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.

<table>
<thead>
<tr>
<th>Citation</th>
<th>Wei Yang, Giuseppe Durisi, Tobias Koch, and Yury Polyanskiy. &quot;Diversity Versus Channel Knowledge at Finite Block-length.&quot; In 2012 IEEE Information Theory Workshop, 3-7 September 2012: Lausanne, Switzerland, Pp. 572–576.</th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dx.doi.org/10.1109/ITW.2012.6404740">http://dx.doi.org/10.1109/ITW.2012.6404740</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Institute of Electrical and Electronics Engineers</td>
</tr>
<tr>
<td>Version</td>
<td>Author’s final manuscript</td>
</tr>
<tr>
<td>Accessed</td>
<td>Sun Dec 30 00:51:19 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/79040">http://hdl.handle.net/1721.1/79040</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Creative Commons Attribution-Noncommercial-Share Alike 3.0</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td><a href="http://creativecommons.org/licenses/by-nc-sa/3.0/">http://creativecommons.org/licenses/by-nc-sa/3.0/</a></td>
</tr>
</tbody>
</table>
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

\textbf{Abstract—}We study the maximal achievable rate $R^\star(n, \epsilon)$ 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, $R^\star(n, \epsilon)$ 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.

\section{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 \textit{coherent setting}) is independent of the fading dynamics \cite{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 \cite{2}–\cite{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 \textit{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) \cite{5}.

In this paper, we model the fading dynamics using the well-known block-fading model \cite{6}–\cite{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 \cite{6}. In \cite{7}–\cite{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 $R^\star(n, \epsilon)$ for a given block-length $n$ and block error probability $\epsilon$. By studying $R^\star(n, \epsilon)$ 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 \cite{10}, because faster fading dynamics yield larger diversity gain.

We expect that the maximal achievable rate $R^\star(n, \epsilon)$ over fading channels in the \textit{noncoherent setting} and in the \textit{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 $R^\star(n, \epsilon)$ 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.

\textbf{Notation:} Uppercase boldface letters denote matrices and lower case boldface letters designate vectors. Uppercase sans-serif letters (e.g., $Q$) 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 $\mathbb{E}_w[\cdot]$ or $\mathbb{E}_{p_x}[\cdot]$ to stress that expectation is taken with respect to $w$ with distribution $P_w$. The relative entropy between two distributions $P$ and $Q$ is denoted by $D(P||Q)$ \cite[Sec. 8.5]{11}. 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 $\limsup_{x \to \infty} |f(x)/g(x)| = 0$. Furthermore, $\mathcal{CN}(0, \mathbf{R})$ 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 $i$th coherence interval, the channel input-output relation can be written as

$$y_i = s_i x_i + w_i$$  \hspace{1cm} (1)

where $x_i$ and $y_i$ are the input and output signals, respectively, $w_i \sim \mathcal{CN}(0, I_T)$ is the additive noise, and $s_i \sim \mathcal{CN}(0, 1)$ models the fading, whose realization we assume is not known at the transmitter and receiver (noncoherent setting). In addition, we assume that $\{s_i\}$ and $\{w_i\}$ 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

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

Since the variance of $s_i$ and of the entries of $w_i$ 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\footnote{The subscript $l$ is omitted whenever immaterial.}

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

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

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

$$\text{and} \quad \int_{0}^{\infty} u^{-2} \frac{\Gamma(T - 1, T \rho u)}{1 + T \rho} \left( 1 + \frac{T}{T \rho} \right)^{T - 1} \times \log(u^{-1} \gamma(T - 1, T \rho u)) du \quad \text{(5)}$$

where

$$\gamma(n, x) \triangleq \frac{1}{\Gamma(n)} \int_{0}^{x} t^{n - 1} e^{-t} dt$$

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 $\mathbb{C}^T$. We shall denote this input distribution as $\mathcal{P}^{(U)}_x$. 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 $\{s_i\}$. Specifically, we have that

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

where $R^*_\epsilon(n, \epsilon)$ 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 $R^*_{\text{coh}}(n, \epsilon)$:

$$R^*_{\text{coh}}(n, \epsilon) = C_{\text{coh}}(\rho) - \sqrt{\frac{V_{\text{coh}}(\rho)}{n}} Q^{-1}(\epsilon)$$

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

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

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

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

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

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

$$R^*_{\text{coh}}(n, \epsilon) \approx C_{\text{coh}}(\rho) - \sqrt{\frac{V_{\text{coh}}(\rho)}{n}} Q^{-1}(\epsilon).$$  \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 $H(\rho) = -x \log x - (1 - x) \log(1 - x)$ 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 $\mathcal{P}_{y|x}$ denote the conditional distribution of $y$ given $x$, and $\mathcal{P}_y$ denote the distribution induced on $y$ by the
The same bounding technique was previously used in [17] to obtain upper bounds on the capacity of the phase-noise AWGN channel’s coherence time $T$. Since $y$ is i.i.d. and $\|y\|^2 \sim \text{Gamma}(1, T(1 + \rho))$. Substituting (14) into $E_{P_y}[\log q_y(y)]$ in (13), we obtain

$$\begin{align*}
- E_{P_y}[\log q_y(y)] &= \log \frac{T(1 + \rho)\pi T}{\Gamma(T)} + T + E[\|x\|^2] \\
&= \log \frac{T(1 + \rho)\pi T}{\Gamma(T)} + T + \mathbb{E}[\|x\|^2] \\
&= \log \frac{T(1 + \rho)\pi T}{\Gamma(T)} + \mathbb{E}[\|x\|^2] \\
&= \log \left( T - 1 \right) \sum_{k=0}^{\infty} \frac{(1 + k + \|x\|^2)^{-k}}{k + T - 1} + \mathbb{E}[\|x\|^2].
\end{align*}$$

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

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

we obtain

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

$$\leq \frac{(a) c_1}{T} + \frac{1}{T} \sup_{\lambda \geq 0} \mathbb{E}[\|x\|^2] \\
= \frac{c_1}{T} + \frac{1}{T} \mathbb{E}[\|x\|^2] + \frac{\|x\|^2}{\rho} + \lambda(T - 1)\psi(T - 1) \leq \bar{U}(\rho)$$

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 $T$.

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

$$q_y^{(U)}(y) = \frac{e^{-\|y\|^2/(1 + \rho T)}}{\pi T (1 + \rho T)} \times \frac{\left(1 + \rho T \|y\|^2\right)^{T - 1}}{\left(1 + \frac{1}{\rho T}\right)^{T - 1}}.$$

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

$$R^*(n, \epsilon) \leq \bar{R}(n, \epsilon) \triangleq \frac{U(\rho) + H(\epsilon)/n}{1 - \epsilon}.$$
where there exists a code with $M$ codewords and average probability of error not exceeding [14, Thm. 17]

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

where

$$
i(x^L; y^L) \triangleq \log \frac{p_{y^L | x^L}(y^L | x^L)}{p_{y^L}(y^L)} \tag{21}
$$

is the information density. Note that, conditioned on $x^L$, the output vectors $y_l, l = 1, \ldots, L$, are independent and Gaussian distributed. The pdf of $y_l$ is given by

$$
p_{y | x}(y_l | x_l) = \frac{\exp(-y_H^T (I_T + x_l x_l^H)^{-1} y_l)}{\pi^L \det(I_T + x_l x_l^H)}
\frac{1}{\pi^L (1 + \|x_l\|^2)} \exp\left(-\|y_l\|^2 + \frac{|x_l^H x_l|^2}{1 + \|x_l\|^2}\right) \tag{22}
$$

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

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

$$
q_{yL}^{(U)}(y^L) = \prod_{l=1}^{L} q_{yL}^{(U)}(y_l)
$$

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

$$
i(x^L; y^L) = \sum_{l=1}^{L} i(x_l; y_l) \tag{23}
$$

where

$$
i(x_l; y_l) = \log \frac{1 + T \rho}{1 + T \rho} + \frac{|x_l^H x_l|^2}{1 + \|x_l\|^2} - T \rho \|y_l\|^2
+ (T - 1) \log \frac{T \rho \|y_l\|^2}{1 + T \rho} - \log(1 + \|x_l\|^2)
- \log \bar{\gamma} \left( T - 1, \frac{T \rho \|y_l\|^2}{1 + T \rho} \right) + T - 1.
$$

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

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

The computation of the DT bound $R(n, \epsilon)$ becomes difficult as the block-length $n$ becomes large. We next provide an approximation for $R(n, \epsilon)$, 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 $R(n, \epsilon)$

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

with $V(\rho)$ given by

$$
V(\rho) \triangleq \frac{1}{T} \mathbb{E}_{P_{xL}^{(U)}} \left[ \text{Var}[i(x_l; y_l) | x_l] \right] = \frac{1}{T} \text{Var}[i(\bar{x}; y_l)]
$$

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

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

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

**D. Numerical Results and Discussions**

In Fig. 2, we plot the upper bound $\bar{R}(n, \epsilon)$ in (19), the lower bound $\bar{R}(n, \epsilon)$ in (24), the approximation of $\bar{R}(n, \epsilon)$ 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 $\bar{R}(n, \epsilon)$.

In Figs. 3 and 4, we plot the upper bound $\bar{R}(n, \epsilon)$ in (19), the lower bound $\bar{R}(n, \epsilon)$ 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
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


