularly-symmetric complex Gaussian random variable with variance σ^2 is denoted by $\mathcal{CN}(0,\sigma^2)$. The superscript $(\cdot)^T$ and $(\cdot)^H$ denote transposition and Hermitian transposition, respectively, and $\odot$ is the Schur product, which is defined, for n-dimensional vectors $\mathbf{a}$ and $\mathbf{b}$, as $\mathbf{a}\odot\mathbf{b} = [a_1b_1,\ldots,a_nb_n]$. Furthermore, $\mathbf{0}_n$ and $\mathbf{1}_n$ stand for the all-zero and all-one vectors of size n, respectively. We write $\log(\cdot)$ and $\log_2(\cdot)$ to denote the natural logarithm and the logarithm to the base 2, respectively. Finally, $[a]^+$ stands for $\max\{0,a\}$, we use $\Gamma(\cdot)$ to denote the Gamma function, $I_\nu(z)$ the modified Bessel function of the first kind, $\|\cdot\|$ the l^2-norm, and $\mathbb{E}[\cdot]$ the expectation operator.

II. System Model

We consider a SISO Rician memoryless block-fading channel. Specifically, the random non-line-of-sight (NLOS) component is assumed to stay constant for n_{c} successive channel uses (which form a coherence block) and to change independently across coherence blocks. Coding is performed across ℓ such blocks; we shall refer to ℓ as the number of available *diversity branches*. The duration of each codeword (packet size) is, hence, $n = n_{\mathrm{c}}\ell$. This setup may be used to model, e.g., frequency-hopping systems, and is relevant for orthogonal frequency-division multiplexing (OFDM)-based systems (such as LTE and 5G), where a packet may consists of several resource blocks separated in frequency by more than the coherence bandwidth of the channel (see [18] for more details). The line-of-sight (LOS) component, i.e., the mean of the Rician fading random variable, which is assumed to be known at the receiver, stays constant over the duration of the entire packet (codeword). No *a priori* knowledge of the NLOS component is available at the receiver, in accordance to the no *a priori* CSI assumption.

Mathematically, the channel input-output relation can be expressed as

$$\boldsymbol{Y}_k = H_k\mathbf{x}_k + \boldsymbol{W}_k, \quad k=1,\ldots,\ell. \tag{2}$$

Here, $\mathbf{x}_k \in \mathbb{C}^{n_{\mathrm{c}}}$ and $\boldsymbol{Y}_k \in \mathbb{C}^{n_{\mathrm{c}}}$ contain the transmitted and received symbols within block k, respectively. The Rician fading is modeled by $H_k \sim \mathcal{CN}(\mu_{\mathrm{H}},\sigma_{\mathrm{H}}^2)$ where $\mu_{\mathrm{H}} = \sqrt{\kappa/(1+\kappa)}$ and $\sigma_{\mathrm{H}}^2 = (1+\kappa)^{-1}$ with κ being the Rician factor. Finally, $\boldsymbol{W}_k \sim \mathcal{CN}(\mathbf{0},\mathsf{I}_{n_{\mathrm{c}}})$ is the AWGN noise. The random variables $\{H_k\}$ and $\{\boldsymbol{W}_k\}$, which are mutually independent, are also independent over k.

We next define a channel code.

Definition 1: An $(\ell, n_{\mathrm{c}}, M, \epsilon, \rho)$-code for the channel (2) consists of

- An encoder $f : \{1,\ldots,M\} \to \mathbb{C}^{n_{\mathrm{c}}\ell}$ that maps the message J, which is uniformly distributed on $\{1,\ldots,M\}$ to a codeword in the set $\{\mathbf{c}_1,\ldots,\mathbf{c}_M\}$. Since each codeword $\mathbf{c}_m$, $m=1\ldots,M$, spans ℓ blocks, it is convenient to express it as a concatenation of ℓ subcodewords of dimension n_{c}

$$\mathbf{c}_m = [\mathbf{c}_{m,1},\ldots,\mathbf{c}_{m,\ell}]. \tag{3}$$

  We require that each subcodeword satisfies the average-power constraint [2]

$$\|\mathbf{c}_{m,k}\|^2 = n_{\mathrm{c}}\rho, \quad k=1,\ldots,\ell. \tag{4}$$

  Since the noise has unit variance, we can think of ρ as the average SNR per symbol.
- A decoder $g : \mathbb{C}^{n_{\mathrm{c}}\ell} \to \{1,\ldots,M\}$ satisfying an average error probability constraint

$$\frac{1}{M}\sum_{j=1}^{M}\Pr\{g(\boldsymbol{Y}^\ell) \neq J | J=j\} \leq \epsilon \tag{5}$$

  where $\boldsymbol{Y}^\ell = [\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_\ell]$ is the channel output induced by the codeword $\mathbf{x}^\ell = [\mathbf{x}_1,\ldots,\mathbf{x}_\ell] = f(j)$.

For given ℓ and n_c, ϵ, and ρ, the maximum coding rate R^*, measured in information bits per channel use, is defined as

$$R^*(\ell,n_{\mathrm{c}},\epsilon,\rho) = \sup\left\{\frac{\log_2 M}{\ell n_{\mathrm{c}}} : \exists(\ell,n_{\mathrm{c}},M,\epsilon,\rho)\text{-code}\right\}. \tag{6}$$

In words, for a fixed blocklength ℓn_{c} and a fixed SNR ρ, we seek the largest number M^* of codewords that can be transmitted with average error probability not exceeding ϵ. The maximum coding rate is then given by $R^* = (\log_2 M^*)/(\ell n_{\mathrm{c}})$.

In practical applications, we are often interested in the problem of minimizing the SNR ρ for a fixed packet error probability, a fixed blocklength ℓn_{c}, and a fixed number of information bits $\log_2 M$. This yields the following alternative optimization problem:

$$\rho^*(\ell,n_{\mathrm{c}},M,\epsilon) = \inf\{\rho : \exists(\ell,n_{\mathrm{c}},M,\epsilon,\rho)\text{-code}\}. \tag{7}$$

Throughout, we will repeatedly use that upper and lower bounds on R^* can be translated into lower and upper bounds on ρ^* and vice versa. Also, we will often express our results in terms of the minimum energy per bit E_{b}^*/N_0, which is related to ρ^* as

$$\frac{E_{\mathrm{b}}^*}{N_0}(\ell,n_{\mathrm{c}},M,\epsilon) = \frac{\ell n_{\mathrm{c}}}{\log_2 M}\,\rho^*(\ell,n_{\mathrm{c}},M,\epsilon). \tag{8}$$

III. Finite-Blocklength Bounds on R^*

We shall next present achievability and converse bounds on R^* obtained by using the nonasymptotic information-theoretic tools developed in [6] and [26]. In Section III-B, we provide an achievability bound that is based on the RCUs [26, Th. 1]. This bound does not require an explicit estimation of the fading channel at the receiver. Rather, it relies on a *noncoherent* transmission technique in which the message is encoded onto the one-dimensional subspace spanned by the input vector $\mathbf{x}_k$ in (2)—a quantity that is not affected by the fading process. This is achieved by using i.i.d. shell codes per coherence block. As we shall see, the optimum ML receiver computes the sum of functions of the inner product between the subcodewords and the corresponding received vectors; it then selects the codeword resulting in the largest sum. Note that this decoder does not require CSI.

[2]The per-subcodeword power constraint (4) implies (i.e., is more stringent than) the per-codeword power constraint $\|\mathbf{c}_m\|^2 = \ell n_{\mathrm{c}}\rho$, which is more commonly used in information-theoretic analyses. As we shall see, the per-subcodeword power constraint facilitates the computation of the converse bound.

In Section III-C, we provide a second achievability bound that relies instead on PAT. We assume that the receiver uses pilot symbols to obtain a ML estimate of the channel fading (we do not assume the fading law to be known at the receiver), which is then fed to a SNN decoder that treats it as perfect. This bound relies once more on the RCUs; furthermore, i.i.d. shell codes across the coherence blocks are used in the channel uses dedicated to the data symbols.

Since both bounds cannot be expressed in closed form and require Monte-Carlo simulation for their numerical evaluations (which may be time consuming for low values of ϵ), we present also easy-to-evaluate relaxations of these two bounds based on the generalized RCEE.

In order to investigate the potential gains attainable by using a PAT scheme in which the receiver is aware of the channel distribution, and accounts for the imperfect nature of the CSI, we develop in Section III-D a PAT-based achievability bound, where knowledge of the joint distribution between the fading process and its (pilot-based) estimate allows the decoder to operate according to the ML principle. This bound tightens the one presented in [1].

Finally, in Section III-E, we present a converse bound on R^* that relies on the min-max converse [6, Th. 27], with auxiliary distribution chosen as the distribution of $\{\boldsymbol{Y}_k\}$ induced by the transmission of independent shell codes over each coherence block. This bound generalizes to Rician-fading channels the one presented in [4] for the Rayleigh-fading case.

A. Achievability Bounds on R^*: Preliminaries

Throughout the paper, we shall assume that the decoder produces an estimate $\widehat{m}$ of the transmitted message as follows:

$$\widehat{m} = \arg\max_{m} q^{\ell}\big(\mathbf{c}_m, \mathbf{y}^{\ell}\big). \tag{9}$$

Here, $\{\mathbf{c}_m\}_{m=1}^{M}$ are the codewords and $\mathbf{y}^{\ell}$ is the received signal. Furthermore,

$$q^{\ell}(\mathbf{x}^{\ell}, \mathbf{y}^{\ell}) = \prod_{k=1}^{\ell} q(\mathbf{x}_k, \mathbf{y}_k) \tag{10}$$

where $q(\mathbf{x}_k, \mathbf{y}_k)$ is a bounded nonnegative function, which we refer to as *decoding metric*. In the next sections we will introduce the decoding metrics that are relevant for our achievability results. Before doing so, we review the RCUs bound and its connections to the generalized RCEE.

Theorem 1 (RCUs Bound [26, Th. 1]): For every input distribution $P_{\boldsymbol{X}^{\ell}}$ for which $\|\mathbf{x}_k\|^2 = n_{\mathrm{c}}\rho$, $k = 1, \ldots, \ell$, almost surely and every decoding metric $q(\cdot,\cdot)$, there exists a $(\ell, n_{\mathrm{c}}, M, \epsilon, \rho)$-code with decoder operating according to (9) and with average error probability upper-bounded as

$$\begin{aligned}\epsilon &\le \mathrm{RCUs}(\ell, n_{\mathrm{c}}, M, \rho)\\ &= \inf_{s\ge 0} \mathbb{E}\Big[e^{-\left[i_s^{\ell}\left(\boldsymbol{X}^{\ell}, \boldsymbol{Y}^{\ell}\right) - \log(M-1)\right]^{+}}\Big]\end{aligned} \tag{11}$$

where

$$i_s^{\ell}\big(\mathbf{x}^{\ell}, \mathbf{y}^{\ell}\big) = \log \frac{q^{\ell}\big(\mathbf{x}^{\ell}, \mathbf{y}^{\ell}\big)^{s}}{\mathbb{E}\big[q^{\ell}(\boldsymbol{X}^{\ell}, \mathbf{y}^{\ell})^{s}\big]} \tag{12}$$

is the *generalized information density*.

Assume now that the input distribution factorizes as

$$P_{\boldsymbol{X}^{\ell}}(\mathbf{x}^{\ell}) = \prod_{k=1}^{\ell} P_{\boldsymbol{X}}(\mathbf{x}_k) \tag{13}$$

i.e., the vector $\boldsymbol{X}^{\ell} = [\boldsymbol{X}_1, \ldots \boldsymbol{X}_{\ell}]$ has i.i.d. n_{c}-dimensional components $\{\boldsymbol{X}_k\}$ all distributed according to $P_{\boldsymbol{X}}$. It follows from (10) that the generalized information density in (12) can be rewritten as

$$i_s^{\ell}\big(\mathbf{x}^{\ell}, \mathbf{y}^{\ell}\big) = \sum_{k=1}^{\ell} \log \frac{q(\mathbf{x}_k, \mathbf{y}_k)^{s}}{\mathbb{E}[q(\boldsymbol{X}_k, \mathbf{y}_k)^{s}]} = \sum_{k=1}^{\ell} i_s(\mathbf{x}_k, \mathbf{y}_k). \tag{14}$$

Let now

$$E_0(\tau, s) = -\log \mathbb{E}\Big[e^{-\tau\, i_s(\boldsymbol{X}, \boldsymbol{Y})}\Big] \tag{15}$$

be the Gallager's function for mismatch decoding [19]. Here, $(\boldsymbol{X}, \boldsymbol{Y}) \sim P_{\boldsymbol{X}} P_{\boldsymbol{Y}|\boldsymbol{X}}$, where $P_{\boldsymbol{Y}|\boldsymbol{X}}$ is the channel law (within a coherence block) corresponding to the input-output relation (2). Furthermore, fix a rate $R > 0$ (measured for convenience in nats per channel use) and let

$$E(n_{\mathrm{c}}, R, \rho) = \sup_{s\ge 0, \tau \in [0,1]} \{E_0(\tau, s) - \tau n_{\mathrm{c}} R\} \tag{16}$$

be the generalized RCEE. It follows from [26] that

$$E(n_{\mathrm{c}}, R, \rho) = \sup_{s\ge 0} \liminf_{\ell\to\infty} -\frac{\log\big(\mathrm{RCUs}\big(\ell, n_{\mathrm{c}}, 2^{\ell n_{\mathrm{c}} R}, \rho\big)\big)}{\ell}. \tag{17}$$

In words, for fixed n_{c}, R, ρ, the RCUs bound decays to zero exponentially fast in ℓ, with exponent given by the generalized RCEE. An application of a Chernoff-type bound yields the following classic achievability bound based on the generalized RCEE. This bound is less tight than the RCUs bound in Theorem 1 but it is often easier to evaluate numerically.

Corollary 1 (Generalized RCEE Bound): For every $P_{\boldsymbol{X}}$ in (13) for which $\|\mathbf{x}\|^2 = n_{\mathrm{c}}\rho$ almost surely and every decoding metric $q(\cdot,\cdot)$ there exists a $(\ell, n_{\mathrm{c}}, M, \epsilon, \rho)$-code with decoder operating according to (9) and with average error probability upper-bounded as

$$\epsilon \le e^{-\ell E(n_{\mathrm{c}}, R, \rho)} \tag{18}$$

where $R = (\log M)/(n_{\mathrm{c}}\ell)$.

Note that the absence of a prefactor in (18) (compared to e.g., [11, eq. (7.3.21)]) is because $P_{\boldsymbol{X}}$ satisfies $\|\mathbf{x}\|^2 = n_{\mathrm{c}}\rho$ almost surely.

B. Noncoherent Achievability Bound on R^*

To derive our noncoherent achievability bound, we set

$$q(\mathbf{x}_k, \mathbf{y}_k) = P_{\boldsymbol{Y}|\boldsymbol{X}}(\mathbf{y}_k|\mathbf{x}_k). \tag{19}$$

Specifically, since $\boldsymbol{Y}_k$ is conditionally Gaussian given $\boldsymbol{X}_k = \mathbf{x}_k$, the ML decoding rule obtained by substituting (19) in (10) and then (10) in (9) can be rewritten, after some algebraic manipulations, as

$$\widehat{m} = \arg\max_{m} \sum_{k=1}^{\ell} \big(|\mathbf{y}_k^H \mathbf{c}_{m,k}|^2 + 2\sigma_{\mathrm{H}}^{-2} \Re\big(\mu_{\mathrm{H}} \mathbf{y}_k^H \mathbf{c}_{m,k}\big)\big). \tag{20}$$

$$S_k^s = (n_{\mathrm{c}}-2)\log(s) - \log\left(\frac{1+\sigma_{\mathrm{H}}^2 n_{\mathrm{c}}\rho}{\sigma_{\mathrm{H}}^2}\right) - \log(\Gamma(n_{\mathrm{c}})) - s\left(\|\boldsymbol{W}_k\|^2 - \|\widetilde{\boldsymbol{W}}_k\|^2\right) + \frac{s|\mu_{\mathrm{H}}|^2}{\sigma_{\mathrm{H}}^2} - \log\int_{\mathbb{R}_+} \frac{\exp\left(-s\left(\rho n_{\mathrm{c}} + \sigma_{\mathrm{H}}^{-2}\right) z\right)}{\left(\|\widetilde{\boldsymbol{W}}_k\|\sqrt{\rho n_{\mathrm{c}} z}\right)^{n_{\mathrm{c}}-1}} I_{n_{\mathrm{c}}-1}\left(2s\|\widetilde{\boldsymbol{W}}_k\|\sqrt{\rho n_{\mathrm{c}} z}\right) I_0\left(2s\sigma_{\mathrm{H}}^{-2}\sqrt{z|\mu_{\mathrm{H}}|^2}\right) \mathrm{d}z \quad (23)$$

Note that no CSI is required to compute (20). Next, we take $P_{\boldsymbol{X}}$ in (13) to be a shell distribution, i.e., the uniform distribution over all vectors $\mathbf{x} \in \mathbb{C}^{n_{\mathrm{c}}}$ satisfying the power constraint $\|\mathbf{x}\|^2 = n_{\mathrm{c}}\rho$ (see (4)). With these choices, the RCUs bound in Theorem 1, applied to the channel (2), takes the following form.

Theorem 2 (RCUs Noncoherent Achievability Bound): The maximum coding rate R^* in (6) achievable over the channel (2) is lower-bounded as

$$R^*(\ell, n_{\mathrm{c}}, \epsilon, \rho) \geq \max\left\{\frac{\log_2(M)}{n_{\mathrm{c}}\ell} : \epsilon_{\mathrm{ub}}(\ell, n_{\mathrm{c}}, M, \rho) \leq \epsilon\right\} \quad (21)$$

where

$$\epsilon_{\mathrm{ub}}(\ell, n_{\mathrm{c}}, M, \rho) = \inf_{s\geq 0} \mathbb{E}\left[e^{-\left[\sum_{k=1}^{\ell} S_k^s - \log(M-1)\right]^+}\right] \quad (22)$$

and S_k^s is given in (23), as shown at the top of this page. The $\{\boldsymbol{W}_k\}$ in (23) are defined as in (2) and

$$\widetilde{\boldsymbol{W}}_k = \begin{bmatrix} \mu_{\mathrm{H}}\sqrt{n_{\mathrm{c}}\rho} \\ \mathbf{0}_{n_{\mathrm{c}}-1} \end{bmatrix} + \begin{bmatrix} \sqrt{\sigma_{\mathrm{H}}^2 n_{\mathrm{c}}\rho + 1} \\ \mathbf{1}_{n_{\mathrm{c}}-1} \end{bmatrix} \odot \boldsymbol{W}_k. \quad (24)$$

Proof: See Appendix B. ∎

The random variables $\{S_k^s\}$ have the same distribution as the generalized information density $i_s(\boldsymbol{X}_k, \boldsymbol{Y}_k)$ in (14) with $q(\mathbf{x}_k, \mathbf{y}_k)$ chosen as in (19). By setting $\mu_{\mathrm{H}} = 0$, $\sigma_{\mathrm{H}}^2 = 1$, and $s = 1$ in (23) and (24), one recovers a SISO version of the achievability bound reported in [4, Th. 1] for the Rayleigh-fading case. The bound in [4, Th. 1] does not involve an optimization over the parameter s because it is based on the DT bound, which is less tight than the RCUs bound and coincides with it when $s = 1$.

Note that the expectation in (22) is not known in closed form, which makes the numerical evaluation of the bound demanding, especially for low values of ϵ. We next present an alternative noncoherent lower bound on R^* obtained by relaxing the RCUs to the RCEE in Corollary 1. Although less tight than the bound in Theorem 2, the resulting bound is easier to evaluate numerically.

Corollary 2 (RCEE Noncoherent Achievability Bound): The maximum coding rate R^* in (6) achievable over the channel (2) is lower-bounded as

$$R^*(\ell, n_{\mathrm{c}}, \epsilon, \rho) \geq \max\left\{\frac{\log_2(M)}{n_{\mathrm{c}}\ell} : \epsilon_{\mathrm{ub}}(\ell, n_{\mathrm{c}}, M, \rho) \leq \epsilon\right\} \quad (25)$$

where

$$\epsilon_{\mathrm{ub}}(\ell, n_{\mathrm{c}}, M, \rho) = e^{-\ell E(n_{\mathrm{c}}, R, \rho)} \quad (26)$$

with $R = (\log M)/(n_{\mathrm{c}}\ell)$ and

$$E(n_{\mathrm{c}}, R, \rho) = \max_{0\leq\tau\leq 1}\left\{E_0(\tau) - \tau n_{\mathrm{c}} R\right\}. \quad (27)$$

Here,

$$E_0(\tau) = -\log\left(c(\tau)\int_0^{\infty} r^{n_{\mathrm{c}}-1} e^{-r} J(r,\tau)^{1+\tau}\mathrm{d}r\right) \quad (28)$$

where

$$c(\tau) = \left(\frac{\left(1+\sigma_{\mathrm{H}}^2\rho n_{\mathrm{c}}\right)}{\Gamma(n_{\mathrm{c}})^{-1}}\right)^{\tau} e^{-|\mu_H|^2/\sigma_H^2}\left[\frac{(1+\tau)^{n_{\mathrm{c}}-2}}{\sigma_H^2}\right]^{1+\tau} \quad (29)$$

and

$$J(r,\tau) = \int_0^{\infty} \frac{e^{-\frac{1}{1+\tau}\left(\sigma_H^{-2}+\rho n_{\mathrm{c}}\right)z}}{\left(\sqrt{r\rho n_{\mathrm{c}} z}\right)^{n_{\mathrm{c}}-1}} \times I_{n_{\mathrm{c}}-1}\left(\frac{2\sqrt{r\rho n_{\mathrm{c}} z}}{1+\tau}\right) I_0\left(\frac{2|\mu_H|\sqrt{z}}{\sigma_H^2(1+\tau)}\right)\mathrm{d}z. \quad (30)$$

Proof: See Appendix C. ∎

By setting $\mu_{\mathrm{H}} = 0$ and $\sigma_{\mathrm{H}}^2 = 1$ in (29) and (30), one recovers a SISO version of the RCEE bound reported in [18, Th. 3] for the Rayleigh-fading case.

C. Pilot-Assisted Nearest-Neighbor Achievability Bound on R^*

We assume that, within each coherence block, n_{p} out of the available n_{c} channel uses are reserved for pilot symbols. The remaining $n_{\mathrm{d}} = n_{\mathrm{c}} - n_{\mathrm{p}}$ channel uses convey the data symbols. We further assume that each pilot symbol is transmitted at power ρ_{p}, and that the data symbol vectors $\mathbf{x}_k^{(\mathrm{d})} \in \mathbb{C}^{n_{\mathrm{d}}}$ satisfy the power constraint $\|\mathbf{x}_k^{(\mathrm{d})}\|^2 = n_{\mathrm{d}}\rho_{\mathrm{d}}$, $k = 1, \ldots, \ell$. We require that $n_{\mathrm{p}}\rho_{\mathrm{p}} + n_{\mathrm{d}}\rho_{\mathrm{d}} = n_{\mathrm{c}}\rho$ so as to fulfill (4).

The receiver uses the n_{p} pilot symbols available in each coherence block to perform an ML estimation of the corresponding fading coefficient. Specifically, for a given pilot vector $\mathbf{x}_k^{(\mathrm{p})}$ and a corresponding received-signal vector $\mathbf{y}_k^{(\mathrm{p})}$, the receiver computes the estimate

$$\widehat{h}_k = \left(\mathbf{x}_k^{(\mathrm{p})}\right)^H \mathbf{y}_k^{(\mathrm{p})} / \left\|\mathbf{x}_k^{(\mathrm{p})}\right\|^2. \quad (31)$$

It follows from (31) that, given $H_k = h_k$, we have $\widehat{H}_k \sim \mathcal{CN}(h_k, 1/(n_{\mathrm{p}}\rho_{\mathrm{p}}))$.

We further assume that the fading estimate $\hat{h}_k$ is fed to a SNN detector that treats it as perfect. Specifically, we consider the decoding metric

$$q(\mathbf{x}_k, \mathbf{y}_k) = e^{-\|\mathbf{y}_k^{(\mathrm{d})} - \widehat{h}_k \mathbf{x}_k^{(\mathrm{d})}\|^2} \quad (32)$$

where $\hat{h}_k$ is computed as in (31). Finally, we take as input distribution $P_{\boldsymbol{X}^{\mathrm{d}}}$ the uniform distribution over all vectors $\mathbf{x} \in \mathbb{C}^{n_{\mathrm{d}}}$ satisfying $\|\mathbf{x}\|^2 = n_{\mathrm{d}}\rho_{\mathrm{d}}$. For convenience, we let $\phi_{\mathrm{d}} = \sqrt{n_{\mathrm{d}}\rho_{\mathrm{d}}}$.

Under these assumptions, the RCUs bound in Theorem 1 takes the following form.

$$T_k^s = s\Big(\|\overline{\boldsymbol{W}}_k\|^2 - \|\widetilde{\boldsymbol{W}}_k\|^2\Big) + s\phi_{\mathrm{d}}^2|\widehat{H}_k|^2 + (n_{\mathrm{d}}-1)\log\Big(s|\widehat{H}_k|\|\overline{\boldsymbol{W}}_k\|\phi_{\mathrm{d}}\Big) - \log\Big(\Gamma(n_{\mathrm{d}})\, I_{n_{\mathrm{d}}-1}\Big(2s|\widehat{H}_k|\|\overline{\boldsymbol{W}}_k\|\phi_{\mathrm{d}}\Big)\Big) \tag{35}$$

$$J\big(r,\tau,s,\widehat{h}\big) = \frac{I_{n_{\mathrm{d}}-1}(2s|\widehat{h}|\phi_{\mathrm{d}}\sqrt{r})^{\tau} e^{|a(\widehat{h})|^2\left(\frac{\phi_{\mathrm{d}}^2}{1+\sigma_{\mathrm{p}}^2\phi_{\mathrm{d}}^2}-\frac{1}{\sigma_{\mathrm{p}}^2}\right)}}{\Gamma(n_{\mathrm{d}})^{-\tau}\,(s|\widehat{h}|\sqrt{r}\phi_{\mathrm{d}})^{\tau(n_{\mathrm{d}}-1)}} \int_0^{\infty} \frac{e^{-\left(\sigma_{\mathrm{p}}^{-2}+\phi_{\mathrm{d}}^2\right)z}}{\left(\sqrt{rz}\phi_{\mathrm{d}}\right)^{n_{\mathrm{d}}-1}} I_{n_{\mathrm{d}}-1}\big(2\sqrt{rz}\phi_{\mathrm{d}}\big)\, I_0\big(2|a(\widehat{h})|\sigma_{\mathrm{p}}^{-2}\sqrt{z}\big)\mathrm{d}z \tag{42}$$

Theorem 3 (RCUs–PAT–SNN Achievability Bound): Fix two nonnegative integers n_{p} ($n_{\mathrm{p}} < n_{\mathrm{c}}$) and $n_{\mathrm{d}} = n_{\mathrm{c}} - n_{\mathrm{p}}$, and two nonnegative real-valued parameters ρ_{p} and ρ_{d} satisfying $n_{\mathrm{p}}\rho_{\mathrm{p}} + n_{\mathrm{d}}\rho_{\mathrm{d}} = n_{\mathrm{c}}\rho$. The maximum coding rate R^* in (6) achievable over the channel (2) is lower-bounded as

$$R^*(\ell, n_{\mathrm{c}}, \epsilon, \rho) \geq \max\left\{\frac{\log_2(M)}{n_{\mathrm{c}}\ell} : \epsilon_{\mathrm{ub}}(\ell, n_{\mathrm{c}}, M, \rho) \leq \epsilon\right\} \tag{33}$$

where

$$\epsilon_{\mathrm{ub}}(\ell, n_{\mathrm{c}}, M, \rho) = \min_{s\geq 0} \mathbb{E}\left[e^{-\left[\sum_{k=1}^{\ell} T_k^s - \log(M-1)\right]^+}\right] \tag{34}$$

where T_k^s is given in (35), as shown at the top of this page, where

$$\overline{\boldsymbol{W}}_k = \begin{bmatrix} H_k\phi_{\mathrm{d}} \\ \mathbf{0}_{n_{\mathrm{d}}-1}\end{bmatrix} + \boldsymbol{W}_k \quad \text{and}$$
$$\widetilde{\boldsymbol{W}}_k = \begin{bmatrix} \sqrt{\phi_{\mathrm{d}}^2/(n_{\mathrm{p}}\rho_{\mathrm{p}})+1} \\ \mathbf{1}_{n_{\mathrm{d}}-1}\end{bmatrix} \odot \boldsymbol{W}_k \tag{36}$$

with $\boldsymbol{W}_k \sim \mathcal{CN}(\mathbf{0}_{n_{\mathrm{d}}}, \mathsf{I}_{n_{\mathrm{d}}})$. The expectation in (34) is with respect to the joint distribution $\prod_{k=1}^{\ell} P_{H_k,\widehat{H}_k,\boldsymbol{W}_k}$ where $P_{H_k,\widehat{H}_k,\boldsymbol{W}_k} = P_{H_k}P_{\widehat{H}_k|H_k}P_{\boldsymbol{W}_k}$ with $P_{H_k} = \mathcal{CN}(\mu_{\mathrm{H}}, \sigma_{\mathrm{H}}^2)$ and $P_{\widehat{H}_k|H_k=h} = \mathcal{CN}(h, 1/(n_{\mathrm{p}}\rho_{\mathrm{p}}))$.

Proof: See Appendix D. ∎

The random variables $\{T_k^s\}$ have the same distribution as the generalized information density $i_s(\boldsymbol{X}_k, \boldsymbol{Y}_k)$ in (14) with $q(\mathbf{x}_k, \mathbf{y}_k)$ chosen as in (32). As in Section III-B, we present an alternative, easier-to-compute achievability bound, which is obtained by relaxing the RCUs used in Theorem 3 to the generalized RCEE in Corollary 1.

Corollary 3 (RCEE–PAT–SNN Achievability Bound): Fix two nonnegative integers n_{p} ($n_{\mathrm{p}} < n_{\mathrm{c}}$) and $n_{\mathrm{d}} = n_{\mathrm{c}} - n_{\mathrm{p}}$, and two nonnegative real-valued parameters ρ_{p} and ρ_{d} satisfying $n_{\mathrm{p}}\rho_{\mathrm{p}} + n_{\mathrm{d}}\rho_{\mathrm{d}} = n_{\mathrm{c}}\rho$. The maximum coding rate R^* in (6) achievable over the channel (2) is lower-bounded as

$$R^*(\ell, n_{\mathrm{c}}, \epsilon, \rho) \geq \max\left\{\frac{\log_2(M)}{n_{\mathrm{c}}\ell} : \epsilon_{\mathrm{ub}}(\ell, n_{\mathrm{c}}, M, \rho) \leq \epsilon\right\} \tag{37}$$

where

$$\epsilon_{\mathrm{ub}}(\ell, n_{\mathrm{c}}, M, \rho) = \mathbb{E}\left[e^{-\ell E(n_{\mathrm{c}}, R, \rho, \widehat{H})}\right] \tag{38}$$

with $R = (\log M)/(n_{\mathrm{c}}\ell)$ and where the expectation is with respect to $P_{\widehat{H}} = \mathcal{CN}\big(\mu_{\mathrm{H}}, \sigma_{\mathrm{H}}^2 + 1/(n_{\mathrm{p}}\rho_{\mathrm{p}})\big)$. The error exponent $E(n_{\mathrm{c}}, R, \rho, \widehat{h})$ is

$$E(n_{\mathrm{c}}, R, \rho, \widehat{h}) = \max_{0\leq\tau\leq 1}\max_{s>0}\Big\{E_0(\tau, s, \widehat{h}) - \tau n_{\mathrm{c}} R\Big\} \tag{39}$$

and the Gallager's function for mismatch decoding $E_0(\tau, s, \widehat{h})$ is

$$E_0(\tau, s, \widehat{h}) = -\log c(\widehat{h}) \int_0^{\infty} r^{n_{\mathrm{d}}-1} e^{-r} J(r, \tau, s, \widehat{h})\mathrm{d}r \tag{40}$$

where $c(\widehat{h}) = \sigma_{\mathrm{p}}^{-2}\exp\Big(-\frac{|\mu_{\mathrm{p}}(\widehat{h})|^2\phi_{\mathrm{d}}^2}{1+\sigma_{\mathrm{p}}^2\phi_{\mathrm{d}}^2}\Big)$ with

$$\mu_{\mathrm{p}}(\widehat{h}) = \frac{\sigma_{\mathrm{H}}^2\widehat{h} + (n_{\mathrm{p}}\rho_{\mathrm{p}})^{-1}\mu_{\mathrm{H}}}{\sigma_{\mathrm{H}}^2 + (n_{\mathrm{p}}\rho_{\mathrm{p}})^{-1}}, \quad \sigma_{\mathrm{p}}^2 = \frac{\sigma_{\mathrm{H}}^2(n_{\mathrm{p}}\rho_{\mathrm{p}})^{-1}}{\sigma_{\mathrm{H}}^2 + (n_{\mathrm{p}}\rho_{\mathrm{p}})^{-1}}. \tag{41}$$

Furthermore, $J\big(r, \tau, s, \widehat{h}\big)$ is given in (42), as shown at the top of this page, with $a(\widehat{h}) = \mu_{\mathrm{p}}(\widehat{h}) - \widehat{h}s\tau\big(1 + \sigma_{\mathrm{p}}^2\phi_{\mathrm{d}}^2\big)$.

Proof: See Appendix E. ∎

D. Pilot-Assisted Maximum-Likelihood Achievability Bound on R^*

To assess the performance loss due to the mismatched SNN decoding metric (32), we present next a PAT-based achievability bound in which this metric is replaced by the ML metric

$$q(\mathbf{x}_k, \mathbf{y}_k) = P_{\boldsymbol{Y}^{(\mathrm{d})}|\boldsymbol{X}^{(\mathrm{d})},\hat{H}}(\mathbf{y}_k^{(\mathrm{d})}|\mathbf{x}_k^{(\mathrm{d})}, \hat{h}_k) \tag{43}$$

where $\hat{h}_k$ is the ML channel estimate (31). As argued in the proof of Corollary 3,

$$P_{\boldsymbol{Y}^{(\mathrm{d})}|\boldsymbol{X}^{(\mathrm{d})},\hat{H}}(\mathbf{y}_k^{(\mathrm{d})}|\mathbf{x}_k^{(\mathrm{d})}, \hat{h}_k) = \mathcal{CN}\Big(\mu_{\mathrm{p}}(\widehat{h}_k)\mathbf{x}_k^{(\mathrm{d})}, \boldsymbol{\Sigma}_k\Big) \tag{44}$$

where $\boldsymbol{\Sigma}_k = \sigma_{\mathrm{p}}^2\mathbf{x}_k^{(\mathrm{d})}(\mathbf{x}_k^{(\mathrm{d})})^H + \mathsf{I}_{n_{\mathrm{d}}}$, and $\mu_{\mathrm{p}}(\widehat{h}_k)$ and σ_{p}^2 are defined in (41). This implies that, given the channel estimate $\hat{h}_k$ and the input vector $\mathbf{x}_k^{(\mathrm{d})}$, the conditional probability density function (pdf) of $\boldsymbol{Y}_k^{(\mathrm{d})}$ coincides with the law of the following channel

$$\boldsymbol{Y}_k^{(\mathrm{d})} = Z_k\mathbf{x}_k^{(\mathrm{d})} + \boldsymbol{W}_k, \quad k = 1, \ldots, \ell. \tag{45}$$

Here, $Z_k \sim \mathcal{CN}\Big(\mu_{\mathrm{p}}(\widehat{h}_k), \sigma_{\mathrm{p}}^2\Big)$ and $\boldsymbol{W}_k \sim \mathcal{CN}(\mathbf{0}_{n_{\mathrm{d}}}, \mathsf{I}_{n_{\mathrm{d}}})$.

We see from (45) that we can account for the availability of the noisy CSI $\{\widehat{H}_k = \widehat{h}_k\}$ simply by transforming the Rician fading channel (2) into the equivalent Rician fading channel (45), whose LOS component is a random variable that depends on the channel estimates $\{\widehat{H}_k\}$. A lower bound on R^* for this setup can be readily obtained by assuming that each n_{d}-dimensional data vector is generated independently from a shell code, by applying Theorem 2 to each realization of $\{\widehat{H}_k\}$, and then by averaging over $\{\widehat{H}_k\}$.

Theorem 4 (RCUs–PAT–ML Achievability Bound): Fix two nonnegative integers n_{p} ($n_{\mathrm{p}} < n_{\mathrm{c}}$) and $n_{\mathrm{d}} = n_{\mathrm{c}} - n_{\mathrm{p}}$, and two nonnegative real-valued parameters ρ_{p} and ρ_{d} satisfying

$n_{\mathrm{p}}\rho_{\mathrm{p}} + n_{\mathrm{d}}\rho_{\mathrm{d}} = n_{\mathrm{c}}\rho$. The maximum coding rate R^* in (6) achievable over the channel (2) is lower-bounded as

$$R^*(\ell, n_{\mathrm{c}}, \epsilon, \rho) \geq \max\left\{ \frac{\log_2(M)}{n_{\mathrm{c}}\ell} : \epsilon_{\mathrm{ub}}(\ell, n_{\mathrm{c}}, M, \rho) \leq \epsilon \right\} \quad (46)$$

where

$$\epsilon_{\mathrm{ub}}(\ell, n_{\mathrm{c}}, M, \rho) = \min_{s\geq 0} \mathbb{E}\left[e^{-\left[\sum_{k=1}^{\ell} \bar{S}_k^s(\widehat{H}_k) - \log(M-1)\right]^+} \right]. \quad (47)$$

The expectation in (47) is with respect to $\prod_{k=1}^{\ell} P_{\widehat{H}_k} P_{\boldsymbol{W}_k}$ where $P_{\widehat{H}_k} = \mathcal{CN}\big(\mu_{\mathrm{H}}, \sigma_{\mathrm{H}}^2 + (n_{\mathrm{p}}\rho_{\mathrm{p}})^{-1}\big)$ and $P_{\boldsymbol{W}_k} \sim \mathcal{CN}(\mathbf{0}_{n_{\mathrm{d}}}, \mathsf{I}_{n_{\mathrm{d}}})$. The random variables $\{\bar{S}_k^s(\widehat{H}_K)\}$ are defined similarly as in (23) with the difference that n_{c}, ρ, μ_{H} and σ_{H}^2 in (23) are replaced by n_{d}, ρ_{d}, $\mu_{\mathrm{p}}(\widehat{H}_k)$ and σ_{p}^2, respectively.

Given $\widehat{H}_k = \widehat{h}_k$, the random variables $\{\bar{S}_k^s(\widehat{h}_k)\}$ have the same conditional distribution as the information density $i_s(\boldsymbol{X}_k, \boldsymbol{Y}_k)$ in (14) with $q(\mathbf{x}_k, \mathbf{y}_k)$ chosen as in (43). For the case $n_{\mathrm{p}} = 0$, the pilot-based achievability bound in Theorem 4 coincides with the noncoherent bound given in Theorem 2. Furthermore, by setting $\rho_{\mathrm{d}} = \rho_{\mathrm{p}}$ and $s = 1$, we recover [1, Th. 3].[3] The bound in Theorem 4 can be relaxed to a generalized-RCEE-type bound by proceeding as in the proof of Corollary 2.

E. A Converse Bound on R^*

We next state our converse bound.[4]

Theorem 5 (Min-Max Converse Bound): The maximum coding rate R^* in (6) achievable over the channel (2) is upper-bounded as

$$R^*(\ell, n_{\mathrm{c}}, \epsilon, \rho) \leq \inf_{\lambda\geq 0} \frac{1}{\ell n_{\mathrm{c}}} \left(\lambda - \log\left[\Pr\left\{ \sum_{k=1}^{\ell} S_k^1 \leq \lambda \right\} - \epsilon \right]^+ \right) \quad (48)$$

where the random variables $\{S_k^1\}_{k=1}^{\ell}$ are obtained by setting $s = 1$ in (23).

Proof: See Appendix F. ∎

By setting $\mu_H = 0$ and $\sigma_H^2 = 1$, one recovers a SISO version of the min-max converse bound obtained in [4] for the Rayleigh-fading case.

As pointed out in Appendix F, imposing the per-subcodeword power constraint (4) is instrumental to obtain the converse bound (48). If we replace (4) with a less stringent per-codeword power constraint, we can obtain, for the choice of auxiliary output distribution discussed in Appendix F, a numerically computable converse bound only if we replace the average error probability constraint in (5) with the more stringent maximum error probability constraint

$$\max_j \Pr\{g(\boldsymbol{Y}^\ell) \neq J | J = j\} \leq \epsilon. \quad (49)$$

The resulting bound differs from (48) in that the random variable S_k^1 depends on the power allocated over the kth subcodeword, $k = 1, \ldots, \ell$, and one has to maximize over all possible power allocations across subcodewords that satisfy the overall per-codeword power constraint. Unfortunately, no closed-form expression is available for the solution of this $\ell - 1$ nonconvex optimization problem, which makes the numerical computation of the resulting bound difficult. Numerical experiments conducted for small values of ℓ and for the SNR and ϵ values detailed in Section IV resulted in an optimal power allocation that is uniform across subcodewords. Note that whenever the optimal power allocation is uniform, this new bound coincides with the one given in (48) under the more restrictive per-subcodeword power constraint. This suggests that the simplifying assumption of a per-subcodeword power constraint used in this paper has a negligible impact.

IV. Numerical Results

A. Dependency of R^* and E_{b}^*/N_0 on the Rician Factor κ

In Fig. 1, we plot the RCUs noncoherent achievability bound (Theorem 2), its RCEE relaxation (Corollary 2), and the min-max converse bound (Theorem 5). We assume a blocklength of $n = 168$ channel uses and a packet error probability of $\epsilon = 10^{-3}$. In Fig. 1a, we set $\rho = 6\,\mathrm{dB}$ and investigate the dependency of R^* on the number of diversity branches ℓ or, equivalently, on the size of each coherence block n_{c}. In Fig. 1b, we investigate instead, for a fixed rate $R = 0.48$ bit/channel use (and, hence, a fixed number of information bits, since $n = 168$), the minimum energy per bit E_{b}^*/N_0 in (8) needed to achieve $\epsilon = 10^{-3}$.

We see from Fig. 1 that the bounds are tight and allow one to identify the optimal number of diversity branches that maximizes R^* or, equivalently, minimizes E_{b}^*/N_0. For $\kappa = 0$ (Rayleigh-fading) this number is $\ell^* \approx 21$. When $\ell < \ell^*$, the performance bottleneck is the limited diversity available. When $\ell > \ell^*$, the limiting factor is instead the fast channel variations (which manifest themselves in a small coherence block n_{c}). We note also that, as κ increases, both R^* and E_{b}^*/N_0 become less sensitive to ℓ. This is expected since, when $\kappa \to \infty$, the Rician channel converges to a nonfading AWGN channel. Indeed, we see that the bounds obtained for the case $\kappa = 10^3$ are in good agreement with the normal approximation (1). Note also that the agreement with the normal approximation is better for smaller values of ℓ. This is because, in the AWGN case, the optimum input distribution involves shell codes over $\mathbb{C}^n$, whereas our bounds rely on shell codes over $\mathbb{C}^{n_{\mathrm{c}}}$.

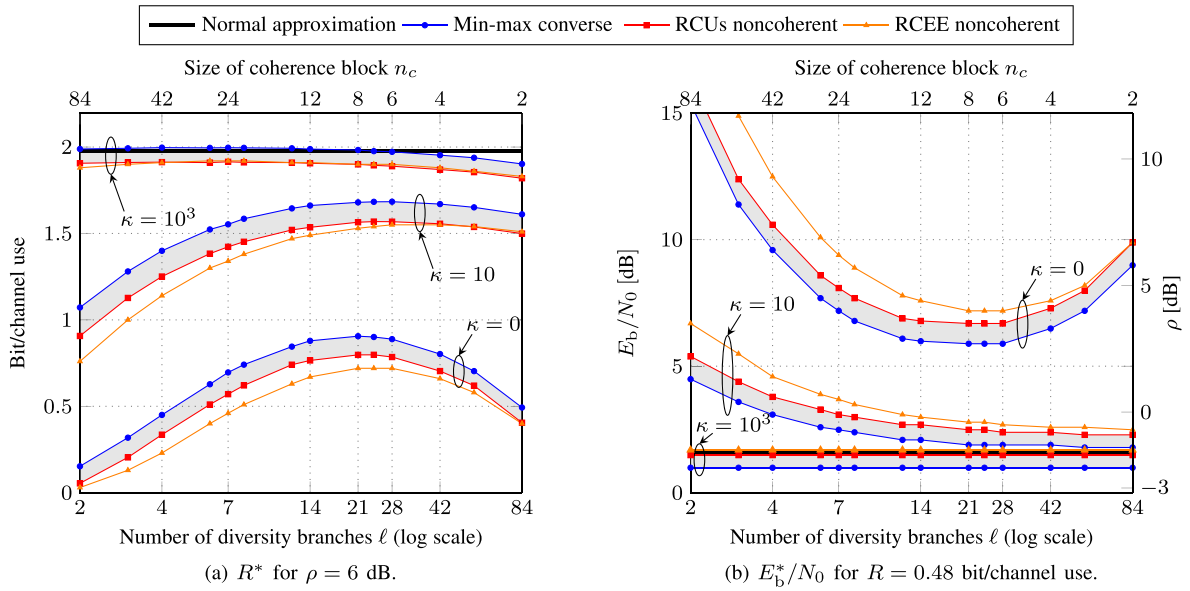
As expected, the RCUs bound is tighter than the RCEE bound, which is however easier to evaluate numerically.

B. PAT or Noncoherent?

In Fig. 2, we compare the RCUs noncoherent achievability bound (Theorem 2) with the RCUs–PAT–SNN achievability bound (Theorem 3). This last bound is computed for different numbers of pilot symbols n_{p}. We consider both the case in which pilot and data symbols are transmitted at the same power ($\rho_{\mathrm{p}} = \rho_{\mathrm{d}}$) and the case in which the power allocation is optimized. The min-max converse (Theorem 5) is also depicted for reference. The parameters are $n = 168$, $\epsilon = 10^{-3}$, $R = 0.48$ bit/channel use, and $\kappa \in \{0, 10\}$. For the case $\rho_{\mathrm{p}} = \rho_{\mathrm{d}}$, we see that the optimum number of pilot symbols decreases as the

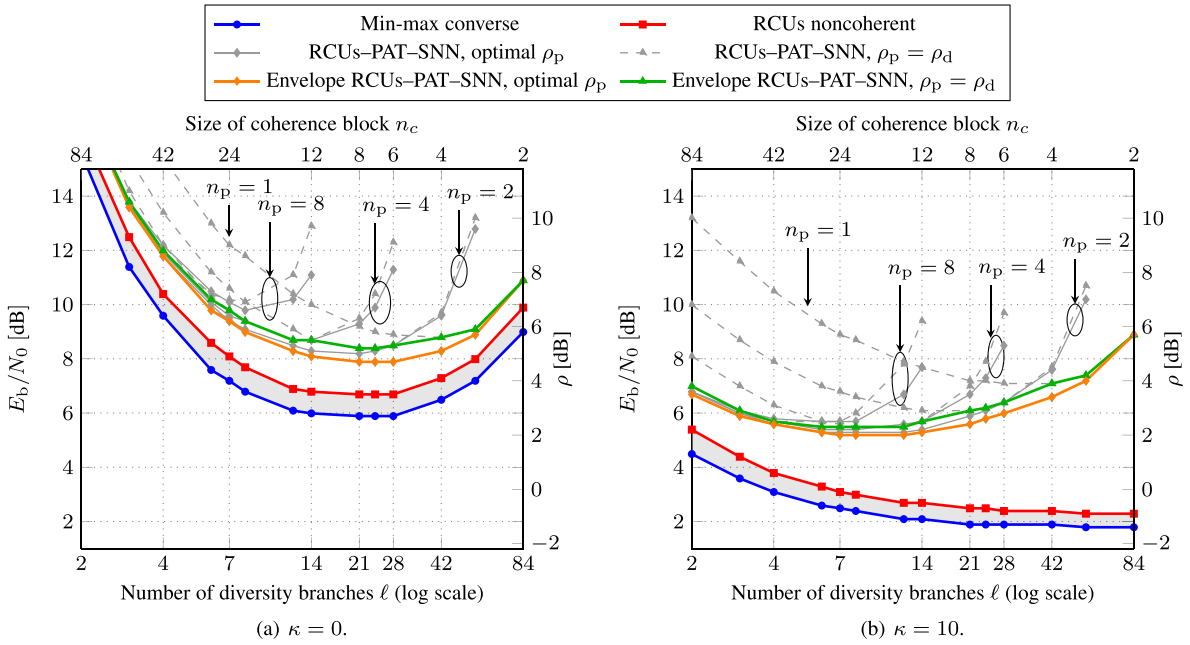
[3] With $(M-1)/2$ replaced by $M-1$.

[4] This bound was first presented in the conference version of this paper [1, Th. 2].



(a) R^* for $\rho = 6$ dB. (b) E_b^*/N_0 for $R = 0.48$ bit/channel use.

Fig. 1. RCUs noncoherent achievability bound (Theorem 2), its RCEE relaxation (Corollary 2), and min-max converse (Theorem 5); $\kappa \in \{0, 10, 1000\}$, $\epsilon = 10^{-3}$ and $n = 168$.

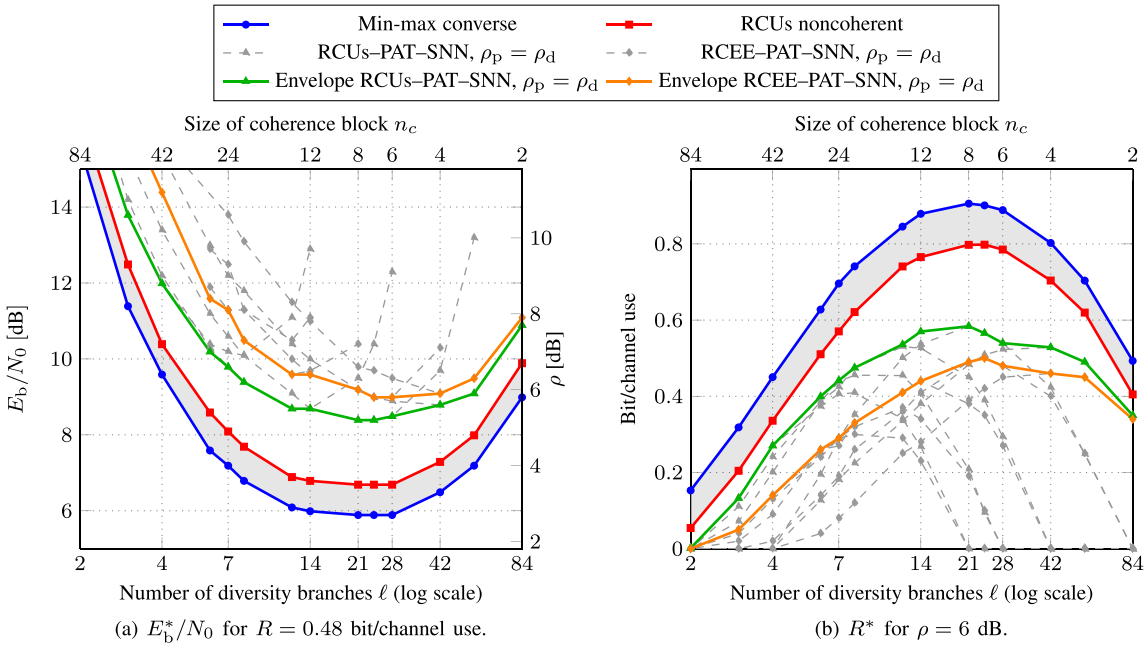


(a) $\kappa = 0$. (b) $\kappa = 10$.

Fig. 2. E_b^*/N_0 for $n = 168$, $\epsilon = 10^{-3}$ and $R = 0.48$ bit/channel use; min-max converse (Theorem 5), RCUs noncoherent achievability bound (Theorem 2), and RCUs–PAT–SNN achievability bound (Theorem 3). The dashed lines are obtained by assuming $\rho_d = \rho_p$; the solid lines are obtained by optimizing over the power allocation.

size n_c of the coherence block decreases, as expected. Indeed, when the coherence block is small, the rate penalty resulting for increasing the number of pilot symbols overcomes the rate gain resulting from the more accurate channel estimation. When one performs an optimization over the power allocation, however, one pilot symbol per coherence block suffices (the curve for $n_p = 1$ overlaps with the corresponding envelope in both Fig. 2a and Fig. 2b). This is in agreement with what was proven in [27, Th. 3] using mutual information as asymptotic performance metric. Furthermore, the optimum power allocation turns out to follow closely the asymptotic rule provided in [27, Th. 3].

We see from both Fig. 2a and Fig. 2b that the noncoherent achievability bound outperforms the RCUs–PAT–SNN achievability bounds. For example, when $\kappa = 0$ and $\ell = 28$, the gap between the RCUs noncoherent achievability bound and the RCUs–PAT–SNN achievability bound with optimum power allocation is about 1.2 dB. This gap increases further by 0.6 dB if the additional constraint $\rho_p = \rho_d$ is imposed. For the same ℓ, when $\kappa = 10$, the gap between the RCUs noncoherent achievability bound and the RCUs–PAT–SNN achievability bound with optimum power allocation is about 3.6 dB, with an additional 0.4 dB if the constraint $\rho_p = \rho_d$ is imposed.



(a) E_b^*/N_0 for $R = 0.48$ bit/channel use. (b) R^* for $\rho = 6$ dB.

Fig. 3. Comparison between RCUs–PAT–SNN (Theorem 3) and RCEE–PAT–SNN (Corollary 3) for $\kappa = 0$, $n = 168$, and $\epsilon = 10^{-3}$ with $\rho_d = \rho_p$. The min-max converse (Theorem 5) and the RCUs noncoherent bound (Theorem 2) are included for reference.

Note also that when $\kappa = 10$ and $n_c = 2$, transmitting even just a single pilot symbol yields a significant performance reduction, because of the loss in spectral efficiency.

In Fig. 3, we compare the PAT–RCUs–SNN achievability bound (Theorem 3) with its RCEE relaxation (Corollary 3) for the case $\rho_d = \rho_p$. We see that for $\ell = 28$, the gap between the bounds is about 0.5 dB.

C. Practical PAT Coding Schemes

We discuss next the design of actual PAT-based coding schemes with moderate decoding complexity. We shall focus for simplicity on the case $\ell = 7$ and $n_c = 24$. Furthermore, we assume that 81 information bits need to be transmitted in each codeword, which yields $R \approx 0.48$ bit/channel use. We allocate n_p channel uses per coherence block to pilot symbols, and use the remaining $(24 - n_p)$ channel uses to carry coded symbols, which belong to a quaternary phase shift keying (QPSK) constellation. Similar to [18], we select a $(324, 81)$ binary QC code and puncture a suitable number of codeword bits to accommodate the pilot symbols within the prescribed 168 channel uses. The code is obtained by tail-biting termination of a rate$-1/4$ nonsystematic convolutional code with memory 14 and generators $[47633\ 57505\ 66535\ 71145]$ in octal notation [35, Table 10.14]. After encoding, a pseudo-random interleaving is applied to the codeword bits, followed by puncturing. For the chosen parameters, the number of punctured bits is $14n_p - 12$ and the blocklength after puncturing (expressed this time in *real* rather than *complex* channel uses) is $336 - 14n_p$. For the $(322, 81)$ punctured code, we evaluated the weight enumerator function (WEF) using the method described in [36]. The lower part of the WEF is $A(X) = 1 + 60X^{36} + 275X^{37} + 421X^{38} + 586X^{39} + \ldots$. Hence, the $(322, 81)$ code has minimum distance 36 with a multiplicity of minimum weight codewords equal to 60. At the receiver side, the pilot symbols are used to perform ML channel estimation according to (31). The bit-wise log-likelihood ratio (LLR) are computed by assuming the estimates $\widehat{h}_k$, $k = 1, \ldots, 7$, to be perfect. Decoding is then performed via OSD [32]. The order of OSD is set to $t = 3$, which provides a reasonable trade-off between performance and decoding complexity. The OSD builds a list $\mathcal{L}$ of $1 + \sum_{i=1}^{t} \binom{81}{i} = 88642$ channel input vectors corresponding to candidate codewords, out of which the decision is obtained as

$$\widehat{\mathbf{x}} = \arg\max_{\mathbf{x} \in \mathcal{L}} \prod_{k=1}^{\ell} \exp\left(-\|\mathbf{y}_k^{(d)} - \widehat{h}_k \mathbf{x}_k\|^2\right) \tag{50}$$

where $\mathbf{x}_k$ denotes the vector of coded QPSK symbols transmitted over the kth coherence interval. We shall refer to the decoder operating according to this rule as OSD–SNN. When the list $\mathcal{L}$ includes all input vectors corresponding to valid codewords, the decoding rule (50) is equivalent to SNN in (32). We also analyze a second scheme, in which a re-estimation of the fading channel is performed by using the initial OSD decision $\widehat{\mathbf{x}}$. Specifically, $\widehat{\mathbf{x}}$ is used to update the ML channel estimates, yielding new bit-wise LLR. A second OSD attempt is then performed with the updated input. We refer to this second scheme as OSD with re-estimation (OSD–REE).

In Fig. 4, we compare the performance of the OSD–SNN coding scheme to what is predicted by the PAT–RCUs–SNN achievability bound (Theorem 3) for different values of n_p, for the case $\rho_p = \rho_d$. We see that the gap is within 1 dB for all values of n_p considered here. This shows that the performance reference provided by the PAT–RCUs–SNN achievability bound is accurate. For the parameters considered in Fig. 4, setting $n_p = 4$

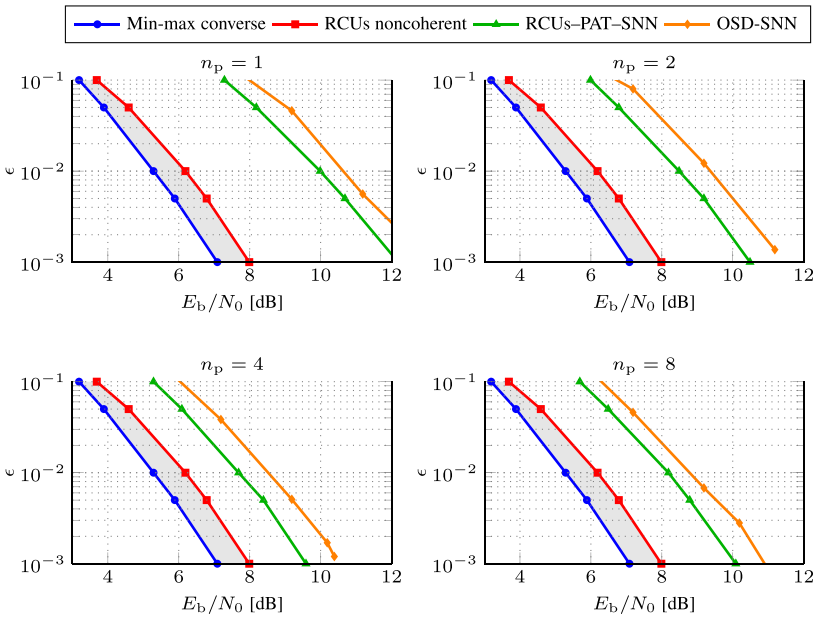


Fig. 4. Performance of the OSD–SNN coding scheme for $n_\mathrm{p} = \{1, 2, 4, 8\}$; the RCUs–PAT–SNN (Theorem 3), the min-max converse (Theorem 5), and the RCUs noncoherent bound (Theorem 2) are also plotted for reference; $n_\mathrm{c} = 24$, $\ell = 7$, $R = 0.48$ bit/channel use, and $\kappa = 0$.

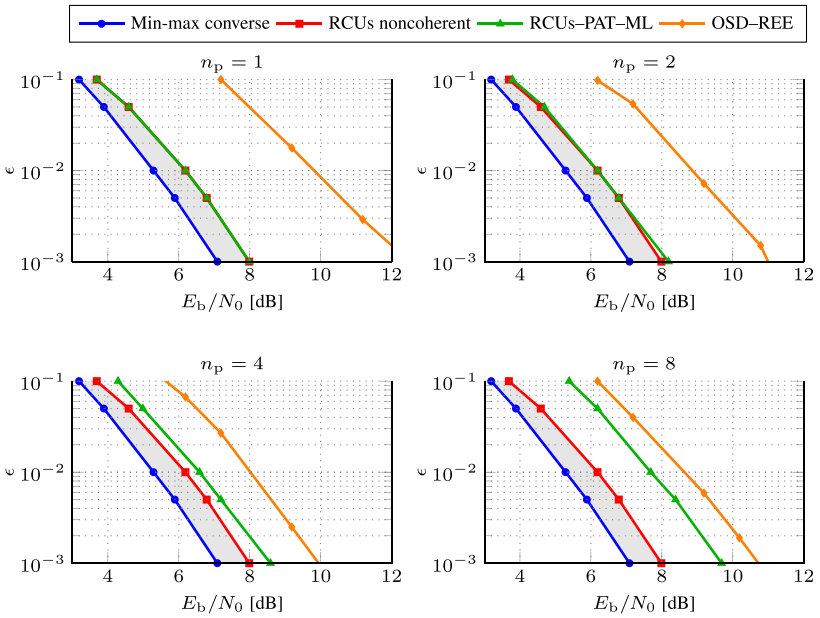


Fig. 5. Performance of the OSD–REE coding scheme for $n_\mathrm{p} = \{1, 2, 4, 8\}$; the RCUs–PAT–ML bound (Theorem 4), the min-max converse (Theorem 5), and the RCUs noncoherent bound (Theorem 2) are plotted for reference; $n_\mathrm{c} = 24$, $\ell = 7$, and $R = 0.48$ bit per channel use, and $\kappa = 0$.

yields the best performance, as predicted by the PAT–RCUs–SNN bound.

In Fig. 5, we compare the performance of the OSD–REE coding scheme with what is predicted by the RCUs–PAT–ML achievability bound in Theorem 4. This bound is relevant since the OSD–REE coding scheme improves on the SNN decoding rule by allowing decision-driven channel re-estimation. The gap between the bound and the code performance is now larger: about 1.3 dB for $\epsilon = 10^{-3}$ and $n_\mathrm{p} = 4$. This is due to the fact that the RCUs–PAT–ML achievability bound assumes ML decoding, which yield too optimistic performance estimates. Comparing Figs. 4 and 5, we see that the performance gains of the OSD–REE coding scheme over the OSD–SNN one are limited to fractions of dBs, e.g., for $n_\mathrm{p} = 4$ and $\epsilon = 10^{-3}$, the gain is about 0.5 dB.

V. Conclusion

We presented bounds on the maximum coding rate achievable over a SISO Rician memoryless block-fading channel under the assumption of no *a priori* CSI. Specifically, we presented converse and achievability bounds on the maximum coding rate that generalize and tighten the bounds previously reported

in [1] and [4]. Our two achievability bounds, built upon the RCUs bound, allow one to compare the performance of noncoherent and PAT schemes. As in [1] and [4] our converse bound relies on the min-max converse.

Through a numerical investigation, we showed that our converse and achievability bounds delimit tightly the maximum coding rate, for a large range of SNR and Rician κ-factor values, and allow one to identify—for given coding rate and packet size—the optimum number of coherence blocks to code over in order to minimize the energy per bit required to attain a target packet error probability.

Furthermore, our achievability bounds reveal that noncoherent transmission is more energy efficient than PAT even when the number of pilot symbols and their power is optimized.[5] When the power of the pilot symbols is optimized, one pilot symbol per coherence block turns out to suffice—a nonasymptotic counterpart of the result obtained in [27].

We finally designed an actual PAT scheme based on punctured tail-biting QC codes and a decoder that, using OSD, performs SNN detection based on ML channel estimates. A comparison between the PAT scheme and our bounds reveals that the bounds provide accurate guidelines on the design of actual PAT schemes. We also discussed how the performance of the decoder can be further improved (without hampering its relatively low computational complexity) by accounting for the inaccuracy of the channel estimates via joint processing of pilot and data symbols. Developing nonasymptotic information theoretic bounds for this setting is an open problem.

An important final remark is that our comparison between noncoherent and PAT schemes is somewhat biased towards the noncoherent case. Indeed, our RCUs noncoherent bound relies on ML decoding (which implies also knowledge of the fading law and requires solving (20), which is unfeasible when the number of information bits is larger than a few tens because of complexity), whereas the performance predicted by the RCUs–PAT–SNN bound can be approached using a low-complexity SNN decoder, which does not require knowledge of the fading law. Designing low-complexity noncoherent coding schemes able to approach our RCUs noncoherent bound is an important open issue.

APPENDIX

A. Auxiliary Lemmas

We state next two lemmas that will be useful for proving our achievability and converse bounds on R^*.

Lemma 1: Let $\boldsymbol{X}$ be an isotropically distributed vector in $\mathbb{C}^{n_\mathrm{c}}$ with norm equal to $\sqrt{\rho n_\mathrm{c}}$, let $H \sim \mathcal{CN}\big(\mu_\mathrm{H}, \sigma_\mathrm{H}^2\big)$, and let $\boldsymbol{W} \sim \mathcal{CN}\big(0, \sigma_\mathrm{w}^2 \mathsf{I}_{n_\mathrm{c}}\big)$ Furthermore, let $\boldsymbol{Y} = H\boldsymbol{X} + \boldsymbol{W}$. The conditional pdf of $\boldsymbol{Y}$ given $H = h$ is

$$P_{\boldsymbol{Y}|H}(\mathbf{y}|h) = \frac{\Gamma(n_\mathrm{c}) \exp\Big(-\frac{\|\mathbf{y}\|^2+|h|^2\rho n_\mathrm{c}}{\sigma_\mathrm{w}^2}\Big)}{\pi^{n_\mathrm{c}}\sigma_\mathrm{w}^2\big(\|\mathbf{y}\||h|\sqrt{\rho n_\mathrm{c}}\big)^{n_\mathrm{c}-1}} \times I_{n_\mathrm{c}-1}\left(\frac{2\|\mathbf{y}\||h|\sqrt{\rho n_\mathrm{c}}}{\sigma_\mathrm{w}^2}\right). \quad (51)$$

[5]We limit our comparison to the two achievability bounds because no tight converse bound for the PAT case is available, even asymptotically.

Proof: Under the assumptions of Lemma 1, the random variable $(\sigma_\mathrm{w}^2/2)\|\mathbf{y}\|^2$ follows (given h) a noncentral χ-squared distribution with $2n_\mathrm{c}$ degrees of freedom and noncentrality parameter $2|h|^2 n_\mathrm{c}\rho/\sigma_\mathrm{w}^2$. Furthermore, the output vector $\mathbf{y}$ is isotropically distributed. We then obtain (51) by recalling that the surface area of an n_c-dimensional complex sphere of radius $\sqrt{n_\mathrm{c}\rho}$ is

$$\frac{2\pi^{n_\mathrm{c}}(\sqrt{n_\mathrm{c}\rho})^{2n_\mathrm{c}-1}}{\Gamma(n_\mathrm{c})}. \quad (52)$$

■

Lemma 2: Under the assumptions of Lemma 1, the pdf of $\boldsymbol{Y}$ is

$$P_{\boldsymbol{Y}}(\mathbf{y}) = \frac{\Gamma(n_\mathrm{c})\, e^{-\frac{\|\mathbf{y}\|^2}{\sigma_\mathrm{w}^2}-\frac{|\mu_\mathrm{H}|^2}{\sigma_\mathrm{H}^2}}}{\pi^{n_\mathrm{c}}\sigma_\mathrm{w}^2\sigma_\mathrm{H}^2} \int_0^\infty \frac{e^{-z\left(\frac{\rho n_\mathrm{c}}{\sigma_\mathrm{w}^2}+\frac{1}{\sigma_\mathrm{H}^2}\right)}}{\big(\|\mathbf{y}\|\sqrt{\rho n_\mathrm{c} z}\big)^{n_\mathrm{c}-1}} \times I_{n_\mathrm{c}-1}\left(\frac{2\|\mathbf{y}\|\sqrt{\rho n_\mathrm{c} z}}{\sigma_\mathrm{w}^2}\right) I_0\left(\frac{2|\mu_\mathrm{H}|\sqrt{z}}{\sigma_\mathrm{H}^2}\right) \mathrm{d}z. \quad (53)$$

Proof: We obtain (53) by averaging (51) over $|H|^2$, which has pdf

$$P_{|H|^2}(z) = \frac{\exp\Big(-\frac{1}{\sigma_\mathrm{H}^2}\big(z+|\mu_\mathrm{H}|^2\big)\Big)}{\sigma_\mathrm{H}^2} I_0\left(\frac{2|\mu_\mathrm{H}|\sqrt{z}}{\sigma_\mathrm{H}^2}\right). \quad (54)$$

■

B. Proof of Theorem 2

We let $\boldsymbol{X}_k = \sqrt{n_\mathrm{c}\rho}\boldsymbol{U}_k$ where $\{\boldsymbol{U}_k\}_{k=1}^{\ell}$ are independent and isotropically distributed unitary vectors in $\mathbb{C}^{n_\mathrm{c}}$. For the chosen decoding metric (19), the generalized information density in (12) can be decomposed as

$$i_s^\ell\big(\mathbf{u}^\ell, \mathbf{y}^\ell\big) = \sum_{k=1}^{\ell} i_s(\mathbf{u}_k, \mathbf{y}_k) = \sum_{k=1}^{\ell} \log \frac{P_{\boldsymbol{Y}|\boldsymbol{U}}(\mathbf{y}_k|\mathbf{u}_k)^s}{\mathbb{E}\big[P_{\boldsymbol{Y}|\boldsymbol{U}}(\mathbf{y}_k|\boldsymbol{U}_k)^s\big]} \quad (55)$$

where

$$P_{\boldsymbol{Y}|\boldsymbol{U}=\mathbf{u}_k} = \mathcal{CN}(\mu_\mathrm{H}\sqrt{n_\mathrm{c}\rho}\mathbf{u}_k, \boldsymbol{\Sigma}_k) \quad (56)$$

with $\boldsymbol{\Sigma}_k = \mathsf{I}_{n_\mathrm{c}} + \sigma_\mathrm{H}^2 n_\mathrm{c}\rho\mathbf{u}_k\mathbf{u}_k^H$. To evaluate the expected value in (55), it is convenient to express $P_{\boldsymbol{Y}|\boldsymbol{U}}(\mathbf{y}_k|\mathbf{u}_k)^s$ as a scalar times a Gaussian pdf as follows:

$$P_{\boldsymbol{Y}|\boldsymbol{U}}(\mathbf{y}_k|\mathbf{u}_k)^s = \big(\pi^{n_\mathrm{c}}\det(\boldsymbol{\Sigma}_k)\big)^{1-s} s^{-n_\mathrm{c}} P_{\widetilde{\boldsymbol{Y}}|\boldsymbol{U}}(\mathbf{y}_k|\mathbf{u}_k) \quad (57)$$

$$= \big(\pi^{n_\mathrm{c}}\big(1+\rho n_\mathrm{c}\sigma_\mathrm{H}^2\big)\big)^{1-s} s^{-n_\mathrm{c}} P_{\widetilde{\boldsymbol{Y}}|\boldsymbol{U}}(\mathbf{y}_k|\mathbf{u}_k) \quad (58)$$

where $P_{\widetilde{\boldsymbol{Y}}|\boldsymbol{U}=\mathbf{u}_k} = \mathcal{CN}\big(\mu_\mathrm{H}\sqrt{n_\mathrm{c}\rho}\mathbf{u}_k, s^{-1}\boldsymbol{\Sigma}_k\big)$. Note now that the conditional pdf $P_{\widetilde{\boldsymbol{Y}}|\boldsymbol{U}}$ describes a channel with input-output relation $\widetilde{\boldsymbol{Y}} = \sqrt{n_\mathrm{c}\rho}\widetilde{H}\boldsymbol{U} + \widetilde{\boldsymbol{W}}$, where $\boldsymbol{U}$ is an n_c-dimensional isotropically distributed unitary vector, $\widetilde{H} \sim \mathcal{CN}\big(\mu_\mathrm{H}, s^{-1}\sigma_\mathrm{H}^2\big)$, and $\widetilde{\boldsymbol{W}} \sim \mathcal{CN}\big(0, s^{-1}\mathsf{I}_{n_\mathrm{c}}\big)$. Applying Lemma 2 in Appendix A to this channel (which entails replacing σ_H^2 in (53) by $s^{-1}\sigma_\mathrm{H}^2$ and σ_w^2 by s^{-1}) we obtain (59), and (60), as shown at the top of the next page, then follows from (58).

$$\mathbb{E}\Big[P_{\widetilde{\boldsymbol{Y}}|\boldsymbol{U}}(\mathbf{y}_k|\boldsymbol{U}_k)\Big] = \frac{s^2\Gamma(n_{\mathrm{c}})\,e^{-s\|\mathbf{y}_k\|^2-s\frac{|\mu_{\mathrm{H}}|^2}{\sigma_{\mathrm{H}}^2}}}{\pi^{n_{\mathrm{c}}}\sigma_{\mathrm{H}}^2}\int_0^\infty \frac{e^{-s\left(\rho n_{\mathrm{c}}+\frac{1}{\sigma_{\mathrm{H}}^2}\right)z}}{\left(\|\mathbf{y}_k\|\sqrt{\rho n_{\mathrm{c}}z}\right)^{n_{\mathrm{c}}-1}} I_{n_{\mathrm{c}}-1}(2s\|\mathbf{y}_k\|\sqrt{\rho n_{\mathrm{c}}z})\, I_0\left(\frac{2s|\mu_{\mathrm{H}}|\sqrt{z}}{\sigma_{\mathrm{H}}^2}\right)\mathrm{d}z \tag{59}$$

$$\mathbb{E}\big[P_{\boldsymbol{Y}|\boldsymbol{U}}(\mathbf{y}_k|\boldsymbol{U}_k)^s\big] = \frac{s^{2-n_{\mathrm{c}}}\Gamma(n_{\mathrm{c}})\,e^{-s\|\mathbf{y}_k\|^2-s\frac{|\mu_{\mathrm{H}}|^2}{\sigma_{\mathrm{H}}^2}}}{\pi^{sn_{\mathrm{c}}}(1+\rho n_{\mathrm{c}}\sigma_{\mathrm{H}}^2)^{s-1}\sigma_{\mathrm{H}}^2}\int_0^\infty \frac{e^{-s\left(\rho n_{\mathrm{c}}+\frac{1}{\sigma_{\mathrm{H}}^2}\right)z}}{\left(\|\mathbf{y}_k\|\sqrt{\rho n_{\mathrm{c}}z}\right)^{n_{\mathrm{c}}-1}} I_{n_{\mathrm{c}}-1}(2s\|\mathbf{y}_k\|\sqrt{\rho n_{\mathrm{c}}z})\, I_0\left(\frac{2s|\mu_{\mathrm{H}}|\sqrt{z}}{\sigma_{\mathrm{H}}^2}\right)\mathrm{d}z \tag{60}$$

Finally, to evaluate the expectation in the RCUs bound (11), we observe that (56) and (60) imply that for every $n_{\mathrm{c}}\times n_{\mathrm{c}}$ unitary matrix V,

$$i_s\big(\mathsf{V}^H\mathbf{u}_k,\mathbf{y}_k\big) = i_s(\mathbf{u}_k,\mathsf{V}\mathbf{y}_k)\,. \tag{61}$$

This in turn implies that when $\boldsymbol{Y}_k \sim P_{\boldsymbol{Y}|\boldsymbol{U}=\mathbf{u}_k}$ the probability distribution of $i_s(\mathbf{u}_k,\boldsymbol{Y}_k)$ does not depend on $\mathbf{u}_k$. Hence, we can set without loss of generality $\mathbf{u}_k=[1,0,\ldots,0]^T$, $k=1,\ldots,\ell$. For this choice of $\{\mathbf{u}_k\}$, it follows from (56) and (60) that $i_s(\mathbf{u}_k,\boldsymbol{Y}_k)$ has the same distribution as the random variable S_k^s defined in (23).

C. Proof of Corollary 2

We evaluate Corollary 1 for $\boldsymbol{X}=\sqrt{n_{\mathrm{c}}\rho}\boldsymbol{U}$ where $\boldsymbol{U}$ is unitary and isotropically distributed. Furthermore, we choose the ML decoding metric (19). For this choice, the maximum over s in the Gallager's function for mismatch decoding (16) is achieved by $s=1/(1+\tau)$ [11, p. 137]. Let now $F_0(\tau)=e^{-E_0\left(\tau,(1+\tau)^{-1}\right)}$, where $E_0\big(\tau,(1+\tau)^{-1}\big)$ is defined in (15). Standard manipulations of the generalized information density reveal that

$$F_0(\tau)=\int_{\mathbb{C}^{n_{\mathrm{c}}}}\mathbb{E}\Big[P_{\boldsymbol{Y}|\boldsymbol{U}}(\mathbf{y}|\boldsymbol{U})^{\frac{1}{1+\tau}}\Big]^{1+\tau}d\mathbf{y}. \tag{62}$$

Note now that the expectation inside the integral in (62) can be computed as in Appendix B; specifically, its value coincides with the right-hand side of (60) provided that one replaces s in (60) with $(1+\tau)^{-1}$. Substituting this expression in (62) and computing the integral in spherical coordinates, we obtain (28).

D. Proof of Theorem 3

We use the PAT scheme described in Section III-C. We let $\boldsymbol{X}_k^{(\mathrm{d})}=\sqrt{\rho_{\mathrm{d}}n_{\mathrm{d}}}\boldsymbol{U}_k^{(\mathrm{d})}$ where $\big\{\boldsymbol{U}_k^{(\mathrm{d})}\big\}_{k=1}^{\ell}$ are n_{d}-dimensional independent and isotropically distributed unitary vectors. The pilot symbols and the corresponding n_{p}-dimensional received vectors are used to obtain a ML estimate of the fading according to (31). We assume that the receiver uses the decoding SNN decoding metric (32). A decoder that operates according to (32) treats the channel estimates $\widehat{h}_k$ as perfect, which is equivalent to assuming that

$$\boldsymbol{Y}_k^{(\mathrm{d})} \sim P_{\widetilde{\boldsymbol{Y}}^{(\mathrm{d})}|\widehat{H}=\widehat{h}_k,\boldsymbol{U}^{(\mathrm{d})}=\mathbf{u}_k^{(\mathrm{d})}} = \mathcal{CN}\Big(\widehat{h}_k\sqrt{\rho_{\mathrm{d}}n_{\mathrm{d}}}\mathbf{u}_k^{(\mathrm{d})},\mathsf{I}_{n_{\mathrm{d}}}\Big)\,. \tag{63}$$

This allows us to rewrite the generalized information density in (14) as

$$\begin{aligned} i_s^{\ell}\big(\mathbf{x}^{\ell},\mathbf{y}^{\ell}\big) &= \sum_{k=1}^{\ell} i_s\Big(\mathbf{u}_k^{(\mathrm{d})},\mathbf{y}_k^{(\mathrm{d})},\widehat{h}_k\Big) \\ &= \sum_{k=1}^{\ell}\log\frac{P_{\widetilde{\boldsymbol{Y}}^{(\mathrm{d})}|\widehat{H},\boldsymbol{U}^{(\mathrm{d})}}\Big(\mathbf{y}_k^{(\mathrm{d})}|\widehat{h}_k,\mathbf{u}_k^{(\mathrm{d})}\Big)^s}{\mathbb{E}\Big[P_{\widetilde{\boldsymbol{Y}}^{(\mathrm{d})}|\widehat{H},\boldsymbol{U}^{(\mathrm{d})}}\Big(\mathbf{y}_k^{(\mathrm{d})}|\widehat{h}_k,\boldsymbol{U}_k^{(\mathrm{d})}\Big)^s\Big]}. \end{aligned} \tag{64}$$

To evaluate the expected value in (64), we proceed similarly as in Appendix B and obtain

$$\begin{aligned} &\mathbb{E}\Big[P_{\widetilde{\boldsymbol{Y}}^{(\mathrm{d})}|\widehat{H},\boldsymbol{U}^{(\mathrm{d})}}\Big(\mathbf{y}_k^{(\mathrm{d})}|\widehat{h}_k,\boldsymbol{U}_k^{(\mathrm{d})}\Big)^s\Big] \\ &\quad= e^{-s\left(\|\mathbf{y}_k\|^2+\rho_{\mathrm{d}}n_{\mathrm{d}}|\widehat{h}_k|^2\right)}\frac{\Gamma(n_{\mathrm{d}})\,I_{n_{\mathrm{d}}-1}\Big(2s\|\mathbf{y}_k\||\widehat{h}_k|\sqrt{\rho_{\mathrm{d}}n_{\mathrm{d}}}\Big)}{\pi^{sn_{\mathrm{d}}}\Big(s\|\mathbf{y}_k\||\widehat{h}_k|\sqrt{\rho_{\mathrm{d}}n_{\mathrm{d}}}\Big)^{n_{\mathrm{d}}-1}}. \end{aligned} \tag{65}$$

Finally, to evaluate the expectation in the RCUs bound (11), we observe that (63) and (65) imply that for every $n_{\mathrm{c}}\times n_{\mathrm{c}}$ unitary matrix V,

$$i_s\Big(\mathsf{V}^H\mathbf{u}_k^{(\mathrm{d})},\mathbf{y}_k^{(\mathrm{d})},\widehat{H}_k\Big) = i_s\Big(\mathbf{u}_k^{(\mathrm{d})},\mathsf{V}\mathbf{y}_k^{(\mathrm{d})},\widehat{H}_k\Big)\,. \tag{66}$$

This in turn implies that when $\boldsymbol{Y}^{(\mathrm{d})}\sim P_{\boldsymbol{Y}^{(\mathrm{d})}|H=h_k,\boldsymbol{U}^{(\mathrm{d})}=\mathbf{u}_k^{(\mathrm{d})}}$ (the actual conditional pdf of the output vector), the probability distribution of $i_s(\mathbf{u}_k^{(\mathrm{d})},\boldsymbol{Y}_k^{(\mathrm{d})},\widehat{H}_k)$ does not depend on $\mathbf{u}_k^{(\mathrm{d})}$. Hence, we can set, without loss of generality, $\mathbf{u}_k^{(\mathrm{d})}=[1,0,\ldots,0]^T$, $k=1,\ldots,\ell$. One can finally show that under this choice of input vector, $i_s(\mathbf{u}_k^{(\mathrm{d})},\boldsymbol{Y}_k^{(\mathrm{d})},\widehat{H}_k)$ has the same distribution as the random variable T_k^s in (35).

E. Proof of Corollary 3

We use the PAT scheme introduced in Section III-C and evaluate Corollary 1 for $\boldsymbol{X}^{(\mathrm{d})}=\sqrt{n_{\mathrm{c}}\rho}\boldsymbol{U}^{(\mathrm{d})}$ where $\boldsymbol{U}^{(\mathrm{d})}$ is an n_{d}-dimensional unitary and isotropically distributed random vector.[6] Furthermore, we choose the SNN decoding metric (32). Assume that ML channel estimation yields the channel estimate $\widehat{H}=\widehat{h}$. Let $F_0\Big(\tau,s,\widehat{h}\Big)=\exp(-E_0(\tau,s,\widehat{h}))$, where $E_0(\tau,s,\widehat{h})$ is defined as in (15) (we indicate explicitly its dependency from the channel estimate $\widehat{h}$). Furthermore, let

$$P_{\widetilde{\boldsymbol{Y}}|\boldsymbol{U}=\mathbf{u},\widehat{H}=\widehat{h}} = \mathcal{CN}\Big(\widehat{h}\sqrt{\rho_{\mathrm{d}}n_{\mathrm{d}}}\mathbf{u},\mathsf{I}_{n_{\mathrm{d}}}\Big)\,. \tag{67}$$

[6]To keep the notation compact, we shall denote $\boldsymbol{U}^{(\mathrm{d})}$ and the corresponding output vector $\boldsymbol{Y}^{(\mathrm{d})}$ simply as $\boldsymbol{U}$ and $\boldsymbol{Y}$.

$$\mathbb{E}_{\boldsymbol{U}'}\left[\left(\frac{P_{\widetilde{\boldsymbol{Y}}|\boldsymbol{U},\widehat{H}}(\mathbf{y}|\boldsymbol{U}',\widehat{h})}{P_{\widetilde{\boldsymbol{Y}}|\boldsymbol{U},\widehat{H},}(\mathbf{y}|\mathbf{u},\widehat{h})}\right)^{s}\right]=\frac{\exp\Big(s\Big(\|\mathbf{y}-\sqrt{\rho_{\mathrm{d}}n_{\mathrm{d}}}\mathbf{u}\widehat{h}\|^2-\|\mathbf{y}\|^2-\rho_{\mathrm{d}}n_{\mathrm{d}}|\widehat{h}|^2\Big)\Big)}{(s\|\mathbf{y}\||\widehat{h}|\sqrt{\rho_{\mathrm{d}}n_{\mathrm{d}}})^{n_{\mathrm{d}}-1}}\Gamma(n_{\mathrm{d}})\,I_{n_{\mathrm{d}}-1}(2s\|\mathbf{y}\||\widehat{h}|\sqrt{\rho_{\mathrm{d}}n_{\mathrm{d}}}) \tag{70}$$

$$F_0\Big(\tau,s,\widehat{h}\Big)=\int_{\mathbb{C}^{n_{\mathrm{d}}}}\frac{I_{n_{\mathrm{d}}-1}\big(2s|\widehat{h}|\sqrt{\|\mathbf{y}\|^2\rho_{\mathrm{d}}n_{\mathrm{d}}}\big)^{\tau}e^{\frac{\rho_{\mathrm{d}}n_{\mathrm{d}}}{u}\left(|a(\widehat{h})|^2-|\mu_{\mathrm{p}}(\widehat{h})|^2\right)}}{\Gamma(n_{\mathrm{d}})^{-\tau}\,\pi^{n_{\mathrm{d}}}\big(1+\sigma_{\mathrm{p}}^2\,\rho_{\mathrm{d}}n_{\mathrm{d}}\big)\,(s|\widehat{h}|\sqrt{\|\mathbf{y}\|^2\rho_{\mathrm{d}}n_{\mathrm{d}}})^{\tau(n_{\mathrm{d}}-1)}}$$
$$\times\,\mathbb{E}_{\boldsymbol{U}}\Big[e^{-\left(\mathbf{y}-\sqrt{\rho_{\mathrm{d}}n_{\mathrm{d}}}a(\widehat{h})\boldsymbol{U}\right)^{H}\boldsymbol{\Sigma}\left(\mathbf{y}-\sqrt{\rho_{\mathrm{d}}n_{\mathrm{d}}}a(\widehat{h})\boldsymbol{U}\right)}\Big]\,\mathrm{d}\mathbf{y} \tag{71}$$

$$F_0\Big(\tau,s,\widehat{h}\Big)=\frac{e^{|a(\widehat{h})|^2\left(\frac{\rho_{\mathrm{d}}n_{\mathrm{d}}}{u}-\frac{1}{\sigma_{\mathrm{p}}^2}\right)-\frac{|\mu_{\mathrm{p}}(\widehat{h})|^2\rho_{\mathrm{d}}n_{\mathrm{d}}}{u}}}{\Gamma(n_{\mathrm{d}})^{-\tau}\,\sigma_{\mathrm{p}}^2}\int_0^{\infty}\frac{\exp(-r)\,r^{n_{\mathrm{d}}-1}}{(s|\widehat{h}|\sqrt{r\rho_{\mathrm{d}}n_{\mathrm{d}}})^{\tau(n_{\mathrm{d}}-1)}}I_{n_{\mathrm{d}}-1}\big(2s|\widehat{h}|\sqrt{r\rho_{\mathrm{d}}n_{\mathrm{d}}}\big)^{\tau}$$
$$\times\int_0^{\infty}\frac{\exp\big(-\big(\sigma_{\mathrm{p}}^{-2}+\rho_{\mathrm{d}}n_{\mathrm{d}}\big)\,z\big)}{\big(\sqrt{rz\rho_{\mathrm{d}}n_{\mathrm{d}}}\big)^{n_{\mathrm{d}}-1}}I_{n_{\mathrm{d}}-1}(2\sqrt{rz\rho_{\mathrm{d}}n_{\mathrm{d}}})\,I_0\big(2|a(\widehat{h})|\sigma_{\mathrm{p}}^{-2}\sqrt{z}\big)\mathrm{d}z\mathrm{d}r \tag{72}$$

Our assumptions imply that

$$F_0\Big(\tau,s,\widehat{h}\Big)=\mathbb{E}\left[\mathbb{E}_{\boldsymbol{U}'}\left[\frac{P_{\widetilde{\boldsymbol{Y}}|\boldsymbol{U},\widehat{H}}(\boldsymbol{Y}|\boldsymbol{U}',\widehat{h})^s}{P_{\widetilde{\boldsymbol{Y}}|\boldsymbol{U},\widehat{H}}(\boldsymbol{Y}|\boldsymbol{U},\widehat{h})^s}\middle|\boldsymbol{U},\boldsymbol{Y}\right]^{\tau}\right] \tag{68}$$

where $P_{\boldsymbol{Y},\boldsymbol{U},\boldsymbol{U}'}(\mathbf{y},\mathbf{u},\mathbf{u}')=P_{\boldsymbol{U}}(\mathbf{u}')P_{\boldsymbol{U}}(\mathbf{u}')P_{\boldsymbol{Y}|\boldsymbol{U},\widehat{H}}(\mathbf{y}|\mathbf{u},\widehat{h})$. Here, $P_{\boldsymbol{Y}|\boldsymbol{U},\widehat{H}}$ is the conditional output distribution of the channel, given the input $\mathbf{u}$ and the channel estimate $\widehat{h}$. Since $P_{H|\widehat{H}=\widehat{h}}=\mathcal{CN}\Big(\mu_{\mathrm{p}}(\widehat{h}),\sigma_{\mathrm{p}}^2\Big)$ where $\mu_{\mathrm{p}}(\widehat{h})$ and σ_{p}^2 are defined in (41), we conclude that

$$P_{\boldsymbol{Y}|\boldsymbol{U},\widehat{H}=\widehat{h}}=\mathcal{CN}\Big(\sqrt{\rho_{\mathrm{d}}n_{\mathrm{d}}}\mu_{\mathrm{p}}(\widehat{h})\mathbf{u},\rho_{\mathrm{d}}n_{\mathrm{d}}\sigma_{\mathrm{p}}^2\mathbf{u}\mathbf{u}^H+\mathsf{I}_{n_{\mathrm{d}}}\Big). \tag{69}$$

We next evaluate the two expectations in (68). Using (67) and (65), we can write the inner expectation as in (70), as shown at the top of this page. Substituting (70) into (68) and using (69), we obtain (71) , as shown at the top of this page, where $a(\widehat{h})=\mu_{\mathrm{p}}(\widehat{h})-\widehat{h}s\tau u$, $u=1+\sigma_{\mathrm{p}}^2\rho_{\mathrm{d}}n_{\mathrm{d}}$, and $\boldsymbol{\Sigma}=\big(\rho_{\mathrm{d}}n_{\mathrm{d}}\sigma_{\mathrm{p}}^2\boldsymbol{U}\boldsymbol{U}^H+\mathsf{I}_{n_{\mathrm{d}}}\big)^{-1}$. Note that the term inside the expectation is proportional to the law of a channel with input-output relation $\widetilde{\boldsymbol{Y}}=\sqrt{\rho_{\mathrm{d}}n_{\mathrm{d}}}\widetilde{H}\boldsymbol{U}+\boldsymbol{W}$, where $\widetilde{H}\sim\mathcal{CN}\Big(a(\widehat{h}),\sigma_{\mathrm{p}}^2\Big)$ and $\boldsymbol{W}\sim\mathcal{CN}(0,\mathsf{I}_{n_{\mathrm{d}}})$. Using Lemma 2 in Appendix A to evaluate this expectation, and computing the outer integral in spherical coordinates, we obtain (72), as shown at the top of this page. Finally, we obtain (38) by using (72) in (16) and by taking an expectation over $\widehat{H}$.

F. Proof of Theorem 5

We use as auxiliary channel in the min-max converse [6, Th. 27], the one for which $\mathbf{y}^\ell$ has pdf

$$Q_{\boldsymbol{Y}^\ell}\big(\mathbf{y}^\ell\big)=\prod_{k=1}^{\ell}P_{\boldsymbol{Y}}(\mathbf{y}_k) \tag{73}$$

where $P_{\boldsymbol{Y}}$ is given in (53). Note now that for every $n_{\mathrm{c}}\times n_{\mathrm{c}}$ unitary matrix V, we have $P_{\boldsymbol{Y}}(\mathsf{V}\mathbf{y}_k)=P_{\boldsymbol{Y}}(\mathbf{y}_k)$ and $P_{\boldsymbol{Y}|\boldsymbol{X}}\big(\mathbf{y}_k|\mathsf{V}^H\mathbf{x}_k\big)=P_{\boldsymbol{Y}|\boldsymbol{X}}(\mathsf{V}\mathbf{y}_k|\mathbf{x}_k)$. Along with (23), and because of the per-codeword power constraint (4), this imply that the Neyman-Pearson function $\beta\big(\mathbf{x}^\ell,Q_{\boldsymbol{Y}^\ell}\big)$ defined in [6, eq. (105)] is independent of $\mathbf{x}^\ell$. Hence, we can use [6, Th. 28] to conclude that R^* is upper-bounded as

$$R^*(\ell,n_{\mathrm{c}},\epsilon,\rho)\le\frac{1}{n_{\mathrm{c}}\ell}\log\frac{1}{\beta_{1-\epsilon}(\mathbf{x}^\ell,Q_{\boldsymbol{Y}^\ell})}. \tag{74}$$

Without loss of generality, we shall set $\mathbf{x}_k=[\sqrt{n_c\rho},0\ldots,0]$, $k=1,\ldots,\ell$. It follows by the Neyman-Pearson lemma [37] that

$$\beta_{1-\epsilon}\big(\mathbf{x}^\ell,q_{\boldsymbol{Y}^\ell}\big)=\Pr\big\{r^\ell\big(\mathbf{x}^\ell,\boldsymbol{Y}^\ell\big)\ge\gamma\big\},\quad\boldsymbol{Y}^\ell\sim Q_{\boldsymbol{Y}^\ell} \tag{75}$$

where γ is the solution to

$$\Pr\big\{r^\ell\big(\mathbf{x}^\ell,\boldsymbol{Y}^\ell\big)\le\gamma\big\}=\epsilon,\quad\boldsymbol{Y}^\ell\sim P_{\boldsymbol{Y}^\ell|\boldsymbol{X}^\ell} \tag{76}$$

and

$$r^\ell\big(\mathbf{x}^\ell,\mathbf{y}^\ell\big)=\sum_{k=1}^{\ell}r(\mathbf{x}_k,\mathbf{y}_k)=\sum_{k=1}^{\ell}\log\frac{P_{\boldsymbol{Y}|\boldsymbol{X}}(\mathbf{y}_k|\mathbf{x}_k)}{P_{\boldsymbol{Y}}(\mathbf{y}_k)}. \tag{77}$$

Finally, we obtain (48) by relaxing (74) using [6, eq. (106)] (which yields a generalized Verdú-Han converse bound, see [38]) and by exploiting that when $\boldsymbol{Y}_k\sim P_{\boldsymbol{Y}|\boldsymbol{X}=\mathbf{x}_k}$ the random variable $r(\mathbf{x}_k,\boldsymbol{Y}_k)$ is distributed as S_k^s in (23) with $s=1$.

REFERENCES

[1] J. Östman, G. Durisi, and E. G. Ström, "Finite-blocklength bounds on the maximum coding rate of Rician fading channels with applications to pilot-assisted transmission," in *Proc. IEEE 18th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Sapporo, Japan, Jul. 2017, pp. 1–5.

[2] *IMT Vision–Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond*, Rec. ITU-R M.2083-0, ITU-R, Geneva, Switzerland, Sep. 2015.

[3] *Technical Specification Group Services and System Aspects; Service Requirements for the 5G Systems; Stage 1; (Release 16), Version 16.2.0*, document TS 22.261, 3GPP, Dec. 2017.

[4] G. Durisi, T. Koch, J. Östman, Y. Polyanskiy, and W. Yang, "Short-packet communications over multiple-antenna Rayleigh-fading channels," *IEEE Trans. Commun.*, vol. 64, no. 2, pp. 618–629, Feb. 2016.

[5] L. Tong, B. M. Sadler, and M. Dong, "Pilot-assisted wireless transmissions: General model, design criteria, and signal processing," *IEEE Signal Process. Mag.*, vol. 21, no. 6, pp. 12–25, Nov. 2004.

[6] 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.

[7] 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.

[8] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.

[9] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits," Ph.D. dissertation, Dept. Elect. Eng., Princeton Univ., Princeton, NJ, USA, Nov. 2010.
[10] V. Y. F. Tan and M. Tomamichel, "The third-order term in the normal approximation for the AWGN channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, pp. 2430–2438, May 2015.
[11] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY, USA: Wiley, 1968.
[12] B. M. Hochwald and T. L. Marzetta, "Unitary space-time modulation for multiple-antenna communications in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 46, no. 2, pp. 543–564, Mar. 2000.
[13] L. Zheng and D. N. C. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Trans. Inf. Theory*, vol. 48, no. 2, pp. 359–383, Feb. 2002.
[14] C. Potter, K. Kosbar, and A. Panagos, "On achievable rates for MIMO systems with imperfect channel state information in the finite length regime," *IEEE Trans. Commun.*, vol. 61, no. 7, pp. 2772–2781, Jul. 2013.
[15] I. Abou-Faycal and B. M. Hochwald, "Coding requirements for multiple-antenna channels with unknown Rayleigh fading," Bell Labs., Lucent Technologies, Murray Hill, NJ, USA, Tech. Rep., 1999.
[16] T. L. Marzetta and B. M. Hochwald, "Capacity of a mobile multiple-antenna communication link in Rayleigh flat fading," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 139–157, Jan. 1999.
[17] M. C. Gursoy, "Error exponents and cutoff rate for noncoherent Rician fading channels," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Istanbul, Turkey, Jun. 2006, pp. 1398–1403.
[18] J. Östman, G. Durisi, E. G. Ström, J. Li, H. Sahlin, and G. Liva, "Low-latency ultra-reliable 5G communications: Finite block-length bounds and coding schemes," in *Proc. 11th Int. ITG Conf. Syst., Commun. Coding (SCC)*, Hamburg, Germany, Feb. 2017, pp. 1–6.
[19] G. Kaplan and S. Shamai (Shitz), "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *AEU. Archiv für Elektronik und Übertragungstechnik*, vol. 47, no. 4, pp. 228–239, Jul. 1993.
[20] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.
[21] A. Ganti, A. Lapidoth, and Ï. E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.
[22] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.
[23] A. Lapidoth and S. Shamai (Shitz), "Fading channels: How perfect need 'perfect side information' be?" *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.
[24] M. Dorpinghaus, A. Ispas, and H. Meyr, "On the gain of joint processing of pilot and data symbols in stationary Rayleigh fading channels," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 2963–2982, May 2012.
[25] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2647–2666, May 2014.
[26] A. Martinez and A. Guillén i Fàbregas, "Saddlepoint approximation of random-coding bounds," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2011, pp. 1–6.
[27] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
[28] M. Godavarti and A. O. Hero, "Training in multiple-antenna Rician fading wireless channels with deterministic specular component," *IEEE Trans. Wireless Commun.*, vol. 6, no. 1, pp. 110–119, Jan. 2007.
[29] H. Weingarten, Y. Steinberg, and S. Shamai (Shitz), "Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1665–1686, Aug. 2004.
[30] G. Liva, L. Gaudio, T. Ninacs, and T. Jerkovits. (2016). "Code design for short blocks: A survey." [Online]. Available: https://arxiv.org/abs/1610.00873
[31] J. V. Wonterghem, A. Alloumf, J. J. Boutros, and M. Moeneclaey, "Performance comparison of short-length error-correcting codes," in *Proc. Symp. Commun. Veh. Technol. (SCVT)*, Nov. 2016, pp. 1–6.
[32] M. P. C. Fossorier and S. Lin, "Soft-decision decoding of linear block codes based on ordered statistics," *IEEE Trans. Inf. Theory*, vol. 41, no. 5, pp. 1379–1396, Sep. 1995.
[33] A. Valembois and M. Fossorier, "Box and match techniques applied to soft-decision decoding," *IEEE Trans. Inf. Theory*, vol. 50, no. 5, pp. 796–810, May 2004.
[34] Y. Wu and C. N. Hadjicostis, "Soft-decision decoding using ordered recodings on the most reliable basis," *IEEE Trans. Inf. Theory*, vol. 53, no. 2, pp. 829–836, Feb. 2007.
[35] R. Johannesson and K. S. Zigangirov, *Fundamentals of Convolutional Coding*, 2nd ed. Hoboken, NJ, USA: Wiley, 2015.
[36] J. K. Wolf and A. J. Viterbi, "On the weight distribution of linear block codes formed from convolutional codes," *IEEE Trans. Commun.*, vol. 44, no. 9, pp. 1049–1051, Sep. 1996.
[37] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philos. Trans. Roy. Soc. London A, Math. Phys. Sci.*, vol. 231, nos. 694–706, pp. 289–337, 1933.
[38] T. S. Han, *Information-Spectrum Methods in Information Theory*. Berlin, Germany: Springer-Verlag, 2003.

Johan Östman (S'16) received the B.Sc. and M.Sc. degrees in electrical engineering from the Chalmers University of Technology, Gothenburg, Sweden, in 2011 and 2014, respectively. From 2014 to 2015, he was with the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology, Hong Kong. He is currently pursuing the Ph.D. degree in electrical engineering with the Chalmers University of Technology. His research interests are in the areas of information- and communications theory.

Giuseppe Durisi (S'02–M'06–SM'12) received the Laurea degree *(summa cum laude)* and the Ph.D. degree from the Politecnico di Torino, Italy, in 2001 and 2006, respectively. From 2006 to 2010, he was a Post-Doctoral Researcher with ETH Zürich, Zürich, Switzerland. In 2010, he joined the Chalmers University of Technology, Gothenburg, Sweden, where he is currently a Professor and the Co-Director of Chalmers information and communication technology area of advance. His research interests are in the areas of communication and information theory and machine learning. He was a recipient of the 2013 IEEE ComSoc Best Young Researcher Award for the Europe, Middle East, and Africa Region. He has co-authored a paper that received the Student Paper Award from the 2012 International Symposium on Information Theory and a paper that received the 2013 IEEE Sweden VT-COM-IT joint chapter Best Student Conference Paper Award. In 2015, he joined the Editorial Board of the IEEE TRANSACTIONS ON COMMUNICATIONS as an Associate Editor. From 2011 to 2014, he served as a Publications Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY.

Erik G. Ström (S'93–M'95–SM'01) received the M.S. degree in electrical engineering from the Royal Institute of Technology (KTH), Stockholm, Sweden, in 1990, and the Ph.D. degree in electrical engineering from the University of Florida, Gainesville, in 1994. He held a post-doctoral position at the Department of Signals, Sensors, and Systems, KTH, in 1995. In 1996, he was appointed an Assistant Professor at KTH. In 1996, he joined the Chalmers University of Technology, Gothenburg, Sweden, where he has been a Professor in communication systems since 2003. He is currently the Head of the Division for Communication and Antenna Systems, the Director of the ChaseOn, a Vinnova Competence Center focused on antenna system, and leads the competence area sensors and communications at the traffic safety center SAFER, which is hosted by Chalmers. His research interests include signal processing and communication theory in general, and constellation labelings, channel estimation, synchronization, multiple access, medium access, multiuser detection, wireless positioning, and vehicular communications in particular. Since 1990, he has been acting as a Consultant for the Educational Group for Individual Development, Stockholm, Sweden. He is a contributing author and an Associate Editor of Roy. Admiralty Publishers FesGas-series. He was a member of the board of the IEEE VT/COM Swedish Chapter from 2000 to 2006. He received the Chalmers Pedagogical Prize in 1998 and the Chalmers Ph.D. Supervisor of the Year Award in 2009. He was a Co-Guest Editor for the PROCEEDINGS OF THE IEEE special issue on Vehicular Communications in 2011 and the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS special issues on Signal Synchronization in Digital Transmission Systems in 2001 and on Multiuser Detection for Advanced Communication Systems and Networks in 2008.

Mustafa C. Coşkun (S'17) received the B.Sc. degree in electrical and electronics engineering from Boğaziçi University, Istanbul, Turkey, and the M.Sc. degree in communications engineering from the Technical University of Munich, Munich, Germany, in 2014 and 2017, respectively, where he is currently pursuing the Ph.D. degree with the Institute for Communications Engineering, funded by the Munich Aerospace Center. He is also a member of the Information Transmission Group, German Aerospace Center (DLR).

Gianluigi Liva (M'08–SM'14) was born in Spilimbergo, Italy, in 1977. He received the M.S. and Ph.D. degrees in electrical engineering from the University of Bologna, Italy, in 2002 and 2006, respectively. Since 2003 he has been investigating channel codes for high-data rate Consultative Committee for Space Data Systems (CCSDS) missions. From 2004 to 2005, he was involved in research at the University of Arizona, Tucson, where he was designing low-complexity error correcting codes for space communications.

Since 2006, he has been with the Institute of Communications and Navigation, German Aerospace Center (DLR), where he currently leads the Information Transmission Group. In 2010, he has been appointed as a Lecturer of channel coding with the Institute for Communications Engineering (LNT), Technical University of Munich (TUM). From 2012 to 2013, he was a Lecturer of channel coding with the Nanjing University of Science and Technology, China. Since 2014, he has been a Lecturer of channel codes with iterative decoding with LNT, TUM. His main research interests include satellite communications, random access techniques, and error control coding. He is/has been active in the DVB-SH, DVB-RCS, and DVB-S2 standardization groups, and in the standardization of error correcting codes for deep-space communications within the CCSDS.

Dr. Liva received the 2007 IST Mobile and Wireless Communication Summit Best Paper Award. He was the Co-Chair of the First DLR Workshop on Random Access and Coding in 2013, the 2014 Sino-German Workshop on Bridging Theory and Practice in Wireless Communications and Networking, in Shenzhen, China, the IEEE ICC 2014 Workshop on Massive Uncoordinated Access Protocols, in Sydney, Australia, the 2015 Munich Workshop on Coding and Modulation, and the 2018 IEEE European School on Information Theory.