Beta-Beta Bounds: Finite-Blocklength Analog
of the Golden Formula

Wei Yang, *Member, IEEE*, Austin Collins, Giuseppe Durisi *Senior Member, IEEE*,
Yury Polyanskiy *Senior Member, IEEE*, and H. Vincent Poor *Fellow, IEEE*

Abstract

It is well known that the mutual information between two random variables can be expressed as
the difference of two relative entropies that depend on an auxiliary distribution, a relation sometimes
referred to as the golden formula. This paper is concerned with a finite-blocklength extension of this
relation. This extension consists of two elements: (i) a finite-blocklength channel-coding converse bound
by Polyanskiy and Verdú (2014), which involves the ratio between two Neyman-Pearson β functions
(beta-beta converse bound), and (ii) a novel beta-beta channel-coding achievability bound, expressed
again as the ratio between two Neyman-Pearson β functions.

To demonstrate the usefulness of this finite-blocklength extension of the golden formula, the beta-
beta achievability and converse bounds are used to obtain a finite-blocklength extension of Verdú's
(2002) wideband-slope approximation. The proof parallels the elegant derivation in Verdú (2002), with
the beta-beta bounds used in place of the golden formula.

W. Yang is with Qualcomm Technologies, Inc., San Diego, CA 92121 USA (email:weiyang@qti.qualcomm.com).

A. Collins and Y. Polyanskiy are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: austinc@mit.edu; yp@mit.edu).

G. Durisi is with the Department of Electrical Engineering, Chalmers University of Technology, Gothenburg 41296, Sweden (e-mail: durisi@chalmers.se).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

The beta-beta (achievability) bound is also shown to be useful in cases where the capacity-achieving output distribution is not a product distribution due to, e.g., a cost constraint or structural constraints on the codebook, such as orthogonality or constant composition. As an example, the bound is used to characterize the channel dispersion of the additive exponential-noise channel and to obtain a finite-blocklength achievability bound (the tightest to date) for multiple-input multiple-output Rayleigh-fading channels with perfect channel state information at the receiver.

I. INTRODUCTION

A. Background

In his landmark 1948 paper [1], Shannon established the noisy channel coding theorem, which expresses the fundamental limit of reliable data transmission over a noisy channel in terms of the *mutual information* $I(X;Y)$ between the input X and the output Y of the channel. More specifically, for stationary memoryless channels, the maximum rate at which data can be transmitted reliably is the channel capacity

$$C = \sup_{P_X} I(X;Y). \tag{1}$$

Here, reliable transmission means that the probability of error can be made arbitrarily small by mapping the information bits into sufficiently long codewords. In the nonasymptotic regime of fixed blocklength (fixed codeword size), there exists a tension between rate and error probability, which is partly captured by the many nonasymptotic (finite-blocklength) bounds available in the literature [2]–[5]. In many of these bounds, the role of the mutual information is taken by the so-called *(mutual) information density*[1]

$$i(X;Y) \triangleq \log \frac{\mathrm{d}P_{XY}}{\mathrm{d}(P_X P_Y)}(X,Y) \tag{2}$$

or information spectrum [6], [7]. While the various properties enjoyed by the mutual information make the evaluation of capacity relatively simple, computing the finite-blocklength bounds that involve the information density (e.g., the information-spectrum bounds [2]–[4]) can be very challenging.

---

[1]In this paper, $\log$ and $\exp$ functions are taken with respect to the same arbitrary basis.

One well-known property of the mutual information is that it can be expressed as a difference of relative entropies involving an arbitrary output distribution Q_Y [8, Eq. (8.7)]:

$$I(X;Y) = D(P_{XY}\|P_X Q_Y) - D(P_Y\|Q_Y). \tag{3}$$

Here, $D(\cdot\|\cdot)$ stands for the relative entropy. This identity—also known as the *golden formula* [9, Th. 3.3] or Topsøe's identity [10]—has found many applications in information theory. One canonical application is to establish upper bounds on channel capacity [11]. Indeed, substituting (3) into (1), we get an alternative expression for channel capacity

$$C = \max_{P_X} \big\{ D(P_{XY}\|P_X Q_Y) - D(P_Y\|Q_Y) \big\} \tag{4}$$

from which an upper bound on C can be obtained by dropping the term $-D(P_Y\|Q_Y)$. The golden formula is also used in the derivation of the capacity per unit cost [12] and the wideband slope [13], in the Blahut-Arimoto algorithm [14], [15], and in Gallager's formula for the minimax redundancy in universal source coding [16]. Furthermore, it is useful for characterizing the properties of good channel codes [17], [18], and it is often used in statistics to prove lower bounds on the minimax risk via Fano's inequality (see [19] and [20]).

The main purpose of this paper is to provide a finite-blocklength analog of (4) that is helpful in deriving nonasymptotic results in the same way in which (4) is helpful in the asymptotic case.[2] Note that a naïve way to derive such a finite-blocklength analog would be to rewrite the information density in the information-spectrum bounds as follows:

$$i(X;Y) = \log \frac{\mathrm{d}P_{XY}}{\mathrm{d}(P_X Q_Y)}(X;Y) - \log \frac{\mathrm{d}P_Y}{\mathrm{d}Q_Y}(X;Y). \tag{5}$$

However, the resulting bounds are not very useful, because it is difficult to decouple the two random variables on the RHS of (5). Instead of tweaking the information-spectrum bounds via (5), we derive a finite-blocklength analog of (4) from first principles.

To summarize our contribution, we need to first introduce some notation. We consider an abstract channel that consists of an input set $\mathcal{A}$, an output set $\mathcal{B}$, and a random transformation $P_{Y|X} : \mathcal{A} \to \mathcal{B}$. An (M, ϵ) code for the channel $(\mathcal{A}, P_{Y|X}, \mathcal{B})$ comprises a message set $\mathcal{M} \triangleq \{1, \ldots, M\}$, an encoder $f : \mathcal{M} \to \mathcal{A}$, and a decoder $g : \mathcal{B} \to \mathcal{M} \cup \{e\}$ (e denotes the error

---

[2]With a slight abuse in terminology, we shall refer to both (3) and (4) as the golden formula.

event) that satisfies the *average* error probability constraint

$$\frac{1}{M} \sum_{j=1}^{M} \Big(1 - P_{Y|X}\big(g^{-1}(j) \,|\, f(j)\big) \Big) \le \epsilon. \tag{6}$$

Here, $g^{-1}(j) \triangleq \{y \in \mathcal{Y} : g(y) = j\}$. In practical applications, we often take $\mathcal{A}$ and $\mathcal{B}$ to be n-fold Cartesian products of two alphabets $\mathcal{X}$ and $\mathcal{Y}$, and the channel to be a sequence of conditional probabilities $P_{Y^n|X^n} : \mathcal{X}^n \to \mathcal{Y}^n$. We shall refer to an (M, ϵ) code for the channel $\{\mathcal{X}^n, P_{Y^n|X^n}, \mathcal{Y}^n\}$ as an (n, M, ϵ) code.

Binary hypothesis testing, which we introduce next, will play an important role. Given two probability measures P and Q on a common measurable space $\mathcal{X}$, we define a randomized test between P and Q as a random transformation $P_{Z|X} : \mathcal{X} \to \{0, 1\}$, where 0 indicates that the test chooses Q. The optimal performance achievable among all such randomized tests is given by the Neyman-Pearson function $\beta_\alpha(P, Q)$, which is defined as

$$\beta_\alpha(P, Q) \triangleq \min \int P_{Z|X}(1 \,|\, x) Q(\mathrm{d}x) \tag{7}$$

where the minimum is over all tests satisfying

$$\int P_{Z|X}(1 \,|\, x) P(\mathrm{d}x) \ge \alpha. \tag{8}$$

The Neyman-Pearson lemma [21] provides the optimal test $P_{Z|X}$ that attains the minimum in (7). This test, which we shall refer to as the Neyman-Pearson test, involves thresholding the Radon-Nikodym derivative of P with respect to Q.

B. *Finite-Blocklength Analog of the Golden Formula*

A first step towards establishing a finite-blocklength analog of the golden formula was recently provided by Polyanskiy and Verdú, who proved that every (n, M, ϵ) code satisfies the following converse bound [18, Th. 15]:

$$M \le \inf_{0 \le \delta < 1-\epsilon} \frac{\beta_{1-\delta}(P_{Y^n}, Q_{Y^n})}{\beta_{1-\epsilon-\delta}(P_{X^nY^n}, P_{X^n}Q_{Y^n})}, \quad \forall Q_{Y^n}. \tag{9}$$

Here, P_{X^n} and P_{Y^n} denote the empirical input and output distributions induced by the code for the case of uniformly distributed messages. The proof of (9) is a refinement of the meta-converse theorem [5, Th. 26], where the decoder is used as a suboptimal test for discriminating $P_{X^nY^n}$ against $P_{X^n}Q_{Y^n}$. We shall refer to the converse bound (9) as the $\beta\beta$ converse bound. Note that

the special case of $\delta = 0$, which is known as the minimax meta-converse bound, has a long history in information theory as surveyed in [5, Sec. II.D] for the classical case and [22] for quantum channels.

In this paper, we provide the following achievability counterpart of (9): for every $0 < \epsilon < 1$ and every input distribution P_{X^n}, there exists an (n, M, ϵ) code such that

$$\frac{M}{2} \geq \sup_{0 < \tau < \epsilon} \frac{\beta_\tau(P_{Y^n}, Q_{Y^n})}{\beta_{1-\epsilon+\tau}(P_{X^n Y^n}, P_{X^n} Q_{Y^n})}, \quad \forall Q_{Y^n} \tag{10}$$

where P_{Y^n} denotes the distribution of Y^n induced by P_{X^n} through $P_{Y^n|X^n}$. The proof of the $\beta\beta$ achievability bound above relies on Shannon's random coding technique and on a suboptimal decoder that is based on the Neyman-Pearson test between $P_{X^n Y^n}$ and $P_{X^n} Q_{Y^n}$. Hypothesis testing is used twice in the proof: to relate the decoding error probability to $\beta_{1-\epsilon+\tau}(P_{X^n Y^n}, P_{X^n} Q_{Y^n})$, and to perform a change of measure from P_{Y^n} to Q_{Y^n}. Fig. 1 gives a pictorial summary of the connections between the $\beta\beta$ bounds and various capacity and nonasymptotic bounds. The analogy between the $\beta\beta$ bounds (9)–(10) and the golden formula follows because, for product distributions P^n and Q^n,

$$-\frac{1}{n} \log \beta_\alpha(P^n, Q^n) = D(P\|Q) + o(1), \quad \forall \alpha \in (0, 1) \tag{11}$$

as $n \to \infty$ by Stein's lemma [23, Th. 11.8.3]. For example, one can prove that (4) is achievable using (10) as follows: set $P_{X^n} = (P_X)^n$ and $Q_{Y^n} = (Q_Y)^n$, take the log of both sides of (10), use Stein's lemma and optimize over P_X.

C. Applications

To demonstrate that the $\beta\beta$ bounds (9) and (10) are the natural nonasymptotic equivalent of the golden formula, we use them to provide a finite-blocklength extension of the *wideband slope* approximation by Verdú [13]. More specifically, we derive a second-order characterization of the minimum energy per bit $E_b^*(k, \epsilon, R)$ required to transmit k bits with error probability ϵ and rate R on additive white Gaussian noise (AWGN) channels and also on Rayleigh-fading channels with perfect channel state information available at the receiver (CSIR). Our result implies that $E_b^*(k, \epsilon, R)$ (expressed in dB) can be approximated by an affine function of the rate R. Furthermore, the slope of this function coincides with the wideband slope by Verdú. Our proof parallels the elegant derivation by Verdú [13], except that the role of the golden formula

Asymptotic   Nonasymptotic

$\beta\beta$ bounds:

Golden formula:

$$I(X;Y) = D(P_{XY}||P_X Q_Y)$$
$$-D(P_Y||Q_Y)$$

$\Longleftarrow$

$$\frac{M}{2} \geq \frac{\beta_\tau(P_{Y^n}, Q_{Y^n})}{\beta_{1-\epsilon+\tau}(P_{X^n Y^n}, P_{X^n} Q_{Y^n})}$$

$$M \leq \frac{\beta_{1-\delta}(P_{Y^n}, Q_{Y^n})}{\beta_{1-\epsilon-\delta}(P_{X^n Y^n}, P_{X^n} Q_{Y^n})}$$

$\Downarrow$ $\Downarrow$

Duality bound [11]:

$$C \leq \max_{P_X} D(P_{XY}||P_X Q_Y)$$

$\Longleftarrow$

Minimax converse bound:

$$M \leq \sup_{P_{X^n}} \frac{1}{\beta_{1-\epsilon}(P_{X^n Y^n}, P_{X^n} Q_{Y^n})}$$
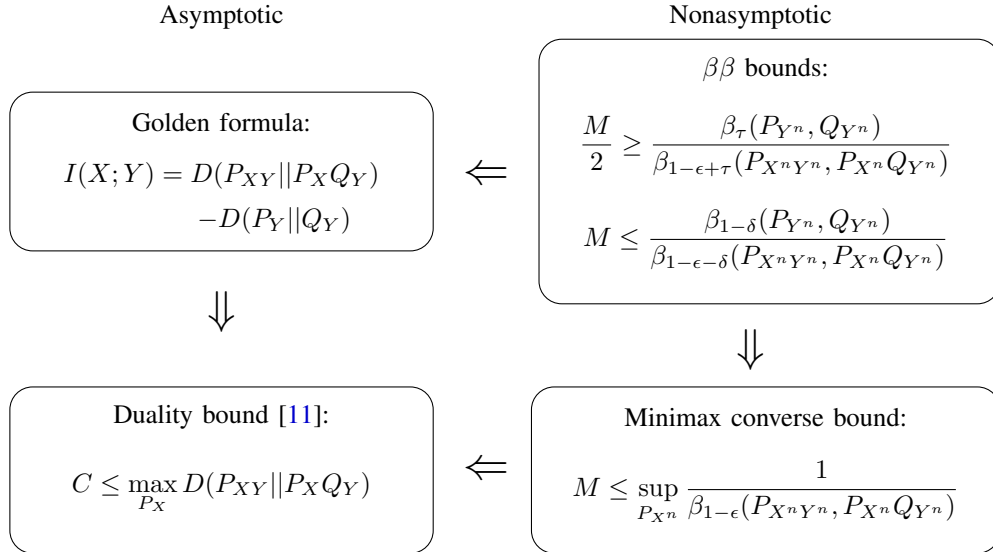
Fig. 1. Connections between the golden formula and nonasymptotic $\beta\beta$ bounds.

is taken by the $\beta\beta$ bounds (9) and (10). Numerical evaluations demonstrate the accuracy of the resulting approximation.

The $\beta\beta$ achievability bound (10) is also useful in situations where P_{Y^n} is not a product distribution (although the underlying channel law $P_{Y^n|X^n}$ is stationary and memoryless), for example due to a cost constraint, or a structural constraint on the channel inputs, such as orthogonality or constant composition. In such cases, traditional achievability bounds such as Feinstein's bound [2] or the dependence-testing (DT) bound [5, Th. 18], which are expressed in terms of the information density, become difficult to evaluate. In contrast, the $\beta\beta$ bound (10) requires the evaluation of $\mathrm{d}P_{Y^n|X^n}/\mathrm{d}Q_{Y^n}$, which factorizes when Q_{Y^n} is chosen to be a product distribution. This allows for an analytical computation of (10). Furthermore, the term $\beta_\tau(P_{Y^n}, Q_{Y^n})$, which captures the cost of the change of measure from P_{Y^n} to Q_{Y^n}, can be evaluated or bounded even when a closed-form expression for P_{Y^n} is not available. To illustrate these points:

- We use the $\beta\beta$ achievability bound to characterize the channel dispersion [5, Def. 1] of the additive exponential noise channel introduced in [24].

- We evaluate (10) for the case of multiple-input multiple-output (MIMO) Rayleigh-fading channels with perfect CSIR. In this case, (10) yields the tightest known achievability bound.

D. Notation

For an input distribution P_X and a channel $P_{Y|X}$, we let $P_X \circ P_{Y|X}$ denote the distribution of Y induced by P_X through $P_{Y|X}$. The distribution of a circularly symmetric complex Gaussian random vector with covariance matrix A is denoted by $\mathcal{CN}(0, \mathsf{A})$. We denote by $\chi_k^2(\lambda)$ the noncentral chi-squared distribution with k degrees of freedom and noncentrality parameter λ; furthermore, $\mathrm{Exp}(\mu)$ stands for the exponential distribution with mean μ. The Frobenius norm of a matrix A is denoted by $\|\mathsf{A}\|_{\mathsf{F}}$. For a vector $\boldsymbol{x} = (x_1, \ldots, x_d)$, we let $\|\boldsymbol{x}\|$, $\|\boldsymbol{x}\|_4$, and $\|\boldsymbol{x}\|_\infty$ denote the ℓ_2, ℓ_4, and ℓ_∞ norms of $\boldsymbol{x}$, respectively. The real part and the complex conjugate of a complex number x are denoted by $\mathrm{Re}(x)$ and x^*, respectively. For a set $\mathcal{A}$, we use $|\mathcal{A}|$ and $\mathcal{A}^c$ to denote the set cardinality and the set complement, respectively. Finally, $\lceil \cdot \rceil$ denotes the ceiling function, and the superscript $^{\mathrm{H}}$ stands for Hermitian transposition.

II. THE $\beta\beta$ ACHIEVABILITY BOUND

In this section, we formally state our $\beta\beta$ achievability bound.

Theorem 1 ($\beta\beta$ achievability bound): For every $0 < \epsilon < 1$ and every input distribution P_X, there exists an (M, ϵ) code for the channel $(\mathcal{A}, P_{Y|X}, \mathcal{B})$ satisfying

$$\frac{M}{2} \geq \sup_{0 < \tau < \epsilon} \sup_{Q_Y} \frac{\beta_\tau(P_Y, Q_Y)}{\beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y)} \tag{12}$$

where $P_Y = P_X \circ P_{Y|X}$.

Proof: Fix $\epsilon \in (0, 1)$, $\tau \in (0, \epsilon)$, and let P_X and Q_Y be two arbitrary probability measures on $\mathcal{A}$ and $\mathcal{B}$, respectively. Furthermore, let

$$M = \left\lceil \frac{2\beta_\tau(P_Y, Q_Y)}{\beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y)} \right\rceil. \tag{13}$$

Finally, let $P_{Z|XY} : \mathcal{A} \times \mathcal{B} \to \{0, 1\}$ be the Neyman-Pearson test that satisfies

$$P_{XY}[Z(X, Y) = 1] \geq 1 - \epsilon + \tau \tag{14}$$

$$P_X Q_Y[Z(X, Y) = 1] = \beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y). \tag{15}$$

In words, (X, Y) may be distributed either according to P_{XY} or according to $P_X Q_Y$, and $Z(X, Y) = 1$ means that the test chooses P_{XY}. For a given codebook $\{c_1, \ldots, c_M\}$ and a received signal y, the decoder computes the values of $Z(c_j, y)$ and returns the smallest index j for which $Z(c_j, y) = 1$. If no such index is found, the decoder declares an error. The average

probability of error of the given codebook $\{c_1, \ldots, c_M\}$, under the assumption of uniformly distributed messages, is given by

$$P_{\text{e}}(c_1, \ldots, c_M) = \mathbb{P}\bigg[\big\{Z(c_W, Y) = 0\big\} \cup \bigg(\bigcup_{m<W}\big\{Z(c_m, Y) = 1\big\}\bigg)\bigg] \tag{16}$$

where W is equiprobable on $\{1, \ldots, M\}$ and $Y \sim P_{Y|W}$.

Following Shannon's random coding idea, we next average (16) over all codebooks $\{C_1, \ldots, C_M\}$ whose codewords are generated as pairwise independent random variables with distribution P_X. This yields

$$\mathbb{E}[P_{\text{e}}(C_1, \ldots, C_M)]$$

$$\leq \mathbb{P}\big[Z(X, Y) = 0\big] + \mathbb{P}\bigg[\max_{m<W} Z(C_m, Y) = 1\bigg] \tag{17}$$

$$\leq \epsilon - \tau + \mathbb{P}\bigg[\max_{m<W} Z(C_m, Y) = 1\bigg]. \tag{18}$$

Here, (17) follows from the union bound and (18) follows from (14).

To conclude the proof of (12), it suffices to show that

$$\mathbb{P}\bigg[\max_{m<W} Z(C_m, Y) = 1\bigg] \leq \tau \tag{19}$$

for the M defined in (13), where the probability is computed with respect to $Y \sim P_Y$. The idea of the proof is to relate this probability to a probability computed with respect to Y Q_Y via binary hypothesis testing. Consider the randomized test $P_{\widetilde{Z}|Y} : \mathcal{Y} \to \{0, 1\}$

$$\widetilde{Z}(y) \triangleq \max_{m<W} Z(C_m, y) \tag{20}$$

where, as before, W is uniformly distributed on $\{1, \ldots, M\}$, and the $\{C_m\}$ are pairwise independent random variables with distribution P_X. Here, the random variable Y may be distributed either according to P_Y or according to Q_Y, and $\widetilde{Z}(Y) = 1$ indicates that the test chooses P_Y. It follows that

$$\beta_{P_Y[\widetilde{Z}=1]}(P_Y, Q_Y) \leq Q_Y[\widetilde{Z}(Y) = 1] \tag{21}$$

$$\leq \frac{1}{M}\sum_{j=1}^{M}(j-1)P_X Q_Y[Z(X, Y) = 1] \tag{22}$$

$$= \frac{M-1}{2}P_X Q_Y[Z(X, Y) = 1] \tag{23}$$

$$= \frac{M-1}{2}\beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y) \tag{24}$$

$$\leq \beta_\tau(P_Y, Q_Y). \tag{25}$$

Here, (21) follows from (7); (22) follows from (20) and from the union bound; (24) follows from (15); and (25) follows from (13). Since the function $\alpha \mapsto \beta_\alpha(P_Y, Q_Y)$ is nondecreasing, we conclude that

$$P_Y[\widetilde{Z} = 1] \leq \tau \tag{26}$$

which is equivalent to (19). ∎

III. RELATION TO EXISTING ACHIEVABILITY BOUNDS

We next discuss the relation between Theorem 1 and the achievability bounds available in the literature.

1) The $\kappa\beta$ bound [5, Th. 25]: The $\kappa\beta$ bound is based on Feinstein's maximal coding approach and on a suboptimal threshold decoder. By further lower-bounding the κ term in the $\kappa\beta$ bound using [25, Lemma 4], we can relax it to the following bound:

$$M \geq \sup_{0 < \tau < \epsilon} \sup_{Q_Y} \frac{\beta_\tau(P_X \circ P_{Y|X}, Q_Y)}{\sup_{x \in \mathcal{F}} \beta_{1-\epsilon+\tau}(P_{Y|X=x}, Q_Y)} \tag{27}$$

which holds under the *maximum* error probability constraint (cf. (6))

$$\max_{j \in \{1,...,M\}} \left\{ 1 - P_{Y|X}\big(g^{-1}(j) \,|\, f(j)\big) \right\} \leq \epsilon. \tag{28}$$

Here, $\mathcal{F} \subset \mathcal{A}$ denotes the permissible set of codewords, and P_X is an arbitrary distribution on $\mathcal{F}$. Because of the similarity between (27) and (12), one can interpret the $\beta\beta$ bound as the average-error-probability counterpart of the $\kappa\beta$ bound.[3] For the case in which $\beta_\alpha(P_{Y|X=x}, Q_Y)$ does not depend on $x \in \mathcal{F}$, by relaxing $M/2$ to M in (12) and by using [5, Lemma 29] we recover (27) under the weaker average-error-probability formalism. However, for the case in which $\beta_\alpha(P_{Y|X=x}, Q_Y)$ does depend on $x \in \mathcal{F}$, the $\beta\beta$ achievability bound (12) can be both easier to analyze and numerically tighter than the $\kappa\beta$ bound (27) (see Section V-B for an example).

---

[3]In fact, the analogy between the $\kappa\beta$ bound and the $\beta\beta$ bound is also clear without applying the relaxation (27). Indeed, the κ term in the $\kappa\beta$ bound [5, Th. 25] defines a relative measure of performance for the composite hypothesis testing between the collection $\{P_{Y|X=x}\}_{x \in \mathcal{F}}$ and Q_Y. The $\beta_\tau(P_Y, Q_Y)$ term in the $\beta\beta$ bound measures the performance of a binary hypothesis testing between P_Y and Q_Y, where the distribution P_Y is an average of the collection of distributions $\{P_{Y|X=x}\}_{x \in \mathcal{F}}$.

2) The dependence-testing (DT) bound [5, Th. 18]: Setting $Q_Y = P_Y$ in (12), using that $\beta_\tau(P_Y, P_Y) = \tau$, and rearranging terms, we obtain

$$\epsilon \leq \inf_{\alpha \in (0,1)} \left\{ 1 - \alpha + \frac{M}{2} \beta_\alpha(P_{XY}, P_X P_Y) \right\}. \tag{29}$$

Further setting $\alpha = P_{XY}[\log(\mathrm{d}P_{XY}/\mathrm{d}(P_X P_Y)) \geq \log(M/2)]$ and using the Neyman-Pearson lemma, we conclude that (29) is equivalent to a weakened version of the DT bound where $(M-1)/2$ is replaced by $M/2$. Since this weakened version of the DT bound implies both Shannon's bound [26] and the bound in [27, Th. 2], the $\beta\beta$ achievability bound (12) implies these two bounds as well.

A cost-constrained version of the DT bound in which all codewords belong to a given set $\mathcal{F}$ can be found in [5, Th. 20]. A slightly weakened version of [5, Th. 20] (with $(M-1)/2$ replaced by $M/2$) is

$$\epsilon \leq Q_{XY}\left[\log \frac{\mathrm{d}Q_{XY}}{\mathrm{d}(Q_X Q_Y)}(X, Y) \leq \log \frac{M}{2}\right]$$
$$+ \frac{M}{2} Q_X Q_Y\left[\log \frac{\mathrm{d}Q_{XY}}{\mathrm{d}(Q_X Q_Y)}(X, Y) \geq \log \frac{M}{2}\right] + Q_X[\mathcal{F}^c]. \tag{30}$$

Here, $Q_{XY} = Q_X P_{Y|X}$, and $Q_Y = Q_X \circ P_{Y|X}$. For $0 < \epsilon < 1/2$, this bound can be derived from (12) by choosing $P_X[\,\cdot\,] = Q_X[\,\cdot\, \cap \mathcal{F}]/Q_X[\mathcal{F}]$ and by using the following bounds:

$$\beta_\tau(P_Y, Q_Y) \geq \beta_\tau(P_X, Q_X) = \tau Q_X[\mathcal{F}] \tag{31}$$

$$\beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y) \leq \frac{1}{Q_X[\mathcal{F}]} \beta_{1-(\epsilon-\tau)Q_X[\mathcal{F}]}(Q_{XY}, Q_X Q_Y). \tag{32}$$

Here, (31) follows from the data-processing inequality for $\beta_\tau(\cdot, \cdot)$ (see, e.g., [28, Sec. V]) and straightforward computations, and (32) follows from [29, Lemma 25].

3) Refinements of the DT/Feinstein bound through change of measure: The idea of bounding the probability $P_Y[\cdot]$ via a simpler-to-analyze $Q_Y[\cdot]$ has been applied previously in the literature to evaluate the DT bound and Feinstein's bound. For example, the following achievability bound is suggested in [30, Sec. II]:

$$\epsilon \leq \inf_{\gamma > 0} \left(P_{XY}\left[\frac{\mathrm{d}P_{XY}}{\mathrm{d}(P_X Q_Y)}(X, Y) \leq \gamma\right] + M \left(\sup_y \frac{\mathrm{d}P_Y}{\mathrm{d}Q_Y}(y)\right) P_X Q_Y\left[\frac{\mathrm{d}P_{XY}}{\mathrm{d}(P_X Q_Y)}(X, Y) \geq \gamma\right]\right) \tag{33}$$

which is equivalent to

$$M \geq \sup_{0 < \tau < \epsilon} \frac{\tau}{\beta_{1-\epsilon+\tau}(P_{XY}, P_X Q_Y)} \left(\sup_y \frac{\mathrm{d}P_Y}{\mathrm{d}Q_Y}(y)\right)^{-1}. \tag{34}$$

This bound can be obtained from the $\beta\beta$ achievability bound (12) by using that

$$\beta_\tau(P_Y, Q_Y) \geq \tau \left(\sup_y \frac{\mathrm{d}P_Y}{\mathrm{d}Q_Y}(y) \right)^{-1}. \tag{35}$$

The following variation of the Feinstein bound is used in [31, Lemma 2] to deal with inputs subject to a cost constraint:

$$\epsilon \leq \inf_{\gamma>0,\eta>0} \left(P_{XY}\left[\frac{\mathrm{d}P_{XY}}{\mathrm{d}(P_X Q_Y)}(X,Y) \leq \gamma\eta \right] + \frac{M}{\gamma} + P_Y\left[\frac{\mathrm{d}P_Y}{\mathrm{d}Q_Y}(Y) \geq \eta \right] \right). \tag{36}$$

This bound can be obtained from (12) by using [5, Eq. (103)] to lower-bound $\beta_\tau(P_Y, Q_Y)$ and by using [5, Eq. (102)] to upper-bound $\beta_\tau(P_{XY}, P_X Q_Y)$.

IV. WIDEBAND SLOPE AT FINITE BLOCKLENGTH

A. Background

Many communication systems (such as deep-space communication and sensor networks) operate in the low-power regime, where both the spectral efficiency and the energy per bit are low. As shown in [13], the key asymptotic performance metrics for additive noise channels in the low-power regime are the normalized minimum energy per bit $\frac{E_b}{N_0}_{\min}$ (where N_0 denotes the noise power per channel use) and the slope[4] S_0 of the function spectral efficiency versus energy per bit (in dB) at $\frac{E_b}{N_0}_{\min}$ (a.k.a. the *wideband slope*). These two quantities determine the first-order behavior of the minimum energy per bit $E_b^*(R)$ as a function of the spectral efficiency[5] R (bits/ch. use) in the limit $R \to 0$. Specifically, [13, Eq. (28)]

$$10 \log_{10} \frac{E_b^*(R)}{N_0} = 10 \log_{10} \frac{E_b}{N_0}_{\min} + \frac{R}{S_0} 10 \log_{10} 2 + o(R), \quad R \to 0. \tag{37}$$

For a wide range of power-constrained channels including the AWGN channel and the fading channel with/without CSIR, it is well known that the minimum energy per bit satisfies [13], [32], [33]

$$\frac{E_b}{N_0}_{\min} = \log_e 2 = -1.59\,\mathrm{dB}. \tag{38}$$

Furthermore, it is shown in [13] that $S_0 = 2$ for AWGN channels, and $S_0 = 2\mathbb{E}[|H|^2]^2 / \mathbb{E}[|H|^4]$ for stationary ergodic fading channels with perfect CSIR, where H is distributed as one of the fading coefficients.

---

[4]The unit of S_0 is $\mathrm{bits/ch.\,use}/(3\,\mathrm{dB})$.

[5]We shall use the terms spectral efficiency and rate interchangeably.

The asymptotic expansion (37) is derived in [13] under the assumption that $n, k \to \infty$ with $k/n \to R$ and $\epsilon \to 0$, where k denotes the number of information bits k (i.e., $k = \log_2 M$). Recently, it was shown in [34] that the minimum energy per bit $E_b^*(k, \epsilon)$ necessary to transmit a finite number k of information bits over an AWGN channel with error probability ϵ and with no constraint on the blocklength satisfies

$$\frac{E_b^*(k, \epsilon)}{N_0} = \frac{E_b}{N_{0\,\min}} + \sqrt{\frac{2\log_e 2}{k}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{k}}\right). \tag{39}$$

Furthermore, it was shown in [35] that the expansion (39) is valid also for block-memoryless Rayleigh-fading channels (perfect CSIR), provided that $0 < \epsilon < 1/2$.

In this section, we study the tradeoff between energy per bit and spectral efficiency in the regime where not only k and ϵ, but also the blocklength n is finite. In particular, we are interested in the minimum energy per bit $E_b^*(k, \epsilon, R)$ necessary to transmit k information bits with rate R bits/ch. use and with error probability not exceeding ϵ. The quantity $E_b^*(R)$ in (37) and $E_b^*(k, \epsilon)$ in (39) can be obtained from $E_b^*(k, \epsilon, R)$ as follows:

$$E_b^*(R) = \lim_{\epsilon \to 0} \lim_{k \to \infty} E_b^*(k, \epsilon, R) \tag{40}$$

$$E_b^*(k, \epsilon) = \lim_{R \to 0} E_b^*(k, \epsilon, R). \tag{41}$$

B. AWGN Channel

We consider the complex-valued AWGN channel

$$Y_i = X_i + Z_i, \quad i = 1, \dots, n \tag{42}$$

where $Z^n \sim \mathcal{CN}(0, N_0 I_n)$. We assume that every codeword x^n satisfies the power constraint

$$\|x^n\|^2 = \sum_{i=1}^{n} x_i^2 \le nP. \tag{43}$$

For notational convenience, we shall set $N_0 = 1$. Hence, P in (43) can be interpreted as the SNR at the receiver.

We next evaluate $E_b^*(k, \epsilon, R)$ in the asymptotic regime $k \to \infty$ and $R \to 0$. Motivated by (37), we shall approximate $E_b^*(k, \epsilon, R)$ (expressed in dB) by an affine function of R. The $\beta\beta$ bounds turn out to be key tools to derive the asymptotic approximation given in Theorem 2 below.

Theorem 2: Consider the AWGN channel (42). The following expansion for $E_\mathrm{b}^*(k, \epsilon, R)$ holds in the asymptotic limit $k \to \infty$, $R \to 0$, and $kR \to \infty$:

$$\frac{E_\mathrm{b}^*(k, \epsilon, R)}{N_0} = \log_e 2 + \sqrt{\frac{2 \log_e 2}{k}} Q^{-1}(\epsilon) + \frac{\log_e^2 2}{2} R + o\left(\frac{1}{\sqrt{k}}\right) + o(R). \qquad (44)$$

Equivalently,

$$10 \log_{10} \frac{E_\mathrm{b}^*(k, \epsilon, R)}{N_0} = \underbrace{10 \log_{10} \frac{E_\mathrm{b}}{N_0}{}_{\min} + \frac{R}{S_0} 10 \log_{10} 2}_{\text{wideband slope approximation}}$$
$$+ \frac{10 \log_{10} e}{\sqrt{\log_e 2}} \sqrt{\frac{2}{k}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{k}}\right) + o(R) \qquad (45)$$

where $10 \log_{10} \frac{E_\mathrm{b}}{N_0}{}_{\min} = -1.59\,\mathrm{dB}$, and $S_0 = 2$. If $k \to \infty$, $R \to 0$, but $kR \to c < \infty$, then we have

$$\frac{E_\mathrm{b}^*(k, \epsilon, R)}{N_0} = \log_e 2 + \sqrt{\frac{2 \log_e 2}{k}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{k}}\right). \qquad (46)$$

Equivalently,

$$10 \log_{10} \frac{E_\mathrm{b}^*(k, \epsilon, R)}{N_0} = 10 \log_{10} \frac{E_\mathrm{b}}{N_0}{}_{\min} + \frac{10 \log_{10} e}{\sqrt{\log_e 2}} \sqrt{\frac{2}{k}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{k}}\right). \qquad (47)$$

Proof: See Appendix I. ∎

Remark 1: The expansion (45) can be seen as the finite-blocklength counterpart of (37). Indeed, comparing (45) with (37), we see that

$$10 \log_{10} \frac{E_\mathrm{b}^*(k, \epsilon, R)}{E_\mathrm{b}^*(R)} = \frac{10 \log_{10} e}{\sqrt{\log_e 2}} \sqrt{\frac{2}{k}} Q^{-1}(\epsilon) + o(R) + o\left(\frac{1}{\sqrt{k}}\right). \qquad (48)$$

Thus, in the low spectral efficiency regime, the gap in dB between $E_\mathrm{b}^*(k, \epsilon, R)$ and the asymptotic limit $E_\mathrm{b}^*(R)$ (obtained as $k \to \infty$ and $\epsilon \to 0$) is—up to first order—proportional to $1/\sqrt{k}$ and independent of R. Furthermore, (44) and (39) imply that

$$10 \log_{10} \frac{E_\mathrm{b}^*(k, \epsilon, R)}{E_\mathrm{b}^*(k, \epsilon)} = \frac{R}{S_0} 10 \log_{10} 2 + o(R) + o\left(\frac{1}{\sqrt{k}}\right). \qquad (49)$$

Thus, in the regime of large k, the gap in dB between $E_\mathrm{b}^*(k, \epsilon, R)$ and the asymptotic limit $E_\mathrm{b}^*(k, \epsilon)$ is—up to first order—proportional to R and is independent of k.

Remark 2: The result (46) implies that the minimum energy per bit $E_\mathrm{b}^*(k, \epsilon)$ in (39) with no blocklength constraint can be achieved with codes of rate $1/k$, or equivalently, with codes of blocklength k^2. For comparison, the code used in [34] to achieve (39) has blocklength 2^k.

We now provide some intuition for the expansion (44) (a rigorous proof is given in Appendix I). The asymptotic expression (37) relies on the following Taylor-series expansion of the channel capacity $C(P)$ as a function of the SNR P when $P \to 0$:

$$C(P) = C'(0)P \log e + \frac{C''(0)}{2}P^2 \log e + o(P^2). \tag{50}$$

Here, $C'(0)$ and $C''(0)$ denote the first and second derivative, respectively, of the function $C(P)$ (in nats per channel use). In particular, the first term in (50) determines the minimum energy per bit $\frac{E_b}{N_0\,\text{min}}$ and the second term in (50) yields the wideband slope S_0 [13, Eq. (35) and Th. 9]:

$$\frac{E_b}{N_0\,\text{min}} = \frac{\log_e 2}{C'(0)}, \qquad S_0 = \frac{2\left[C'(0)\right]^2}{-C''(0)}. \tag{51}$$

Both $\frac{E_b}{N_0\,\text{min}}$ and S_0 in (51) can be computed directly, without the knowledge of $C(P)$. Indeed, set $Q_Y = P_{Y|X=0}$ in the golden formula (4). One can show that [13, Eq. (41)]

$$C'(0) = \frac{1}{\log e} \lim_{P \to 0} \max_{P_X : \mathbb{E}[|X|^2]=P} \frac{D(P_{Y|X}\|Q_Y\|P_X)}{P} \tag{52}$$

and that for both AWGN channels and for fading channels with perfect CSIR,

$$C''(0) = -\frac{2}{\log e} \lim_{P \to 0} \min_{P_X} \frac{D(P_Y\|Q_Y)}{P^2} \tag{53}$$

where the minimization in (53) is over all P_X that achieve $C'(0)$ in (52) and that satisfy $\mathbb{E}[|X|^2] = P$. In other words, $C'(0)$ is determined solely by $D(P_{Y|X}\|Q_Y|P_X)$, and $C''(0)$ is determined solely by $D(P_Y\|Q_Y)$.

Moving to the nonasymptotic case, let $R^*(n, \epsilon, P)$ be the maximum coding rate for a given blocklength n, error probability ϵ, and SNR P. Then, (44) turns out to be equivalent to the following asymptotic expression (see Appendix I-A for the proof of this equivalence):

$$\frac{R^*(n, \epsilon, P)}{\log e} = P - \sqrt{\frac{2P}{n}}Q^{-1}(\epsilon) - \frac{1}{2}P^2 + o\left(\sqrt{\frac{P}{n}}\right) + o(P^2) \tag{54}$$

as $n \to \infty$, $P \to 0$, and $nP^2 \to \infty$. In view of (52) and (53), it is natural to use the $\beta\beta$ bounds (9) and (12), since they are nonasymptotic versions of the golden formula. Indeed, we obtain from (9) and (12) that

$$R^*(n, \epsilon, P) \approx \max_{P_{X^n}} \frac{1}{n}\left(-\log \beta_{1-\epsilon}(P_{X^nY^n}, P_{X^n}Q_{Y^n}) + \log \beta_\alpha(P_{Y^n}, Q_{Y^n})\right). \tag{55}$$

Next, we choose $Q_{Y^n} = (P_{Y|X=0})^n$. The analysis in [34, pp. 4882–4883] implies that

$$\max_{P_{X^n}}\left(-\frac{1}{n}\log \beta_{1-\epsilon}(P_{X^nY^n}, P_{X^n}Q_{Y^n})\right) \approx P \log e - \sqrt{\frac{2P}{n}}Q^{-1}(\epsilon)\log e \tag{56}$$

which yields the first two terms in (54).

Furthermore, one can show through a large-deviation analysis that,

$$\max_{P_{X^n}} \frac{1}{n} \log \beta_\alpha(P_{Y^n}, Q_{Y^n}) \approx -D(P_Y^* \| Q_Y) \approx -\frac{\log e}{2} P^2. \tag{57}$$

Here, the maximization in (57) is taken with respect to all input distributions P_{X^n} for which $-\frac{1}{n} \log \beta_{1-\epsilon}(P_{X^n Y^n}, P_{X^n} Q_{Y^n})$ is close to the RHS of (56). Substituting (56) and (57) into (55), we recover the dominant terms in (54).

One may attempt to establish (54) by using the normal approximation [5]

$$R^*(n, \epsilon, P) = C(P) - \sqrt{\frac{V(P)}{n}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \tag{58}$$

and then by Taylor-expanding $C(P)$ and $V(P)$ for $P \to 0$. However, there are two major drawbacks with this approach. First, establishing the normal approximation (58) is challenging for fading channels (see [36] and the remarks after Theorem 3). So this approach would work only in the AWGN case. Second, one needs to verify that the $o(1/\sqrt{n})$ term in (58) is uniform in P, which is nontrivial.

C. Rayleigh-Fading Channels With Perfect CSIR

We next consider the Rayleigh-fading channel

$$Y_i = H_i X_i + Z_i, \quad i = 1, \ldots, n \tag{59}$$

where both $\{H_i\}$ and $\{Z_i\}$ are independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ random variables. We assume that the channel coefficients $\{H_i\}$ are known to the receiver but not to the transmitter. Furthermore, we assume that every codeword x^n satisfies the power constraint (43). The wideband slope of this channel is [13, Eq. (208)]

$$S_0 = \frac{2\mathbb{E}[|H|^2]^2}{\mathbb{E}[|H|^4]} = 1 \tag{60}$$

where $H \sim \mathcal{CN}(0, 1)$.

Theorem 3 below characterizes the minimum energy per bit $E_b^*(k, \epsilon, R)$ for the Rayleigh-fading channel in the asymptotic limit $k \to \infty$ and $R \to 0$.

Theorem 3: Consider the Rayleigh block-fading channel (59). The following expansion for $E_b^*(k, \epsilon, R)$ holds in the asymptotic limit $k \to \infty$, $R \to 0$, and $kR \to \infty$:

$$\frac{E_b^*(k, \epsilon, R)}{N_0} = \log_e 2 + \sqrt{\frac{2 \log_e 2}{k}} Q^{-1}(\epsilon) + (\log_e^2 2)R + o\left(\frac{1}{\sqrt{k}}\right) + o(R) \tag{61}$$

or, equivalently,

$$10 \log_{10} \frac{E_{\mathrm{b}}^*(k, \epsilon, R)}{N_0} = \underbrace{10 \log_{10} \frac{E_{\mathrm{b}}}{N_0}_{\min} + \frac{R}{S_0} 10 \log_{10} 2}_{\text{wideband slope approximation}}$$

$$+ \frac{10 \log_{10} e}{\sqrt{\log_e 2}} \sqrt{\frac{2}{k}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{k}}\right) + o(R) \qquad (62)$$

where $10 \log_{10} \frac{E_{\mathrm{b}}}{N_0}_{\min} = -1.59\,\mathrm{dB}$, and $S_0 = 1$. If $k \to \infty$, $R \to 0$, but $kR \to c < \infty$, then we have

$$\frac{E_{\mathrm{b}}^*(k, \epsilon, R)}{N_0} = \log_e 2 + \sqrt{\frac{2 \log_e 2}{k}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{k}}\right) \qquad (63)$$

or, equivalently,

$$10 \log_{10} \frac{E_{\mathrm{b}}^*(k, \epsilon, R)}{N_0} = 10 \log_{10} \frac{E_{\mathrm{b}}}{N_0}_{\min} + \frac{10 \log_{10} e}{\sqrt{\log_e 2}} \sqrt{\frac{2}{k}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{k}}\right). \qquad (64)$$

Proof: See Appendix II. ∎

A few remarks are in order:

- As in the AWGN case, the minimum energy per bit $E_{\mathrm{b}}^*(k, \epsilon, R)$ over the Rayleigh-fading channel (59) with perfect CSIR satisfies

$$10 \log_{10} \frac{E_{\mathrm{b}}^*(k, \epsilon, R)}{E_{\mathrm{b}}^*(R)} = \frac{10 \log_{10} e}{\sqrt{\log_e 2}} \sqrt{\frac{2}{k}} Q^{-1}(\epsilon) + o(R) + o\left(\frac{1}{\sqrt{k}}\right). \qquad (65)$$

  Again we observe that, in the low spectral efficiency regime, the gap in dB between $E_{\mathrm{b}}^*(k, \epsilon, R)$ and the asymptotic limit $E_{\mathrm{b}}^*(R)$ is—up to first order—proportional to $1/\sqrt{k}$ and is independent of R. Furthermore, the gap in the fading case coincides with that in the AWGN case up to $o(R) + o(1/\sqrt{k})$ terms.

- Unlike the asymptotic wideband approximation (37), which holds for all fading distributions (see [13]), our result in Theorem 3 relies on the Gaussianity of the fading coefficients, and does not necessarily hold for other fading distributions. In fact, as observed in [35, Sec. III.D], there are fading distributions for which the minimum energy per bit $E_{\mathrm{b}}^*(k, \epsilon, R)$ does not converge to $-1.59\,\mathrm{dB}$ when $k \to \infty$, $R \to 0$, and ϵ is fixed.

- For the case of nonvanishing rate (or, equivalently, nonvanishing SNR P), a normal approximation for the maximum rate $R^*(n, P, \epsilon)$ achievable over the channel (59) when CSIR is available is reported in [36]. This approximation relies on the additional constraint that
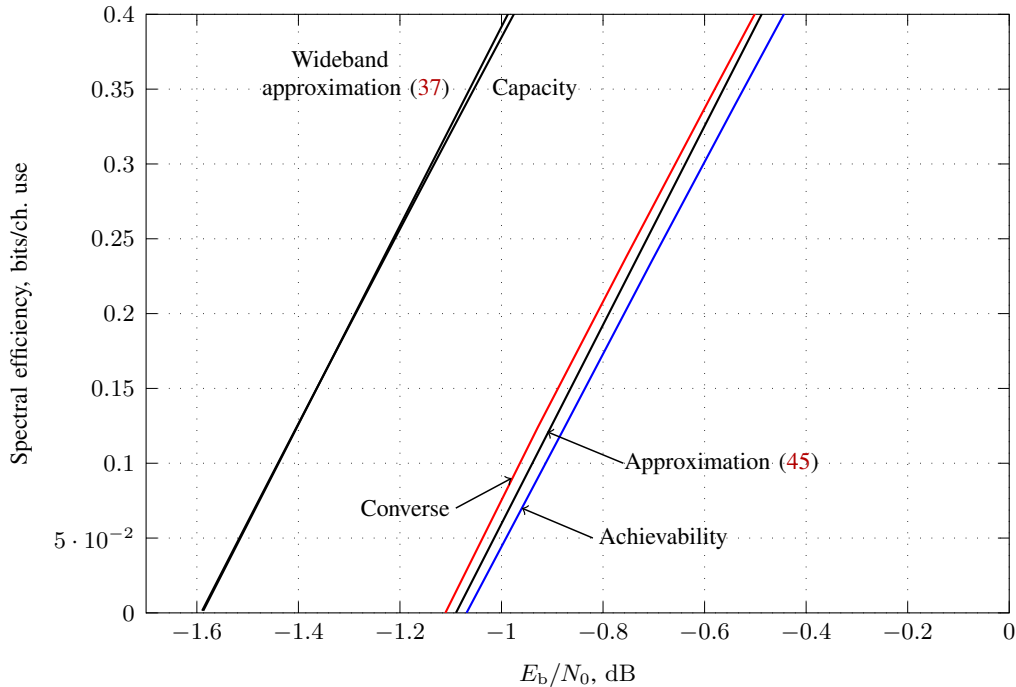
Fig. 2.   Minimum energy per bit versus spectral efficiency of the AWGN channel; here, $k = 2000$ bits, and $\epsilon = 10^{-3}$.

every codeword x^n satisfies $\|x^n\|_\infty = o(n^{1/4})$. In contrast, Theorem 3 does not require this additional constraint.

- One of the challenges that one has to address when establishing a nonasymptotic converse bound on $E_{\mathrm{b}}^*(k, \epsilon, R)$ is that the variance of the information density $i(x^n; Y^n H^n)$ depends on $\|x^n\|_4$ (see [25, Eqs. (47)–(52)]). In order to obtain a tight converse bound on $E_{\mathrm{b}}^*(k, \epsilon, R)$ for fixed R, one needs to expurgate the codewords whose ℓ_4 norm $\|x^n\|_4$ is far from that of the codewords of a Gaussian code (see [25] and [36]). However, in the limit $R \to 0$ of interest in this paper, this expurgation procedure is not needed since the dominant term in the asymptotic expansion of the variance of $i(x^n; Y^n H^n)$ does not depend on $\|x^n\|_4$. Furthermore, the wideband slope is also insensitive to $\|x^n\|_4$. Indeed, to achieve the wideband slope of a fading channel with perfect CSIR, QPSK inputs are as good as Gaussian inputs [13].
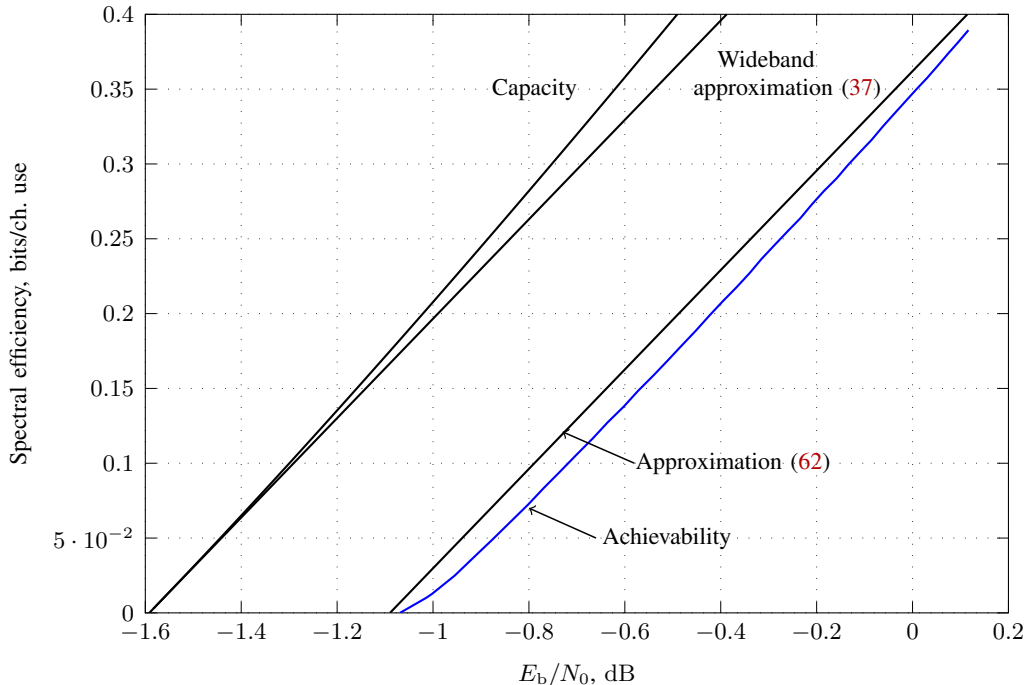
Fig. 3. Minimum energy per bit versus spectral efficiency of the Rayleigh-fading channel (59) with perfect CSIR; here, $k = 2000$ bits, $\epsilon = 10^{-3}$.

D. Numerical results

In Fig. 2, we present a comparison between the approximation (45) (with the two $o(\cdot)$ terms omitted), the $\beta\beta$ achievability bound (12), and the $\beta\beta$ converse bound (9). In both the achievability and the converse bound, Q_{Y^n} is chosen to be the product distribution obtained from the capacity-achieving output distribution of the channel (42) (i.e., $Q_{Y^n} = \mathcal{CN}(0, (1+P)\mathsf{I}_n)$). For the parameters considered in Fig. 2, i.e., $k = 2000$ bits and $\epsilon = 10^{-3}$, the approximation (45) is accurate. Fig. 3 provides a similar comparison for the Rayleigh fading channel (59). In this case, the $\beta\beta$ converse bound is difficult to compute (due to the need to perform an optimization over all input distributions), and is not plotted.

V. OTHER APPLICATIONS OF THEOREM 1

A. The Additive Exponential-Noise Channel

Consider the additive exponential-noise channel

$$Y_i = X_i + Z_i, \quad i = 1, \ldots, n \tag{66}$$

where $\{Z_i\}$ are i.i.d. $\mathrm{Exp}(1)$-distributed. As in [24], we require each codeword $x^n \in \mathbb{R}^n$ to satisfy

$$x_i \geq 0, \; i = 1, \ldots, n, \quad \text{and} \quad \sum_{i=1}^{n} x_i \leq n\sigma. \tag{67}$$

The additive exponential-noise channel (66) can be used to model a communication system where information is conveyed through the arrival time of packets, and where each packet goes through a single-server queue with exponential service time [37]. It also models a fast-varying phase-noise channel combined with an energy detector at the receiver [38].

The capacity of the channel (66) under the input constraints specified in (67) is given by [24, Th. 3]

$$C(\sigma) = \log(1 + \sigma) \tag{68}$$

and is achieved by the input distribution P_X^* according to which X takes the value zero with probability $1/(1 + \sigma)$ and follows an $\mathrm{Exp}(1 + \sigma)$ distribution conditioned on it being positive. Furthermore, the capacity-achieving output distribution is $\mathrm{Exp}(1 + \sigma)$. A discrete counterpart of the exponential-noise channel is studied in [39], where a lower bound on the maximum coding rate is derived.

Theorem 4 below characterizes the dispersion of the channel (66).

Theorem 4: Consider the additive exponential-noise channel (66) subject to the constraint (67). For every $0 < \epsilon < 1$, the maximum coding rate $R^*(n, \epsilon)$ admits the following expansion:

$$R^*(n, \epsilon) = \log(1 + \sigma) - \sqrt{\frac{V(\sigma)}{n}} Q^{-1}(\epsilon) + \mathcal{O}\left(\frac{\log n}{n}\right) \tag{69}$$

where

$$V(\sigma) = \frac{\sigma^2}{(1 + \sigma)^2} \log^2 e. \tag{70}$$

Proof: We first prove that (69) is achievable using the $\beta\beta$ achievability bound in Theorem 1. Let $Q_{X^n} = (P_X^*)^n$, where P_X^* is the capacity-achieving input distribution. Let P_{X^n} be the conditional distribution of $X^n \sim Q_{X^n}$ given that X^n belongs to the set $\mathcal{F}$ specified below:

$$\mathcal{F} = \left\{ x^n \in \mathbb{R}^n : x_i \geq 0, \; n\sigma - \log n \leq \sum_{i=1}^{n} x_i \leq n\sigma \right\}. \tag{71}$$

By construction, $X^n \sim P_{X^n}$ satisfies the constraint (67). Finally, let $P_{Y^n} \triangleq P_{X^n} \circ P_{Y^n|X^n}$ and let $Q_{Y^n} \triangleq Q_{X^n} \circ P_{Y^n|X^n}$. We apply the $\beta\beta$ bound in Theorem 1 with $\tau = 1/\sqrt{n}$ and with P_{X^n}

and Q_{Y^n} chosen as above. The term $\beta_\tau(P_{Y^n}, Q_{Y^n})$ can be evaluated as follows:

$$\log \beta_\tau(P_{Y^n}, Q_{Y^n}) \geq \log Q_{X^n}[\mathcal{F}] + \log \tau \tag{72}$$

$$= \log\left(Q\left(\frac{-\log n}{\sqrt{n \mathbb{V}\mathrm{ar}_{P_X^*}[X]}} \right) - Q(0) - \mathcal{O}\left(\frac{1}{\sqrt{n}} \right) \right) + \log \tau \tag{73}$$

$$= \mathcal{O}(\log n). \tag{74}$$

Here, (72) follows from (31), and (73) follows from the Berry-Esséen theorem (see, e.g., [40, Sec. XVI.5]).

We next evaluate $\beta_{1-\epsilon+\tau}(P_{X^n Y^n}, P_{X^n} Q_{Y^n})$. It follows from [5, Eq. (103)] that

$$\beta_{1-\epsilon+\tau}(P_{X^n Y^n}, P_{X^n} Q_{Y^n}) \leq \exp(-\gamma_n) \tag{75}$$

where γ_n satisfies

$$P_{X^n Y^n}\left[\log \frac{\mathrm{d}P_{X^n Y^n}}{\mathrm{d}(P_{X^n} Q_{Y^n})}(X^n, Y^n) \geq \gamma_n \right] \geq 1 - \epsilon + \tau. \tag{76}$$

Note that, under $P_{X^n Y^n}$, the random variable $\log \frac{\mathrm{d}P_{X^n Y^n}}{\mathrm{d}(P_{X^n} Q_{Y^n})}(X^n, Y^n)$ has the same distribution as

$$n \log(1 + \sigma) + \frac{\log e}{1 + \sigma} \sum_{i=1}^n X_i - \frac{\sigma \log e}{1 + \sigma} \sum_{i=1}^n Z_i \tag{77}$$

where Z_i are i.i.d. $\mathrm{Exp}(1)$-distributed. This random variable depends on the codeword X^n only through $\sum_{i=1}^n X_i$. Furthermore, given $\sum_{i=1}^n X_i$, this random variable is the sum of n i.i.d. random variables. Using the Berry-Esséen theorem and (71) to evaluate the left-hand side (LHS) of (76), we conclude that

$$P_{X^n Y^n}\left[\log \frac{\mathrm{d}P_{X^n Y^n}}{\mathrm{d}P_{X^n} Q_{Y^n}}(X^n, Y^n) \geq \gamma_n \right]$$

$$\geq \mathbb{P}\left[\frac{\sigma \log e}{1 + \sigma} \sum_{i=1}^n (1 - Z_i) \geq \gamma_n - n \log(1 + \sigma) - \frac{\log e}{1 + \sigma} \log n \right] \tag{78}$$

$$\geq Q\left(\frac{\gamma_n + (\log e)(\log n)/(1 + \sigma) - n \log(1 + \sigma)}{\sqrt{n V(\sigma)}} \right) - \mathcal{O}\left(\frac{1}{\sqrt{n}} \right). \tag{79}$$

Equating the RHS of (79) to $1 - \epsilon + \tau$, and solving it for γ_n, we conclude that

$$\gamma_n = n \log(1 + \sigma) - \sqrt{n V(\sigma)} Q^{-1}(\epsilon) + \mathcal{O}(\log n). \tag{80}$$

Substituting (80) into (75), and then (75) and (74) into (12), we establish that (69) is achievable.

To prove the converse part of Theorem 4, we first notice that by [5, Lemma 39], we can assume without loss of generality that all codewords x^n belong to the simplex

$$\mathcal{F}_n \triangleq \Big\{ x^n \in \mathbb{R}^n : \sum_{i=1}^n x_i = n\sigma, \ x_i \geq 0 \Big\}. \tag{81}$$

Let $Q_{Y^n} = Q_{X^n} \circ P_{Y^n|X^n}$, where, as before, $Q_{X^n} = (P_X^*)^n$. By the meta-converse theorem [5, Th. 27], every (n, M, ϵ) code for the channel (66) satisfies

$$M \leq \sup_{P_{X^n}} \frac{1}{\beta_{1-\epsilon}(P_{X^nY^n}, P_{X^n}Q_{Y^n})} \tag{82}$$

where the supremum is over all probability distributions supported on $\mathcal{F}_n$. We next note that the function $\beta_{1-\epsilon}(P_{Y^n|X^n=x^n}, Q_{Y^n})$ takes the same value for all $x^n \in \mathcal{F}_n$. Indeed, by using the Neyman-Pearson lemma we observe that

$$\beta_{1-\epsilon}(P_{Y^n|X^n=x^n}, Q_{Y^n}) = e^{-n\sigma/(1+\sigma)} \mathbb{P}\Big[\sum_{i=1}^n Z_i \leq \frac{n-\xi_n}{1+\sigma} \Big], \quad \forall x^n \in \mathcal{F}_n \tag{83}$$

where ξ_n satisfies

$$\mathbb{P}\Big[\sum_{i=1}^n Z_i \leq n - \xi_n \Big] = 1 - \epsilon. \tag{84}$$

Using [5, Lemma 29], we conclude that for every $x^n \in \mathcal{F}_n$,

$$\sup_{P_{X^n}} \beta_{1-\epsilon}(P_{X^nY^n}, P_{X^n}Q_{Y^n}) = \beta_{1-\epsilon}(P_{Y^n|X^n=x^n}, Q_{Y^n}). \tag{85}$$

For convenience, we shall set $x^n = \bar{x}^n \triangleq [\sigma, \ldots, \sigma]$. This choice makes $P_{Y^n|X^n=x^n}$ a distribution of i.i.d. random variables. Using [5, Lemma 58] and performing straightforward algebraic manipulations, we obtain

$$-\log \beta_{1-\epsilon}(P_{Y^n|X^n=\bar{x}^n}, Q_{Y^n}) = n\log(1+\sigma) - \sqrt{nV(\sigma)}Q^{-1}(\epsilon) + \mathcal{O}(\log n). \tag{86}$$

We conclude the proof by substituting (85) and (86) into (82). ■

B. MIMO Block-Fading Channel With Perfect CSIR

In this section, we use the $\beta\beta$ achievability bound (12) to characterize the maximum coding rate achievable over an $m_\mathrm{t} \times m_\mathrm{r}$ Rayleigh MIMO block-fading channel. The channel is assumed to stay constant over n_c channel uses (a coherence interval) and to change independently across coherence intervals. The input-output relation within the kth coherence interval is given by

$$\mathbb{Y}_k = \mathbb{X}_k \mathbb{H}_k + \mathbb{Z}_k. \tag{87}$$

Here, $\mathbb{X}_k \in \mathbb{C}^{n_c \times m_t}$ and $\mathbb{Y}_k \in \mathbb{C}^{n_c \times m_r}$ are the transmitted and received matrices, respectively; the entries of the fading matrix $\mathbb{H}_k \in \mathbb{C}^{m_t \times m_r}$ and of the noise matrix $\mathbb{Z}_k \in \mathbb{C}^{n_c \times m_r}$ are i.i.d. $\mathcal{CN}(0,1)$. We assume that $\{\mathbb{H}_k\}$ and $\{\mathbb{Z}_k\}$ are independent, that they take on independent realizations over successive coherence intervals, and that they do not depend on the matrices $\{\mathbb{X}_k\}$. The channel matrices $\{\mathbb{H}_k\}$ are assumed to be known to the receiver but not to the transmitter. We shall also assume that each codeword spans $\ell \in \mathbb{N}$ coherence intervals, i.e., that the blocklength of the code is $n = \ell n_c$. Finally, each codeword X^ℓ is constrained to satisfy

$$\left\| \mathsf{X}^\ell \right\|_\mathsf{F}^2 \leq nP. \tag{88}$$

1) Capacity and dispersion: In the asymptotic limit $\ell \to \infty$ for fixed n_c, the capacity of (87) is given by [41]

$$C(P) = \mathbb{E}_\mathbb{H}\Big[\log\det\Big(\mathsf{I}_{m_t} + \sqrt{P/m_t}\,\mathbb{H}\mathbb{H}^\mathrm{H}\Big)\Big]. \tag{89}$$

If either $m_t = m_r = 1$ or $m_r \geq 2$, the capacity is achieved by a unique input distribution, under which the matrix $\mathbb{X}$ has i.i.d. $\mathcal{CN}(0, P/m_t)$ entries [41]. In this case, we denote the capacity-achieving input distribution by $P_\mathbb{X}^*$. If $m_t > 1$ and $m_r = 1$ (i.e., multiple-input single-output channel), the capacity-achieving input distribution is not unique [42]. The capacity-achieving output distribution is always unique and is denoted by $P_{\mathbb{YH}}^*$.[6] More specifically, $P_{\mathbb{YH}}^* = P_\mathbb{H} P_{\mathbb{Y}|\mathbb{H}}^*$, where under $P_{\mathbb{Y}|\mathbb{H}=\mathsf{H}}^*$, the column vectors of $\mathbb{Y}$ are i.i.d. $\mathcal{CN}(\mathbf{0}, \mathsf{H}_i^\mathrm{H}\mathsf{H}_i + P/\sqrt{m_t}\mathsf{I}_{m_r})$ distributed.

The channel dispersion for the single-antenna case with perfect CSIR was derived in [25]. This result was extended to multiple-antenna block-fading channels in [42] and [36]. In particular, it was shown in [36] that[7] for every $0 < \epsilon < 1/2$

$$R^*(n, \epsilon) \geq C(P) - \sqrt{\frac{V(P)}{n}} Q^{-1}(\epsilon) + o(1/\sqrt{n}). \tag{90}$$

Here,

$$V(P) = \inf \mathbb{E}\left[\mathbb{V}\mathrm{ar}\left[\log \frac{\mathrm{d}P_{\mathbb{YH}|\mathbb{X}}}{\mathrm{d}P_{\mathbb{YH}}^*}(\mathbb{X}, \mathbb{Y}, \mathbb{H}) \,\Big|\, \mathbb{X}\right]\right] \tag{91}$$

where the infimum is over the set of capacity-achieving input distributions. For the case $m_t = m_r = 1$ or $m_r \geq 2$, the infimum is over the singleton $\{P_\mathbb{X}^*\}$.

---

[6]Since the channel matrix $\mathbb{H}$ is known at the receiver, the channel output consists of the pair $(\mathbb{Y}, \mathbb{H})$.

[7]It is also shown in [36] that the bound (90) is tight under the additional constraint $\|\mathsf{X}^\ell\|_\infty = o(n^{1/4})$ on each codeword X^ℓ.

2) Evaluation of the $\beta\beta$ achievability bound (12)*:* We now turn our attention to the computation of the $\beta\beta$ achievability bound (12) for the channel (87). For simplicity, we shall focus on the case when the capacity-achieving input distribution is unique. To compute (12), we choose $P_{\mathbb{X}^\ell}$ as the uniform distribution on $\mathcal{S}_n \triangleq \{\mathsf{X}^\ell : \|\mathsf{X}^\ell\|_{\mathsf{F}}^2 = nP\}$, and set $Q_{\mathbb{Y}^\ell\mathbb{H}^\ell} = (P_{\mathbb{YH}}^*)^\ell$. With these choices, we have

$$R^*(n,\epsilon) \geq \frac{1}{n} \sup_{0<\tau<\epsilon} \log \frac{\beta_\tau(P_{\mathbb{Y}^\ell\mathbb{H}^\ell}, Q_{\mathbb{Y}^\ell\mathbb{H}^\ell})}{\beta_{1-\epsilon+\tau}(P_{\mathbb{X}^\ell\mathbb{Y}^\ell\mathbb{H}^\ell}, P_{\mathbb{X}^\ell}Q_{\mathbb{Y}^\ell\mathbb{H}^\ell})}. \tag{92}$$

The denominator $\beta_{1-\epsilon+\tau}(P_{\mathbb{X}^\ell\mathbb{Y}^\ell\mathbb{H}^\ell}, P_{\mathbb{X}^\ell}Q_{\mathbb{Y}^\ell\mathbb{H}^\ell})$ in (92) can be computed using the Neyman-Pearson lemma and standard Monte Carlo techniques. However, computing $\beta_\tau(P_{\mathbb{Y}^\ell\mathbb{H}^\ell}, Q_{\mathbb{Y}^\ell\mathbb{H}^\ell})$ in the numerator is more involved, since there is no simple expression for $P_{\mathbb{Y}^\ell\mathbb{H}^\ell}$. To circumvent this, we further lower-bound $\beta_\tau(P_{\mathbb{Y}^\ell\mathbb{H}^\ell}, Q_{\mathbb{Y}^\ell\mathbb{H}^\ell})$ using the data-processing inequality for $\beta_\alpha(\cdot,\cdot)$ as follows. Let $\widetilde{\mathbb{X}}^\ell$ be a sequence of i.i.d. random matrices with $\widetilde{\mathbb{X}}_k \sim P_{\mathbb{X}}^*$, $k = 1, \ldots, \ell$. Then, $P_{\mathbb{X}^\ell}$ can be obtained via $\widetilde{\mathbb{X}}^\ell$ through $\mathbb{X}^\ell = \sqrt{nP}\widetilde{\mathbb{X}}^\ell/\|\widetilde{\mathbb{X}}^\ell\|_{\mathsf{F}}$. Let $P_{\mathbb{Y}^\ell\mathbb{H}^\ell|\mathbb{X}^\ell}^{(\mathrm{s})} \triangleq P_{\mathbb{H}^\ell}P_{\mathbb{Y}^\ell|\mathbb{X}^\ell\mathbb{H}^\ell}^{(\mathrm{s})}$, where $P_{\mathbb{Y}^\ell|\mathbb{X}^\ell\mathbb{H}^\ell}^{(\mathrm{s})}$ denotes the channel law defined by

$$\mathbb{Y}_k = \mathbb{X}_k\mathbb{H}_k\frac{\sqrt{nP}}{\|\mathbb{X}^\ell\|_{\mathsf{F}}} + \mathbb{Z}_k, \qquad k = 1, \ldots, \ell. \tag{93}$$

We have that $P_{\mathbb{Y}^\ell\mathbb{H}^\ell} = P_{\mathbb{X}^\ell} \circ P_{\mathbb{Y}^\ell\mathbb{H}^\ell|\mathbb{X}^\ell} = (P_{\mathbb{X}}^*)^\ell \circ P_{\mathbb{Y}^\ell\mathbb{H}^\ell|\mathbb{X}^\ell}^{(\mathrm{s})}$. Furthermore, $Q_{\mathbb{Y}^\ell\mathbb{H}^\ell} = (P_{\mathbb{X}}^*)^\ell \circ P_{\mathbb{Y}^\ell\mathbb{H}^\ell|\mathbb{X}^\ell}$. Now, by the data-processing inequality,

$$\beta_\tau(P_{\mathbb{Y}^\ell\mathbb{H}^\ell}, Q_{\mathbb{Y}^\ell\mathbb{H}^\ell}) \geq \beta_\tau\big((P_{\mathbb{X}}^*)^\ell P_{\mathbb{Y}^\ell\mathbb{H}^\ell|\mathbb{X}^\ell}^{(\mathrm{s})}, (P_{\mathbb{X}}^*)^\ell P_{\mathbb{Y}^\ell\mathbb{H}^\ell|\mathbb{X}^\ell}\big). \tag{94}$$

Since the Radon-Nikodym derivative $\frac{\mathrm{d}P_{\mathbb{Y}^\ell\mathbb{H}^\ell|\mathbb{X}^\ell}^{(\mathrm{s})}}{\mathrm{d}P_{\mathbb{Y}^\ell\mathbb{H}^\ell|\mathbb{X}^\ell}}$ can be computed in closed form, the RHS of (94) can be computed using the Neyman-Pearson lemma and Monte Carlo techniques.

The resulting bound[8] is shown in Fig. 4 for a 4×4 Rayleigh block-fading channel with $n_{\mathrm{c}} = 4$, $\epsilon = 0.001$, and SNR = 0 dB. For comparison, we have also plotted the normal approximation (90) (with the $o(1/\sqrt{n})$ term omitted), the cost-constrained DT bound (30), and the variation of Feinstein's bound provided in (36). More specifically, (30) is computed with

$$\mathcal{F} = \{\mathsf{X}^\ell : \|\mathsf{X}^\ell\|_{\mathsf{F}}^2 \leq nP\} \tag{95}$$

and with $\mathbb{X}^\ell \sim Q_{\mathbb{X}^\ell}$ having i.i.d. $\mathcal{CN}(0, P'/m_{\mathrm{t}})$ entries. Here, P' is chosen such that $Q_{\mathbb{X}^\ell}[\mathcal{F}^c] = \epsilon/2$. To compute (36), we chose the same $P_{\mathbb{X}^\ell}$ and $Q_{\mathbb{Y}^\ell\mathbb{H}^\ell}$ that we used to compute (92). Not

---

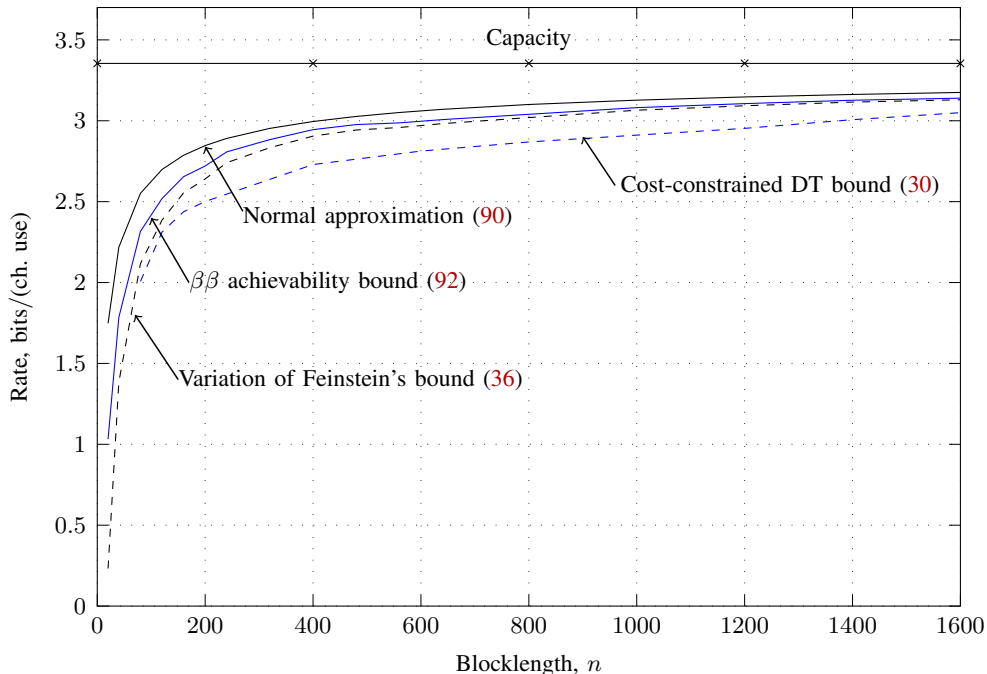[8]The numerical routines used to obtain these results are available at https://github.com/yp-mit/spectre

Fig. 4. Bounds on the maximum rate for a 4×4 MIMO Rayleigh block-fading channel; here SNR=0 dB, $\epsilon = 0.001$, and $n_\mathrm{c} = 4$.

surprisingly, the $\beta\beta$ achievability bound (12) is uniformly tighter than both (30) and (36). Note also that (36) is better than the cost-constrained DT bound (30) mainly because it uses a better input distribution.

We also note that the $\kappa\beta$ bound [5, Th. 25] (with $\mathcal{F} = \mathcal{S}_n$) is much more difficult to compute because one needs to maximize over all codewords $\mathsf{X}^\ell \in \mathcal{S}_n$. Furthermore, for blocklength values of practical interest, we expect that

$$\max_{\mathsf{X}^\ell \in \mathcal{S}_n} \beta_{1-\epsilon+\tau}(P_{\mathbb{Y}^\ell\mathbb{H}^\ell \mid \mathbb{X}^\ell = \mathsf{X}^\ell}, Q_{\mathbb{Y}^\ell\mathbb{H}^\ell}) \gg \beta_{1-\epsilon+\tau}(P_{\mathbb{X}^\ell\mathbb{Y}^\ell\mathbb{H}^\ell}, P_{\mathbb{X}^\ell}Q_{\mathbb{Y}^\ell\mathbb{H}^\ell}) \tag{96}$$

which means that the resulting $\kappa\beta$ bound may be much looser than (92). To validate this claim, we evaluate $\beta_{1-\epsilon+\tau}(P_{\mathbb{Y}^\ell\mathbb{H}^\ell \mid \mathbb{X}^\ell = \mathsf{X}^\ell}, Q_{\mathbb{Y}^\ell\mathbb{H}^\ell})$ for a specific codeword $\hat{\mathsf{X}}^\ell \in \mathcal{S}_n$ constructed as follows: the entries $\hat{x}_{j,k}^{(i)}$ of the matrix $\hat{\mathsf{X}}_i$ are $\hat{x}_{1,1}^{(1)} = \sqrt{nP}$, and $\hat{x}_{j,k}^{(i)} = 0$, $\forall (i, j, k) \neq (1, 1, 1)$. By construction, $\beta_{1-\epsilon+\tau}(P_{\mathbb{Y}^\ell\mathbb{H}^\ell \mid \mathbb{X}^\ell = \hat{\mathsf{X}}^\ell})$ is a lower bound on the LHS of (96). It can be shown that the ratio between $\beta_{1-\epsilon+\tau}(P_{\mathbb{Y}^\ell\mathbb{H}^\ell \mid \mathbb{X}^\ell = \hat{\mathsf{X}}^\ell}, Q_{\mathbb{Y}^\ell\mathbb{H}^\ell})$ and $\beta_{1-\epsilon+\tau}(P_{\mathbb{X}^\ell\mathbb{Y}^\ell\mathbb{H}^\ell}, P_{\mathbb{X}^\ell}Q_{\mathbb{Y}^\ell\mathbb{H}^\ell})$ grows exponentially in ℓ. Numerically, we observe that for the parameters in Fig. 4 and the choices $\tau = \epsilon/2 = 5 \times 10^{-4}$

and $n = n_\mathrm{c} l = 400$,

$$-\log_2 \beta_{1-\epsilon+\tau}(P_{\mathbb{Y}^\ell \mathbb{H}^\ell \mid \mathbb{X}^\ell = \hat{\mathbb{x}}^\ell}, Q_{\mathbb{Y}^\ell \mathbb{H}^\ell}) \approx -\log_2 \beta_{1-\epsilon+\tau}(P_{\mathbb{X}^\ell \mathbb{Y}^\ell \mathbb{H}^\ell}, P_{\mathbb{X}^\ell} Q_{\mathbb{Y}^\ell \mathbb{H}^\ell}) + 735 \text{ bits}. \qquad (97)$$

VI. Concluding Remarks

We have developed a novel channel coding achievability bound (which we refer to as $\beta\beta$ achievability bound), that involves two binary hypothesis tests, one for the joint distribution of the input and the output, and the other for the output distribution. Our bound together with a $\beta\beta$ converse bound established earlier by Polyanskiy and Verdú [18, Th. 15] yield a nonasymptotic version of the golden formula (3). Connections between the $\beta\beta$ bounds and various other nonasymptotic bounds as well as their asymptotic counterparts are summarized in Fig. 1.

The analogy between the $\beta\beta$ bounds and the golden formula allows us to extend to the nonasymptotic regime asymptotic analyses in which the golden formula plays a key role . To demonstrate this point, we have used the $\beta\beta$ bounds to obtain a finite-blocklength extension of Verdú's wideband-slope approximation [13]. Our proof parallels the elegant derivation by Verdú, except that the beta-beta bounds are used in place of the golden formula. We have also used the $\beta\beta$ achievability bound to characterize the channel dispersion of the additive exponential-noise channel and to obtain a finite-blocklength achievability bound for MIMO Rayleigh-fading channels with perfect CSIR. In both cases, the $\beta\beta$ achievability bound proves to be very useful: in the former case, it simplifies the asymptotic analysis; in the latter case, it yields the tightest known achievability bound to date.

A crucial step in evaluating both the $\beta\beta$ achievability and converse bounds is to choose an appropriate output distribution Q_Y. A good choice of Q_Y yields a bound that is both analytically tractable and numerically tight. Some general principles for choosing Q_Y are discussed in [43, Ch. I.3.4] and [44].

Appendix I

Proof of Theorem 2

The proof is divided into four parts. (i) We first show in Appendix I-A that, when $k \to \infty$, $R \to 0$, and $kR \to \infty$, the expansion (54) implies (44). We then prove (54) by providing (ii)

achievability and (iii) converse results in Appendices I-B and I-C, respectively. (iv) Finally, we prove (46) in Appendix I-D.

A. The asymptotic expansion (54) implies (44)

Observe that the blocklength n, the transmit power P, the data rate R, the energy per bit E_{b}, and the number of information bits k are related as follows:

$$k = nR \quad \text{and} \quad P = E_{\mathrm{b}}R \tag{98}$$

where R is measured in bits per channel use. Hence, $E_{\mathrm{b}}^*(k, \epsilon, R)$ can be obtained from $R^*(n, \epsilon, P)$ by solving the equation

$$R = R^*(k/R, \epsilon, E_{\mathrm{b}}^*(k, \epsilon, R)R). \tag{99}$$

Substituting (54) into (99) after setting the base of the $\log$ in (54) to 2, we obtain

$$R \log_e 2 = E_{\mathrm{b}}^* \cdot R - \sqrt{\frac{2E_{\mathrm{b}}^* \cdot R^2}{k}} Q^{-1}(\epsilon) - \frac{1}{2}(E_{\mathrm{b}}^* \cdot R)^2$$
$$+ o\left(\sqrt{\frac{E_{\mathrm{b}}^* R^2}{k}}\right) + o\left((E_{\mathrm{b}}^* \cdot R)^2\right) \tag{100}$$

in the asymptotic limit $n \to \infty$, $P \to 0$, and $nP^2 \to \infty$. Here, we set $E_{\mathrm{b}}^* = E_{\mathrm{b}}^*(k, \epsilon, R)$ for notational convenience. Dividing both sides of (100) by R, and rearranging, we obtain (recall that $N_0 = 1$)

$$E_{\mathrm{b}}^* = \log_e 2 - \sqrt{\frac{2E_{\mathrm{b}}^*}{k}} Q^{-1}(\epsilon) - \frac{(E_{\mathrm{b}}^*)^2 R}{2} + o\left(\sqrt{\frac{E_{\mathrm{b}}^*}{k}}\right) + o\left((E_{\mathrm{b}}^*)^2 R\right). \tag{101}$$

This implies that

$$E_{\mathrm{b}}^* = \log_e 2 + o(1). \tag{102}$$

Substituting (102) into the RHS of (101), we recover (44).

In the next two sections we shall establish that (54) holds in the asymptotic limit $n \to \infty$, $P \to 0$, and $nP^2 \to \infty$. Note that by (98) and (102), these limits are equivalent to the limits $k \to \infty$, $R \to 0$, and $kR \to \infty$ stated in Theorem 2. Unless otherwise specified, all asymptotic expansions in the next two sections hold in the limit $n \to \infty$, $P \to 0$, and $nP^2 \to \infty$.

B. Proof of (54): Achievability

We shall apply the $\beta\beta$ achievability bound (12) with P_{X^n} chosen as the uniform distribution over the power sphere

$$\mathcal{F}_n \triangleq \{x^n \in \mathbb{C}^n : \|x^n\|^2 = nP\} \tag{103}$$

and with $Q_{Y^n} = \mathcal{CN}(\mathbf{0}, \mathsf{I}_n)$. Furthermore, we shall set

$$\tau = (nP^2)^{-1/2}. \tag{104}$$

Since we are interested in the asymptotic regime $n \to \infty$, we can assume without loss of generality that $\tau < \epsilon$. By (12),

$$R^*(n, \epsilon, P) \geq \frac{1}{n} \log \frac{\beta_\tau(P_{Y^n}, Q_{Y^n})}{\beta_{1-\epsilon+\tau}(P_{X^nY^n}, P_{X^n}Q_{Y^n})}. \tag{105}$$

The denominator on the RHS of (105) can be computed as follows. Due to the spherical symmetry of $\mathcal{F}_n$ and Q_{Y^n}, we have that $\beta_\alpha(P_{Y^n|X^n=x^n}, Q_{Y^n})$ takes the same value for all $x^n \in \mathcal{F}_n$. Let $\bar{x}^n \triangleq [\sqrt{nP}, 0, \ldots, 0]$. We next apply [5, Lemma 29] to conclude that for every P_{X^n} supported on $\mathcal{F}_n$,

$$\beta_{1-\epsilon+\tau}(P_{X^nY^n}, P_{X^n}Q_{Y^n}) = \beta_{1-\epsilon+\tau}(P_{Y^n \mid X^n = \bar{x}^n}, Q_{Y^n}) \tag{106}$$

$$= \beta_{1-\epsilon+\tau}(\mathcal{CN}(\sqrt{nP}, 1), \mathcal{CN}(0, 1)). \tag{107}$$

It follows from the Neyman-Pearson lemma [21] that

$$\beta_{1-\epsilon+\tau}(P_{X^nY^n}, P_{X^n}Q_{Y^n}) = Q\left(\sqrt{2nP} + Q^{-1}(1 - \epsilon + \tau)\right). \tag{108}$$

To evaluate the RHS of (108) we use that [45, Eq. (3.53)]

$$\log Q(x) = -\frac{x^2 \log e}{2} - \log x - \frac{1}{2}\log(2\pi) + o(1), \quad x \to \infty \tag{109}$$

and obtain

$$\log \beta_{1-\epsilon+\tau}(P_{X^nY^n}, P_{X^n}Q_{Y^n})$$
$$= -nP \log e + \sqrt{2nP}Q^{-1}(\epsilon) \log e + o\left(\sqrt{nP}\right). \tag{110}$$

For latter use, we notice that the expansion (110) holds not only for the uniform distribution over the power sphere $\mathcal{F}_n$; rather it holds for every probability distribution P_{X^n} supported on $\mathcal{F}_n$.

To evaluate $\beta_\tau(P_{Y^n}, Q_{Y^n})$, we shall make use of the following lemma, which provides a variational characterization of the β function.

Lemma 5: For every pair of probability measures P and Q on $\mathcal{X}$ and every $\alpha \in (0,1)$, we have that

$$\beta_\alpha(P, Q) = \max_R \beta_{\beta_\alpha(P,R)}(R, Q) \tag{111}$$

where the maximum is over all probability measures R on $\mathcal{X}$. Furthermore, the maximum is achieved by all probability measures contained in the one-parameter exponential family $\{R_\lambda : \lambda \in (0,1)\}$ connecting P and Q. Specifically, R_λ is defined as

$$\frac{\mathrm{d}R_\lambda}{\mathrm{d}\mu}(x) \triangleq e^{-\lambda D_{1-\lambda}(P\|Q)} \left(\frac{\mathrm{d}P}{\mathrm{d}\mu}(x)\right)^{1-\lambda} \left(\frac{\mathrm{d}Q}{\mathrm{d}\mu}(x)\right)^\lambda. \tag{112}$$

Here, $D_{1-\lambda}(P\|Q)$ denotes the Rényi divergence of order $1-\lambda$ [46], and μ is a measure on $\mathcal{X}$ that satisfies $P \ll \mu$ and $Q \ll \mu$.

Proof: See Appendix I-E. ∎

Let $(P_Y^*)^n = \mathcal{CN}(\mathbf{0}, (1+P)\mathsf{I}_n)$ be the product distribution obtained from the capacity-achieving output distribution of an AWGN channel with SNR P. It follows from Lemma 5 that

$$\beta_\tau(P_{Y^n}, Q_{Y^n}) \geq \beta_{\beta_\tau(P_{Y^n},(P_Y^*)^n)}((P_Y^*)^n, Q_{Y^n}). \tag{113}$$

By (35) and [30, Prop. 2],

$$\beta_\tau(P_{Y^n}, (P_Y^*)^n) \geq c_1 \tau \frac{\sqrt{1+2P}}{1+P} \triangleq \hat{\tau} \tag{114}$$

where c_1 is a positive constant independent of P. Since $\tau \mapsto \beta_\tau$ is nondecreasing, it follows from (113) and (114) that

$$\beta_\tau(P_{Y^n}, Q_{Y^n}) \geq \beta_{\hat{\tau}}((P_Y^*)^n, Q_{Y^n}). \tag{115}$$

To evaluate the RHS of (115), we use the Neyman-Pearson lemma and that both $(P_Y^*)^n$ and Q_{Y^n} are product distributions. Specifically, under $(P_Y^*)^n$, the random variable $\log \frac{\mathrm{d}(P_Y^*)^n}{\mathrm{d}Q_{Y^n}}(Y^n)$ has the same distribution as

$$\sum_{i=1}^n \left(|Z_i|^2 P \log e - \log(1+P)\right) \tag{116}$$

where the $\{Z_i\}$ are i.i.d. $\mathcal{CN}(0,1)$ random variables. Furthermore,

$$\mathbb{E}\left[|Z_i|^2 P \log e - \log(1+P)\right] = \frac{1}{2}P^2 \log e + \mathcal{O}(P^3) \tag{117}$$

$$\mathbb{V}\mathrm{ar}\left[|Z_i|^2 P \log e - \log(1+P)\right] = P^2 \log^2 e. \tag{118}$$

Using [5, Lemma 59], and (116)–(118), we conclude that

$$\log \beta_{\hat{\tau}}((P_Y^*)^n, Q_{Y^n}) \geq -\frac{1}{2}nP^2 \log e - \sqrt{\frac{2nP^2 \log^2 e}{\hat{\tau}}} + \log \frac{\hat{\tau}}{2} + \mathcal{O}(nP^3) \tag{119}$$

$$= -\frac{1}{2}nP^2 \log e + o(nP^2). \tag{120}$$

Here, the second and third terms on the RHS of (119) behave as $o(nP^2)$ due to (104), (115), and the assumption that $nP^2 \to \infty$. Substituting (120) into (115), and then (110) and (115) into (105), we conclude that (54) is achievable.

C. Proof of (54): Converse

As in [5, Section III.J], we can assume without loss of generality that each codeword x^n satisfies the power constraint (43) with equality, i.e., $x^n \in \mathcal{F}_n$. To prove the converse, we shall use the $\beta\beta$ converse bound (9) with $Q_{Y^n} = \mathcal{CN}(\mathbf{0}, \mathsf{I}_n)$ and with

$$\delta = 24\sqrt{2}n^{-1/2} + e^{-\sqrt{nP^2}}. \tag{121}$$

This yields

$$R^*(n, \epsilon, P) \leq \sup_{P_{X^n}} \frac{1}{n} \log \frac{\beta_{1-\delta}(P_{Y^n}, Q_{Y^n})}{\beta_{1-\epsilon-\delta}(P_{X^nY^n}, P_{X^n}Q_{Y^n})} \tag{122}$$

where the supremum is over all probability distributions P_{X^n} supported on $\mathcal{F}_n$, and $P_{Y^n} = P_{X^n} \circ P_{Y^n|X^n}$.

For this choice of parameters, the denominator on the RHS of (122) admits the same asymptotic expansion as (110). Next, we evaluate the numerator $\beta_{1-\delta}(P_{Y^n}, Q_{Y^n})$. Consider the following test between P_{Y^n} and Q_{Y^n}:

$$T(y^n) = \mathbb{1}\{2\|y^n\|^2 \geq \gamma\} \tag{123}$$

where γ is chosen so that

$$P_{Y^n}\Big[2\|Y^n\|^2 \geq \gamma\Big] = 1 - \delta. \tag{124}$$

By the Neyman-Pearson lemma,

$$\beta_{1-\delta}(P_{Y^n}, Q_{Y^n}) \leq Q_{Y^n}\Big[2\|Y^n\|^2 \geq \gamma\Big]. \tag{125}$$

Note that under P_{Y^n} the random variable $2\|Y^n\|^2$ has the same distribution as $\sum_{i=1}^{2n}(Z_i + \sqrt{P})^2$ with $\{Z_i\}$ i.i.d. and $\mathcal{N}(0, 1)$-distributed, and regardless of the choice of P_{X^n} (provided that P_{X^n}

is supported on $\mathcal{F}_n$). Next, we estimate $P_{Y^n}[2\|Y^n\|^2 \geq \gamma]$ using the Berry-Esséen theorem (see, e.g., [40, Ch. XVI.5]) as follows:

$$\left| P_{Y^n}\left[2\|Y^n\|^2 \geq \gamma\right] - Q\left(\frac{\gamma - 2n(1+P)}{\sqrt{4n(1+2P)}}\right) \right| \leq \frac{6\mathbb{E}\left[|(Z_1 + \sqrt{P})^2 - 1 - P|^3\right]}{(2+4P)^{3/2}\sqrt{2n}} \tag{126}$$

$$\leq 24\sqrt{2}n^{-1/2}. \tag{127}$$

Here, the last step follows because

$$\mathbb{E}\left[|(Z_1 + \sqrt{P})^2 - 1 - P|^3\right] \leq \left(\mathbb{E}\left[|(Z_1 + \sqrt{P})^2 - 1 - P|^4\right]\right)^{3/4} \tag{128}$$

$$= (12(1+2P)^2 + 48(1+4P))^{3/4} \tag{129}$$

$$\leq 64^{3/4}(1+2P)^{3/2}. \tag{130}$$

Using (124) and (127), we obtain

$$\gamma \geq 2n(1+P) - \sqrt{4n(1+2P)}Q^{-1}(\delta - 24\sqrt{2}n^{-1/2}) \triangleq \widetilde{\gamma}. \tag{131}$$

Using (121) and the expansion

$$Q^{-1}(t) = \sqrt{2\log_e(1/t)}(1 + o(1)), \quad t \to 0 \tag{132}$$

which follows from (109), we conclude that the threshold $\widetilde{\gamma}$ behaves as

$$\widetilde{\gamma} = 2n(1+P) + \sqrt{8n(1+2P)}(nP^2)^{1/4}(1 + o(1)) \tag{133}$$

$$= 2n(1+P) + o(nP) \tag{134}$$

in the limit $n \to \infty$. Under Q_{Y^n}, the random variable $2\|Y^n\|$ has the same distribution as $\sum_{i=1}^{2n} Z_i^2$ with $\{Z_i\}$ i.i.d. and $\mathcal{N}(0,1)$-distributed. Next, we use the moderate-deviation bound [47, Th. 3.7.1] to evaluate $Q_{Y^n}[2\|Y^n\|^2 \geq \gamma]$ as follows:

$$\limsup_{n\to\infty} \frac{2n}{(\widetilde{\gamma} - 2n)^2} \log Q_{Y^n}[2\|Y^n\|^2 \geq \gamma]$$

$$\leq \limsup_{n\to\infty} \frac{2n}{(\widetilde{\gamma} - 2n)^2} \log \mathbb{P}\left[\sum_{i=1}^{2n} Z_i^2 \geq \widetilde{\gamma}\right] \tag{135}$$

$$= \limsup_{n\to\infty} \frac{2n}{(\widetilde{\gamma} - 2n)^2} \log \mathbb{P}\left[\frac{1}{\widetilde{\gamma} - 2n}\left(\sum_{i=1}^{2n} Z_i^2 - 2n\right) \geq 1\right] \tag{136}$$

$$\leq -\frac{1}{4}\log e. \tag{137}$$

Here, (135) follows from (131), and (137) follows by using [47, Th. 3.7.1] with $a_n = 2n(\widetilde{\gamma}_n - 2n)^{-2}$ and $\Gamma = [1, \infty]$. Combining (125), (134), and (137), we obtain that

$$\log \beta_{1-\delta}(P_{Y^n}, Q_{Y^n}) \leq -\frac{\log e}{4} \frac{(\widetilde{\gamma} - 2n)^2}{2n}(1 + o(1)) \tag{138}$$

$$= -\frac{1}{2}nP^2 \log e + o(nP^2). \tag{139}$$

We conclude the converse proof of (54) by substituting (110) and (137) into (122).

D. Proof of (46)

Note that the converse part of (46) follows directly from (39) since, by definition,

$$E_{\mathrm{b}}^*(k, \epsilon, R) \geq E_{\mathrm{b}}^*(k, \epsilon). \tag{140}$$

Thus, it remains to show that (46) is achievable under the conditions $k \to \infty$, $R \to 0$, and $kR \to c < \infty$, which implies that (46) is achievable with codes of blocklength k^2/c. Without loss of generality, we can assume that $c > 0$. Indeed, we can always transform a code that satisfies $kR \to \tilde{c} > 0$ into a code satisfying $kR \to 0$ by appending k^3 zero symbols to each codeword, without changing the total energy of each codeword.

Applying the $\beta\beta$ bound (12) with the same P_{X^n} and the same Q_{Y^n} as in Appendix I-B but with $\tau = (nP)^{-1/2}$, we conclude that there exists a sequence of (n, M, ϵ) codes that satisfy

$$M \geq \frac{\beta_\tau(P_{Y^n}, Q_{Y^n})}{\beta_{1-\epsilon+\tau}(P_{X^nY^n}, P_{X^n}Q_{Y^n})}. \tag{141}$$

This time, we shall study the asymptotic behavior of (141) in the asymptotic limit $n \to \infty$, $P \to 0$, and $nP^2 \to c/\log_2^2 e$ (which is equivalent to the limit $k \to \infty$, $R \to 0$, and $kR \to c$). The denominator satisfies the same expansion as in (110), whereas the numerator satisfies

$$\log \beta_\tau(P_{Y^n}, Q_{Y^n}) \geq -\frac{1}{2}nP^2 \log e - \sqrt{\frac{2nP^2 \log^2 e}{\hat{\tau}}} + \log \frac{\hat{\tau}}{2} + \mathcal{O}(nP^3) \tag{142}$$

$$= \mathcal{O}((nP)^{1/4}). \tag{143}$$

Here, $\hat{\tau}$ is defined in (114), and (142) follows from (119). Substituting (110) and (143) into (141), taking the logarithm (with base 2) on both sides, we conclude that

$$k \geq nP \log_2 e - \sqrt{2nP}Q^{-1}(\epsilon) \log_2 e + o\left(\sqrt{nP}\right). \tag{144}$$

Using the relations (98) and the bound (144) and proceeding as in (99)–(102), we conclude that (46) is achievable.

E. Proof of Lemma 5

Let $T : \mathcal{X} \to \{0, 1\}$ be the (possibly randomized) Neyman-Pearson test that achieves $\beta_\alpha(P, Q)$. For every probability measure R on $\mathcal{X}$, it follows that

$$\beta_\alpha(P, Q) = Q[T = 1] \tag{145}$$

$$\geq \beta_{R[T=1]}(R, Q) \tag{146}$$

$$\geq \beta_{\beta_\alpha(P,R)}(R, Q). \tag{147}$$

Here, (146) follows from the definition of $\beta_\alpha(R, Q)$, and (147) follows because $R[T = 1] \geq \beta_{P[T=1]}(P, R)$, because $P[T = 1] = \alpha$ by definition of T, and because $\alpha \mapsto \beta_\alpha(R, Q)$ is monotonically nondecreasing. Maximizing the RHS of (147) over all probability measures R on $\mathcal{X}$, we obtain

$$\beta_\alpha(P, Q) \geq \sup_R \beta_{\beta_\alpha(P,R)}(R, Q). \tag{148}$$

It remains to show that for the $\{R_\lambda\}$, $\lambda \in (0, 1)$, defined in (112), we have

$$\beta_\alpha(P, Q) = \beta_{\beta_\alpha(P,R_\lambda)}(R_\lambda, Q). \tag{149}$$

Indeed, we observe that[9]

$$\log \frac{\mathrm{d}P}{\mathrm{d}R_\lambda}(x) = \lambda D_{1-\lambda}(P\|Q) + \lambda \log \frac{\mathrm{d}P}{\mathrm{d}Q}(x) \tag{150}$$

and

$$\log \frac{\mathrm{d}R_\lambda}{\mathrm{d}Q}(x) = -\lambda D_{1-\lambda}(P\|Q) + (1 - \lambda) \log \frac{\mathrm{d}P}{\mathrm{d}Q}(x) \tag{151}$$

for every x in the support of Q. The identities (150) and (151) imply that the test T in (145)–(147) coincides with the Neyman-Pearson test for distinguishing between P and R_λ and between R_λ and Q. This in turn implies, by the Neyman-Pearson lemma, that both (146) and (147) hold with equality. This establishes (149).

---

[9]In the case where P is not absolutely continuous with respect to Q, we set $\mathrm{d}P/\mathrm{d}Q = +\infty$ on the singular set.

APPENDIX II

PROOF OF THEOREM 3

As in Appendix I-A, to prove (61) it is sufficient to show that the maximum coding rate $R^*(n, \epsilon, P)$ satisfies

$$\frac{R^*(n, \epsilon, P)}{\log e} = P - \sqrt{\frac{2P}{n}} Q^{-1}(\epsilon) - P^2 + o\left(\sqrt{\frac{P}{n}}\right) + o(P^2) \tag{152}$$

in the asymptotic limit $n \to \infty$, $P \to 0$, and $nP^2 \to \infty$. The achievability and the converse part of (152) are proved in Appendices II-A and II-B, respectively. The proof of (63) follows similar steps as the ones reported in Appendix I-D, and is thus omitted. Unless otherwise specified, all asymptotic expansions in Appendices II-A and II-B hold in the limit $n \to \infty$, $P \to 0$, and $nP^2 \to \infty$.

A. Proof of (152): Achievability

We shall apply the $\beta\beta$ achievability bound (12) to the channel (59) with input X^n and output (Y^n, H^n) (recall that we assumed perfect CSIR) with the same choices of P_{X^n} and Q_{Y^n} as in Section I-B. Namely, P_{X^n} is chosen as the uniform distribution over the power sphere $\mathcal{F}_n$ defined in (103), and $Q_{Y^n} = \mathcal{CN}(\mathbf{0}, \mathsf{I}_n)$. Furthermore, we set

$$\tau = P + (nP^2)^{-1/2}. \tag{153}$$

Since $\tau \to 0$ as $P \to 0$ and $nP^2 \to \infty$, and since we are interested in the asymptotic behavior of $R^*(n, \epsilon, P)$ as $n \to \infty$, $P \to 0$, and $nP^2 \to \infty$, we can assume without loss of generality that $\tau < \epsilon$. It follows from (12) that

$$nR^*(n, \epsilon, P) \geq \log \beta_\tau(P_{Y^n H^n}, Q_{Y^n} P_{H^n}) - \log \beta_{1-\epsilon+\tau}(P_{X^n Y^n H^n}, P_{X^n} Q_{Y^n} P_{H^n}) \tag{154}$$

where $P_{Y^n H^n} = P_{X^n} \circ P_{Y^n H^n | X^n}$.

To evaluate the second term on the RHS of (154), we use [5, Eq. (103)] and obtain

$$- \log \beta_{1-\epsilon+\tau}(P_{X^n Y^n H^n}, P_{X^n} Q_{Y^n} P_{H^n}) \geq \gamma_0 \tag{155}$$

where γ_0 satisfies

$$P_{X^n Y^n H^n}\left[\log \frac{\mathrm{d}P_{X^n Y^n H^n}}{\mathrm{d}(P_{X^n} Q_{Y^n} P_{H^n})} \leq \gamma_0\right] = \epsilon - \tau. \tag{156}$$

Observe now that, under $P_{X^n Y^n H^n}$, the random variable $\log \frac{\mathrm{d}P_{X^n Y^n H^n}}{\mathrm{d}(P_{X^n} Q_{Y^n} P_{H^n})}(X^n, Y^n, H^n)$ has the same distribution as

$$\log e \sum_{i=1}^{n} \left(|H_i X_i|^2 + 2\mathrm{Re}(H_i X_i Z_i^*) \right). \tag{157}$$

Next, we use the central limit theorem for functions [48, Prop. 1] (see also [30, Prop. 1]) to derive an asymptotic expansion for γ_0 in (156). Specifically, let $\widetilde{X}^n \sim \mathcal{CN}(\mathbf{0}, P\mathsf{I}_n)$. It follows that $X^n \sim P_{X^n}$ has the same distribution as $\sqrt{nP}\widetilde{X}^n / \|\widetilde{X}^n\|$. Let

$$A_{1,i} \triangleq |H_i \widetilde{X}_i|^2 - P \tag{158}$$

$$A_{2,i} \triangleq |\widetilde{X}_i|^2 - P \tag{159}$$

$$A_{3,i} \triangleq 2\mathrm{Re}(H_i \widetilde{X}_i Z_i^*), \quad i = 1, \ldots, n. \tag{160}$$

The random vectors $\{[A_{1,i}, A_{2,i}, A_{3,i}]\}$ are i.i.d. with zero mean and covariance matrix

$$\mathsf{V} = \begin{bmatrix} 3P^2 & P^2 & 0 \\ P^2 & P^2 & 0 \\ 0 & 0 & 2P \end{bmatrix}. \tag{161}$$

Let $g : \mathbb{R}^3 \to \mathbb{R}$ be defined as

$$g(a_1, a_2, a_3) \triangleq \frac{(a_1 + P)P}{P + a_2} + \frac{a_3 \sqrt{P}}{\sqrt{P + a_2}} \tag{162}$$

and observe that

$$(n \log e) \cdot g\left(\frac{1}{n} \sum_{i=1}^{n} A_{1,i}, \frac{1}{n} \sum_{i=1}^{n} A_{2,i}, \frac{1}{n} \sum_{i=1}^{n} A_{3,i} \right) \tag{163}$$

has the same distribution as (157). Finally, let $\boldsymbol{j}$ denote the gradient of g at $(0, 0, 0)$. It follows that

$$\boldsymbol{j}\mathsf{V}\boldsymbol{j}^{\mathrm{T}} = 2P^2 + 2P. \tag{164}$$

We are now ready to invoke [48, Prop. 1] and conclude that for every $\gamma \in \mathbb{R}$

$$\mathbb{P}\left[\sum_{i=1}^{n} \left(|H_i X_i|^2 + 2\mathrm{Re}(H_i X_i Z_i^*) \right) \leq \gamma \right] \leq Q\left(\frac{nP - \gamma}{\sqrt{2n(P + P^2)}} \right) + \mathcal{O}\left(n^{-1/2} \right). \tag{165}$$

Setting the RHS of (165) equal to $\epsilon - \tau$ and solving for γ, we obtain an asymptotic lower bound on γ_0, which we use to further lower-bound (155) as follows:

$$-\log \beta_{1-\epsilon+\tau}(P_{X^n} P_{Y^n H^n | X^n}, P_{X^n} Q_{Y^n} P_{H^n})$$

$$\geq nP \log e - \sqrt{2nP(1+P)} Q^{-1}\left(\epsilon - \tau - \mathcal{O}(n^{-1/2})\right) \log e \qquad (166)$$

$$= nP \log e - \sqrt{2nP} Q^{-1}(\epsilon) \log e + o\left(\sqrt{nP}\right). \qquad (167)$$

Here, (167) follows by Taylor-expanding $\sqrt{1+P}$ for $P \to 0$ and by Taylor-expanding $Q^{-1}(\cdot)$ around ϵ for $\tau \to 0$ and $n \to \infty$.

To evaluate $\beta_\tau(P_{Y^n H^n}, Q_{Y^n} P_{H^n})$ on the RHS of (154), we again use Lemma 5 in Appendix I-B. Let $(P_{YH}^*)^n$ be the product distribution obtained from the capacity-achieving output distribution of the channel (59) with SNR P. Then, by Lemma 5,

$$\beta_\tau(P_{Y^n H^n}, Q_{Y^n} P_{H^n}) \geq \beta_{\beta_\tau(P_{Y^n H^n}, (P_{YH}^*)^n)}((P_{YH}^*)^n, Q_{Y^n} P_{H^n}). \qquad (168)$$

We lower-bound $\beta_\tau(P_{Y^n H^n}, (P_{YH}^*)^n)$ by following similar steps as the ones reported in Section V-B. Specifically, let $P_{Y^n H^n | X^n}^{(s)}$ be the transition probability of the following channel:

$$Y_i = \frac{\sqrt{nP} H_i X_i}{\|X^n\|} + Z_i, \qquad i = 1, \ldots, n. \qquad (169)$$

Let $(P_X^*)^n = \mathcal{CN}(\mathbf{0}, P\mathsf{I}_n)$ be the product distribution obtained from the capacity-achieving input distribution for the channel (59) under perfect CSIR. Then, we have that $P_{Y^n H^n} = (P_X^*)^n \circ P_{Y^n H^n | X^n}^{(s)}$ and $(P_{YH}^*)^n = (P_X^*)^n \circ P_{Y^n H^n | X^n}$. By the data-processing inequality,

$$\beta_\tau(P_{Y^n H^n}, (P_{YH}^*)^n) \geq \beta_\tau((P_X^*)^n P_{Y^n H^n | X^n}^{(s)}, (P_X^*)^n P_{Y^n H^n | X^n}). \qquad (170)$$

To lower-bound the RHS of (170), we shall use the following bound [49] (see also [5, Eqs. (154)–(156)]):

$$\log \beta_\alpha(P, Q) \geq -\frac{D(P\|Q) + h_{\mathrm{b}}(\alpha)}{\alpha} \qquad (171)$$

where $h_{\mathrm{b}}(\alpha) \triangleq -\alpha \log \alpha - (1-\alpha) \log(1-\alpha)$ denotes the binary entropy function. This yields

$$\beta_\tau((P_X^*)^n P_{Y^n H^n | X^n}^{(s)}, (P_X^*)^n P_{Y^n H^n | X^n})$$

$$\geq \exp\left(-\frac{D((P_X^*)^n P_{Y^n H^n | X^n}^{(s)} \| (P_X^*)^n P_{Y^n H^n | X^n}) + h_{\mathrm{b}}(\tau)}{\tau}\right). \qquad (172)$$

Note now that

$$D\Big((P_X^*)^n P_{Y^n H^n|X^n}^{(s)} \big\| (P_X^*)^n P_{Y^n H^n|X^n}\Big)$$

$$= D\Big(P_{Y^n|H^n X^n}^{(s)} \big\| P_{Y^n|H^n X^n} \big| P_{H^n}(P_X^*)^n\Big) \tag{173}$$

$$= \mathbb{E}_{(P_X^*)^n}\left[\sum_{i=1}^n \left|\frac{\sqrt{nP}H_i X_i}{\|X^n\|} - H_i X_i\right|^2\right]\log e \tag{174}$$

$$= \mathbb{E}_{(P_X^*)^n}\left[\big(\sqrt{nP} - \|X^n\|\big)^2\right]\log e \tag{175}$$

$$= 2nP\left(1 - \frac{\Gamma(n+1/2)}{\sqrt{n}\Gamma(n)}\right)\log e \tag{176}$$

$$\leq 2nP\left(1 - \sqrt{1 - \frac{1}{2n+1}}\right)\log e \tag{177}$$

$$\leq P\frac{2n}{2n+1}\log e \tag{178}$$

$$\leq P\log e. \tag{179}$$

Here, (176) follows because $\mathbb{E}_{(P_X^*)^n}[\|X^n\|] = \sqrt{P}\Gamma(n+1/2)/\Gamma(n)$; (177) follows from Wendel's inequality [50, Eq. (7)]; and (178) follows because $\sqrt{1-x} \geq 1-x$ for every $x \in [0,1]$. Substituting (179) in (172), we obtain

$$\beta_\tau(P_{Y^n H^n}, (P_{YH}^*)^n) \geq \exp\left(-\frac{P\log e + h_{\mathrm{b}}(\tau)}{\tau}\right) \tag{180}$$

$$\geq e^{-2}\tau \triangleq \hat{\tau}. \tag{181}$$

In the first step we used that $\tau \geq P$ and that

$$\frac{h_{\mathrm{b}}(\tau)}{\tau} = -\log\tau + \frac{1-\tau}{\tau}\log\frac{1}{1-\tau} \leq -\log\tau + \log e. \tag{182}$$

Since $\alpha \mapsto \beta_\alpha(P, Q)$ is nondecreasing, we conclude from (168) and (181) that

$$\beta_\tau(P_{Y^n H^n}, Q_{Y^n}P_{H^n}) \geq \beta_{\hat{\tau}}((P_{YH}^*)^n, Q_{Y^n}P_{H^n}). \tag{183}$$

We next lower-bound the RHS of (183) by using the Neyman-Pearson lemma and that both $(P_{YH}^*)^n$ and $Q_{Y^n}P_{H^n}$ are product distributions. Specifically, under $(P_{YH}^*)^n$, the random variable $\log\frac{\mathrm{d}(P_{YH}^*)^n}{\mathrm{d}(Q_{Y^n}P_{H^n})}(Y^n, H^n)$ has the same distribution as

$$\sum_{i=1}^n \underbrace{|Z_i|^2|H_i|^2 P\log e - \log(1 + |H_i|^2 P)}_{\triangleq B_i} \tag{184}$$

where the random variables $\{B_i\}$ defined above are i.i.d.. Let

$$I_n \triangleq \mathbb{E}[B_i], \quad V_n \triangleq \mathbb{V}\mathrm{ar}[B_i]. \tag{185}$$

A straightforward computation reveals that

$$I_n = \frac{\mathbb{E}[|H|^4]}{2} P^2 \log e + o(P^2) \tag{186}$$

and that V_n can be bounded as follows:

$$3P^2 \log^2 e \leq V_n \leq 11 P^2 \log^2 e. \tag{187}$$

By [5, Lemma 59],

$$\log \beta_{\hat{\tau}}((P_{YH}^*)^n, Q_{Y^n} P_{H^n}) \geq -nI_n - \sqrt{\frac{2nV_n}{\hat{\tau}}} + \frac{1}{2} \log \frac{\hat{\tau}}{2} \tag{188}$$

$$= -\frac{\mathbb{E}[|H|^4]}{2} n P^2 \log e + o(nP^2). \tag{189}$$

Here, in (189) we used (153), (186), and (187). Finally, substituting (189) into (183), then (167) and (183) into (154), and using that $\mathbb{E}[|H|^4] = 2$, we conclude that (152) is achievable.

B. Proof of (152): Converse

It follows from [5, Lemma 39] that we can assume without loss of generality that each codeword x^n of a given (n, M, ϵ) code satisfies the power constraint (43) with equality. Furthermore, by arguing as in [5, Eqs. (284)–(286)], we can assume without loss of generality that the *maximum* probability of error of the code is upper-bounded by ϵ. This allows us to use the maximum-error-probability version [18, Eq. (222)] of the $\beta\beta$ converse bound (9). Particularizing this bound to the channel (59), we conclude that every (n, M, ϵ) code $\mathcal{C}$ (maximum probability of error) satisfies

$$\beta_\alpha(P_{Y^n H^n}, Q_{Y^n H^n}) \geq \frac{\delta M}{1 - \alpha + \delta} \inf_{x^n \in \mathcal{C}} \beta_{\alpha - \epsilon - \delta}(P_{Y^n H^n | X^n = x^n}, Q_{Y^n H^n}) \tag{190}$$

where $\epsilon + \delta \leq \alpha \leq 1$, $\delta > 0$, and $P_{Y^n H^n}$ denotes the output distribution induced by the code.

To evaluate the bound (190), we shall choose $Q_{Y^n H^n} = Q_{Y^n} P_{H^n}$ with $Q_{Y^n} = \mathcal{CN}(\mathbf{0}, \mathsf{I}_n)$, and set $\delta = \delta_n$ and $\alpha = 1 - \delta_n$ with δ_n chosen such that $\delta_n \to 0$ as $n \to \infty$. With these choices, we obtain

$$\log M \leq \log \beta_{1 - \delta_n}(P_{Y^n H^n}, Q_{Y^n} P_{H^n})$$

$$- \inf_{x^n \in \mathcal{C}} \log \beta_{1 - \epsilon - 2\delta_n}(P_{Y^n H^n | X^n = x^n}, Q_{Y^n} P_{H^n}) + \log 2. \tag{191}$$

The second term on the RHS of (191) is upper-bounded by

$$- \inf_{x^n \in \mathcal{C}} \log \beta_{1-\epsilon-2\delta_n}(P_{Y^n H^n | X^n = x^n}, Q_{Y^n} P_{H^n})$$

$$\leq - \inf_{x^\infty : \|x^\infty\|^2 = nP} \log \beta_{1-\epsilon-2\delta_n}(P_{Y^\infty H^\infty | X^\infty = x^\infty}, P_{Y^\infty | X^\infty = \mathbf{0}} P_{H^\infty}) \tag{192}$$

$$\leq nP \log e - \sqrt{2nP} Q^{-1}(\epsilon - 2\delta_n) \log e + \mathcal{O}(\log(nP)) \tag{193}$$

$$= nP \log e - \sqrt{2nP} Q^{-1}(\epsilon) \log e + o(\sqrt{nP}). \tag{194}$$

Here, x^∞ denotes the infinite-dimensional sequence $(x_1, x_2, ...)$, $P_{Y^\infty H^\infty | X^\infty} = \prod_{i=1}^\infty P_{Y_i H_i | X_i}$, and $P_{Y^\infty | X^\infty = \mathbf{0}} = \prod_{i=1}^n P_{Y_i | X_i = 0}$. The inequality (192) follows because the feasible region of the optimization problem on the LHS of (192) is contained in the feasible region of the optimization problem on the RHS of (192); (193) follows from [35, App. IV] (in particular, see [35, Eqs. (243) and (267)]); and in (194) we have used that $\delta_n \to 0$ as $n \to \infty$. Note that, the RHS of (194) holds for all (n, M, ϵ) codes.

To conclude the proof, it is sufficient to show that there exists a vanishing sequence $\{\delta_n\}$ such that for every (n, M, ϵ) code

$$\log \beta_{1-\delta_n}(P_{Y^n H^n}, Q_{Y^n} P_{H^n}) \leq -\frac{\mathbb{E}[|H|^4]}{2} nP^2 \log e + o(nP^2). \tag{195}$$

The idea is to construct a suboptimal test to distinguish between $P_{Y^n H^n}$ and $Q_{Y^n} P_{H^n}$. Coarsely speaking, our test is constructed as follows: if $P_{Y^n H^n}$ is induced by a code whose codewords have suitably small ℓ_4 norm, then we use as test the optimal Neyman-Pearson test between the capacity-achieving output distribution $(P_{YH}^*)^n$ and $Q_{Y^n} P_{H^n}$. Indeed, we expect the output distribution induced by such a code to resemble the capacity-achieving output distribution $(P_{YH}^*)^n$. If instead $P_{Y^n H^n}$ is induced by a code whose codewords are peaky in an ℓ_4 sense, we distinguish between $P_{Y^n H^n}$ and $Q_{Y^n} P_{H^n}$ simply by testing the peakiness of Y^n.

We proceed now with the proof. We start by dividing the codebook $\mathcal{C}$ into two subcodebooks $\mathcal{C}_1$ and $\mathcal{C}_2$ according to the peakiness of the codewords. More specifically, we set

$$\mathcal{C}_1 \triangleq \{x^n \in \mathcal{C} : \|x^n\|_4^4 \leq n\eta_n\} \tag{196}$$

$$\mathcal{C}_2 \triangleq \mathcal{C} \setminus \mathcal{C}_1 = \{x^n \in \mathcal{C} : \|x^n\|_4^4 > n\eta_n\} \tag{197}$$

where the threshold η_n is defined as[10]

$$\eta_n \triangleq \max\left\{nP^{5/2}, (nP^2)^{3/4}\right\}. \tag{198}$$

Note that $\eta_n \to \infty$ as $n \to \infty$. Let $\lambda \triangleq |\mathcal{C}_1|/|\mathcal{C}|$. Furthermore, let $P_{Y^nH^n}^{(1)}$ and $P_{Y^nH^n}^{(2)}$ denote the output probability distributions induced by the subcodes $\mathcal{C}_1$ and $\mathcal{C}_2$, respectively. It follows that

$$P_{Y^nH^n} = \lambda P_{Y^nH^n}^{(1)} + (1-\lambda)P_{Y^nH^n}^{(2)}. \tag{199}$$

Lemma 6 below allows us to upper-bound $\beta_{1-\delta_n}(P_{Y^nH^n}, Q_{Y^n}P_{H^n})$ by $\beta_{1-\delta_{1,n}}(P_{Y^nH^n}^{(1)}, Q_{Y^n}P_{H^n})+$ $\beta_{1-\delta_{2,n}}(P_{Y^nH^n}^{(2)}, Q_{Y^n}P_{H^n})$ for some suitably chosen $\delta_{1,n}$ and $\delta_{2,n}$.

Lemma 6: Let $P = \lambda P_1 + (1-\lambda)P_2$ be a convex combination of P_1 and P_2. Then, for every probability measure Q, and every $\delta_1, \delta_2 \in (0,1)$, we have

$$\beta_{1-\lambda\delta_1-(1-\lambda)\delta_2}(P, Q) \leq \beta_{1-\delta_1}(P_1, Q) + \beta_{1-\delta_2}(P_2, Q). \tag{200}$$

Proof: See Appendix II-C. ∎

Set now

$$\delta_{1,n} = \max\left\{P^{1/4}, (nP^2)^{-1/8}\right\} \tag{201}$$

$$\delta_{2,n} = 1 - \exp\left(-4\tilde{\xi}_n^{1/2}P/\eta_n\right)\left(1 - \frac{2n}{\tilde{\xi}_n}\right) \tag{202}$$

with $Z_1 \sim \mathcal{N}(0,1)$, and

$$\tilde{\xi}_n \triangleq \max\left\{n^2P^{15/4}, n^{9/8}P^{1/4}\right\}. \tag{203}$$

Furthermore, set

$$\delta_n = \lambda\delta_{1,n} + (1-\lambda)\delta_{2,n}. \tag{204}$$

It can be shown that the sequences $\{\delta_{1,n}\}$, $\{\delta_{2,n}\}$, and $\{\delta_n\}$ all vanish as $n \to \infty$, $P \to 0$, and $nP^2 \to \infty$. We shall prove that

$$\log\beta_{1-\delta_{1,n}}(P_{Y^nH^n}^{(1)}, Q_{Y^n}P_{H^n}) \leq -\frac{\mathbb{E}[|H|^4]}{2}nP^2\log e + o(nP^2) \tag{205}$$

and that

$$\log\beta_{1-\delta_{2,n}}(P_{Y^nH^n}^{(2)}, Q_{Y^n}P_{H^n}) \leq -\frac{\mathbb{E}[|H|^4]}{2}nP^2\log e + o(nP^2). \tag{206}$$

The desired bound (195) then follows from (199), (205), (206) and Lemma 6. The proof of (205) and (206) are provided in Appendices II-B1 and II-B2, respectively.

---

[10]When the entries of X^n are independent Gaussian random variables, we expect to have $\|X^n\|_4^4 \approx \mathcal{O}(nP^2)$. This means that for a random Gaussian code $|\mathcal{C}_1| \gg |\mathcal{C}_2|$.

1) Proof of (205)*:* To upper-bound $\beta_{1-\delta_{1,n}}(P^{(1)}_{Y^n H^n}, Q_{Y^n} P_{H^n})$, we consider the following test between $P^{(1)}_{Y^n H^n}$ and $Q_{Y^n} P_{H^n}$:

$$T(y^n, h^n) \triangleq \mathbb{1}\left\{ \sum_{i=1}^{n} \frac{|h_i|^2(|y_i|^2 - 1)}{1 + |h_i|^2\sqrt{P}} \geq \xi_n \right\} \tag{207}$$

where the threshold ξ_n satisfies

$$P^{(1)}_{Y^n H^n}[T = 1] \geq 1 - \delta_{1,n}. \tag{208}$$

As mentioned earlier, this test is related to the Neyman-Pearson test between the capacity-achieving output distribution $(P^*_{YH})^n$ and $Q_{Y^n} P_{H^n}$. The term $(1 + |h_i|^2\sqrt{P})$ in the denominator of (207) is included because the moment generating function of the random variable $|H_i|^2(|Y_i|^2 - 1)$ (with $Y^n \sim Q_{Y^n}$) does not exist. It follows from (208) and from the definition of the β function that

$$\beta_{1-\delta_n}(P^{(1)}_{Y^n H^n}, Q_{Y^n} P_{H^n}) \leq Q_{Y^n} P_{H^n}[T = 1]. \tag{209}$$

To evaluate the RHS of (209), we first determine ξ_n. Let

$$A_i \triangleq \frac{|H_i(H_i X_i + Z_i)|^2 - |H_i|^2}{1 + |H_i|^2\sqrt{P}}. \tag{210}$$

Then

$$P^{(1)}_{Y^n H^n}[T = 1] = \mathbb{E}_{X^n}\left[\mathbb{P}\left[\sum_{i=1}^{n} A_i \geq \xi_n \Big| X^n \right] \right]. \tag{211}$$

Set now

$$\xi_n = \min_{x^n \in \mathcal{C}_1} \left\{ \sum_{i=1}^{n} \mathbb{E}[A_i | X^n = x^n] - \sqrt{\delta_{1,n}^{-1} \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}[A_i | X^n = x^n]} \right\}. \tag{212}$$

We have that for every $x^n \in \mathcal{C}_1$

$$\mathbb{P}\left[\sum_{i=1}^{n} A_i \leq \xi_n \Big| X^n = x^n \right]$$

$$\leq \mathbb{P}\left[\sum_{i=1}^{n} A_i \leq \sum_{i=1}^{n} \mathbb{E}[A_i | X^n = x^n] - \sqrt{\delta_{1,n}^{-1} \sum_{i=1}^{n} \mathbb{V}\mathrm{ar}[A_i | X^n = x^n]} \Big| X^n = x^n \right] \tag{213}$$

$$\leq \delta_{1,n}. \tag{214}$$

Here, (213) follows from (212), and (214) follows from Chebyshev's inequality. Note that (211) and (214) imply that ξ_n defined in (212) indeed satisfies (208).

To characterize the asymptotic behavior of ξ_n, we make use of the following estimates of the conditional mean and the conditional variance of A_i given $X^n = x^n$:

$$\sum_{i=1}^{n} \mathbb{E}[A_i | X^n = x^n] = nP\mathbb{E}\left[\frac{|H|^4}{1 + |H|^2\sqrt{P}}\right] \tag{215}$$

$$= nP\mathbb{E}\big[|H|^4\big] + o(nP) \tag{216}$$

and

$$\sum_{i=1}^{n} \mathbb{V}\mathrm{ar}[A_i | X^n = x^n] \leq 3 \sum_{i=1}^{n} \left(|x_i|^4 \mathbb{V}\mathrm{ar}\left[\frac{|H_i|^4}{1 + |H_i|^2\sqrt{P}}\right]\right.$$
$$\left. + \mathbb{V}\mathrm{ar}\left[\frac{2|H_i|^2\mathrm{Re}(H_i x_i Z_i^*)}{1 + |H_i|^2\sqrt{P}}\right] + \mathbb{V}\mathrm{ar}\left[\frac{|H_i|^2(Z_i^2 - 1)}{1 + |H_i|^2\sqrt{P}}\right]\right) \tag{217}$$

$$\leq 3\mathbb{E}\big[|H|^8\big]\,\|x^n\|_4^4 + \mathcal{O}(n) \tag{218}$$

$$\leq 3\mathbb{E}\big[|H|^8\big]\,n\eta_n + \mathcal{O}(n) \tag{219}$$

for every $x^n \in \mathcal{C}_1$. Here, (217) follows because

$$\mathbb{V}\mathrm{ar}\left[\sum_{i=1}^{K} B_i\right] \leq K \sum_{i=1}^{K} \mathbb{V}\mathrm{ar}[B_i] \tag{220}$$

for all random variables $B_1, \ldots, B_K$, and in (219) we have used (196). Substituting (216) and (219) into (212), we obtain

$$\xi_n \geq nP\mathbb{E}\big[|H|^4\big] - \sqrt{3\mathbb{E}[|H|^8]\left(\delta_{1,n}^{-1}n\eta_n\right)} + o\left(\sqrt{\delta_{1,n}^{-1}n\eta_n}\right). \tag{221}$$

To conclude this part of the proof, we next characterize the asymptotic behavior of $\delta_{1,n}^{-1}n\eta_n$ as follows:

$$\delta_{1,n}^{-1}n\eta_n = n\frac{\max\{nP^{5/2}, (nP)^{3/4}\}}{\max\{P^{1/4}, (nP)^{-1/8}\}} \tag{222}$$

$$\leq n\max\left\{\frac{nP^{5/2}}{P^{1/4}}, \frac{(nP^2)^{3/4}}{(nP^2)^{-1/8}}\right\} \tag{223}$$

$$= \max\{n^2 P^{9/4}, n(nP^2)^{7/8}\} = o(n^2 P^2). \tag{224}$$

Here, (223) follows from the inequality

$$\frac{\max\{a, b\}}{\max\{c, d\}} \leq \max\left\{\frac{a}{c}, \frac{b}{d}\right\}, \quad \forall a, b, c, d > 0. \tag{225}$$

Substituting (224) into (221), we obtain

$$\xi_n \geq nP\mathbb{E}\big[|H|^4\big] + o(nP). \tag{226}$$

Since $\xi_n \leq nP\mathbb{E}[|H|^4] + o(nP)$ as can be inferred from (212) and (216), we conclude that

$$\xi_n = nP\mathbb{E}\big[|H|^4\big] + o(nP). \tag{227}$$

Next, we evaluate $Q_{Y^n}P_{H^n}[T=1]$. The idea is to use the Gärtner-Ellis theorem [47, Th. 2.3.6], which characterizes the probability of large deviations of a random variable from its mean. Let

$$D_i \triangleq \frac{|H_i|^2(|Z_i|^2 - 1)}{1 + |H_i|^2\sqrt{P}} \tag{228}$$

where $Z_i \sim \mathcal{CN}(0,1)$, $i = 1, \ldots, n$, are independent of $\{H_i\}$. Note that

$$Q_{Y^n}P_{H^n}[T=1] = \mathbb{P}\left[\sum_{i=1}^n D_i \geq \xi_n\right]. \tag{229}$$

Let $\Lambda_n(\cdot)$ denote the logarithmic moment-generating function [47, Eq. (2.2.1)] of the random variable $\xi_n^{-1}\sum_{i=1}^n D_i$. We shall prove the following result: for every $c \in \mathbb{R}$:

$$\lim_{n\to\infty} \frac{n}{\xi_n^2}\Lambda_n\left(\frac{c\xi_n^2}{n}\right) = \frac{c^2\mathbb{E}[|H|^4]}{2} \triangleq \Lambda(c). \tag{230}$$

Let us assume that (230) holds; then we have

$$\limsup_{n\to\infty} \frac{n}{\xi_n^2}\log Q_{Y^n}P_{H^n}[T=1] \tag{231}$$

$$= \limsup_{n\to\infty} \frac{n}{\xi_n^2}\log\mathbb{P}\left[\xi_n^{-1}\sum_{i=1}^n D_i \geq 1\right] \tag{232}$$

$$\leq -\inf_{t\geq 1}\frac{t^2\log e}{2\mathbb{E}[|H|^4]} \tag{233}$$

$$= -\frac{\log e}{2\mathbb{E}[|H|^4]}. \tag{234}$$

Here, (233) follows from the Gärtner-Ellis theorem,[11] and because the Fenchel-Legendre transformation [47, Def. 2.2.2] of $\Lambda(c)$ defined in (230) is

$$\Lambda^*(t) = \sup_{\alpha\in\mathbb{R}}\left\{t\alpha - \Lambda(\alpha)\right\} = \frac{t^2}{2\mathbb{E}[|H|^4]}. \tag{235}$$

Using (234) in (209), we obtain

$$\log\beta_{1-\delta_{1,n}}(P^{(1)}_{Y^nH^n}, Q_{Y^n}P_{H^n}) \leq -\frac{\log e}{2\mathbb{E}[|H|^4]}\frac{\xi_n^2}{n}(1 + o(1)). \tag{236}$$

The desired bound (205) follows by substituting (227) into (236).

---

[11]Note that we used the Gärtner-Ellis theorem with $1/n$ replaced by the vanishing sequence n/ξ_n^2 (see [47, Remark (a), p. 44]).

It remains to prove (230). We have

$$\Lambda_n\left(\frac{c\xi_n^2}{n}\right) = \log_e \mathbb{E}\left[\exp\left(\log e \frac{c\xi_n^2}{n\xi_n}\sum_{i=1}^n D_i\right)\right] \tag{237}$$

$$= n\log_e \mathbb{E}_H\left[\mathbb{E}_Z\left[\exp\left(\frac{c\xi_n}{n}\frac{|H|^2\log e}{1+|H|^2\sqrt{P}}(|Z|^2-1)\right)\Big|H\right]\right]. \tag{238}$$

Observe now that the following bound holds for every $c > 0$, every $h \in \mathbb{C}$, and for all sufficiently large n:

$$\frac{c\xi_n}{n}\frac{|h|^2\log e}{1+|h|^2\sqrt{P}} \le \frac{c\xi_n\log e}{n\sqrt{P}} < 1. \tag{239}$$

Here, the second inequality follows from (227). Since $|Z|^2 \sim \text{Exp}(1)$, we have that for all sufficiently large n,

$$\mathbb{E}_H\left[\mathbb{E}_Z\left[\exp\left(\frac{c\xi_n}{n}\frac{|H|^2\log e}{1+|H|^2\sqrt{P}}(|Z|^2-1)\right)\Big|H\right]\right]$$

$$= \mathbb{E}_H\left[\left(1-\frac{c\xi_n}{n}\frac{|H|^2\log e}{1+|H|^2\sqrt{P}}\right)^{-1}\underbrace{\exp\left(-\frac{c\xi_n}{n}\frac{|H|^2\log e}{1+|H|^2\sqrt{P}}\right)}_{\le 1}\right] \tag{240}$$

$$\le \left(1-\frac{c\xi_n\log e}{n\sqrt{P}}\right)^{-1} < \infty. \tag{241}$$

Using (241), the dominated convergence theorem (see, e.g., [51, Th. 1.34]), and the Taylor series expansion

$$e^x = 1 + x + \frac{x^2}{2} + \mathcal{O}(x^3), \quad x \to 0 \tag{242}$$

we conclude that

$$\mathbb{E}\left[\exp\left(\log e\frac{c\xi_n}{n}\frac{|H|^2(|Z|^2-1)}{1+|H|^2\sqrt{P}}\right)\right]$$

$$= 1 + \frac{c\xi_n}{n}\underbrace{\mathbb{E}\left[\frac{|H|^2(|Z|^2-1)}{1+|H|^2\sqrt{P}}\right]}_{=0} + \frac{c^2}{2}\frac{\xi_n^2}{n^2}\underbrace{\mathbb{E}\left[\left(\frac{|H|^2(|Z|^2-1)}{1+|H|^2\sqrt{P}}\right)^2\right]}_{=\mathbb{E}[|H|^4]+\mathcal{O}(\sqrt{P})} + \mathcal{O}\big((\xi_n/n)^3\big) \tag{243}$$

$$= 1 + \frac{c^2}{2}\frac{\xi_n^2}{n^2}\Big(\mathbb{E}\big[|H|^4\big] + \mathcal{O}(\sqrt{P})\Big) + \mathcal{O}\big((\xi_n/n)^3\big). \tag{244}$$

Substituting (244) into (238), and then using that $\log_e(1+x) = x + o(x)$, $x \to 0$, we conclude that

$$\frac{n}{\xi_n^2}\Lambda_n\big(c\xi_n^2/n\big) = \frac{n^2}{\xi_n^2}\left(\frac{c^2\mathbb{E}[|H|^4]}{2}\frac{\xi_n^2}{n^2} + o\left(\frac{\xi_n^2}{n^2}\right)\right) = \frac{c^2\mathbb{E}[|H|^4]}{2} + o(1) \tag{245}$$

which implies (230).

2) Proof of (206)*:* Consider the test

$$T(y^n, h^n) \triangleq \mathbb{1}\left\{ \|y^n\|_4^4 \geq \tilde{\xi}_n \right\} \tag{246}$$

where $\tilde{\xi}_n$ is defined in (203). Note that

$$\frac{\tilde{\xi}_n}{n} \geq (nP^2)^{1/8} \to \infty, \quad n \to \infty. \tag{247}$$

We evaluate the probability that $T = 1$ under $P^{(2)}_{Y^n H^n}$ and under $Q_{Y^n} P_{H^n}$ by following closely the proof of [52, Th. 9]. We start by noting that

$$P^{(2)}_{Y^n H^n}[T = 1] = \mathbb{P}[\|H^n X^n + Z^n\|_4 \geq \tilde{\xi}_n^{1/4}] \tag{248}$$

$$\geq \mathbb{P}[\|H^n X^n\|_4 - \|Z^n\|_4 \geq \tilde{\xi}_n^{1/4}] \tag{249}$$

$$\geq \mathbb{P}[\|H^n X^n\|_4 \geq 2\tilde{\xi}_n^{1/4}] \cdot \mathbb{P}[\|Z^n\|_4 \leq \tilde{\xi}_n^{1/4}] \tag{250}$$

$$\geq \mathbb{P}[|H|\|X^n\|_\infty \geq 2\tilde{\xi}_n^{1/4}] \cdot \mathbb{P}[\|Z^n\|_4 \leq \tilde{\xi}_n^{1/4}]. \tag{251}$$

Here, (249) follows by the triangle inequality, and (251) follows because $\|\cdot\|_4 \geq \|\cdot\|_\infty$. The first term in the product on the RHS of (251) can be bounded as

$$\mathbb{P}[|H|\|X^n\|_\infty \geq 2\tilde{\xi}_n^{1/4}] \geq \mathbb{P}\left[|H|\sqrt{\eta_n/P} \geq 2\tilde{\xi}_n^{1/4}\right] \tag{252}$$

$$= \exp\left(-4\tilde{\xi}_n^{1/2}P/\eta_n\right) \tag{253}$$

$$\geq 1 - o(1). \tag{254}$$

Here, (252) follows because, by Hölder's inequality and by (197), $\|x^n\|_\infty \geq \|x^n\|_4^2/\|x^n\|_2 \geq \sqrt{\eta_n/P}$ for every $x^n \in \mathcal{C}_2$; (254) follows because

$$\frac{\tilde{\xi}_n^{1/2}P}{\eta_n} = \frac{\max\{nP^{23/8}, (nP^2)^{5/8}\}}{\max\{nP^{5/2}, (nP^2)^{3/4}\}} \tag{255}$$

$$\leq \max\left\{ \frac{nP^{23/8}}{nP^{5/2}}, \frac{(nP^2)^{5/8}}{(nP^2)^{3/4}} \right\} \tag{256}$$

$$= \max\{P^{3/8}, (nP^2)^{-1/8}\} = o(1). \tag{257}$$

The second term in the product on the RHS of (251) can be bounded using Markov's inequality as follows:

$$\mathbb{P}[\|Z^n\|_4 \leq \tilde{\xi}_n^{1/4}] = 1 - \mathbb{P}[\|Z^n\|_4^4 \geq \tilde{\xi}_n] \tag{258}$$

$$\geq 1 - \frac{n\mathbb{E}[|Z_1|^4]}{\tilde{\xi}_n} \tag{259}$$

$$= 1 - o(1). \tag{260}$$

Here, the last step follows from (247). Substituting (253) and (259) into (251), we obtain

$$P^{(2)}_{Y^n H^n}[T = 1] \geq \exp\left(-4\tilde{\xi}_n^{1/2} P/\eta_n\right)\left(1 - \frac{n\mathbb{E}[|Z_1|^4]}{\tilde{\xi}_n}\right) \tag{261}$$

$$= 1 - \delta_{2,n}. \tag{262}$$

Next, we evaluate $Q_{Y^n} P_{H^n}[T = 1]$. Since $Y^n \sim \mathcal{CN}(\mathbf{0}, \mathsf{I}_n)$ under Q_{Y^n}, by following the same steps as the ones reported in [52, Eqs. (117)–(122)] and by using the Gaussian isoperimetric inequality [53, Eq. (3.4.8)], we obtain

$$\log Q_{Y^n}[\|Y^n\|_4^4 \geq \tilde{\xi}_n] \leq -\left(\tilde{\xi}_n^{1/4} - 3^{1/4} n^{1/4}\right)^2 \log e - \log 2 \tag{263}$$

$$= -\tilde{\xi}_n^{1/2} \log e + o(\tilde{\xi}_n^{1/2}) \tag{264}$$

$$\leq -\frac{\mathbb{E}[|H|^4]}{2} nP^2 \log e + o(nP^2). \tag{265}$$

Here, in (264) we used (247), and (265) holds for all sufficiently large n, since

$$\tilde{\xi}_n^{1/2}/(nP^2) \geq P^{-1/8} \to \infty, \quad n \to \infty. \tag{266}$$

Combining (262) and (265), we conclude (206).

C. Proof of Lemma 6

Let T_1 and T_2 be the Neyman-Pearson tests that achieve $\beta_{1-\delta_1}(P_1, Q)$ and $\beta_{1-\delta_2}(P_2, Q)$, respectively. Consider the following test for distinguishing between P and Q:

$$T \triangleq \mathbb{1}\{T_1 = 1 \cup T_2 = 1\}. \tag{267}$$

Under P, we have

$$P[T = 1] = \lambda P_1[T = 1] + (1 - \lambda)P_2[T = 1] \tag{268}$$

$$\geq \lambda P_1[T_1 = 1] + (1 - \lambda)P_2[T_2 = 1] \tag{269}$$

$$= 1 - \lambda\delta_1 - (1 - \lambda)\delta_2. \tag{270}$$

Under Q, we have

$$Q[T = 1] \leq Q[T_1 = 1] + Q[T_2 = 1] \tag{271}$$

$$= \beta_{1-\delta_1}(P_1, Q) + \beta_{1-\delta_2}(P_2, Q) \tag{272}$$

where in the first step we used the union bound. Note that (270) and (272) imply (200).

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423 and 623–656, Jul./Oct. 1948.

[2] A. Feinstein, "A new basic theorem of information theory," *IRE Trans. Inform. Theory*, vol. 4, no. 4, pp. 2–22, 1954.

[3] C. E. Shannon, "Probability of error for optimal codes in a Gaussian channel," *Bell Syst. Tech. J.*, vol. 38, no. 3, pp. 611–656, May 1959.

[4] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.

[5] 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.

[6] M. Pinsker, *Information and information stability of random variables and processes*. San Francisco: Holden-Day, 1964.

[7] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[8] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[9] Y. Polyanskiy and Y. Wu, *Lecture Notes on Information Theory*. MIT (6.441), UIUC (ECE 563), Yale (STAT 664), 2012–2017.

[10] F. Topsøe, "An information theoretical identity and a problem involving capacity," *Studia Scientiarum Math. Hung.*, vol. 2, pp. 291-292, 1967.

[11] 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.

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

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

[14] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.

[15] R. E. Blahut, "Computation of channel capacity and rate-distortion function," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, Jul. 1972.

[16] R. G. Gallager, "Source coding with side information and universal coding," 1979, unpublished manuscript. [Online]. Available: http://web.mit.edu/gallager/www/papers/paper5.pdf

[17] S. Shamai (Shitz) and S. Verdú, "The empirical distribution of good codes," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 836–846, May 1997.

[18] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with non-vanishing error probability," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 5–21, Jan. 2014.

[19] Y. Yang and A. Barron, "Information-theoretic determination of minimax rates of convergence," *Ann. Stat.*, vol. 27, no. 5, pp. 1564–1599, 1999.

[20] D. Haussler and M. Opper, "Mutual information, metric entropy and cumulative relative entropy risk," *Ann. Stat.*, vol. 25, no. 6, pp. 2451–2492, 1997.

[21] 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.

[22] H. Nagaoka, "Strong converse theorems in quantum information theory," in *Proc. ERATO Conference on Quantum Information Science (EQIS) 2001*, p. 33, Tokyo, Japan, Sep. 2001.

[23] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. New Jersey: Wiley, 2006.

[24] S. Verdú, "The exponential distribution in information theory," *Probl. Inf. Transm.*, vol. 32, no. 1, pp. 86–95, 1996.

[25] 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.

[26] C. E. Shannon, "Certain results in the coding theory for noisy channels," *Inf. Contr.*, vol. 1, pp. 6–25, 1957.

[27] L. Wang, R. Colbeck, and R. Renner, "Simple channel coding bounds," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Seoul, Korea, Jul. 2009.

[28] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," in *Proc. 48th Allerton Conf. Commun., Contr., Comp.*, Monticello, IL, USA, Sep. 2010.

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

[30] E. MolavianJazi and J. N. Laneman, "A second-order achievable rate region for Gaussian multi-access channels via a central limit theorem for functions," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6719–6733, 2015.

[31] J. Hoydis, R. Couillet, and P. Piantanida, "The second-order coding rate of the MIMO quasi-static Rayleigh fading channel," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, pp. 6591–6622, Dec 2015.

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

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

[34] 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.

[35] W. Yang, G. Durisi, and Y. Polyanskiy, "Minimum energy to send k bits over multiple-antenna fading channels," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 6831–6853, Dec. 2016.

[36] A. Collins and Y. Polyanskiy, "Dispersion of the coherent MIMO block-fading channel," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Barcelona, Spain, Jul. 2016.

[37] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 4–18, Jan. 1996.

[38] A. Martinez, "Communication by energy modulation: The additive exponential noise channel," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, pp. 3333–3351, Jun. 2011.

[39] T. J. Riedl, T. P. Coleman, and A. C. Singer, "Finite block-length achievable rates for queuing timing channels," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Paraty, Oct. 2011, pp. 200–204.

[40] W. Feller, *An Introduction to Probability Theory and Its Applications*. New York, NY, USA: John Wiley & Sons, 1970, vol. 1.

[41] İ. E. Telatar, "Capacity of multi-antenna Gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, pp. 585–595, Nov. 1999.

[42] 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.

[43] W. Yang, "Fading channels: Capacity and channel coding rate in the finite-blocklength regime," Ph.D. dissertation, Chalmers University of Technology, Gothenburg, Sweden, Aug. 2015.

[44] Y. Polyanskiy, "Finite blocklength methods in channel coding," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Istanbul, Turkey, Jul. 2013, tutorial. [Online]. Available: http://people.lids.mit.edu/yp/homepage/data/ISIT13_tutorial.pdf

[45] S. Verdú, *Multiuser Detection*. Cambridge, UK: Cambridge University Press, 1998.

[46] A. Rényi, "On measures of entropy and information," *Proc. 4th Berkeley Symp. Math. Statist. and Prob.*, vol. 1, pp. 547–561, 1961.

[47] A. Dembo and O. Zeitouni, *Large deviations techniques and applications.* New York: Springer Verlag, 1998.

[48] N. Iri and O. Kosut, "Third-order coding rate for universal compression of Markov sources," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, China, Jun. 2015, pp. 1996–2000.

[49] E. A. Haroutunian, "Bounds for the exponent of the probability of error for a semicontinuous memoryless channel," *Probl. Peredachi Inf.*, vol. 4, no. 4, pp. 3748, 1968.

[50] J. G. Wendel, "Note on the Gamma function," *Amer. Math. Monthly*, vol. 55, no. 9, pp. 563–564, Nov. 1948.

[51] W. Rudin, *Real and Complex Analysis*, 3rd ed. New York, NY, U.S.A.: McGraw-Hill, 1987.

[52] Y. Polyanskiy, "ℓ_p-norms of codewords from capacity- and dispersion-achieveing Gaussian codes," in *Proc. 50th Allerton Conf. Commun., Contr., Comp.*, Monticello, IL, USA, Oct. 2012.

[53] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications and coding," in *Foundations and Trends in Communications and Information Theory.* now Publishers, 2013, vol. 10, no. 1–2, pp. 1–246.