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

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

| Citation       | Yang, Wei et al. “Beta–Beta Bounds: Finite-Blocklength Analog of the Golden Formula.” IEEE Transactions on Information Theory 64, 9 (September 2018): 6236 - 6256 © 1963-2012 IEEE. |
|----------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| As Published   | http://dx.doi.org/10.1109/tit.2018.2837104                                                                                                                                                       |
| Publisher      | Institute of Electrical and Electronics Engineers (IEEE)                                                                                                                                          |
| Version        | Author’s final manuscript                                                                                                                                                                       |
| Citable link   | https://hdl.handle.net/1721.1/124314                                                                                                                                                            |
| Terms of Use   | Creative Commons Attribution-Noncommercial-Share Alike                                                                                                                                            |
| Detailed Terms | http://creativecommons.org/licenses/by-nc-sa/4.0/                                                                                                                                                |
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: 1) a finite-blocklength channel-coding converse bound by Polyanskiy and Verdú (2014), which involves the ratio of two Neyman-Pearson $\beta$ functions (beta-beta converse bound); and 2) a novel beta-beta channel-coding achievability bound, expressed again as the ratio of two Neyman-Pearson $\beta$ 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 derivation of the latter, with the beta-beta bounds used in place of the golden formula.

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.

Index Terms—Channel coding, achievability bound, hypothesis testing, golden formula, finite-blocklength regime.

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

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

$$i(X;Y) \triangleq \log \frac{dP_{XY}}{dP_X P_Y}(X,Y)$$  \hspace{1cm} (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.

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)]:

$$I(X;Y) = D(P_{XY}||P_X Q_Y) - D(P_Y||Q_Y).$$  \hspace{1cm} (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

$$C = \max_{P_X} \left\{ D(P_{XY}||P_X Q_Y) - D(P_Y||Q_Y) \right\}$$  \hspace{1cm} (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

---

1In this paper, log and exp functions are taken with respect to the same arbitrary basis.
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.\footnote{With a slight abuse in terminology, we shall refer to both (3) and (4) as the golden formula.} 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:

\[
i(X;Y) = \log \frac{dP_{XY}}{d(P_Q Y)}(X;Y) - \log \frac{dP_Y}{dQ_Y}(X;Y).
\] (5)

However, the resulting bounds are not very useful, because it is difficult to decouple the two random variables on the right-hand side (RHS) of (5). Instead of tweaking the information-spectrum bounds (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 \(P_{Y|X}: \mathcal{A} \rightarrow \mathcal{B}\). An \((M, \epsilon)\) code for the channel \((\mathcal{A}, P_{Y|X}, \mathcal{B})\) comprises a message set \(\mathcal{M} \triangleq \{1,\ldots, M\}\), an encoder \(f: \mathcal{M} \rightarrow \mathcal{A}\), and a decoder \(g: \mathcal{B} \rightarrow \mathcal{M} \cup \{\epsilon\}\) (\(\epsilon\) denotes the error event) that satisfies the average error probability constraint

\[
\frac{1}{M} \sum_{j=1}^{M} \left(1 - P_{Y|X}(g^{-1}(j) | f(j))\right) \leq \epsilon.
\] (6)

Here, \(g^{-1}(j) \triangleq \{y \in Y: g(y) = j\}\). 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 \(P_{Y^n|X^n}: \mathcal{X}^n \rightarrow \mathcal{Y}^n\). We shall refer to an \((M, \epsilon)\) code for the channel \(\{\mathcal{X}^n, P_{Y^n|X^n}, \mathcal{Y}^n\}\) as an \((n, M, \epsilon)\) 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 \(P_Z|X: \mathcal{X} \rightarrow \{0,1\}\), 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

\[
\beta_\alpha(P, Q) \triangleq \min \int P_Z|X(1|x)Q(dx)
\] (7)

where the minimum is over all tests satisfying

\[
\int P_Z|X(1|x)P(dx) \geq \alpha.
\] (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 \((n, M, \epsilon)\) code satisfies the following converse bound [18, Th. 15]:

\[
M \leq \inf_{0 \leq \delta < 1 - \epsilon} \frac{\beta_{1-\delta}(P_{Y^n}, Q_{Y^n})}{\beta_{1-\epsilon-\delta}(P_{X^n Y^n}, P_{X^n Q_{Y^n}})}, \quad \forall Q_{Y^n}.
\] (9)

Here, \(P_{Y^n}\) and \(P_{X^n}\) 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], in which the decoder is used as a suboptimal test for discriminating \(P_{X^n Y^n}\) 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 \((n, M, \epsilon)\) code such that

\[
\frac{M}{2} \geq \sup_{P_{Y^n|X^n}, \beta_{1-\epsilon}(P_{Y^n}, Q_{Y^n})} \beta_{1-\epsilon}(P_{X^n Y^n}, P_{X^n Q_{Y^n}}), \quad \forall Q_{Y^n}.
\] (10)

where \(P_{X^n}\) denotes the distribution of \(Y^n\) induced by \(P_{X^n}\) through \(P_{Y^n|X^n}\). The proof of the \(\beta\beta\) achievable 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 \(\beta_{1-\epsilon}(P_{X^n Y^n}, P_{X^n Q_{Y^n}})\), 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 \(Q^n\),

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

as \(n \rightarrow \infty\) 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 wide-band slope approximation of Verdú [13]. More specifically, we derive a second-order characterization of the minimum energy per bit \(E_0^k(k, \epsilon, R)\) required to transmit \(k\) bits with error probability \(\epsilon\) 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 \(E_0^k(k, \epsilon, R)\) (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 derivation
of the latter in [13], except that the role of the golden formula is taken by the $ββ$ bounds (9) and (10). Numerical evaluations demonstrate the accuracy of the resulting approximation.

The $ββ$ 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 $ββ$ bound (10) requires the evaluation of $dP_{Y^n|X^n}/dQ_{Y^n}$, which factorizes when $Q_{Y^n}$ is chosen to be a product distribution. This allows for an analytical computation of (10). Furthermore, the term $β_τ(P_{Y^n}, Q_{Y^n})$, 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 $ββ$ 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 $CN(0, A)$. We denote by $χ^2_k(λ)$ the noncentral chi-squared distribution with $k$ degrees of freedom and noncentrality parameter $λ$; furthermore, $Exp(μ)$ stands for the exponential distribution with mean $μ$. The Frobenius norm of a matrix $A$ is denoted by $\|A\|_F$. For a vector $x = (x_1, \ldots, x_d)$, we let $\|x\|_2$, $\|x\|_4$, and $\|x\|_∞$ denote the $ℓ_2$, $ℓ_4$, and $ℓ_∞$ norms of $x$, respectively. The real part and the complex conjugate of a complex number $x$ are denoted by $\text{Re}(x)$ and $x^*$, respectively. For a set $A$, we use $|A|$ and $A^c$ to denote the set cardinality and the set complement, respectively. Finally, $[\cdot]$ denotes the ceiling function, and the superscript $H$ stands for Hermitian transposition.

## II. The $ββ$ Achievability Bound

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

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

$$\frac{M}{2} ≥ \sup_{0 < τ < ϵ} \frac{β_τ(P_Y, Q_Y)}{β_1-ϵ}(P_{XY}, P_{X|Y})$$

(12)

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

**Proof:** Fix $ϵ ∈ (0, 1)$, $τ ∈ (0, ϵ)$, and let $P_X$ and $Q_Y$ be two arbitrary probability measures on $A$ and $B$, respectively. Furthermore, let

$$M = \left[ \frac{2β_τ(P_Y, Q_Y)}{β_1-ϵ}(P_{XY}, P_{X|Y}) \right].$$

(13)

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

$$P_{XY}[Z(X, Y) = 1] ≥ 1 - ϵ + τ$$

(14)

$$P_{XQ_Y}[Z(X, Y) = 1] = β_1-ϵ(P_{XY}, P_{X|Y})$$

(15)

In words, $(X, Y)$ may be distributed either according to $P_{XY}$ or according to $P_{X|Y}$, 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

\[
P_e(c_1, \ldots, c_M) = P \left[ \{Z(c_W, Y) = 0\} \cup \left( \bigcup_{m<W} \{Z(c_m, Y) = 1\} \right) \right]
\]  

(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

\[
E[P_e(C_1, \ldots, C_M)] 
\leq P[Z(X, Y) = 0] + P \left[ \max_{m<W} Z(C_m, Y) = 1 \right]
\]  

(17)

\[
\leq \epsilon - \tau + P \left[ \max_{m<W} Z(C_m, Y) = 1 \right].
\]  

(18)

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

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

\[
P \left[ \max_{m<W} Z(C_m, Y) = 1 \right] \leq \tau
\]  

(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 \( Y \sim Q_Y \) via binary hypothesis testing. Consider the randomized test \( \tilde{Z} : \mathcal{Y} \rightarrow \{0, 1\} \)

\[
\tilde{Z}(y) \doteq \max_{m<W} Z(C_m, y)
\]  

(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 \( \tilde{Z}(Y) = 1 \) indicates that the test chooses \( P_Y \). It follows that

\[
\beta_{P_Y|Z=1}(P_Y, Q_Y) 
\leq Q_Y(\tilde{Z}(Y) = 1)
\]  

(21)

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

(22)

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

(23)

\[
= \frac{M-1}{2} \beta_{1-\epsilon+\epsilon}(P_{XY}, P_X Q_Y)
\]  

(24)

\[
\leq \beta_{\epsilon}(P_Y, Q_Y).
\]  

(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

\[
P_Y[\tilde{Z} = 1] \leq \tau
\]  

(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 \( \kappa \) term in the \( \kappa\beta \) bound using [25, Lemma 4], we can relax it to the following bound:

\[
M \geq \sup_{0<\tau<\epsilon} \sup_{Q_Y} \beta_{\tau}(P_X \circ P_Y|X, Q_Y) \sup_{x \in \mathcal{F}} (\tau \circ Q_Y)(P_X|X = x, Q_Y)
\]  

(27)

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

\[
\max_{j \in \{1, \ldots, M\}} \left\{ 1 - P_Y|X(g^{-1}(j) | f(j)) \right\} \leq \epsilon.
\]  

(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 \( \beta_{\epsilon} \) bound as the average-error-probability counterpart of the \( \kappa\beta \) bound.3 For the case in which \( \beta_{\epsilon}(P_Y|X=x, Q_Y) \) 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 \( \beta_{\epsilon}(P_Y|X=x, Q_Y) \) does depend on \( x \in \mathcal{F} \), the \( \beta_{\epsilon} \) 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).

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

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

(29)

Further setting \( \alpha = P_{XY} \log \frac{dP_{XY}}{dP_X P_Y} \geq \log(M/2) \) 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 \( \kappa\beta \) 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

\[
\epsilon \leq Q_{XY} \left[ \log \frac{dQ_{XY}}{dQ_{X} Q_Y}(X, Y) \leq \log \frac{M}{2} \right] + Q_X[\mathcal{F}]
\]  

\[
+ \frac{M}{2} Q_X Q_Y \left[ \log \frac{dQ_{XY}}{dQ_{X} Q_Y}(X, Y) \geq \log \frac{M}{2} \right].
\]  

(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 \( P_X[\cdot] = 3\)In fact, the analogy between the \( \kappa\beta \) bound and the \( \beta_{\epsilon} \) bound is also clear without applying the relaxation (27). Indeed, the \( \kappa \) term in the \( \kappa\beta \) bound [5, Th. 25] defines a relative measure of performance for composite hypothesis testing between the collection \( \{P_Y|X=x\}_{x \in \mathcal{F}} \) and \( Q_Y. \) The \( \beta_{\epsilon}(P_Y, Q_Y) \) term in the \( \beta_{\epsilon} \) bound measures the performance of a binary hypothesis test between \( P_Y \) and \( Q_Y, \) where the distribution \( P_Y \) is an average of the collection of distributions \( \{P_Y|X=x\}_{x \in \mathcal{F}}. \)
\( Q_X[\cdot \cap F]/Q_X[F] \) and by using the following bounds:

\[
\begin{align*}
\beta_\tau(P_X, Q_Y) & \geq \beta_\tau(P_X, Q_X) = \tau Q_X[F] \\
\beta_{1-\tau}(P_{XY}, P_X Q_Y) & \leq \frac{1}{Q_X[F]} \beta_{1-(\tau-\epsilon)}(Q_{XY}, Q_X Q_Y).
\end{align*}
\]

Here, (31) follows from the data-processing inequality for \( \beta_\tau(\cdot, \cdot) \) (see, e.g., [28, Sec. VI]) 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]:

\[
\epsilon \leq \inf_{\gamma > 0} \left\{ P_{XY} \left[ \frac{dP_{XY}}{d(P_{XY} Q_Y)}(X, Y) \leq \gamma \right] + M \sup_y \frac{dP_y}{dQ_Y}(y) P_X Q_Y \left[ \frac{dP_{XY}}{d(P_{XY} Q_Y)}(X, Y) \geq \gamma \right] \right\}
\]

which is equivalent to

\[
M \geq \sup_{0 < \tau < 1} \beta_{1-\tau}(P_{XY}, P_X Q_Y) \left( \sup_y \frac{dP_y}{dQ_Y}(y) \right)^{-1}.
\]

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

\[
\beta_\tau(P_Y, Q_Y) \geq \tau \left( \sup_y \frac{dP_y}{dQ_Y}(y) \right)^{-1}.
\]

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

\[
\epsilon \leq \inf_{\gamma > 0, \eta > 0} \left\{ P_{XY} \left[ \frac{dP_{XY}}{d(P_{XY} Q_Y)}(X, Y) \leq \eta \gamma \right] + \frac{M}{\gamma} \right. \\
\left. + P_X Q_Y \left[ \frac{dP_y}{dQ_Y}(Y) \geq \eta \right] \right\}.
\]

This bound can be obtained from (12) by using [5, Eq. (103)] to lower-bound \( \beta_\tau(P_X, Q_Y) \) 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 \( E_{b_{\min}} \) (where \( N_0 \)) denotes the noise power per channel use) and the slope \( S_0 \) of the function spectral efficiency versus energy per bit (in dB) at \( E_{b_{\min}} \) (known as 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 \( R \) (bits/ch. use) in the limit \( R \to 0 \). Specifically, [13, Eq. (28)]

\[
10 \log_{10} \frac{E_b^*(R)}{N_0} = 10 \log_{10} \frac{E_b}{N_0 S_0} + R S_0 - 10 \log_{10} 2 + o(R),
\]

\( R \to 0 \). (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]

\[
E_b \frac{N_0}{N_0_{\min}} = \log_2 e - 1.59 \text{ dB}.
\]

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

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 \( \epsilon \) and with no constraint on the blocklength satisfies

\[
\frac{E_b^*(k, \epsilon)}{N_0} = \frac{E_b}{N_0 S_0} + \frac{1}{k} \left( \frac{2 \log_2 e}{k} \right)^{1/2} + o\left( \frac{1}{k} \right).
\]

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 \( \epsilon \), 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 \( \epsilon \). 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:

\[
E_b^*(R) = \lim_{\epsilon \to 0, k \to \infty} E_b^*(k, \epsilon, R)
\]

\( R \to 0 \). (40)

\[
E_b^*(k, \epsilon) = \lim_{R \to 0} E_b^*(k, \epsilon, R).
\]

B. AWGN Channel

We consider the complex-valued AWGN channel

\[
Y_i = X_i + Z_i, \quad i = 1, \ldots, n
\]

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

\[
||x^n||^2 = \sum_{i=1}^n x_i^2 \leq nP.
\]

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

\[
E_b^*(k, \epsilon, R) = a(R) + b(R) \epsilon + c(R).
\]

We shall use the terms spectral efficiency and rate interchangeably.
achieved with codes of rate of blocklength \( k \) per bit \( E_k \). The asymptotic limit of \( R \) as \( k \rightarrow \infty \) is—up to first order—proportional to \( \sqrt{b} \) and is independent of \( \epsilon \).

**Remark 2:** The result (46) implies that the minimum energy per bit \( E_k^* \) 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 \( P \rightarrow 0 \):

\[
C(P) = C'(0)P \log e + \frac{C''(0)}{2} P^2 \log e + o(P^2).
\]

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 \( E_k^* \) and the second term in (50) yields the wideband slope \( S_0 \) \cite[Eq. (35) and Th. 9]{13}:

\[
\frac{E_b}{N_0} = \log e \frac{2}{C'(0)}, \quad S_0 = \frac{2[C'(0)]^2}{-C''(0)}.
\]

Both \( E_b/N_0 \) 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 \cite[Eq. (41)]{13}:

\[
C'(0) = \frac{1}{\log e} \lim_{P \rightarrow 0} \max_{P_X: \mathbb{E}[|X|^2] = P} \frac{D(P_Y|X||Q_Y||P_X)}{P},
\]

and that for both AWGN channels and for fading channels with perfect CSIR, \( C''(0) = -\frac{2}{\log e} \lim_{P \rightarrow 0} \frac{\mathbb{E}[|X|^2]}{P^2} D(P_Y||Q_Y), \) (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 \( \epsilon \), 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):

\[
\frac{R^*(n, \epsilon, P)}{\log e} = P - \frac{2P}{n} Q^{-1}(\epsilon) - \frac{1}{2} P^2 + o\left(\frac{P}{n}\right) + o(P^2).
\]

As \( n \rightarrow \infty \), \( P \rightarrow 0 \), and \( nP^2 \rightarrow \infty \). 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:

\[
R^*(n, \epsilon, P) \approx \max_{P_X^n} \left\{ -\frac{1}{n} \log \beta_{1-\epsilon}(P_X^n, Y^n, P_X^n, Q_Y^n) + \log \beta_{0}(P_Y^n, Q_Y^n) \right\}.
\]

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

\[
\max_{P_X^n} \left\{ -\frac{1}{n} \log \beta_{1-\epsilon}(P_X^n, Y^n, P_X^n, Q_Y^n) \right\} \approx P \log e - \sqrt{\frac{2P}{n}} Q^{-1}(\epsilon) \log e
\]

which yields the first two terms in (54).
Furthermore, one can show through a large-deviation analysis that
\[ \max_{P_X} \frac{1}{n} \log \beta_n(P_Y^n, Q_Y^n) \approx -D(P_Y^n \| Q_Y) \approx - \frac{\log e}{2} P^2. \]
\[ (57) \]
Here, the maximization in (57) is taken with respect to all input distributions \( P_X \) for which \( \frac{1}{n} \log \beta_n(P_X, P_Y^n, Q_Y^n) \) 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]
\[ R^*(n, \epsilon, P) = C(P) - \sqrt{\frac{V(P)}{n}} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{n}}\right) \]
\[ (58) \]
and then by Taylor-expanding \( C(P) \) and \( V(P) \) for \( P \to 0 \). However, there are two major drawbacks to 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
\[ Y_i = H_i X_i + Z_i, \quad i = 1, \ldots, n \]
\[ (59) \]
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)]
\[ S_0 = \frac{2E[|H|^2]}{E[|H|^4]} = 1 \]
\[ (60) \]
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 \( kR \to \infty \):
\[ \frac{E^*_b(k, \epsilon, R)}{N_0} = \log_2 e + \frac{2 \log_2 e}{k} Q^{-1}(\epsilon) + (\log_2 e^2)R \]
\[ + o\left(\frac{1}{\sqrt{k}}\right) + o(R) \]
\[ (61) \]
or, equivalently,
\[ 10 \log_{10} \frac{E^*_b(k, \epsilon, R)}{N_0} = 10 \log_{10} \frac{E_b}{N_0 \min} + \frac{R}{S_0} 10 \log_{10} 2 \]
\[ \text{wideband slope approximation} \]
\[ + \frac{10 \log_{10} e}{\sqrt{\log_2 2}} \frac{\sqrt{2} Q^{-1}(\epsilon)}{k} + o\left(\frac{1}{\sqrt{k}}\right) + o(R) \]
\[ (62) \]
where \( 10 \log_{10} \frac{E_b}{N_0 \min} = -1.59 \text{ dB} \), and \( S_0 = 1 \). If \( k \to \infty \), \( R \to 0 \), but \( kR \to \epsilon < \infty \), then we have
\[ \frac{E^*_b(k, \epsilon, R)}{N_0} = \log_2 e + \frac{2 \log_2 e}{k} Q^{-1}(\epsilon) + o\left(\frac{1}{\sqrt{k}}\right) \]
\[ (63) \]
or, equivalently,
\[ 10 \log_{10} \frac{E^*_b(k, \epsilon, R)}{N_0} = 10 \log_{10} \frac{E_b}{N_0 \min} + \frac{10 \log_{10} e}{\sqrt{\log_2 2}} \frac{\sqrt{2} Q^{-1}(\epsilon)}{k} + o\left(\frac{1}{\sqrt{k}}\right). \]
\[ (64) \]

Proof: See Appendix II.
A few remarks are in order:

- As in the AWGN case, the minimum energy per bit \( E^*_b(k, \epsilon, R) \) over the Rayleigh-fading channel (59) with perfect CSIR satisfies
\[ 10 \log_{10} \frac{E^*_b(k, \epsilon, R)}{E_b(R)} = \frac{10 \log_{10} e}{\sqrt{\log_2 2}} \frac{\sqrt{2} Q^{-1}(\epsilon)}{k} \]
\[ + o(R) + o\left(\frac{1}{\sqrt{k}}\right). \]
\[ (65) \]
Again we observe that, in the low spectral efficiency regime, the gap in dB between \( E^*_b(k, \epsilon, R) \) and the asymptotic limit \( E_b(R) \) 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 \( E^*_b(k, \epsilon, R) \) does not converge to \(-1.59 \text{ dB} \) when \( k \to \infty \), \( R \to 0 \), and \( \epsilon \) is fixed.

- For the case of nonvanishing rate (or, equivalently, nonvanishing SNR \( P \)), a normal approximation for the maximum rate \( R^*(n, \epsilon, P) \) achievable over the channel (59) when CSIR is available is reported in [36]. This approximation relies on the additional constraint that 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 \( E^*_b(k, \epsilon, R) \) 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 \( E^*_b(k, \epsilon, R) \) for fixed \( R \), one needs to expurgate the codewords whose \( \ell_4 \) 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].
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., $Q_{Y^n} = \mathcal{CN}(0, (1 + P)I_n)$). 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

$$Y_i = X_i + Z_i, \quad i = 1, \ldots, n$$

(66)

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

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

(67)

The additive exponential-noise channel (66) can be used to model a communication system where information is conveyed through the arrival times of packets, and where each packet goes through a single-server queue with exponential service time [37]. It also models a rapidly-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]

$$C(\sigma) = \log(1 + \sigma)$$  

(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 Exp(1 + $\sigma$) distribution conditioned on it being positive. Furthermore, the capacity-achieving output distribution is 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:

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

(69)

where

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

(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:

$$\mathcal{F} = \{x^n \in \mathbb{R}^n : x_i \geq 0, n\sigma - \log n \leq \sum_{i=1}^{n} x_i \leq n\sigma\}.  \quad (71)$$

By construction, $X^n \sim P_{X^n}$ satisfies the constraint (67). Finally, let $P_{Y^n} \triangleq P_{X^n} \circ Y^n | X^n$ and let $Q_{Y^n} \triangleq Q_{X^n} \circ Y^n | X^n$. Therefore, $Q_{Y^n}$ is a probability distribution on $\mathbb{R}^n$ and $Q_{Y^n}$ is a probability distribution on $\mathbb{R}^n$.

Fig. 2. Minimum energy per bit versus spectral efficiency of the AWGN channel; here, $k = 2000$ bits, and $\epsilon = 10^{-3}$. 

\[ 
\text{Spectral efficiency, bits/ch. use} \\
\begin{tabular}{lll}
$E_b/N_0$, dB & Achievability & Converse \\
\hline
-1.6 & - & - \\
-1.4 & - & - \\
-1.2 & - & - \\
-1.0 & - & - \\
-0.8 & - & - \\
-0.6 & - & - \\
-0.4 & - & - \\
-0.2 & - & - \\
0 & - & - \\
\end{tabular} 
\]
We apply the $\beta \beta$ bound in Theorem 1 with $\tau = 1/\sqrt{m}$ 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:

$$\log \beta \tau(P_{Y^n}, Q_{Y^n}) \geq \log Q_{X^n}[F] + \log \tau \geq \log Q \left( \frac{-\log n}{\sqrt{n \text{Var}_{P_X} [X]}} \right) - Q(0) - O\left( \frac{1}{\sqrt{n}} \right) + \log \tau = O(\log n).$$

Here, (72) follows from (31), and (73) follows from the Berry-Esseen 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

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

where $\gamma_n$ satisfies

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

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

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

where $Z_i$ are i.i.d. 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-Esseen theorem and (71) to evaluate the left-hand side (LHS) of (76), we conclude that

$$P_{X^n}Y^n \left[ \log \frac{dP_{X^n}Y^n}{dP_{X^n}Q_{Y^n}}(X^n, Y^n) \geq \gamma_n \right] \geq P \left[ \frac{\sigma \log e}{1+\sigma} \sum_{i=1}^{n} (1 - Z_i) \geq \gamma_n - n \log(1+\sigma) - \frac{\log e}{1+\sigma} \log n \right] \geq Q \left( \frac{\gamma_n + (\log e)(\log n)/(1+\sigma) - n \log(1+\sigma)}{\sqrt{n \text{Var} [\gamma]}}, \frac{\gamma_n}{\sqrt{n \text{Var} [\gamma]}} \right) - O\left( \frac{1}{\sqrt{n}} \right). \quad (79)$$

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

$$\gamma_n = n \log(1+\sigma) - \sqrt{n \text{Var} [\gamma]}Q^{-1}(\epsilon) + O(\log n). \quad (80)$$

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

$$\mathcal{F}_n \triangleq \left\{ x^n \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = n \sigma, \quad x_i \geq 0 \right\}. \quad (81)$$

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 $(n, M, \epsilon)$ code for the channel (66) satisfies

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

where the supremum is over all probability distributions supported on $\mathcal{F}_n$. We next note that the function

Fig. 3. Minimum energy per bit versus spectral efficiency of the Rayleigh-fading channel (59) with perfect CSIR; here, $k = 2000$ bits, $\epsilon = 10^{-3}$. 

- $P_{Y^n|X^n}$. We apply the $\beta \beta$ bound in Theorem 1 with $\tau = 1/\sqrt{m}$ 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:

$$\log \beta \tau(P_{Y^n}, Q_{Y^n}) \geq \log Q_{X^n}[F] + \log \tau \geq \log Q \left( \frac{-\log n}{\sqrt{n \text{Var}_{P_X} [X]}} \right) - Q(0) - O\left( \frac{1}{\sqrt{n}} \right) + \log \tau = O(\log n).$$

Here, (72) follows from (31), and (73) follows from the Berry-Esseen 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

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

where $\gamma_n$ satisfies

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

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

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

where $Z_i$ are i.i.d. 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-Esseen theorem and (71) to evaluate the left-hand side (LHS) of (76), we conclude that

$$P_{X^n}Y^n \left[ \log \frac{dP_{X^n}Y^n}{dP_{X^n}Q_{Y^n}}(X^n, Y^n) \geq \gamma_n \right] \geq P \left[ \frac{\sigma \log e}{1+\sigma} \sum_{i=1}^{n} (1 - Z_i) \geq \gamma_n - n \log(1+\sigma) - \frac{\log e}{1+\sigma} \log n \right] \geq Q \left( \frac{\gamma_n + (\log e)(\log n)/(1+\sigma) - n \log(1+\sigma)}{\sqrt{n \text{Var} [\gamma]}}, \frac{\gamma_n}{\sqrt{n \text{Var} [\gamma]}} \right) - O\left( \frac{1}{\sqrt{n}} \right). \quad (79)$$

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

$$\gamma_n = n \log(1+\sigma) - \sqrt{n \text{Var} [\gamma]}Q^{-1}(\epsilon) + O(\log n). \quad (80)$$

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

$$\mathcal{F}_n \triangleq \left\{ x^n \in \mathbb{R}^n : \sum_{i=1}^{n} x_i = n \sigma, \quad x_i \geq 0 \right\}. \quad (81)$$

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 $(n, M, \epsilon)$ code for the channel (66) satisfies

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

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, for every \(x^n \in \mathcal{F}_n\),

\[
\beta_{1-\epsilon}(P_{Y^n|X^n=x^n}, Q_{Y^n}) = e^{-n\log(1+\sigma)}\mathbb{P}\left[\sum_{i=1}^{n} Z_i \leq n - \xi_n \right]
\]

(83)

where \(\xi_n\) satisfies

\[
\mathbb{P}\left[\sum_{i=1}^{n} Z_i \leq n - \xi_n \right] = 1 - \epsilon.
\]

(84)

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

\[
\sup_{P_{X^n}} \beta_{1-\epsilon}(P_{X^n,Y^n}, P_{X^n}Q_{Y^n}) = \beta_{1-\epsilon}(P_{Y^n|X^n=x^n}, Q_{Y^n}).
\]

(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

\[
-\log \beta_{1-\epsilon}(P_{Y^n|X^n=x^n}, Q_{Y^n}) = n \log(1+\sigma) - n\sqrt{nV(\sigma)}Q^{-1}(\epsilon) + O(\log n).
\]

(86)

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

\[
\text{B. MIMO Block-Fading Channel With Perfect CSIR}
\]

In this section, we use the \(\beta\beta\) achievability bound (12) to characterize the maximum coding rate achievable over an \(m_t \times m_t\) 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 \(k\)th coherence interval is given by

\[
Y_k = X_k H_k + Z_k.
\]

(87)

Here, \(X_k \in \mathbb{C}^{m_t \times m_t}\) and \(Y_k \in \mathbb{C}^{n \times m_t}\) are the transmitted and received matrices, respectively; the entries of the fading matrix \(H_k \in \mathbb{C}^{m_t \times m_t}\) and of the noise matrix \(Z_k \in \mathbb{C}^{n \times m_t}\) are i.i.d. \(\mathcal{CN}(0,1)\). We assume that \(\{H_k\}\) and \(\{Z_k\}\) are independent, that they are distributed on independent sequences over successive coherence intervals, and that they do not depend on the matrices \(\{X_k\}\). The channel matrices \(\{H_k\}\) 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 \(X_\ell\) is constrained to satisfy

\[
\|X_\ell\|_{\mathbb{F}}^2 \leq nP.
\]

(88)

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

\[
C(P) = \mathbb{E}_H \left[ \log \det \left( I_{m_t} + \sqrt{P/m_t}HH^H \right) \right].
\]

(89)

If either \(m_t = m_r = 1\) or \(m_t \geq 2\), the capacity is achieved by a unique input distribution, under which the matrix \(X\) has i.i.d. \(\mathcal{CN}(0, P/m_t)\) entries [41]. In this case, we denote the capacity-achieving input distribution by \(P_X^*\). If \(m_t > 1\) and \(m_r = 1\) (i.e., a 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 \(P_{Y^n|X^n}^*\). More specifically, \(P_{Y^n|X^n}^* = P_{Y^n|X^n}^*H^H\), where \(H^H\) is the column vectors of \(Y\) are i.i.d. \(\mathcal{CN}(0, H^H H + P/\sqrt{m_t m_r})\) 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 [36] and [42]. In particular, it was shown in [36] that\(^7\) for every \(0 < \epsilon < 1/2\)

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

(90)

Here,

\[
V(P) = \inf_{\mathbb{E}} \left[ \mathbb{E} \left[ \left( \log \frac{dP_{Y^n|X}^*}{dP_{Y^n|X}^*} \right)^2 \right] \right].
\]

(91)

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

2) Table 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 in which the capacity-achieving input distribution is unique. To compute (12), we choose \(P_{X|Y}\) as the uniform distribution on \(S_n \triangleq \{X^n : \|X^n\|_{\mathbb{F}}^2 = nP\}\), and set \(Q_{Y^n|X^n}^* = (P_{Y^n|X^n}^*)H^H\). With these choices, we have

\[
R^*(n, \epsilon) \geq 1 \sup_{0 < \epsilon < 1} \mathbb{E}_H \beta_{1-\epsilon}(P_{Y^n|X^n}^*, P_{Y^n|X^n}^*)
\]

(92)

The denominator \(\beta_{1-\epsilon}(P_{Y^n|X^n}^*, P_{Y^n|X^n}^*)\) in (92) can be computed using the Neyman-Pearson lemma and standard Monte Carlo techniques. However, computing \(\beta_{1-\epsilon}(P_{Y^n|X^n}^*, P_{Y^n|X^n}^*)\) in the numerator is more involved, since there is no simple expression for \(P_{Y^n|X^n}^*H^H\). To circumvent this, we further lower-bound \(\beta_{1-\epsilon}(P_{Y^n|X^n}^*, P_{Y^n|X^n}^*)\) using the data-processing inequality for \(\beta_{1-\epsilon}(\cdot, \cdot)\) as follows.

Let \(X^{\ell}_k\) be a sequence of i.i.d. random matrices with \(X^{\ell}_k \sim P_X^*, k = 1, \ldots, \ell\). Then, \(P_{X^{\ell}_k}\) can be obtained via \(X^{\ell}_k\) through \(X^{\ell}_k = \sqrt{nP}X^{\ell}_k/\|X^{\ell}_k\|_{\mathbb{F}}\). Let \(P_{Y^{\ell}_k|X^{\ell}_k}^* = P_{Y^{\ell}_k|X^{\ell}_k}^*\), where \(P_{Y^{\ell}_k|X^{\ell}_k}^*\) denotes the channel law defined by

\[
Y_k = X_k H_k + Z_k, \quad k = 1, \ldots, \ell.
\]

(93)

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

\[
\beta_{1-\epsilon}(P_{Y^{\ell}_k|X^{\ell}_k}^*, Q_{Y^{\ell}_k|X^{\ell}_k}^*) \geq \beta_{1-\epsilon}(P_X^{\ell})^s \circ P_{Y^{\ell}_k|X^{\ell}_k}^*, (P_X^{\ell})^s \circ P_{Y^{\ell}_k|X^{\ell}_k}^*.
\]

(94)

Since the Radon-Nikodym derivative \(\frac{dP_{Y^{\ell}_k|X^{\ell}_k}^*}{dP_{X^{\ell}_k}}\) can be computed in closed form, the RHS of (94) can be computed using the Neyman-Pearson lemma and Monte Carlo techniques.

\(^7\)Since the channel matrix \(H\) is known at the receiver, the channel output consists of the pair \((Y, H)\).

\(^8\)It is also shown in [36] that the bound (90) is tight under the additional constraint \(\|X_\ell\|_{\mathbb{F}} = o(n^{1/4})\) on each codeword \(X_\ell\).
The resulting bound\(^8\) is shown in Fig. 4 for a 4×4 Rayleigh block-fading channel with \(n_c = 4\), \(\epsilon = 0.001\), and SNR = 0 dB. For comparison, we have also plotted the normal approximation (90) (with the \(o(1/\sqrt{m})\) term omitted), the cost-constrained DT bound (30), and the variation of Feinstein’s bound provided in (36). More specifically, (30) is computed

\[ F = \{X': ||X'||^2 \leq nP \} \quad (95) \]

and with \(X' \sim Q_{X'}\) having i.i.d. \(CN(0, P'/m_i)\) entries. Here, \(P'\) is chosen such that \(Q_{X'}[F'] = \epsilon/2\). To compute (36), we chose the same \(P_{X'}\) and \(Q_{Y|X'}\) that we used to compute (92). Not 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 \(F = S_n\)) is much more difficult to compute because one needs to maximize over all codewords \(X' \in S_n\). Furthermore, for blocklength values of practical interest, we expect that

\[ \max_{X' \in S_n} \beta_{1-\epsilon+\tau}(P_{Y|X'} | X' = \hat{X'}, Q_{Y|X'}) \geq \beta_{1-\epsilon+\tau}(P_{X'|Y|X'} | X' = \hat{X'}, Q_{Y|X'}) \quad (96) \]

which means that the resulting \(\kappa\beta\) bound may be much looser than (92). To validate this claim, we evaluate \(\beta_{1-\epsilon+\tau}(P_{Y|X'} | X' = \hat{X'}, Q_{Y|X'})\) for a specific codeword \(\hat{X'} \in S_n\) constructed as follows: the entries \(\hat{x}_{j,k}^{(i)}\) of the matrix \(\hat{X}_i\) are \(\hat{x}_{j,k}^{(i)} = \sqrt{n}P\), and \(\hat{x}_{j,k}^{(i)} = 0\), \(\forall(i, j, k) \neq (1, 1, 1)\). By construction, \(\beta_{1-\epsilon+\tau}(P_{Y|X'} | X' = \hat{X'})\) is a lower bound on the LHS of (96). It can be shown that the ratio between \(\beta_{1-\epsilon+\tau}(P_{Y|X'} | X' = \hat{X'}, Q_{Y|X'})\) and \(\beta_{1-\epsilon+\tau}(P_{X'|Y|X'} | X' = \hat{X'}, Q_{Y|X'})\) grows exponentially in \(\ell\). Numerically, we observe that for the parameters in Fig. 4 and the choices \(\tau = \epsilon/2 = 5 \times 10^{-4}\) and \(n = n_c l = 400\),

\[-\log_2 \beta_{1-\epsilon+\tau}(P_{Y|X'} | X' = \hat{X'}, Q_{Y|X'}) \approx -\log_2 \beta_{1-\epsilon+\tau}(P_{X'|Y|X'} | X' = \hat{X'}, Q_{Y|X'}) + 735 \text{ bits.} \quad (97)\]

VI. CONCLUDING REMARKS

We have developed a novel channel coding achievability bound (which we refer to as the \(\beta\beta\) 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] yields 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 derivation of the latter, 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 achievable bound known to date.

\(^8\)The numerical routines used to obtain these results are available at https://github.com/yp-mit/spectre.
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:

\[
k = nR \quad \text{and} \quad P = E_b R
\]

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

\[
R = R^* (k/R, \epsilon, E_b^* (k, \epsilon, R) R).
\]

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

\[
R \log_2 e = E_b^* - R - \sqrt{\frac{2E_b^* \cdot R^2}{k} Q^{-1}(\epsilon) - \frac{1}{2} (E_b^* \cdot R)^2} + o \left( \sqrt{\frac{E_b^* R^2}{k}} \right) + o ((E_b^*)^2 R).
\]

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

\[
E_b^* = \log_2 e - \sqrt{\frac{2E_b^* \cdot R^2}{k} Q^{-1}(\epsilon) - \frac{(E_b^*)^2 R}{2}} + o \left( \sqrt{\frac{E_b^* R^2}{k}} \right) + o ((E_b^*)^2 R).
\]

This implies that

\[
E_b^* = \log_2 e + o(1).
\]

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

\[
F_n \triangleq \{ x^n \in \mathbb{C}^n : \| x^n \|^2 = nP \}
\]

and with \( Q_{Y^n} = \mathcal{C}N(0, I_n) \). Furthermore, we shall set

\[
\tau = \frac{(nP^2)^{-1/2}}{.}
\]

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

\[
R^*(n, \epsilon, P) \geq \frac{1}{n} \log \frac{\beta_{\epsilon}(P_{X^n}, Q_{Y^n})}{\beta_{1-\epsilon+r}(P_{X^n} Q_{Y^n}, P_{X^n} Q_{Y^n})}.
\]

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

\[
\beta_{1-\epsilon+r}(P_{X^n} Q_{Y^n}, P_{X^n} Q_{Y^n}) = \beta_{1-\epsilon+r}(P_{X^n} | X^n = x^n, Q_{Y^n})
\]

which implies

\[
= \beta_{1-\epsilon+r}(P_{X^n} | X^n = x^n, Q_{Y^n})
\]

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

\[
\beta_{1-\epsilon+r}(P_{X^n} Q_{Y^n}, P_{X^n} Q_{Y^n}) = Q \left( \sqrt{2nP} + Q^{-1}(1 - \epsilon + \tau) \right).
\]

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

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

and obtain

\[
\log \beta_{1-\epsilon+r}(P_{X^n} Q_{Y^n}, P_{X^n} Q_{Y^n}) = -nP \log e + \sqrt{2nP} Q^{-1}(\epsilon) \log e + o \left( \sqrt{nP} \right).
\]

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

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

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

\[
\beta_{\alpha}(P, Q) = \max_R \beta_{\beta_{\alpha}(P, R)}(R, Q)
\]

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, \( R_{\lambda} \) is defined as

\[
\frac{d R_{\lambda}}{d \mu}(x) \triangleq e^{-\lambda D_{1-\lambda}(P||Q)} \left( \frac{d P}{d \mu}(x) \right)^{1-\lambda} \left( \frac{d Q}{d \mu}(x) \right)^{\lambda}.
\]

Here, \( D_{1-\lambda}(P||Q) \) denotes the Rényi divergence of order \( 1 - \lambda \) [46], and \( \mu \) is a measure on \( \mathcal{X} \) that satisfies \( P \ll \mu \) and \( Q \ll \mu \).
This yields
\[
R^*(n, e, P) \leq \sup_{P_{X^n}} \frac{1}{n} \log \frac{\beta_{1-\delta}(P_{X^n}, Q_{Y^n})}{\beta_{1-\delta}(P_{X^n} \cdot P_{Y^n} \mid Y^n)}
\] (122)
where the supremum is over all probability distributions \(P_{X^n}\) supported on \(\mathcal{F}_n\), and \(P_{X^n} = P_{X^n} \cdot P_{Y^n} \mid Y^n\).

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

Consider the following test between \(P_{Y^n}\) and \(Q_{Y^n}\):
\[
T(y^n) = \mathbb{1}\{\|y^n\|^2 \geq \gamma\}
\] (123)
where \(\gamma\) is chosen so that
\[
P_{Y^n}[\|Y^n\|^2 \geq \gamma] = 1 - \delta.
\] (124)
By the Neyman-Pearson lemma,
\[
\beta_{1-\delta}(P_{Y^n}, Q_{Y^n}) \leq Q_{Y^n}[\|Y^n\|^2 \geq \gamma].
\] (125)
Note that under \(P_{Y^n}\) the random variable \(\|Y^n\|^2\) 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 \(P_{Y^n}[\|Y^n\|^2 \geq \gamma]\) using the Berry-Essén theorem (see, e.g., [40, Ch. XVI.5]) as follows:
\[
P_{Y^n}[\|Y^n\|^2 \geq \gamma] - Q\left(\frac{\gamma - 2\sqrt{P} \sqrt{n}}{\sqrt{4n(1 + 2P)}}\right) \\
\leq \frac{64}{(2 + 4P)^{3/2} \sqrt{2n}}
\] (126)
\[
\leq 24 \sqrt{2n} \log n.
\] (127)
Here, the last step follows because
\[
\mathbb{E}\left[\left(\|Z_i + \sqrt{P}\|^2 - 1 - P\right)^3\right] \\
\leq \left(\mathbb{E}\left[\left(\|Z_i + \sqrt{P}\|^2 - 1 - P\right)^4\right]\right)^{3/4} \\
= \left(12(1 + 2P)^2 + 48(1 + 4P)\right)^{3/4} \\
\leq 64^{3/4}(1 + 2P)^{3/2}.
\] (128)
Using (124) and (127), we obtain
\[
\gamma \geq 2n(1 + P) - \sqrt{4n(1 + 2P)Q^{-1}(\delta - 24 \sqrt{2n}^{-1/2})} \triangleq \tilde{\gamma}.
\] (131)
Using (121) and the expansion
\[
Q^{-1}(t) = \sqrt{2 \log_2(1/t)}(1 + o(1)), \quad t \to 0
\] (132)
which follows from (109), we conclude that the threshold \(\tilde{\gamma}\) behaves as
\[
\tilde{\gamma} = 2n(1 + P) + \sqrt{8n(1 + 2P)(nP^2)^{1/4}(1 + o(1))}
\] (133)
\[
= 2n(1 + P) + o(nP),
\] (134)
in the limit \(n \to \infty\). Under \(Q_{Y^n}\), the random variable \(\|Y^n\|^2\) 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 \(Q_{Y^n}[\|Y^n\|^2 \geq \gamma]\) as follows:
\[
\limsup_{n \to \infty} \frac{2n}{\gamma - 2n} \log Q_{Y^n}[\|Y^n\|^2 \geq \gamma] \\
\leq \limsup_{n \to \infty} \frac{2n}{\gamma - 2n} \log \mathbb{P}\left[\sum_{i=1}^{2n} Z_i^2 \geq \gamma\right]
\] (135)
\[
= \limsup_{n \to \infty} \frac{2n}{\gamma - 2n} \log \mathbb{P}\left[\frac{1}{\gamma - 2n} \sum_{i=1}^{2n} Z_i^2 - 2n \geq 1\right]
\] (136)
\[
\leq - \frac{1}{4} \log e.
\] (137)
Here, (135) follows from (131), and (137) follows by using [47, Th. 3.7.1] with \(a_n = 2n(\widehat{\gamma}_n - 2n)^{-2}\) and \(\Gamma = [1, \infty]\). Combining (125), (134), and (137), we obtain that
\[
\log \beta_{1-\delta}(P_{Y^n}, Q_{Y^n}) \leq -\frac{\log e (\widehat{\gamma} - 2n)^2}{4} \frac{1}{2n} (1 + o(1))
\]
(138)
\[
= -\frac{1}{2} n P^2 \log e + o(n P^2).
\]
(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,
\[
E'_{n}(k, \epsilon, R) \geq E'_{n}(k, \epsilon).
\]
(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 block length \(k^2/c\). Without loss of generality, we can assume that \(c > 0\). Indeed, we can always transform a code that satisfies \(kR \to c > 0\) 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 \(\tau = (n P)^{-1/2}\), we conclude that there exists a sequence of \((n, M, \epsilon)\) codes that satisfy
\[
M \geq \frac{\beta_{\tau}(P_{Y^n}, Q_{Y^n})}{\beta_{1-\epsilon+\tau}(P_{X^n Y^n}, Q_{X^n Y^n})}.
\]
(141)
This time, we shall study the asymptotic behavior of (141) in the asymptotic limit \(n \to \infty, P \to 0\), and \(nP \to c/\log e\) 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
\[
\log \beta_{\tau}(P_{Y^n}, Q_{Y^n}) \geq -\frac{1}{2} n P^2 \log e - \sqrt{\frac{2n P^2 \log e}{\tau}}
\]
\[
+ \log \frac{\tau}{2} + O(n P^3)
\]
(142)
\[
= O((nP)^{1/4}).
\]
(143)
Here, \(\tau\) 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
\[
k \geq n P \log_2 e - \sqrt{2nPQ^{-1}(1)} \log_2 e + o\left(\sqrt{nP}\right).
\]
(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
\[
\beta_\alpha(P, Q) = Q[T = 1]
\]
(145)
\[
\geq R[T=1](R, Q)
\]
(146)
\[
\geq \beta_{\beta_\alpha(P, R)}(R, Q).
\]
(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
\[
\beta_\alpha(P, Q) \geq \sup_R \beta_{\beta_\alpha(P, R)}(R, Q).
\]
(148)
It remains to show that for the \(\{R_\lambda\}, \lambda \in (0, 1)\), defined in (112), we have
\[
\beta_\alpha(P, Q) = \beta_{\beta_\alpha(P, R_\lambda)}(R_\lambda, Q).
\]
(149)
Indeed, we observe that\(^9\)
\[
\log \frac{dP}{dR_\lambda}(x) = \lambda D_{1-\lambda}(P||Q) + \lambda \log \frac{dP}{dQ}(x)
\]
(150)
and
\[
\log \frac{dR_\lambda}{dQ}(x) = -\lambda D_{1-\lambda}(P||Q) + (1 - \lambda) \log \frac{dP}{dQ}(x)
\]
(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 \(R_\lambda\) and between \(R_\lambda\) and \(Q\). This in turn implies, by the Neyman-Pearson lemma, that both (146) and (147) hold with equality. This establishes (149).

**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
\[
\frac{R^*(n, \epsilon, P)}{\log e} = P - \frac{2P}{n} Q^{-1}(\epsilon) - P^2 + o\left(\sqrt{\frac{P}{n}}\right) + o(P^2)
\]
(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 \(F_n\) defined in (103), and \(Q_{Y^n} = CN(0, I_n)\). Furthermore, we set
\[
\tau = P + (nP^2)^{-1/2}.
\]
(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 \(n \to \infty\), we set \(dP/dQ = +\infty\) on the singular set.

\(^9\)In the case in which \(P\) is not absolutely continuous with respect to \(Q\), we set \(dP/dQ = +\infty\) on the singular set.
∞, \( P \to 0 \), and \( nD^2 \to \infty \), we can assume without loss of generality that \( \tau < \epsilon \). It follows from (12) that
\[
nR^*(n, \epsilon, P) \geq \log \beta_\tau(P_{Y^*H^n}, Q_{Y^n}P_{H^n}) - \log \beta_{1-\epsilon+\tau}(P_{X^nY^nH^n}, P_{X^nQ_{Y^n}P_{H^n}})
\]
where \( P_{Y^*H^n} = P_{X^n} \circ P_{Y^*H^n|X^n} \).

To evaluate the second term on the RHS of (154), we use [5, Eq. (103)] and obtain
\[
- \log \beta_{1-\epsilon+\tau}(P_{X^nY^nH^n}, P_{X^nQ_{Y^n}P_{H^n}}) \geq \gamma_0
\]
where \( \gamma_0 \) satisfies
\[
P_{X^nY^nH^n}\left[ \log \frac{dP_{X^nY^nH^n}}{dP_{X^nQ_{Y^n}P_{H^n}}} \right] \leq \gamma_0 = \epsilon - \tau.
\]
Observe now that, under \( P_{X^nY^nH^n} \), the random variable \( \log \frac{dP_{X^nY^nH^n}}{dP_{X^nQ_{Y^n}P_{H^n}}}(X^n, Y^n, H^n) \) has the same distribution as
\[
\log \left( \sum_{i=1}^{n} \left| H_iX_i \right|^2 + 2 \text{Re}(H_iX_iZ_i^\star) \right).
\]

Next, we use the central limit theorem for functions [48, Prop. 1] (see also [30, Prop. 1]) to derive an asymptotic expansion for \( \gamma_0 \) in (156). Specifically, let \( X^n \sim C N(0, P_{H^n}) \). It follows that \( X^n \sim P_{X^n} \) has the same distribution as \( \sqrt{n}P X^n / ||X^n|| \). Let
\[
A_{1,i} \triangleq |H_iX_i|^2 - P \quad A_{2,i} \triangleq |X_i|^2 - P \quad A_{3,i} \triangleq 2 \text{Re}(H_iX_iZ_i^\star), \quad i = 1, \ldots, n.
\]
The random vectors \( \{[A_{1,i}, A_{2,i}, A_{3,i}]\} \) are i.i.d. with zero mean and covariance matrix
\[
V = \begin{bmatrix}
3P^2 & P^2 & 0 \\
P^2 & P^2 & 0 \\
0 & 0 & 2P
\end{bmatrix}.
\]
Let \( g : \mathbb{R}^3 \to \mathbb{R} \) be defined as
\[
g(a_1, a_2, a_3) \triangleq \frac{(a_1 + P)P}{P + a_2} + \frac{a_3\sqrt{P}}{P + a_2}
\]
and observe that
\[
(n \log e) \cdot g \left( \frac{1}{n} \sum_{i=1}^{n} A_{1,i}, \frac{1}{n} \sum_{i=1}^{n} A_{2,i}, \frac{1}{n} \sum_{i=1}^{n} A_{3,i} \right)
\]
has the same distribution as (157). Finally, let \( j \) denote the gradient of \( g \) at \( (0, 0, 0) \). It follows that
\[
jV j^T = 2P^2 + 2P.
\]
We are now ready to invoke [48, Prop. 1] and conclude that for every \( \gamma \in \mathbb{R} \)
\[
P \left[ \sum_{i=1}^{n} \left( |H_iX_i|^2 + 2 \text{Re}(H_iX_iZ_i^\star) \right) \leq \gamma \right] \leq Q \left( \frac{nP - \gamma}{\sqrt{2nP(P + P^2)}} \right) + O(n^{-1/2}).
\]
Setting the RHS of (165) equal to \( \epsilon - \tau \) and solving for \( \gamma \), we obtain an asymptotic lower bound on \( \gamma_0 \), which we use to further lower-bound (155) as follows:
\[
- \log \beta_{1-\epsilon+\tau}(P_{X^nY^nH^n|X^n}, P_{X^nQ_{Y^n}P_{H^n}}) \geq \frac{(n-1)P}{2} \log e - \sqrt{2nP(1 + P)}Q^{-1}(\epsilon - \tau - O(n^{-1/2}))
\]
\[
= nP \log e - \sqrt{2nPQ^{-1}(\epsilon - \tau)} + o\left( \sqrt{nP} \right).
\]
Here, (167) follows by Taylor-expanding \( \sqrt{1 + P} \) for \( P \to 0 \) and by Taylor-expanding \( Q^{-1}(\cdot) \) around \( \epsilon \) for \( \tau \to 0 \) and \( n \to \infty \).

To evaluate \( \beta_\tau(P_{Y^*H^n}, Q_{Y^n}P_{H^n}) \) on the RHS of (154), we again use Lemma 5 in Appendix I-B. Let \( (P_X^n)^n \) be the product distribution obtained from the capacity-achieving output distribution of the channel (59) with SNR \( P \). Then, by Lemma 5,
\[
\beta_\tau(P_{Y^*H^n}, Q_{Y^n}P_{H^n}) \geq \beta_\tau(P_{Y^*H^n}, P_{Y^nH^n|X^n})(P_X^n)^n, Q_{Y^n}P_{H^n}).
\]
We lower-bound \( \beta_\tau(P_{Y^*H^n}, (P_X^n)^nP_{Y^nH^n|X^n}) \) by following steps similar to those reported in Section V-B. Specifically, let \( (P_X^n)^n \) be the transition probability of the following channel:
\[
Y_i = \frac{\sqrt{nP}H_iX_i}{\|X^n\|} + Z_i, \quad i = 1, \ldots, n.
\]
Let \( (P_X^n)^n \) be the product distribution obtained from the capacity-achieving input distribution for the channel (59) under perfect CSIR. Then, we have that \( P_{Y^*H^n} = (P_X^n)^nP_{Y^nH^n|X^n} \) and \( (P_X^n)^n = (P_X^n)^nP_{Y^*H^n|X^n} \). By the data-processing inequality,
\[
\beta_\tau((P_X^n)^nP_{Y^nH^n|X^n}, (P_X^n)^nP_{Y^nH^n|X^n}) \geq \beta_\tau(P_{Y^nH^n|X^n}, P_{Y^nH^n|X^n})
\]
To lower-bound the RHS of (170), we shall use the following bound [49] (see also [5, Eqs. (154)–(156)]):
\[
\log \beta_\alpha(P, Q) \geq -D(P||Q) + h_b(\alpha)
\]
where \( h_b(\alpha) \triangleq -\alpha \log \alpha - (1 - \alpha) \log(1 - \alpha) \) denotes the binary entropy function. This yields
\[
\beta_\tau((P_X^n)^nP_{Y^nH^n|X^n}, (P_X^n)^nP_{Y^nH^n|X^n}) \geq \exp \left( -D((P_X^n)^nP_{Y^nH^n|X^n})(P_X^n)^nP_{Y^nH^n|X^n} + h_b(\tau) \right)
\]

Note now that
\[
D \left( (P_X^n)^nP_{Y^nH^n|X^n} \| (P_X^n)^nP_{Y^nH^n|X^n} \right) = D \left( P_{Y^nH^n|X^n} \| P_{Y^nH^n|X^n} \| P_{H^n} \right)
\]
\[
= \mathbb{E}(P_X^n) \sum_{i=1}^{n} \left( \sqrt{nP}H_iX_i \right)^2 / ||X^n|| - H_iX_i^2 \right) \log e
\]
\[
\begin{align*}
\mathbb{E}_{(P_X^n)^n} \left[ (\sqrt{nP} - \|X^n\|)^2 \right] & \leq 2nP \left( 1 - \frac{\Gamma(n + 1/2)}{\sqrt{n}\Gamma(n)} \right) \log e \\
& \leq 2nP \left( 1 - \sqrt{1 - \frac{1}{2n+1}} \right) \log e \\
& \leq P \frac{2n}{2n+1} \log e \\
& \leq P \log e.
\end{align*}
\]  

Here, (176) follows because \(\mathbb{E}_{(P_X^n)^n}[\|X^n\|] = \sqrt{\mathbb{P}\Gamma(n + 1/2)/\Gamma(n)}\); (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

\[
\beta_\tau(P_{Y^nH^n}, (P_{Y^nH^n})^n) \geq \exp \left( - \frac{P \log e + h_b(\tau)}{\tau} \right) \geq e^{-2\tau} \frac{1}{n}, \quad (181)
\]

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

\[
\frac{h_b(\tau)}{\tau} = - \log \tau + 1 - \log \left( 1 - \frac{1}{\tau} \right) \leq - \log \tau + \log e. \quad (182)
\]

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

\[
\beta_\tau(P_{Y^nH^n}, Q_{Y^nH^n}) \geq \beta_\tau((P_{Y^nH^n})^n, Q_{Y^nH^n}). \quad (183)
\]

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

\[
\sum_{i=1}^n |Z_i|^2 |H_i|^2 P \log e - \log(1 + |H_i|^2 P) \quad (184)
\]

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

\[
I_n \triangleq \mathbb{E}[B_i], \quad V_n \triangleq \text{Var}[B_i]. \quad (185)
\]

A straightforward computation reveals that

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

and that \(V_n\) can be bounded as follows:

\[
3P^2 \log^2 e \leq V_n \leq 11P^2 \log^2 e. \quad (187)
\]

By [5, Lemma 59],

\[
\log \beta_\tau((P_{Y^nH^n})^n, Q_{Y^nH^n}) \geq -nP_\tau - \frac{2nV_n}{\tau} + \frac{1}{2} \log \frac{e}{2} \quad (188)
\]

where \(I_n \triangleq \mathbb{E}[|H|^4] = 2\), 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 \((n, M, \epsilon)\) 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 \(\epsilon\). 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 \((n, M, \epsilon)\) code \(C\) (maximum probability of error) satisfies

\[
\beta_\alpha(P_{Y^nH^n}, Q_{Y^nH^n}) \geq \frac{\delta M}{1 - \alpha + \delta \exp(\log(nP^2))} \inf_{x^n \in C} \beta_{\alpha-\epsilon-\delta}(P_{Y^nH^n} | X^n=x^n, Q_{Y^nH^n}) \quad (190)
\]

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

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

\[
\log M \leq \log(1 - \delta_n) + \log \beta_{1-\delta_n}(P_{Y^nH^n}, Q_{Y^nH^n}) - \log \beta_{1-\epsilon-2\delta_n}(P_{Y^nH^n} | X^n=x^n, Q_{Y^nH^n}) + \log 2. \quad (191)
\]

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

\[
\inf_{x^n \in C} \log \beta_{1-\epsilon-2\delta_n}(P_{Y^nH^n} | X^n=x^n, Q_{Y^nH^n}) \leq -nP_\epsilon - \frac{1}{2} \log \left( 1 - \epsilon - 2\delta_n \right) \quad (192)
\]

Here, \(x^n\) denotes the infinite-dimensional sequence \((x_1, x_2, \ldots)\), \(P_{Y^nH^n} | X^n=x^n = \prod_{i=1}^n P_{Y^nH^n}[x_i, x_i=n] \), and the supremum in (192) is taken over all \(x^n\) that satisfy \(|x|^2 = n^2 P\). 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 \((n, M, \epsilon)\) codes.

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

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

The idea is to construct a suboptimal test to distinguish between \(P_{Y^nH^n}\) and \(Q_{Y^nH^n}\). Coarsely speaking, our test is constructed as follows: if \(P_{Y^nH^n}\) is induced by a code whose codewords have suitably small \(\ell_1\) norm, then we use as the
test the optimal Neyman-Pearson test between the capacity-achieving output distribution \((P_Y^n)^n\) 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 \((P_Y^n)^n\). If instead \(P_Y^n P_H^n\) is induced by a code whose codewords are peaky in an \(\ell_4\) sense, we distinguish between \(P_Y^n P_H^n\) 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 \(C\) into two subcodebooks \(C_1\) and \(C_2\) according to the peakiness of the codewords. More specifically, we set

\[
C_1 \triangleq \{ x^n \in C : \| x^n \|_4^4 \leq n \eta_n \} \quad (196)
\]

\[
C_2 \triangleq C \setminus C_1 = \{ x^n \in C : \| x^n \|_4^4 > n \eta_n \} \quad (197)
\]

where the threshold \(\eta_n\) is defined as

\[
\eta_n \triangleq \max \left\{ n P^{5/2}, (n P^2)^{3/4} \right\}. \quad (198)
\]

Note that \(\eta_n \to \infty\) as \(n \to \infty\). Let \(\lambda \triangleq |C_1|/|C|\). Furthermore, let \(P_Y^n P_H^n\) and \(P_Y^n P_H^n\) denote the output probability distributions induced by the subcodes \(C_1\) and \(C_2\), respectively. It follows that

\[
P_Y^n P_H^n = \lambda P_Y^n P_H^n + (1 - \lambda) P_Y^n P_H^n. \quad (199)
\]

**Lemma 6** below allows us to upper-bound \(\beta_{1 - \delta_1}(P_Y^n P_H^n, Q_Y^n P_H^n)\) by \(\beta_{1 - \delta_1}(P_Y^n P_H^n, Q_Y^n P_H^n) + \beta_{1 - \delta_2}(P_Y^n P_H^n, Q_Y^n P_H^n)\) 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

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

**Proof:** See Appendix II-C.

Set now

\[
\delta_{1,n} = \max \left\{ P^{1/4}, (n P^2)^{-1/8} \right\} \quad (201)
\]

\[
\delta_{2,n} = 1 - \exp \left( -4 \xi_n^3/2 P/\eta_n \right) \left( 1 - \frac{2n}{\xi_n} \right) \quad (202)
\]

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

\[
\xi_n \triangleq \max \left\{ n^2 P^{15/4}, n^{9/8} P^{1/4} \right\}. \quad (203)
\]

Furthermore, set

\[
\delta_n = \lambda \delta_{1,n} + (1 - \lambda) \delta_{2,n}. \quad (204)
\]

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 \(n P^2 \to \infty\). We shall prove that

\[
\log \beta_{1 - \delta_{1,n}}(P_Y^n P_H^n, Q_Y^n P_H^n) \leq - \frac{E[|H|^4]}{2} n P^2 \log e + o(n P^2). \quad (205)
\]

10When the entries of \(X^n\) are independent Gaussian random variables, we expect to have \(\|X^n\|_4^4 \approx O(n P^2)\). This means that for a random Gaussian code \(|C_1| \gg |C_2|\).
To characterize the asymptotic behavior of $\xi_n$, we make use of the following estimates of the conditional mean and the conditional variance of $A_i$ given $X^n = x^n$:

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

and

$$
\sum_{i=1}^{n} \text{Var}[A_i | X^n = x^n] \leq 3 \sum_{i=1}^{n} \left( |x_i|^4 \text{Var} \left[ \frac{|H|^4}{1 + |H|^2 \sqrt{P}} \right] + \text{Var} \left[ \frac{2|H|^2 \text{Re}(H,x_i Z_i)}{1 + |H|^2 \sqrt{P}} \right] + \text{Var} \left[ \frac{|H|^2 (Z_i^2 - 1)}{1 + |H|^2 \sqrt{P}} \right] \right) \tag{218}
$$

for every $x^n \in C_1$. Here, (217) follows because

$$
\text{Var} \left[ \sum_{i=1}^{K} B_i \right] \leq K \sum_{i=1}^{K} \text{Var}[B_i] \tag{220}
$$

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

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

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

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

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

$$
= \max \left\{ n^{2/3} P^{3/4}, n (P/2)^{1/3} \right\} = o(n^2 P^2). \tag{224}
$$

Here, (223) follows from the inequality

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

Substituting (224) into (221), we obtain

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

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

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

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

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

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 $c \in \mathbb{R}$,

$$
\lim_{n \to \infty} n \xi_n^{-1} \Lambda_n \left( \frac{c}{n} \right) = \frac{c^2 \mathbb{E} \left[ |H|^4 \right]}{2} \approx \Lambda(c). \tag{230}
$$

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

$$
\lim_{n \to \infty} \frac{n}{\xi_n^2} \log Q_{Y^n} P_{H^n} [T = 1] = \lim_{n \to \infty} \frac{n}{\xi_n^2} \log \mathbb{P} \left[ \xi_n^{-1} \sum_{i=1}^{n} D_i \geq 1 \right] \tag{231}
$$

$$
= \lim_{n \to \infty} \frac{n}{\xi_n^2} \log \mathbb{P} \left[ \xi_n^{-1} \sum_{i=1}^{n} D_i \geq 1 \right] \geq - \inf_{t \geq 1} \frac{t^2 \log e}{2 \mathbb{E} \left[ |H|^4 \right]} \tag{232}
$$

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

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

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

Using (234) in (209), we obtain

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

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

It remains to prove (230). We have

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

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

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

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

\footnote{Note that we have used the Gärtner-Ellis theorem with $1/n$ replaced by the vanishing sequence $n/\xi_n^2$ (see [47, Remark (a), p. 44]).}
Here, the second inequality follows from (227). Since $|Z|^2 \sim \text{Exp}(1)$, we have that for all sufficiently large $n$,

$$E_H \left[ E_Z \left[ \exp \left( \frac{c_{\xi_n}}{n} \left| H \right|^2 \log e \right) \left( |Z|^2 - 1 \right) \right] \right]$$

$$= E \left[ \left( 1 - \frac{c_{\xi_n}}{n} \left| H \right|^2 \log e \right)^{-1} \left( 1 - \frac{c_{\xi_n}}{n} \left| H \right|^2 \log e \right) \right] \leq 1$$

$$\leq \left( 1 - \frac{c_{\xi_n}}{n} \log e \right)^{-1} < \infty.$$  

(240)

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

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

we conclude that

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

$$= 1 + \frac{c_{\xi_n}}{n} E \left[ \left| H \right|^2 \left( (|Z|^2 - 1) \right) \right]$$

$$+ \frac{c_{\xi_n}^2}{2} \frac{\xi_n^2}{n^2} E \left[ \left( \frac{\left| H \right|^2 \left( (|Z|^2 - 1) \right) \right)}{2} \right] + O((\xi_n/n)^3)$$

$$= 1 + \frac{c_{\xi_n}^2}{2} \frac{\xi_n^2}{n^2} \left( E \left[ \left| H \right|^4 \right] + O(\sqrt{n}) \right) + O((\xi_n/n)^3).$$

(243)

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

$$\frac{n}{\xi_n} A_n (\xi_n^2/n) = \frac{n^2}{\xi_n^2} \left( \frac{c_{\xi_n}^2}{2} \frac{\xi_n^2}{n^2} + o \left( \frac{c_{\xi_n}^2}{n^2} \right) \right)$$

$$= \frac{\xi_n^2}{2} E \left[ \left| H \right|^4 \right] + o(1)$$

(245)

which implies (230).

2) Proof of (206): Consider the test

$$T(y^n, h^n) \triangleq 1 \left\{ \| y^n \|_4 \geq \hat{\xi}_n \right\}$$

(246)

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

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

(247)

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

$$P_{Y^n H^n}^{(2)} [T = 1]$$

$$= \mathbb{P}[\| H^n X^n + Z^n \|_4 \geq \hat{\xi}_n^{1/4}]$$

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

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

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

(248)

(249)

(250)

(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

$$\mathbb{P}[\| H \|_4 \| X^n \|_\infty \geq 2^{1/4} \cdot \mathbb{P}\left[\| H\|_\infty \sqrt{\eta}/P \geq 2^{1/4} \right] \geq \hat{\xi}_n^{1/4} \right]$$

$$\geq \exp(-4\hat{\xi}_n^{1/2} P/\eta_{\xi_n})$$

$$\geq 1 - o(1).$$

(252)

(253)

(254)

Here, (252) follows because, by Hölder’s inequality and by (197), $\| x^n \|_\infty \geq \| x^n \|_2/\| x^n \|_2 \geq \sqrt{\eta_n/P}$ for every $x^n \in C_2$; (245) follows because

$$\frac{\xi_n^{1/2} P}{\eta_{\xi_n}} = \frac{\max \{nP^{23/8}, (nP^2)^{5/8} \}}{\max \{nP^{23/8}, (nP^2)^{5/8} \}}$$

$$\leq \max \{nP^{23/8}, (nP^2)^{5/8} \}$$

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

(255)

(256)

(257)

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

$$\mathbb{P}\left[ \| Z^n \|_4 \leq \xi_n^{1/4} \right] = 1 - \mathbb{P}[\| Z^n \|_4 \geq \hat{\xi}_n]$$

$$\geq 1 - \frac{n\mathbb{E}[\| Z_1 \|_4]}{\xi_n}$$

$$= 1 - o(1).$$

(258)

(259)

(260)

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

$$P_{Y^n H^n}^{(2)} [T = 1] \geq \exp(-4\hat{\xi}_n^{1/2} P/\eta_{\xi_n}) \left( 1 - \frac{n\mathbb{E}[\| Z_1 \|_4]}{\xi_n} \right)$$

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

(261)

(262)

Next, we evaluate $Q_{Y^n} P_{H^n} [T = 1]$. Since $Y^n \sim \mathcal{CN}(0, I_n)$ 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

$$\log Q_{Y^n} \left[ \| Y^n \|_4 \geq \hat{\xi}_n \right]$$

$$\leq - \left( \hat{\xi}_n^{1/4} - 3^{1/4} n^{1/4} \right)^2 \log e - \log 2$$

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

$$\leq - \frac{1}{2} \frac{\xi_n^2}{n^2} P \log e + o(nP^2).$$

(263)

(264)

(265)

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

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

(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 $Q$:

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

(267)
Under $P$, we have

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

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

$$= 1 - \lambda \delta_1 - (1 - \lambda) \delta_2 \quad (270)$$

Under $Q$, we have

$$Q[T = 1] \leq Q[T_1 = 1] + Q[T_2 = 1] \quad (271)$$

$$= \beta_1 - \delta_1 (P_1, Q) + \beta_1 - \delta_2 (P_2, Q) \quad (272)$$

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] A. Lapidoth and S. M. Moser, “Capacity bounds via duality with Rényi divergence,” in Proc. 48th Allerton Conf. Commun., Contr., Comp., Monticello, IL, USA, Sep. 2010.

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

[7] Y. Polyanskiy, “Finite block-length 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/PDF/ISIT13_tutorial.pdf

[8] E. A. Haroutunian, “Bounds for the exponent of the probability of error in finite-memoryless channels,” 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.