Minimum Energy to Send k Bits over Rayleigh-Fading Channels

Wei Yang[1], Giuseppe Durisi[1], and Yury Polyanskiy[2]
[1]Chalmers University of Technology, 41296 Gothenburg, Sweden
[2]Massachusetts Institute of Technology, Cambridge, MA, 02139 USA

***Abstract*—This paper investigates the minimum energy required to transmit, with a given reliability, k information bits over a stationary memoryless Rayleigh-fading channel, under the assumption that neither the transmitter nor the receiver have *a priori* channel state information (CSI). It is well known that the ratio between the minimum energy per bit and the noise level converges to -1.59 dB as k goes to infinity, regardless of whether CSI is available at the receiver or not. This paper shows that lack of CSI at the receiver causes a slowdown in the speed of convergence to -1.59 dB as $k \to \infty$ compared to the case of perfect receiver CSI. Specifically, we show that in the no-CSI case, the gap to -1.59 dB is proportional to $((\log k)/k)^{1/3}$, whereas when perfect CSI is available at the receiver, this gap is proportional to $1/\sqrt{k}$. Numerically, we observe that to achieve an energy per bit of -1.5 dB in the no-CSI case, one needs to transmit at least 7×10^7 information bits, whereas 6×10^4 bits suffice for the case of perfect CSI at the receiver (same number of bits as for nonfading AWGN channels). Interestingly, all results (asymptotic and numerical) are unchanged if multiple transmit antennas and/or block fading is assumed.**

I. Introduction

A classic result in information theory is that, for a wide class of channels including AWGN channels and fading channels, the minimum energy per information bit required for reliable communication satisfies [1], [2]

$$\frac{E_{\mathrm{b}}}{N_0}_{\min} = \log_e 2 = -1.59\,\mathrm{dB} \tag{1}$$

where E_{b} is the energy per bit, and N_0 is the noise power per *complex* degree of freedom. For fading channels with unitary average channel gain, this result holds regardless of whether the instantaneous fading realizations are known to the receiver or not [2, Th. 1], [3].[1]

The expression in (1) is asymptotic in several aspects:

- the number of degrees of freedom n is infinite;
- the number of information bits k, or equivalently, the number of messages $M = 2^k$ is infinite;
- the error probability ϵ vanishes;
- the total energy E is infinite;
- the rate k/n vanishes.

The limit in (1) does not change if we allow the error probability to be positive. However, keeping any of the other parameters fixed results in a backoff from (1) [2], [4]–[7].

In this paper, we study the maximum number of information bits k that can be transmitted with a finite energy E and a fixed error probability $\epsilon > 0$ over a stationary memoryless Rayleigh-fading channel, when there is no constraint on the number of degrees of freedom n. Equivalently, we determine the minimum energy E required to transmit k information bits with error probability ϵ. Our analysis is targeted to sensor networks, where energy constraints are often more stringent than bandwidth constraints, and where data packets are usually short. Furthermore, we assume that neither the transmitter nor the receiver have *a priori* knowledge of the realizations of the fading channel, but both know its statistics perfectly. The assumption of no *a priori* channel state information (CSI) allows us to assess the cost of learning the fading channel in a fundamental fashion.

Related work: For nonfading AWGN channels with unlimited number of degrees of freedom, Polyanskiy, Poor, and Verdú [7] showed that the maximum number of codewords $M^*(E, \epsilon)$ that can be transmitted with energy E and error probability ϵ satisfies[2]

$$\log M^*(E,\epsilon) = \frac{E}{N_0}\log e - \sqrt{\frac{2E}{N_0}}Q^{-1}(\epsilon)\log e + \frac{1}{2}\log\frac{E}{N_0} + \mathcal{O}(1), \qquad E \to \infty. \tag{2}$$

Here, $Q^{-1}(\cdot)$ denotes the inverse of the Gaussian Q-function. The first term on the right-hand side (RHS) of (2) gives the -1.59 dB limit. The second term captures the penalty due to the stochastic variations of the channel. This term plays the same role as the *channel dispersion* in finite-blocklength analyses [6], [8]. In terms of minimum energy per bit $E_{\mathrm{b}}^*(k,\epsilon)$, the asymptotic expansion (2) implies that, for large E,

$$\frac{E_{\mathrm{b}}^*(k,\epsilon)}{N_0} \approx \log_e 2 + \sqrt{\frac{2\log_e 2}{k}}Q^{-1}(\epsilon) \tag{3}$$

i.e., the gap to -1.59 dB is proportional to $1/\sqrt{k}$.

Moving to fading channels, for the case of no receiver CSI (no-CSIR), *flash signalling* [2, Def. 2] must be used to reach the -1.59 dB limit [2]. In the presence of a finite peak-power constraint, (1) can not be achieved [9]–[12]. Verdú [2] studied the rate of convergence of the minimum energy per bit to -1.59 dB as the signal-to-noise ratio vanishes. He showed that, differently from the perfect CSIR setup, in the no-CSIR case, the -1.59 dB limit is approached with zero *wideband slope*. This implies that operating close to the -1.59 dB limit is expensive in terms of bandwidth. For the case of finite blocklength n, fixed energy budget E, and fixed probability

This work was supported in part by the Swedish Research Council (VR) under grant no. 2012-4571, and in part by the National Science Foundation CAREER award under grant agreement CCF-12-53205.

[1]Knowledge of the fading realizations at the transmitter may improve (1), because it enables the transmitter to signal along the channel maximum-eigenvalue eigenspace [2].

[2]Unless otherwise indicated, the log and the exp functions in this paper are taken with respect to an arbitrary fixed base.

of error ϵ, bounds and approximations for the maximum channel coding rate over fading channels (under various CSI assumptions) are reported in [13]–[16].

Contributions: Focusing on the regime of unlimited number of degrees of freedom n but finite energy E, and nonzero error probability ϵ, we provide upper and lower bounds on the maximum number of codewords $M^*(E,\epsilon)$ for the case of stationary memoryless Rayleigh-fading channels with no CSI at transmitter and receiver. An asymptotic analysis of these bounds reveals that

$$\log M^*(E,\epsilon) = \frac{E}{N_0}\log e - V_0(\epsilon)\left(\frac{E}{N_0}\right)^{\frac{2}{3}} \log^{\frac{1}{3}}\frac{E}{N_0} + \mathcal{O}\left(\frac{E^{\frac{2}{3}}\log\log E}{\log^{\frac{2}{3}} E}\right), \qquad E\to\infty \quad (4)$$

where $V_0(\epsilon) > 0$ is given in (39). In terms of minimum energy per bit, (4) implies that, for large E,

$$\frac{E_{\mathrm{b}}^*(k,\epsilon)}{N_0} \approx \log_e 2 + V_0(\epsilon)\log_e^{\frac{4}{3}} 2\cdot\left(\frac{\log_e k}{k}\right)^{\frac{1}{3}} \qquad (5)$$

i.e., the gap to -1.59 dB is proportional to $((\log_e k)/k)^{1/3}$.

For the case of perfect CSIR, we prove that the asymptotic expansion of $M^*(E,\epsilon)$ coincides with the one given in (2) for the AWGN case up to the third-order term. By comparing (2) with (4), we see that, although the minimum energy per bit approaches (1) as E increases regardless of whether CSIR is available or not, the convergence is slower for the no-CSIR case. Our nonasymptotic bounds reveal that to achieve an energy per bit of -1.5 dB, one needs to transmit at least 7×10^7 information bits in the no-CSIR case, whereas 6×10^4 bits suffice in the perfect CSIR case. Furthermore, we observe that it takes 2 dB more of energy to transmit 1000 information bits in the no-CSIR case compared to the perfect CSIR case.

The results (asymptotic and numerical) provided in this paper continue to hold for multiple-input multiple-output Rayleigh *block-fading* channels (up to a suitable normalization of the received energy) [17]. Due to space limitations, we omit the proofs of some of the results provided in this paper. The proofs can be found in [17].

II. Problem Formulation

We consider a single-antenna stationary memoryless Rayleigh-fading channel with input-output relation

$$V_i = H_i u_i + Z_i, \quad i = 1,2,\dots \qquad (6)$$

Here, $u_i \in \mathbb{C}$ and $V_i \in \mathbb{C}$ are the input and output of the channel; both the $\{H_i\}$ and the $\{Z_i\}$ are i.i.d. $\mathcal{CN}(0,N_0)$-distributed; furthermore, $\{H_i\}$ and $\{Z_i\}$ are independent. We assume that neither the transmitter nor the receiver have *a priori* knowledge of the realizations of $\{H_i\}$, but both know the statistics of $\{H_i\}$. In the remainder of the paper, we shall assume that $N_0 = 1$, for notational convenience. Moreover, we shall denote infinite-dimensional vectors such as $[A_1, A_2, \dots]$ and $[b_1, b_2, \dots]$ by $\boldsymbol{A}$ and $\boldsymbol{b}$, respectively.

Given $\boldsymbol{U} = \boldsymbol{u}$, the output $\boldsymbol{V}$ of the channel (6) follows a circularly symmetric Gaussian distribution

$$P_{\boldsymbol{V}\,|\,\boldsymbol{U}=\boldsymbol{u}} = \prod_{i=1}^{\infty}\mathcal{CN}(0,(1+|u_i|^2)). \qquad (7)$$

Since $\boldsymbol{V}$ depends on the input symbols $\{u_i\}$ only through their squared magnitude $\{|u_i|^2\}$, without loss of generality (w.l.o.g.) we can reduce the input space to $\mathbb{R}_+^\infty$. We also note that the $\{|V_i|^2\}$ are a sufficient statistics for the detection of $\boldsymbol{u}$ from $\boldsymbol{V}$. Letting $x_i \triangleq |u_i|^2$ and $Y_i \triangleq |V_i|^2$, $i=1,2,\dots$, we obtain the following equivalent input-output relation

$$Y_i = (1+x_i)S_i, \quad i=1,2,\dots \qquad (8)$$

where x_i and Y_i are nonnegative real numbers, and $\{S_i\}$ are i.i.d. $\mathrm{Exp}(1)$-distributed.[3] In the remainder of the paper, we shall focus on the equivalent channel (8). Since $\|\boldsymbol{x}\|_1 = \|\boldsymbol{u}\|_2^2$ and $\|\boldsymbol{x}\|_\infty = \|\boldsymbol{u}\|_\infty^2$, we shall measure the energy and the peakiness of an input codeword $\boldsymbol{x}$ for the channel (8) by its ℓ_1-norm $\|\boldsymbol{x}\|_1$ and ℓ_∞-norm $\|\boldsymbol{x}\|_\infty$, respectively.

Definition 1: An (E,M,ϵ)-code for the channel (8) consists of a set of codewords $\{\boldsymbol{c}_1,\dots,\boldsymbol{c}_M\}\in(\mathbb{R}_+^\infty)^M$ satisfying the energy constraint

$$\|\boldsymbol{c}_j\|_1 \le E, \quad j\in\{1,\dots,M\} \qquad (9)$$

and a decoder $g:\mathbb{R}_+^\infty \to \{1,\dots,M\}$ satisfying the maximum error probability constraint

$$\max_{j\in\{1,\dots,M\}} \mathbb{P}[g(\boldsymbol{Y})\neq j\,|\,\boldsymbol{X}=\boldsymbol{c}_j] \le \epsilon \qquad (10)$$

where $\boldsymbol{Y}$ is the output vector induced by the codeword $\boldsymbol{X}=\boldsymbol{c}_j$ according to (8). The maximum number of messages that can be transmitted with energy E and error probability ϵ is

$$M^*(E,\epsilon) \triangleq \max\bigl\{M : \exists\,(E,M,\epsilon)\text{-code}\bigr\}. \qquad (11)$$

III. Main Results

A. General Achievability Bounds

Below, we present the $\kappa\beta$ and the random-coding union (RCU) versions of Verdú's achievability bound [4, pp. 1023–1024] on the capacity per unit cost of memoryless channels with a zero-cost input symbol (which we label as "0"). As in [4], we use $b[x]$ to denote the cost of the symbol x in the input alphabet $\mathcal{X}$.

Theorem 1 ($\kappa\beta$, capacity per unit cost): Consider a stationary memoryless channel $P_{Y|X}$. For every $N\in\mathbb{N}$, $0<\epsilon<1$, and every input symbol $x_0\in\mathcal{X}$ such that $b[x_0]>0$, there exists an (E,M,ϵ)-code with $E=b[x_0]N$ and

$$M-1 \ge \sup_{0<\tau<\epsilon}\frac{\tau}{\beta_{1-\epsilon+\tau}(P_{Y|X=x_0}^{\otimes N}, P_{Y|X=0}^{\otimes N})} \qquad (12)$$

where $\beta_{(\cdot)}(\cdot,\cdot)$ is the function given in [6, Eq. (100)], and

$$P_{Y|X=x}^{\otimes N} \triangleq \underbrace{P_{Y|X=x}\times\cdots\times P_{Y|X=x}}_{N\text{ times}}. \qquad (13)$$

Proof: As in [4, p. 1023], we choose the set of codewords $\{\boldsymbol{c}_1,\dots,\boldsymbol{c}_M\}\in\left(\mathcal{X}^\infty\right)^M$ as follows:

$$\boldsymbol{c}_j \triangleq [\underbrace{0,\dots,0}_{(j-1)N},\underbrace{x_0,\dots,x_0}_{N},0,\dots], \quad j=1,\dots,M. \qquad (14)$$

Fix $0<\tau<\epsilon$. For a given received signal $\boldsymbol{Y}\in\mathcal{Y}^\infty$, the decoder runs M parallel binary hypothesis tests Z_j between

[3]We use $\mathrm{Exp}(\mu)$ to denote the exponential distribution with mean μ.

$P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{0}}$ and $P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{c}_j}$, where $Z_j = 1$ indicates that the test selects $P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{c}_j}$. The tests $\{Z_j\}$ are chosen to satisfy[4]

$$\mathbb{P}[Z_j = 1\,|\,\boldsymbol{X} = \boldsymbol{c}_j] \geq 1 - \epsilon + \tau \tag{15}$$

$$\mathbb{P}[Z_j = 1\,|\,\boldsymbol{X} = \boldsymbol{0}] = \beta_{1-\epsilon+\tau}(P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{c}_j}, P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{0}}) \tag{16}$$

for all $j = 1, \dots, M$. The decoder outputs the index m if both $Z_m = 1$ and $Z_j = 0$ for all $j \neq m$. It outputs 1 if no such index can be found.

By construction, the maximum probability of error ϵ of the code just defined is upper-bounded by

$$\begin{aligned}\epsilon &\leq \mathbb{P}[Z_1 = 0\,|\,\boldsymbol{X} = \boldsymbol{c}_1] + (M-1)\mathbb{P}[Z_1 = 1\,|\,\boldsymbol{X} = \boldsymbol{0}] \quad (17)\\ &\leq \epsilon - \tau + (M-1)\beta_{1-\epsilon+\tau}(P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{c}_1}, P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{0}}) \quad (18)\end{aligned}$$

where (17) follows because for each optimal test Z_j, $j \neq 1$,

$$\begin{aligned}\mathbb{P}[Z_j = 1|\boldsymbol{X} = \boldsymbol{c}_1] &= \mathbb{P}[Z_j = 1|\boldsymbol{X} = \boldsymbol{0}] \quad (19)\\ &= \mathbb{P}[Z_1 = 1|\boldsymbol{X} = \boldsymbol{0}] \quad (20)\end{aligned}$$

and (18) follows from (15) and (16). From (18), we conclude that

$$M - 1 \geq \frac{\tau}{\beta_{1-\epsilon+\tau}(P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{c}_1}, P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{0}})}. \tag{21}$$

The proof is completed by noting that, for every $\alpha \in (0,1)$,

$$\beta_\alpha(P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{c}_1}, P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{0}}) = \beta_\alpha(P^{\otimes N}_{Y|X=x_0}, P^{\otimes N}_{Y|X=0}) \tag{22}$$

and by maximizing the RHS of (21) over $0 < \tau < \epsilon$. ∎

Remark: A slightly weakened version of (21), with $M-1$ replaced by M, follows from the $\kappa\beta$ bound [6, Th. 25] upon setting $Q_{\boldsymbol{Y}} = P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{0}}$ and choosing F_N as

$$\mathsf{F}_N = \Big\{\boldsymbol{x} \in \mathcal{X}^\infty : \boldsymbol{x} = [\underbrace{0, \dots, 0}_{kN}, \underbrace{x_0, \dots, x_0}_{N}, 0, \dots] \text{ for some } k \in \mathbb{N}\Big\}. \tag{23}$$

Indeed, it suffices to observe that $\kappa_\tau(\mathsf{F}_N, P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{0}}) = \tau$.

Using the same codebook as in Theorem 1 together with a maximum likelihood decoder, we obtain the following achievability bound.

Theorem 2 (RCU, capacity per unit cost): Consider a stationary memoryless channel $P_{Y|X}$. For every $N \in \mathbb{N}$, $0 < \epsilon < 1$, and every input symbol $x_0 \in \mathcal{X}$ such that $b[x_0] > 0$, there exists an (E, M, ϵ)-code satisfying $E = b[x_0]N$ and

$$\epsilon \leq \mathbb{E}\Big[\min\Big\{1, (M-1)\mathbb{P}\big[\imath_N(x_0; Y^N) \leq \imath_N(x_0; \bar{Y}^N)\,|\,Y^N\big]\Big\}\Big] \tag{24}$$

where $P_{Y^N\bar{Y}^N}(a^N, b^N) \triangleq P^{\otimes N}_{Y|X=x_0}(a^N)P^{\otimes N}_{Y|X=0}(b^N)$ and

$$\imath_N(x; y^N) \triangleq \log \frac{dP^{\otimes N}_{Y|X=x}}{dP^{\otimes N}_{Y|X=0}}(y^N). \tag{25}$$

In the AWGN case, the bound (2) takes the same value for all $N \in \mathbb{N}$. Hence, it suffices to set $N = 1$ (see [7]).

[4]The Neyman-Pearson lemma [18] implies the existence of tests that satisfy (15) and (16).

B. Nonasymptotic Bounds

Particularizing Theorems 1 and 2 to the channel (8) and setting $x_0 = E/N$, we obtain the following achievability bounds.

Corollary 3: For every $E > 0$, and $0 < \epsilon < 1$, there exists an (E, M, ϵ)-code for the channel (8) satisfying

$$M - 1 \geq \sup_{0<\tau<\epsilon, N\in\mathbb{N}} \frac{\tau}{\mathbb{P}[G_N \geq (1 + E/N)\xi_{N,\tau}]} \tag{26}$$

where $G_N \sim \mathrm{Gamma}(N, 1)$ and $\xi_{N,\tau}$ satisfies

$$\mathbb{P}[G_N \leq \xi_{N,\tau}] = \epsilon - \tau. \tag{27}$$

Corollary 4: For every $M > 0$ and $0 < \epsilon < 1$, there exists an (E, M, ϵ)-code for the channel (8) satisfying

$$\epsilon \leq \min_{N\in\mathbb{N}} \mathbb{E}\Big[\min\Big\{1, (M-1)\mathbb{P}\Big[\bar{G}_N \geq (1 + E/N)G_N \Big| G_N\Big]\Big\}\Big] \tag{28}$$

where G_N and $\bar{G}_N$ are i.i.d. $\mathrm{Gamma}(N, 1)$ random variables.

Numerical evidence (provided in Section III-E) suggests that (28) is tighter than (26). However, (26) turns out to be more suitable for asymptotic analyses.

On the converse side, we have the following result, which is based on the meta-converse theorem [6, Th. 31].

Theorem 5: Let $\{S_i\}$ be i.i.d. $\mathrm{Exp}(1)$ random variables. Then, every (E,M,ϵ)-code for the channel (8) satisfies

$$\frac{1}{M} \geq \sup_{\eta\in\mathbb{R}} \frac{\inf_{\boldsymbol{x}} \mathbb{P}\Big[\sum_{i=1}^{\infty}\Big(x_i S_i \log e - \log(1 + x_i)\Big) \leq \eta\Big] - \epsilon}{\exp(\eta)} \tag{29}$$

where the infimum is over the subset of $\boldsymbol{x} \in \mathbb{R}^\infty_+$ taking one of the following two forms:

$$\boldsymbol{x} = [q_3, \underbrace{q_2, \dots, q_2}_{N}, q_1, 0, 0, \dots] \tag{30}$$

$$\boldsymbol{x} = [\underbrace{\tilde{q}_2, \dots, \tilde{q}_2}_{\widetilde{N}_2}, \underbrace{\tilde{q}_1, \dots, \tilde{q}_1}_{\widetilde{N}_1}, 0, 0, \dots]. \tag{31}$$

Here, $N \in \mathbb{N}$ and $0 < q_1 < q_2 < q_3$ satisfy $q_1 + Nq_2 + q_3 = E$; furthermore, $\widetilde{N}_1, \widetilde{N}_2 \in \mathbb{N}$ and $0 \leq \tilde{q}_1 \leq \tilde{q}_2$ satisfy $\widetilde{N}_1\tilde{q}_1 + \widetilde{N}_2\tilde{q}_2 = E$.

Proof: We assume w.l.o.g. that each codeword $\boldsymbol{c}_j$ satisfies the energy constraint (9) with equality. We use the meta-converse bound [6, Th. 31] with the auxiliary output distribution $Q_{\boldsymbol{Y}} = P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{0}}$. This results in

$$\frac{1}{M} \geq \inf_{\boldsymbol{x}\in\mathbb{R}^\infty_+ : \|\boldsymbol{x}\|_1 = E} \beta_{1-\epsilon}(P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{x}}, P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{0}}). \tag{32}$$

Next, we lower-bound $\beta_{1-\epsilon}$ using [6, Eq. (102)]. Specifically, we fix an arbitrary $\eta \in \mathbb{R}$ and obtain

$$\begin{aligned}&\beta_{1-\epsilon}(P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{x}}, P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{0}})\\ &\quad\geq \exp(-\eta)\Big(P_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{x}}[\imath(\boldsymbol{x}, \boldsymbol{Y}) \leq \eta] - \epsilon\Big) \quad (33)\end{aligned}$$

where

$$\imath(\boldsymbol{x}, \boldsymbol{y}) \triangleq \log \frac{dP_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{x}}}{dP_{\boldsymbol{Y}\,|\,\boldsymbol{X}=\boldsymbol{0}}}(\boldsymbol{y}). \tag{34}$$

Under $P_{\boldsymbol{Y} \mid \boldsymbol{X}=\boldsymbol{x}}$, the random variable $\imath(\boldsymbol{x},\boldsymbol{Y})$ has the same distribution as

$$\sum_{i=1}^{\infty}\Big(x_i S_i \log e - \log(1+x_i)\Big). \tag{35}$$

Substituting (35) into (33), and then (33) into (32), we obtain

$$\frac{1}{M} \geq \frac{\inf_{\boldsymbol{x}} \mathbb{P}\left[\sum_{i=1}^{\infty}\Big(x_i S_i \log e - \log(1+x_i)\Big) \leq \eta\right] - \epsilon}{\exp(\eta)} \tag{36}$$

where the infimum is over all $\boldsymbol{x} \in \mathbb{R}_{+}^{\infty}$ that satisfy $\|\boldsymbol{x}\|_1 = E$. Lemma 6 below provides a partial characterization of the solution to the minimization problem on the RHS of (36).

Lemma 6: Let $\boldsymbol{x}^*$ be a minimizer of

$$\inf_{\boldsymbol{x}\in\mathbb{R}_{+}^{\infty}:\|\boldsymbol{x}\|_1=E} \mathbb{P}\left[\sum_{i=1}^{\infty}\Big(x_i S_i \log e - \log(1+x_i)\Big) \leq \eta\right]. \tag{37}$$

Assume w.l.o.g. that the entries of $\boldsymbol{x}^*$ are in nonincreasing order. Then, $\boldsymbol{x}^*$ must take the form (30) or (31).

The proof of (29) is concluded by using Lemma 6 in (36) and by maximizing the RHS of (36) over η. ∎

Remark: Lemma 6 reduces the infinite-dimensional optimization problem (37) into a 3-dimensional one, which can be solved numerically. The proof of Lemma 6 relies on an elegant argument of Abbe, Huang, and Telatar [19], used in the proof of Telatar's minimum outage probability conjecture for multiple-input single-output Rayleigh-fading channels. Indeed, both [19] and Lemma 6 deal with optimization of quantiles of convolutions of exponential distributions.

C. Asymptotic Analysis

Evaluating the bounds in Corollary 3 and Theorem 5 for large E, we obtain the following closed-form characterization of $M^*(E,\epsilon)$ for large E.

Theorem 7: The maximum number of messages $M^*(E,\epsilon)$ that can be transmitted with energy E and error probability $\epsilon \in (0,1/2)$ over the channel (8) admits the following expansion

$$\log M^*(E,\epsilon) = E\log e - V_0(\epsilon) E^{\frac{2}{3}} \log^{\frac{1}{3}} E + \mathcal{O}\left(\frac{E^{\frac{2}{3}}\log\log E}{\log^{\frac{2}{3}} E}\right) \tag{38}$$

where

$$V_0(\epsilon) = \left(12^{-\frac{1}{3}} + \left(\frac{2}{3}\right)^{\frac{1}{3}}\right)\left(Q^{-1}(\epsilon)\log e\right)^{\frac{2}{3}}. \tag{39}$$

The intuition behind (38) is as follows. It is well known that in the no-CSIR case, to achieve the asymptotic limit $-1.59\,\mathrm{dB}$, it is necessary to use flash signalling, i.e., signals with unbounded peak power [2]. Indeed, if all codewords satisfy a peak-power constraint $\|\boldsymbol{x}\|_\infty \leq A$ in addition to (9), then $\log M(E,\epsilon)/E$ converges to (see [11] and [12, Eq. (59)])

$$\log e - A^{-1}\log(1+A) \tag{40}$$

as $E\to\infty$. The second term in (40) can be interpreted as the penalty due to bounded peakiness, which vanishes as $A\to\infty$. When the energy E is finite,

$$\frac{\log M(E,\epsilon)}{E} \approx \log e - \frac{\log(1+A)}{A} - \sqrt{\frac{A}{E}}Q^{-1}(\epsilon)\log e. \tag{41}$$

The second term on the RHS of (41) captures the fact that codewords that satisfy (9) for a finite E are necessarily peak-power limited (we denote again the peak power by A). The third term captures the penalty resulting from the stochastic variations of the fading channel and the noise processes, which cannot be averaged out for finite E. Since the channel is multiplicative (see (8)), this penalty increases with the peak power. Indeed, peakier codewords result in less channel averaging. To summarize, peakiness in the codewords reduces the second term on the RHS of (41) but increases the third term. The optimal peak power A^* that minimizes the sum of these two penalties in (41) turns out to be

$$A^* = \left(\frac{3}{2}Q^{-1}(\epsilon)\log e\right)^{-\frac{2}{3}} E^{\frac{1}{3}} \log^{\frac{2}{3}} E + o(E^{\frac{1}{3}}). \tag{42}$$

Substituting (42) into (41) we obtain (38). See [17] for a rigorous proof.

D. Perfect CSIR

To assess the penalty due to lack of CSIR, we provide in this section achievability and converse bounds for the case of perfect CSIR.

Theorem 8: Every (E,M,ϵ)-code for the AWGN channel can be converted into an (E,M,ϵ)-code for the fading channel (6) with perfect CSIR.

Remark: Theorem 8 continues to hold also if the fading is not Rayleigh, provided that the coefficients $\{H_i\}$ are i.i.d. and satisfy $\mathbb{E}\big[|H_i|^2\big] = 1$.

Proof: Take an arbitrary (E,M,ϵ)-code for the AWGN channel. We can assume w.l.o.g. that only the first M coordinates of each codeword are nonzero . This is because for the AWGN channel, the error probability of the optimal decoder depends only on the Euclidean distance between codewords. Therefore, we can embed the M codewords in an M-dimensional space while preserving the distance between codewords. We assume that the decoder is maximum likelihood. Then, the decoding regions are M-dimensional Voronoi regions, and can be further assumed open.

Now, we simulate an AWGN channel via a fading channel with perfect CSIR as follows: fix an $N > 0$; for every codeword $\boldsymbol{u} = [u_1,\ldots,u_M,0,\ldots]$ for the AWGN channel, we generate the following codeword $\tilde{\boldsymbol{u}}$ for the fading channel (6)

$$\tilde{\boldsymbol{u}} \triangleq \frac{1}{\sqrt{N}}[\underbrace{u_1,\ldots,u_1}_{N},\underbrace{u_2,\ldots,u_2}_{N},\ldots]. \tag{43}$$

By construction, $\boldsymbol{u}$ and $\tilde{\boldsymbol{u}}$ have the same energy. For a given channel output $\boldsymbol{V}$ (see (6)), the receiver computes

$$\widetilde{V}_j = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\bar{H}_{(j-1)N+i}V_{(j-1)N+i} \tag{44}$$

$$= \frac{u_j\sum_{i=1}^{N}|H_{(j-1)N+i}|^2}{N} + \frac{\bar{H}_j Z_j}{\sqrt{N}}, \quad j=1,\ldots,M \tag{45}$$

where $\bar{\cdot}$ stands for the complex conjugate. As $N\to\infty$, the first term in (45) converges in distribution to u_j by the law of large numbers. The second term converges in distribution to $Z_j \sim \mathcal{CN}(0,1)$ by the central limit theorem. Therefore, $\widetilde{V}_j$ converges in distribution to $u_j + Z_j$. Thus, our progressively

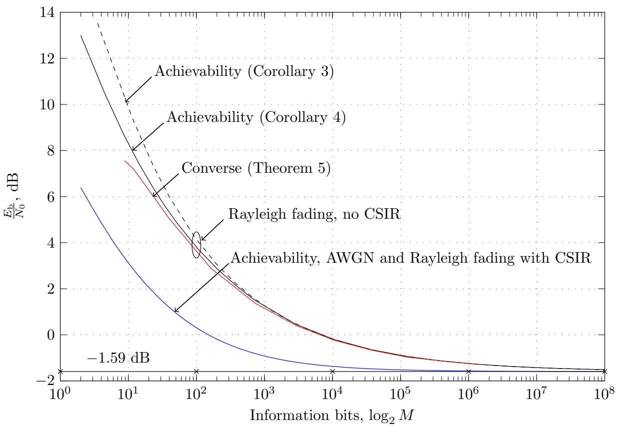

Fig. 1. Minimum energy per bit versus number of information bits; here $\epsilon = 10^{-3}$.

TABLE I
MINIMUM ENERGY E AND OPTIMAL NUMBER OF CHANNEL USES N^* VS. NUMBER OF INFORMATION BITS k; HERE, $\epsilon = 10^{-3}$.

	Cor. 3		Cor. 4		Asymptotics	
k	E/N_0	N^*	E/N_0	N^*	E/N_0	N^*
10^1	98	25	67	18	120	50
10^2	2.6×10^2	39	2.4×10^2	38	2.9×10^2	63
10^3	1.3×10^3	96	1.3×10^3	96	1.4×10^3	124
10^4	9.6×10^3	304	9.6×10^3	304	9.7×10^3	336
10^5	8.2×10^4	1089	8.2×10^4	1090	8.2×10^4	1137

improving simulations are such that $P_{\widetilde{V}^M | U^M = u^M}$ converges (in distribution) to an AWGN channel law. Since the decoding regions are open, the probability of correct decoding for the fading case converges to that for the AWGN case. ■

Theorem 8 implies that the asymptotic expansion (2) is achievable also for fading channels with perfect CSIR. Theorem 9 below establishes that, for $0 < \epsilon < 1/2$, the converse holds as well.

Theorem 9: In the presence of perfect CSIR, the maximum number of messages $M^*(E, \epsilon)$ that can be transmitted with energy E and error probability $0 < \epsilon < 1/2$ over the channel (6) satisfies

$$\log M^*(E,\epsilon) = E \log e - \sqrt{2E} Q^{-1}(\epsilon) \log e + \frac{1}{2}\log E + \mathcal{O}(\sqrt{\log E}), \qquad E \to \infty. \quad (46)$$

Note that, unlike Theorem 8, Theorem 9 requires the fading to be Rayleigh distributed.

E. Numerical Results

Fig. 1 shows the achievability bounds (Corollaries 3 and 4) and the converse bound (Theorem 5) for the channel (8) (no CSIR) for the case $\epsilon = 10^{-3}$. Specifically, the energy per bit $E_\mathrm{b} = E/\log_2 M^*(E,\epsilon)$ (for the case $N_0 = 1$) is plotted against the number of information bits $\log_2 M^*(E,\epsilon)$. For reference, we also plot the achievability bound corresponding to the AWGN case [7, Eq. (15)]. As proved in Theorem 8, this bound is also achievable in the Rayleigh-fading case when perfect CSIR is available. As expected, as the number of information bits increases, the minimum energy per bit converges to the same limit (-1.59 dB) regardless of whether CSIR is available or not. However, for a fixed number of information bits, it is more costly to communicate in the no-CSIR case than in the perfect-CSIR case. For example, it takes 2 dB more of energy to transmit 1000 information bits in the no-CSIR case compared to the perfect-CSIR case. Additionally, to achieve an energy per bit of -1.5 dB, we need to transmit 7×10^7 information bits in the no-CSI case, but only 6×10^4 bits when perfect CSIR is available.

The codebook used in Corollaries 3 and 4 uses only one symbol in the input alphabet in addition to 0. In Table I we list the number of channel uses $N^* = E/x_0^*$ that the optimal input symbol x_0^* occupies to minimize the energy per bit, as a function of the number of information bits k. For comparison, we also list the number of channel uses N^* used in the asymptotic analysis in Theorem 7.

REFERENCES

[1] C. E. Shannon, "Communication in the presense of noise," *Proc. IRE*, vol. 37, pp. 10–21, 1949.

[2] S. Verdú, "Spectral efficiency in the wideband regime," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, pp. 1319–1343, Jun. 2002.

[3] R. S. Kennedy, *Fading Dispersive Communication Channels*. New York: Willy-Interscience, 1969.

[4] S. Verdú, "On channel capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 36, no. 5, pp. 1019–1030, Sep. 1990.

[5] L. Zheng, D. N. C. Tse, and M. Médard, "Channel coherence in the low-SNR regime," *IEEE Trans. Inf. Theory*, vol. 53, no. 3, pp. 976–997, Mar. 2007.

[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] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Minimum energy to send k bits through the Gaussian channel with and without feedback," *IEEE Trans. Inf. Theory*, vol. 57, no. 8, pp. 4880–4902, Aug. 2011.

[8] V. Y. F. Tan, "Asymptotic estimates in information theory with non-vanishing error probabilities," in *Foundations and Trends in Communications and Information Theory*. Delft, The Netherlands: now Publishers, 2014, vol. 11, no. 1–2, pp. 1–184.

[9] İ. E. Telatar, "Capacity and mutual information of wideband multipath fading channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 4, pp. 1384–1400, Jul. 2000.

[10] M. Médard, "Bandwidth scaling for fading multipath channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 4, pp. 840–852, 2002.

[11] V. Sethuraman and B. Hajek, "Capacity per unit energy of fading channels with peak constraint," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3102–3120, Sep. 2005.

[12] G. Durisi, U. G. Schuster, H. Bölcskei, and S. Shamai (Shitz), "Noncoherent capacity of underspread fading channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 367–395, Jan. 2010.

[13] Y. Polyanskiy and S. Verdú, "Scalar coherent fading channel: dispersion analysis," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Saint Petersburg, Russia, Aug. 2011, pp. 2959–2963.

[14] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Diversity versus channel knowledge at finite block-length," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Lausanne, Switzerland, Sep. 2012, pp. 577–581.

[15] ——, "Quasi-static multiple-antenna fading channels at finite block-length," *IEEE Trans. Inf. Theory*, vol. 60, no. 7, pp. 4232–4265, Jul. 2014.

[16] A. Collins and Y. Polyanskiy, "Orthogonal designs optimize achievable dispersion for coherent MISO channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Honolulu, HI, USA, Jul. 2014.

[17] W. Yang, G. Durisi, and Y. Polyanskiy, "Optimal energy and spectral efficiency for the transmission of k bits over multi-antenna fading channels," 2015, in preparation.

[18] J. Neyman and E. S. Pearson, "On the problem of the most efficient tests of statistical hypotheses," *Philosophical Trans. Royal Soc. A*, vol. 231, pp. 289–337, 1933.

[19] E. Abbe, S.-L. Huang, and İ. E. Telatar, "Proof of the outage probability conjecture for MISO channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2596–2602, May 2013.