



MIT Open Access Articles

Diversity versus Channel Knowledge at Finite Block-Length

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

Citation	Wei Yang, Giuseppe Durisi, Tobias Koch, and Yury Polyanskiy. "Diversity Versus Channel Knowledge at Finite Block-length." In 2012 IEEE Information Theory Workshop, 3-7 September 2012 : Lausanne, Switzerland , Pp. 572-576.
As Published	http://dx.doi.org/10.1109/ITW.2012.6404740
Publisher	Institute of Electrical and Electronics Engineers
Version	Author's final manuscript
Citable link	http://hdl.handle.net/1721.1/79040
Terms of Use	Creative Commons Attribution-Noncommercial-Share Alike 3.0
Detailed Terms	http://creativecommons.org/licenses/by-nc-sa/3.0/

Diversity versus Channel Knowledge at Finite Block-Length

Wei Yang¹, Giuseppe Durisi¹, Tobias Koch², and Yury Polyanskiy³

¹Chalmers University of Technology, 41296 Gothenburg, Sweden

²Universidad Carlos III de Madrid, 28911 Leganés, Spain

³Massachusetts Institute of Technology, Cambridge, MA, 02139 USA

Abstract—We study the maximal achievable rate $R^*(n, \epsilon)$ for a given block-length n and block error probability ϵ over Rayleigh fading channels in the noncoherent setting and in the finite block-length regime. Our results show that for a given block-length and error probability, $R^*(n, \epsilon)$ is not monotonic in the channel’s coherence time, but there exists a rate maximizing coherence time that optimally trades between diversity and cost of estimating the channel.

I. INTRODUCTION

It is well known that the capacity of the single-antenna Rayleigh-fading channel with perfect channel state information (CSI) at the receiver (the so-called *coherent setting*) is independent of the fading dynamics [1]. In practical wireless systems, however, the channel is usually not known *a priori* at the receiver and must be estimated, for example, by transmitting training symbols. An important observation is that the training overhead is a function of the channel dynamics, because the faster the channel varies, the more training symbols are needed in order to estimate the channel accurately [2]–[4]. One way to determine the training overhead, or more generally, the capacity penalty due to lack of channel knowledge, is to study capacity in the *noncoherent setting*, where neither the transmitter nor the receiver are assumed to have *a priori* knowledge of the realizations of the fading channel (but both are assumed to know its statistics perfectly) [5].

In this paper, we model the fading dynamics using the well-known block-fading model [6]–[8] according to which the channel coefficients remain constant for a period of T symbols, and change to a new independent realization in the next period. The parameter T can be thought of as the channel’s coherence time. Unfortunately, even for this simple model, no closed-form expression for capacity is available to date. A capacity lower bound based on the *isotropically distributed (i.d.)* unitary distribution is reported in [6]. In [7]–[9], it is shown that capacity in the high signal-to-noise ratio (SNR) regime grows logarithmically with SNR, with the *pre-log* (defined as the asymptotic ratio between capacity and the logarithm of SNR as SNR goes to infinity) being $1 - 1/T$. This agrees with the intuition that the capacity penalty due to lack of a priori channel knowledge at the receiver is small when the channel’s coherence time is large.

In order to approach capacity, the block-length n of the codewords must be long enough to average out the fading effects (i.e., $n \gg T$). Under practical delay constraints, however, the actual performance metric is the maximal achievable rate $R^*(n, \epsilon)$ for a given block-length n and block error probability ϵ . By studying $R^*(n, \epsilon)$ for the case of fading channels and in the coherent setting, Polyanskiy and Verdú recently showed that faster fading dynamics are advantageous in the finite block-length regime when the channel is known to the receiver [10], because faster fading dynamics yield larger diversity gain.

We expect that the maximal achievable rate $R^*(n, \epsilon)$ over fading channels in the *noncoherent setting* and in the *finite block-length regime* is governed by two effects working in opposite directions: when the channel’s coherence time decreases, we can code the information over a larger number of independent channel realizations, which provides higher diversity gain, but we need to transmit training symbols more frequently to learn the channel accurately, which gives rise to a rate loss.

In this paper, we shed light on this fundamental tension by providing upper and lower bounds on $R^*(n, \epsilon)$ in the noncoherent setting. For a given block-length and error probability, our bounds show that there exists indeed a rate-maximizing channel’s coherence time that optimally trades between diversity and cost of estimating the channel.

Notation: Uppercase boldface letters denote matrices and lowercase boldface letters designate vectors. Uppercase sans-serif letters (e.g., \mathbf{Q}) denote probability distributions, while lowercase sans-serif letters (e.g., \mathbf{r}) are reserved for probability density functions (pdf). The superscripts T and H stand for transposition and Hermitian transposition, respectively. We denote the identity matrix of dimension $T \times T$ by \mathbf{I}_T ; the sequence of vectors $\{\mathbf{a}_1, \dots, \mathbf{a}_n\}$ is written as \mathbf{a}^n . We denote expectation and variance by $\mathbb{E}[\cdot]$ and $\text{Var}[\cdot]$, respectively, and use the notation $\mathbb{E}_{\mathbf{x}}[\cdot]$ or $\mathbb{E}_{\mathbf{P}_{\mathbf{x}}}[\cdot]$ to stress that expectation is taken with respect to \mathbf{x} with distribution $\mathbf{P}_{\mathbf{x}}$. The relative entropy between two distributions \mathbf{P} and \mathbf{Q} is denoted by $D(\mathbf{P}||\mathbf{Q})$ [11, Sec. 8.5]. For two functions $f(x)$ and $g(x)$, the notation $f(x) = \mathcal{O}(g(x))$, $x \rightarrow \infty$, means that $\limsup_{x \rightarrow \infty} |f(x)/g(x)| < \infty$, and $f(x) = o(g(x))$, $x \rightarrow \infty$, means that $\lim_{x \rightarrow \infty} |f(x)/g(x)| = 0$. Furthermore, $\mathcal{CN}(\mathbf{0}, \mathbf{R})$ stands for the distribution of a circularly-symmetric com-

Tobias Koch has received funding from the European Community’s Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 252663.

plex Gaussian random vector with covariance matrix \mathbf{R} , and $\text{Gamma}(\alpha, \beta)$ denotes the gamma distribution [12, Ch. 17] with parameters α and β . Finally, $\log(\cdot)$ indicates the natural logarithm, $\Gamma(\cdot)$ denotes the gamma function [13, Eq. (6.1.1)], and $\psi(\cdot)$ designates the digamma function [13, Eq. (6.3.2)].

II. CHANNEL MODEL AND FUNDAMENTAL LIMITS

We consider a single-antenna Rayleigh block-fading channel with *coherence time* T . Within the l th coherence interval, the channel input-output relation can be written as

$$\mathbf{y}_l = s_l \mathbf{x}_l + \mathbf{w}_l \quad (1)$$

where \mathbf{x}_l and \mathbf{y}_l are the input and output signals, respectively, $\mathbf{w}_l \sim \mathcal{CN}(0, \mathbf{I}_T)$ is the additive noise, and $s_l \sim \mathcal{CN}(0, 1)$ models the fading, whose realization we assume is not known at the transmitter and receiver (noncoherent setting). In addition, we assume that $\{s_l\}$ and $\{\mathbf{w}_l\}$ take on independent realizations over successive coherence intervals.

We consider channel coding schemes employing codewords of length $n = LT$. Therefore, each codeword spans L independent fading realizations. Furthermore, the codewords are assumed to satisfy the following power constraint

$$\sum_{l=1}^L \|\mathbf{x}_l\|^2 \leq LT\rho. \quad (2)$$

Since the variance of s_l and of the entries of \mathbf{w}_l is normalized to one, ρ in (2) can be interpreted as the SNR at the receiver.

Let $R^*(n, \epsilon)$ be the maximal achievable rate among all codes with block-length n and decodable with probability of error ϵ . For every fixed T and ϵ , we have¹

$$\lim_{n \rightarrow \infty} R^*(n, \epsilon) = C(\rho) = \frac{1}{T} \sup_{\mathbf{P}_{\mathbf{x}}} I(\mathbf{x}; \mathbf{y}) \quad (3)$$

where $C(\rho)$ is the capacity of the channel in (1), $I(\mathbf{x}; \mathbf{y})$ denotes the mutual information between \mathbf{x} and \mathbf{y} , and the supremum in (3) is taken over all input distributions $\mathbf{P}_{\mathbf{x}}$ that satisfy

$$\mathbb{E}[\|\mathbf{x}\|^2] \leq T\rho. \quad (4)$$

No closed-form expression of $C(\rho)$ is available to date. The following lower bound $L(\rho)$ on $C(\rho)$ is reported in [6, Eq. (12)]

$$\begin{aligned} L(\rho) = & \frac{1}{T} \left((T-1) \log(T\rho) - \log \Gamma(T) - T + \frac{T(1+\rho)}{1+T\rho} \right) \\ & - \frac{1}{T} \int_0^\infty e^{-u} \tilde{\gamma}(T-1, T\rho u) \left(1 + \frac{1}{T\rho} \right)^{T-1} \\ & \times \log(u^{1-T} \tilde{\gamma}(T-1, T\rho u)) du \quad (5) \end{aligned}$$

where

$$\tilde{\gamma}(n, x) \triangleq \frac{1}{\Gamma(n)} \int_0^x t^{n-1} e^{-t} dt$$

denotes the *regularized incomplete gamma function*. The input distribution used in [6] to establish (5) is the i.d. unitary distribution, where the input vector takes on the form $\mathbf{x} = \sqrt{T\rho} \mathbf{u}_{\mathbf{x}}$

with $\mathbf{u}_{\mathbf{x}}$ uniformly distributed on the unit sphere in \mathbb{C}^T . We shall denote this input distribution as $\mathbf{P}_{\mathbf{x}}^{(U)}$. It can be shown that $L(\rho)$ is asymptotically tight at high SNR (see [7, Thm. 4]), i.e.,

$$C(\rho) = L(\rho) + o(1), \quad \rho \rightarrow \infty.$$

III. BOUNDS ON $R^*(n, \epsilon)$

A. Perfect-Channel-Knowledge Upper Bound

We establish a simple upper bound on $R^*(n, \epsilon)$ by assuming that the receiver has perfect knowledge of the realizations of the fading process $\{s_l\}$. Specifically, we have that

$$R^*(n, \epsilon) \leq R_{\text{coh}}^*(n, \epsilon) \quad (6)$$

where $R_{\text{coh}}^*(n, \epsilon)$ denotes the maximal achievable rate for a given block-length n and probability of error ϵ in the coherent setting.

By generalizing the method used in [10] for stationary ergodic fading channels to the present case of block-fading channels, we obtain the following asymptotic expression for $R_{\text{coh}}^*(n, \epsilon)$:

$$\begin{aligned} R_{\text{coh}}^*(n, \epsilon) = & C_{\text{coh}}(\rho) - \sqrt{\frac{V_{\text{coh}}(\rho)}{n}} Q^{-1}(\epsilon) \\ & + o\left(\frac{1}{\sqrt{n}}\right), \quad n \rightarrow \infty. \quad (7) \end{aligned}$$

Here, $C_{\text{coh}}(\rho)$ is the capacity of the block-fading channel in the coherent setting, which is given by [1, Eq. (3.3.10)]

$$C_{\text{coh}}(\rho) = \mathbb{E}_s[\log(1 + |s|^2 \rho)] \quad (8)$$

$Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$ denotes the Q -function, and

$$V_{\text{coh}}(\rho) = T \text{Var}[\log(1 + \rho|s|^2)] + 1 - \mathbb{E}^2 \left[\frac{1}{1 + \rho|s|^2} \right]$$

is the *channel dispersion*. Neglecting the $o(1/\sqrt{n})$ term in (7), we obtain the following approximation for $R_{\text{coh}}^*(n, \epsilon)$

$$R_{\text{coh}}^*(n, \epsilon) \approx C_{\text{coh}}(\rho) - \sqrt{\frac{V_{\text{coh}}(\rho)}{n}} Q^{-1}(\epsilon). \quad (9)$$

It was reported in [14], [15] that approximations similar to (9) are accurate for many channels for block-lengths and error probabilities of practical interest. Hence, we will use (9) to evaluate $R_{\text{coh}}^*(n, \epsilon)$ in the remainder of the paper.

B. Upper Bound through Fano's inequality

Our second upper bound follows from Fano's inequality [11, Thm. 2.10.1]

$$R^*(n, \epsilon) \leq \frac{C(\rho) + H(\epsilon)/n}{1 - \epsilon} \quad (10)$$

where $H(x) = -x \log x - (1-x) \log(1-x)$ is the binary entropy function. Since no closed-form expression is available for $C(\rho)$, we will further upper-bound the right-hand side (RHS) of (10) by replacing $C(\rho)$ with the capacity upper bound we shall derive below.

Let $\mathbf{P}_{\mathbf{y}|\mathbf{x}}$ denote the conditional distribution of \mathbf{y} given \mathbf{x} , and $\mathbf{P}_{\mathbf{y}}$ denote the distribution induced on \mathbf{y} by the

¹The subscript l is omitted whenever immaterial.

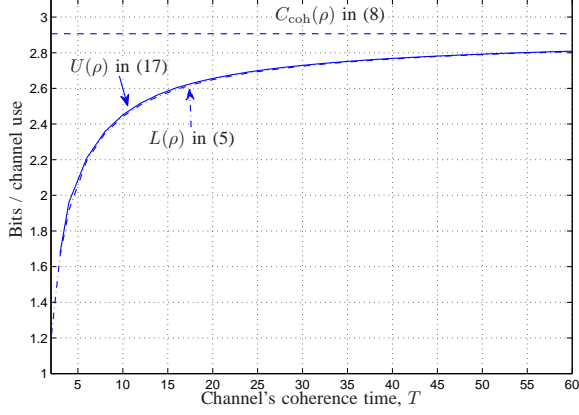


Fig. 1. $U(\rho)$ in (17), $L(\rho)$ in (5) and $C_{\text{coh}}(\rho)$ in (8) as a function of the channel's coherence time T , $\rho = 10$ dB.

input distribution $P_{\mathbf{x}}$ through (1). Furthermore, let $Q_{\mathbf{y}}$ be an arbitrary distribution on \mathbf{y} with pdf $q_{\mathbf{y}}(\mathbf{y})$. We can upper-bound $I(\mathbf{x}; \mathbf{y})$ in (3) by duality as follows [16, Thm. 5.1]:

$$\begin{aligned} I(\mathbf{x}; \mathbf{y}) &\leq \mathbb{E}[D(P_{\mathbf{y}|\mathbf{x}}\|Q_{\mathbf{y}})] \\ &= -\mathbb{E}_{P_{\mathbf{y}}}[\log q_{\mathbf{y}}(\mathbf{y})] - h(\mathbf{y}|\mathbf{x}). \end{aligned} \quad (11)$$

Since

$$T\rho - \mathbb{E}[\|\mathbf{x}\|^2] \geq 0 \quad (12)$$

for every $P_{\mathbf{x}}$ satisfying (4), we can upper bound $C(\rho)$ in (3) by using (11) and (12) to obtain

$$\begin{aligned} C(\rho) &\leq \frac{1}{T} \inf_{\lambda \geq 0} \sup_{P_{\mathbf{x}}} \left\{ -\mathbb{E}_{P_{\mathbf{y}}}[\log q_{\mathbf{y}}(\mathbf{y})] \right. \\ &\quad \left. - h(\mathbf{y}|\mathbf{x}) + \lambda(T\rho - \mathbb{E}[\|\mathbf{x}\|^2]) \right\}. \end{aligned} \quad (13)$$

The same bounding technique was previously used in [17] to obtain upper bounds on the capacity of the phase-noise AWGN channel (see also [18]).

We next evaluate the RHS of (13) for the following pdf

$$q_{\mathbf{y}}(\mathbf{y}) = \frac{\Gamma(T)\|\mathbf{y}\|^{2(1-T)}}{\pi^T T(\rho+1)} e^{-\|\mathbf{y}\|^2/[T(\rho+1)]}, \quad \mathbf{y} \in \mathbb{C}^T. \quad (14)$$

Thus, \mathbf{y} is i.i.d. and $\|\mathbf{y}\|^2 \sim \text{Gamma}(1, T(1+\rho))$. Substituting (14) into $\mathbb{E}_{P_{\mathbf{y}}}[\log q_{\mathbf{y}}(\mathbf{y})]$ in (13), we obtain

$$\begin{aligned} &-\mathbb{E}_{P_{\mathbf{y}}}[\log q_{\mathbf{y}}(\mathbf{y})] \\ &= \log \frac{T(1+\rho)\pi^T}{\Gamma(T)} + \frac{T + \mathbb{E}[\|\mathbf{x}\|^2]}{T(\rho+1)} \\ &\quad + (T-1)\mathbb{E}[\log((1+\|\mathbf{x}\|^2)z_1 + z_2)] \\ &= \log \frac{T(1+\rho)\pi^T}{\Gamma(T)} + \frac{1}{\rho+1} + (T-1)\psi(T-1) \\ &\quad + \mathbb{E} \left[(T-1) \sum_{k=0}^{\infty} \frac{(1+1/\|\mathbf{x}\|^2)^{-k}}{k+T-1} + \frac{\|\mathbf{x}\|^2}{T(1+\rho)} \right]. \end{aligned} \quad (15)$$

The first equality in (15) follows because the random variable $\|\mathbf{y}\|^2$ is conditionally distributed as $(1+\|\mathbf{x}\|^2)z_1 + z_2$ given \mathbf{x} , where $z_1 \sim \text{Gamma}(1, 1)$ and $z_2 \sim \text{Gamma}(T-1, 1)$.

Substituting (15) into (13), and using that the differential entropy $h(\mathbf{y}|\mathbf{x})$ is given by

$$h(\mathbf{y}|\mathbf{x}) = \mathbb{E}_{\mathbf{x}}[\log(1+\|\mathbf{x}\|^2)] + T \log(\pi e)$$

we obtain

$$\begin{aligned} C(\rho) &\leq \frac{c_1}{T} + \frac{1}{T} \inf_{\lambda \geq 0} \sup_{P_{\mathbf{x}}} \left\{ \mathbb{E} \left[\sum_{k=0}^{\infty} \frac{(T-1)(1+1/\|\mathbf{x}\|^2)^{-k}}{k+T-1} \right. \right. \\ &\quad \left. \left. - \log(1+\|\mathbf{x}\|^2) + \frac{\|\mathbf{x}\|^2}{T(1+\rho)} + \lambda(T\rho - \|\mathbf{x}\|^2) \right] \right\} \\ &\stackrel{(a)}{\leq} \frac{c_1}{T} + \frac{1}{T} \inf_{\lambda \geq 0} \sup_{\|\mathbf{x}\|} \left\{ \sum_{k=0}^{\infty} \frac{(T-1)(1+1/\|\mathbf{x}\|^2)^{-k}}{k+T-1} \right. \\ &\quad \left. - \log(1+\|\mathbf{x}\|^2) + \frac{\|\mathbf{x}\|^2}{T(1+\rho)} + \lambda(T\rho - \|\mathbf{x}\|^2) \right\} \\ &\triangleq U(\rho) \end{aligned} \quad (16)$$

where

$$c_1 \triangleq \log \frac{T(1+\rho)}{\Gamma(T)} - T + \frac{1}{\rho+1} + (T-1)\psi(T-1).$$

To obtain (a), we upper-bounded the second term on the RHS of (16) by replacing the expectation over $\|\mathbf{x}\|$ by the supremum over $\|\mathbf{x}\|$.

The bounds $L(\rho)$ and $U(\rho)$ are plotted in Fig. 1 as a function of the channel's coherence time T for SNR equal to 10 dB. For reference, we also plot the capacity in the coherent setting [$C_{\text{coh}}(\rho)$ in (8)]. We observe that $U(\rho)$ and $L(\rho)$ are surprisingly close for all values of T .

At low SNR, the gap between $U(\rho)$ and $L(\rho)$ increases. In this regime, $U(\rho)$ can be tightened by replacing $q_{\mathbf{y}}(\mathbf{y})$ in (13) by the output pdf induced by the i.i.d. unitary input distribution $P_{\mathbf{x}}^{(U)}$, which is given by

$$\begin{aligned} q_{\mathbf{y}}^{(U)}(\mathbf{y}) &= \frac{e^{-\|\mathbf{y}\|^2/(1+T\rho)}\|\mathbf{y}\|^{2(1-T)}\Gamma(T)}{\pi^T(1+T\rho)} \\ &\quad \times \tilde{\gamma} \left(T-1, \frac{T\rho\|\mathbf{y}\|^2}{1+T\rho} \right) \left(1 + \frac{1}{T\rho} \right)^{T-1}. \end{aligned} \quad (18)$$

Substituting (17) into (10), we obtain the following upper bound on $R^*(n, \epsilon)$:

$$R^*(n, \epsilon) \leq \bar{R}(n, \epsilon) \triangleq \frac{U(\rho) + H(\epsilon)/n}{1 - \epsilon}. \quad (19)$$

C. Dependence Testing (DT) Lower Bound

We next present a lower bound on $R^*(n, \epsilon)$ that is based on the DT bound recently proposed by Polyanskiy, Poor, and Verdú [14]. The DT bound uses a threshold decoder that sequentially tests all messages and returns the first message whose likelihood exceeds a pre-determined threshold. With this approach, one can show that for a given input distribution

$P_{\mathbf{x}^L}$, there exists a code with M codewords and average probability of error not exceeding [14, Thm. 17]

$$\epsilon \leq \mathbb{E}_{P_{\mathbf{x}^L}} \left[P_{\mathbf{y}^L | \mathbf{x}^L} \left(i(\mathbf{x}^L; \mathbf{y}^L) \leq \log \frac{M-1}{2} \right) + \frac{M-1}{2} P_{\mathbf{y}^L} \left(i(\mathbf{x}^L; \mathbf{y}^L) > \log \frac{M-1}{2} \right) \right] \quad (20)$$

where

$$i(\mathbf{x}^L; \mathbf{y}^L) \triangleq \log \frac{P_{\mathbf{y}^L | \mathbf{x}^L}(\mathbf{y}^L | \mathbf{x}^L)}{P_{\mathbf{y}^L}(\mathbf{y}^L)} \quad (21)$$

is the *information density*. Note that, conditioned on \mathbf{x}^L , the output vectors \mathbf{y}_l , $l = 1, \dots, L$, are independent and Gaussian distributed. The pdf of \mathbf{y}_l is given by

$$\begin{aligned} P_{\mathbf{y} | \mathbf{x}}(\mathbf{y}_l | \mathbf{x}_l) &= \frac{\exp(-\mathbf{y}_l^H (\mathbf{I}_T + \mathbf{x}_l \mathbf{x}_l^H)^{-1} \mathbf{y}_l)}{\pi^T \det(\mathbf{I}_T + \mathbf{x}_l \mathbf{x}_l^H)} \\ &\stackrel{(a)}{=} \frac{1}{\pi^T (1 + \|\mathbf{x}_l\|^2)} \exp\left(-\|\mathbf{y}_l\|^2 + \frac{|\mathbf{y}_l^H \mathbf{x}_l|^2}{1 + \|\mathbf{x}_l\|^2}\right) \end{aligned} \quad (22)$$

where (a) follows from Woodbury's matrix identity [19, p. 19].

To evaluate (20), we choose \mathbf{x}_l , $l = 1, \dots, L$, to be independently and identically distributed according to the i.i.d. unitary distribution $P_{\mathbf{x}}^{(U)}$. The pdf of the corresponding output distribution is equal to

$$q_{\mathbf{y}^L}^{(U)}(\mathbf{y}^L) = \prod_{l=1}^L q_{\mathbf{y}}^{(U)}(\mathbf{y}_l)$$

where $q_{\mathbf{y}}^{(U)}(\cdot)$ is given in (18). Substituting (22) and (18) into (21), we obtain

$$i(\mathbf{x}^L; \mathbf{y}^L) = \sum_{l=1}^L i(\mathbf{x}_l; \mathbf{y}_l) \quad (23)$$

where

$$\begin{aligned} i(\mathbf{x}_l; \mathbf{y}_l) &= \log \frac{1 + T\rho}{\Gamma(T)} + \frac{|\mathbf{y}_l^H \mathbf{x}_l|^2}{1 + \|\mathbf{x}_l\|^2} - \frac{T\rho \|\mathbf{y}_l\|^2}{1 + T\rho} \\ &+ (T-1) \log \frac{T\rho \|\mathbf{y}_l\|^2}{1 + T\rho} - \log(1 + \|\mathbf{x}_l\|^2) \\ &- \log \tilde{\gamma} \left(T-1, \frac{T\rho \|\mathbf{y}_l\|^2}{1 + T\rho} \right) + T-1. \end{aligned}$$

Due to the isotropy of both the input distribution $P_{\mathbf{x}^L}^{(U)}$ and the output distribution $Q_{\mathbf{y}^L}^{(U)}$, the distribution of the information density $i(\mathbf{x}^L; \mathbf{y}^L)$ depends on $P_{\mathbf{x}^L}^{(U)}$ only through the distribution of the norm of the inputs \mathbf{x}_l . Furthermore, under $P_{\mathbf{x}^L}^{(U)}$, we have that $\|\mathbf{x}_l\| = \sqrt{T\rho}$ with probability 1, $l = 1, \dots, L$. This allows us to simplify the computation of (20) by choosing an arbitrary input sequence $\mathbf{x}_l = \bar{\mathbf{x}} \triangleq [\sqrt{T\rho}, 0, \dots, 0]^T$, $l = 1, \dots, L$. Substituting (23) into (20), we obtain the desired lower bound on $R^*(n, \epsilon)$ by solving numerically the following maximization problem

$$\underline{R}(n, \epsilon) \triangleq \max \left\{ \frac{1}{n} \log M : M \text{ satisfies (20)} \right\}. \quad (24)$$

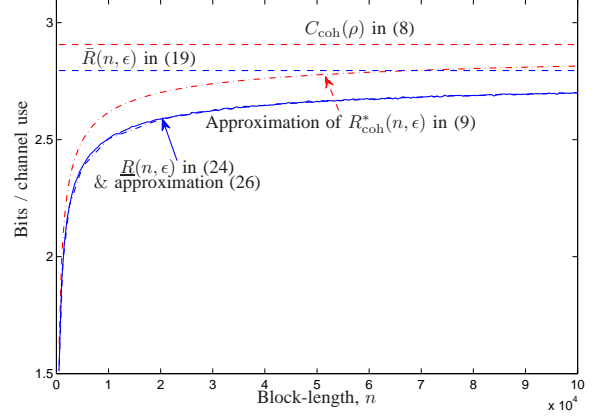


Fig. 2. Bounds on maximal achievable rate $R^*(n, \epsilon)$ for noncoherent Rayleigh block-fading channels; $\rho = 10$ dB, $T = 50$, $\epsilon = 10^{-3}$.

The computation of the DT bound $\underline{R}(n, \epsilon)$ becomes difficult as the block-length n becomes large. We next provide an approximation for $\underline{R}(n, \epsilon)$, which is much easier to evaluate. As in [15, App. A], applying *Berry-Esseen inequality* [14, Thm. 44] to the first term on the RHS of (20), and applying [20, Lemma 20] to the second term on the RHS of (20), we get the following asymptotic expansion for $\underline{R}(n, \epsilon)$

$$\underline{R}(n, \epsilon) = L(\rho) - \sqrt{\frac{\underline{V}(\rho)}{n}} Q^{-1}(\epsilon) + \mathcal{O}\left(\frac{1}{n}\right), n \rightarrow \infty \quad (25)$$

with $\underline{V}(\rho)$ given by

$$\underline{V}(\rho) \triangleq \frac{1}{T} \mathbb{E}_{P_{\mathbf{x}}^{(U)}} [\text{Var}[i(\mathbf{x}; \mathbf{y}) | \mathbf{x}]] = \frac{1}{T} \text{Var}[i(\bar{\mathbf{x}}; \mathbf{y})]$$

where, as in the DT bound, we can choose $\bar{\mathbf{x}} = [\sqrt{T\rho}, 0, \dots, 0]^T$. By neglecting the $\mathcal{O}(1/n)$ term in (25), we arrive at the following approximation for $\underline{R}(n, \epsilon)$

$$\underline{R}(n, \epsilon) \approx L(\rho) - \sqrt{\frac{\underline{V}(\rho)}{n}} Q^{-1}(\epsilon). \quad (26)$$

Although the term $\underline{V}(\rho)$ in (26) needs to be computed numerically, the computational complexity of (26) is much lower than that of the DT bound $\underline{R}(n, \epsilon)$.

D. Numerical Results and Discussions

In Fig. 2, we plot the upper bound $\bar{R}(n, \epsilon)$ in (19), the lower bound $\underline{R}(n, \epsilon)$ in (24), the approximation of $\underline{R}(n, \epsilon)$ in (26), and the approximation of $R_{\text{coh}}^*(n, \epsilon)$ in (9) as a function of the block-length n for $T = 50$, $\epsilon = 10^{-3}$ and $\rho = 10$ dB. For reference, we also plot the coherent capacity $C_{\text{coh}}(\rho)$ in (8). As illustrated in the figure, (26) gives an accurate approximation of $\underline{R}(n, \epsilon)$.

In Figs. 3 and 4, we plot the upper bound $\bar{R}(n, \epsilon)$ in (19), the lower bound $\underline{R}(n, \epsilon)$ in (24), the approximation of $R_{\text{coh}}^*(n, \epsilon)$ in (9), and the coherent capacity $C_{\text{coh}}(\rho)$ in (8) as a function of the channel's coherence time T for block-lengths $n = 4 \times 10^3$ and $n = 4 \times 10^4$, respectively. We see that, for a given

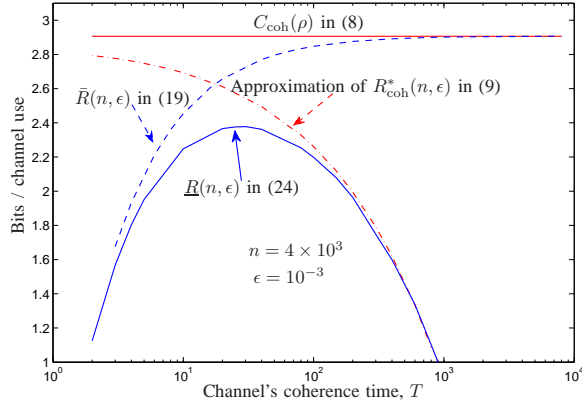


Fig. 3. $\bar{R}(n, \epsilon)$ in (19), $\underline{R}(n, \epsilon)$ in (24), approximation of $R_{\text{coh}}^*(n, \epsilon)$ in (9), and $C_{\text{coh}}(\rho)$ in (8) at block-length $n = 4 \times 10^3$ as a function of the channel's coherence time T for the noncoherent Rayleigh block-fading channel; $\rho = 10$ dB, $\epsilon = 10^{-3}$.

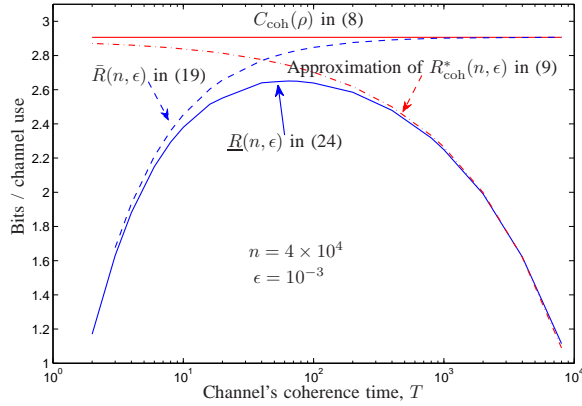


Fig. 4. $\bar{R}(n, \epsilon)$ in (19), $\underline{R}(n, \epsilon)$ in (24), approximation of $R_{\text{coh}}^*(n, \epsilon)$ in (9), and $C_{\text{coh}}(\rho)$ in (8) at block-length $n = 4 \times 10^4$ as a function of the channel's coherence time T for the noncoherent Rayleigh block-fading channel; $\rho = 10$ dB, $\epsilon = 10^{-3}$.

block-length and error probability, $R^*(n, \epsilon)$ is not monotonic in the channel's coherence time, but there exists a channel's coherence time T^* that maximizes $R^*(n, \epsilon)$. This confirms the claim we made in the introduction that there exists a tradeoff between the diversity gain and the cost of estimating the channel when communicating in the noncoherent setting and in the finite block-length regime. A similar phenomenon was observed in [15] for the Gilbert-Elliott channel with no state information at the transmitter and receiver.

From Figs. 3 and 4, we also observe that T^* decreases as

we shorten the block-length. For example, the rate-maximizing channel's coherence time T^* for block-length $n = 4 \times 10^4$ is roughly 64, whereas for block-length $n = 4 \times 10^3$, it is roughly 28.

REFERENCES

- [1] E. Biglieri, J. Proakis, and S. Shamai (Shitz), "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [2] 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.
- [3] 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.
- [4] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "On the capacity of frequency-selective channels in training-based transmission schemes," *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 2572–2583, Sep. 2004.
- [5] A. Lapidoth, "On the asymptotic capacity of stationary gaussian fading channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 2, pp. 437–446, Feb. 2005.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] G. Durisi and H. Bölcskei, "High-SNR capacity of wireless communication channels in the noncoherent setting: A primer," *Int. J. Electron. Commun. (AEÜ)*, vol. 65, no. 8, pp. 707–712, Aug. 2011.
- [10] Y. Polyanskiy and S. Verdú, "Scalar coherent fading channel: dispersion analysis," in *IEEE Int. Symp. Inf. Theory (ISIT)*, Saint Petersburg, Russia, Aug. 2011, pp. 2959–2963.
- [11] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New Jersey: Wiley, 2006.
- [12] N. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed. New York: Wiley, 1995, vol. 1.
- [13] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed. New York: Dover: Government Printing Office, 1972.
- [14] 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.
- [15] —, "Dispersion of the Gilbert-Elliott channel," *IEEE Trans. Inf. Theory*, vol. 57, pp. 1829–1848, 2011.
- [16] A. Lapidoth and S. M. Moser, "Capacity bounds via duality with applications to multiple-antenna systems on flat-fading channels," *IEEE Trans. Inf. Theory*, vol. 49, no. 10, pp. 2426–2467, Oct. 2003.
- [17] M. Katz and S. Shamai (Shitz), "On the capacity-achieving distribution of the discrete-time noncoherent and partially coherent AWGN channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2257–2270, Oct. 2004.
- [18] A. Martinez, "Spectral efficiency of optical direct detection," *J. Opt. Soc. Am.- B*, vol. 24, no. 4, pp. 739–749, Apr. 2007.
- [19] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [20] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Ph.D. dissertation, Princeton University, 2010.