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.
Diversity versus Channel Knowledge
at Finite Block-Length

Wei Yang\textsuperscript{1}, Giuseppe Durisi\textsuperscript{1}, Tobias Koch\textsuperscript{2}, and Yury Polyanskiy\textsuperscript{3}

\textsuperscript{1}Chalmers University of Technology, 41296 Gothenburg, Sweden
\textsuperscript{2}Universidad Carlos III de Madrid, 28911 Leganés, Spain
\textsuperscript{3}Massachusetts 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 \(\epsilon\) over Rayleigh block-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 \textit{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 \textit{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 \textit{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 \textit{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 \(\epsilon\). 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., \(\mathbb{Q}\)) denote probability distributions, while lowercase sans-serif letters (e.g., \(r\)) 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 \(I_T\); the sequence of vectors \(\{a_1, \ldots, a_n\}\) is written as \(a^n\). We denote expectation and variance by \(\mathbb{E}[\cdot]\) and \(\text{Var}[\cdot]\), respectively, and use the notation \(\mathbb{E}_a[\cdot]\) or \(\mathbb{E}_{P_x}[\cdot]\) to stress that expectation is taken with respect to \(x\) with distribution \(P_x\). The relative entropy between two distributions \(P\) and \(Q\) is denoted by \(D(P||Q)\) [11, Sec. 8.5]. For two functions \(f(x)\) and \(g(x)\), the notation \(f(x) = \mathcal{O}(g(x)), \ x \to \infty\), means that \(\limsup_{x \to \infty} |f(x)/g(x)| < \infty\), and \(f(x) = o(g(x)), \ x \to \infty\), means that \(\lim_{x \to \infty} |f(x)/g(x)| = 0\). Furthermore, \(\mathcal{C}(0, \mathbb{R})\) stands for the distribution of a circularly-symmetric com-
plex Gaussian random vector with covariance matrix $R$, and Gamma($\alpha, \beta$) denotes the gamma distribution [12, Ch. 17] with parameters $\alpha$ and $\beta$. 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

$$y_l = s_l x_t + w_l$$  \hspace{1cm} (1)

where $x_t$ and $y_t$ are the input and output signals, respectively, $w_t \sim \mathcal{C}N(0, I_T)$ is the additive noise, and $s_l \sim \mathcal{C}N(0, 1)$ 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 $\{w_l\}$ 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

$$\sum_{l=1}^{L} \|x_l\|^2 \leq LT\rho.$$  \hspace{1cm} (2)

Since the variance of $s_l$ and of the entries of $w_t$ is normalized to one, $\rho$ 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 $\epsilon$. For every fixed $T$ and $\epsilon$, we have

$$\lim_{n \to \infty} R^*(n, \epsilon) = C(\rho) = \frac{1}{T} \sup_{P_x} I(x; y)$$  \hspace{1cm} (3)

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

$$\mathbb{E}[\|x\|^2] \leq T\rho.$$  \hspace{1cm} (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)]

$$L(\rho) = \frac{1}{T} \left( (T - 1) \log(T\rho) - \log(T) - T + \frac{T(1 + \rho)}{1 + T\rho} \right)$$

$$- \frac{1}{T} \int_0^\infty e^{-u} \tilde{\gamma}(T - 1, T\rho u) \left( 1 + \frac{1}{T\rho} \right)^{T-1} x \log(u^{1-T} \tilde{\gamma}(T - 1, T\rho u)) du$$  \hspace{1cm} (5)

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

$$C(\rho) = L(\rho) + o(1), \quad \rho \to \infty.$$  \hspace{1cm} (6)

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

$$R^*(n, \epsilon) \leq R^\text{coh}(n, \epsilon)$$  \hspace{1cm} (6)

where $R^\text{coh}(n, \epsilon)$ denotes the maximal achievable rate for a given block-length $n$ and probability of error $\epsilon$ 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 $R^\text{coh}(n, \epsilon)$:

$$R^\text{coh}(n, \epsilon) = C(\rho) - \sqrt{\frac{V(\rho)}{n}} Q^{-1}(\epsilon)$$

$$+ o\left( \frac{1}{\sqrt{n}} \right), \quad n \to \infty.$$  \hspace{1cm} (7)

Here, $C(\rho)$ is the capacity of the fading channel in the coherent setting, which is given by [1, Eq. (3.3.10)]

$$C(\rho) = \mathbb{E}_x \left[ \log(1 + |s|^2) \right]$$  \hspace{1cm} (8)

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

denotes the $Q$-function, and

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

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

$$R^\text{coh}(n, \epsilon) \approx C(\rho) - \sqrt{\frac{V(\rho)}{n}} Q^{-1}(\epsilon).$$  \hspace{1cm} (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]

$$R^*(n, \epsilon) \leq \frac{C(\rho) + H(\epsilon)/n}{1 - \epsilon}$$  \hspace{1cm} (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 $P_{y|x}$ denote the conditional distribution of $y$ given $x$, and $P_y$ the distribution induced on $y$ by the
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 $[C_{coh}(\rho)$ 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 $q_\nu(y)$ in (13) by the output pdf induced by the i.d. unitary input distribution $P_x^{(U)}$, which is given by

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

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

$$-\mathbb{E}_{P_y}[\log q_\nu(y)] = \log \frac{T(1+\rho)T}{\Gamma(T)} + \frac{T+\mathbb{E}[||y||^2]}{\rho+1} + \mathbb{E}[\log((1+||y||^2)z_1 + z_2)]$$

$$= \log \left( \frac{T(1+\rho)T}{\Gamma(T)} \right) + \frac{T}{\rho+1} + (T-1)\psi(T-1)$$

$$+ \mathbb{E} \left[ \sum_{k=0}^{\infty} \frac{1+(1+||y||^2)^{-k}}{k+T-1} + \frac{||y||^2}{T(1+\rho)} \right]. \quad (15)$$

The first equality in (15) follows because the random variable $||y||^2$ is conditionally distributed as $(1+||x||^2)z_1 + z_2$ given $x$, 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 $h(y \mid x)$ is given by

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

we obtain

$$C(\rho) \leq \frac{c_1}{T} + \frac{1}{T} \inf_{\lambda \geq 0} \sup_{P_x} \left\{ \mathbb{E} \left[ \sum_{k=0}^{\infty} \frac{(T-1)(1+||y||^2)^{-k}}{k+T-1} \right] - \log(1+||x||^2) + \frac{||x||^2}{T(1+\rho)} + \lambda(T\rho - ||x||^2) \right\} \quad (16)$$

$$\leq \frac{c_1}{T} + \frac{1}{T} \inf_{\lambda \geq 0} \sup_{||x||} \left\{ \sum_{k=0}^{\infty} \frac{(T-1)(1+||y||^2)^{-k}}{k+T-1} \right\} - \log(1+||x||^2) + \frac{||x||^2}{T(1+\rho)} + \lambda(T\rho - ||x||^2) \right\}$$

$$\triangleq U(\rho) \quad (17)$$

where

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

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

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 $[C_{coh}(\rho)$ in (8)]. We observe that $U(\rho)$ and $L(\rho)$ are surprisingly close for all values of $T$.

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
where \( \mathbf{x} \) is the information density. Note that, conditioned on \( \mathbf{x} \), the output vectors \( \mathbf{y}_l \), \( l = 1, \ldots, L \), are independent and Gaussian distributed. The pdf of \( \mathbf{y}_1 \) is given by

\[
p_{\mathbf{y} | \mathbf{x}}(\mathbf{y} | \mathbf{x}_1) = \frac{\exp(-\mathbf{y}^H (\mathbf{1}_T + \mathbf{x}_1 \mathbf{x}_1^H)^{-1} \mathbf{y} \mathbf{1}_T)}{\pi^T \det(\mathbf{1}_T + \mathbf{x}_1 \mathbf{x}_1^H)}
\]

(22)

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

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

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

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

\[
i(\mathbf{x}^L; \mathbf{y}^L) = \sum_{l=1}^{L} i(\mathbf{x}_l; \mathbf{y}_l)
\]

(23)

where

\[
i(\mathbf{x}_l; \mathbf{y}_l) = \log \frac{1 + T \rho}{\Gamma(T)} + \log \frac{\mathbf{y}_l^H \mathbf{x}_l^2}{1 + \mathbf{x}_l^H \mathbf{x}_l} - T \rho \mathbf{y}_l^2 - T \rho \log(1 + T \rho)
\]

\[
+ \log \frac{T \rho \mathbf{y}_l^2}{1 + T \rho} - T \rho \log(1 + \mathbf{x}_l^H \mathbf{x}_l)
\]

\[
- \log \left( T - 1, \frac{T \rho \mathbf{y}_l^2}{1 + T \rho} \right) + T - 1.
\]

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

\[
\hat{R}(n, \epsilon) \triangleq \max \left\{ \frac{1}{n} \log M : M \text{ satisfies (20)} \right\}.
\]

D. Numerical Results and Discussions

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

In Figs. 3 and 4, we plot the upper bound \( \hat{R}(n, \epsilon) \) in (19), the lower bound \( \hat{R}(n, \epsilon) \) in (24), the approximation of \( R_{coh}^*(n, \epsilon) \) in (9), and the coherent capacity \( C_{coh}(\rho) \) 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
coherence time $T$ in the channel’s coherence time, but there exists a channel’s 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


