

# Soft Covering via Hypothesis Testing: Typical-Code Exponents and Mismatched Detection

Neri Merhav

The Viterbi Faculty of Electrical and Computer Engineering  
Technion – Israel Institute of Technology  
Technion City, Haifa 3200003, Israel  
E-mail: merhav@technion.ac.il

June 5, 2026

## Abstract

We study the typical-code (quenched) behavior of the false-alarm (FA) and missed-detection (MD) error exponents of the Neyman–Pearson test associated with soft covering, complementing the average-code (annealed) analysis that has been carried out in a companion paper [1]. We prove that, as the block-length tends to infinity, for almost every randomly selected fixed-composition codebook, the negative normalized logarithms of both error probabilities converge to their respective average-code exponents. In other words, the error exponents are *self-averaging*. We then extend the scope and study a *mismatched* likelihood ratio test that assumes the wrong channel model. Here, we derive the mismatched error exponents, show that self-averaging persists under mismatch, and characterize the degradation. In particular, we characterize the coding rate beyond which the two kinds of error exponents cannot be both positive at the same time, which in the matched case, is given by the channel input-output mutual information rate.

**Index Terms:** soft covering, hypothesis testing, self-averaging, error exponents, mismatched detection, almost-sure convergence.

## 1 Introduction

The soft covering lemma [2–6] asserts that a random codebook of rate  $R$  larger than the channel input-output per-letter mutual information,  $I(X; Y)$ , causes the output mixture distribution  $P_{Y^n|C}$  to be nearly indistinguishable from the i.i.d. distribution,  $P_Y^{\otimes n}$ , induced by the input distribution and the channel. Introduced by Wyner [2] as a tool for proving the common information theorem and systematically developed by Han and Verdú [3] under the name of channel resolvability, soft covering underpins wiretap secrecy, identification coding, common randomness generation, and channel synthesis [4].

A substantial body of work characterizes the *exact exponent* at which the soft covering approximation improves for  $R > I(X; Y)$ , under various distance measures. Hayashi [7] obtained a lower bound on the exponent under the Kullback-Leibler (KL) divergence. Parizi, Telatar and Merhav [8] derived the exact exponent under KL divergence with application to the wiretap

channel. Yu and Tan [9] characterized the exact exponent under Rényi divergence of order  $\alpha \in (0, 2)$ . Yagli and Cuff [10] derived the exact exponent under total variation distance. Li, Li and Yu [11] recently derived a strong-converse exponent under KL divergence. Cuff [5, 6] showed that soft covering holds with doubly exponential probability over the random codebook, enabling applications via the union bound.

In a companion paper [1], we studied soft covering through the lens of Neyman–Pearson hypothesis testing and derived the false-alarm (FA) and missed-detection (MD) error exponents,  $E_{\text{FA}}^a(\tau, R)$  and  $E_{\text{MD}}^a(\tau, R)$ , as functions of the coding rate  $R$  and the log-likelihood ratio (LLR) threshold  $\tau$ . The analysis in [1] was carried out for the *average code*, namely, it concerns the long-block limits of the normalized logarithms of the expected error probabilities of both kinds, where the expectation is with respect to (w.r.t.) the random selection of a fixed-composition code. In the jargon of statistical physics, these exponents can be referred to as *annealed* error exponents, in analogy to annealed free energies (and other physical quantities) associated with disordered physical systems. The analysis in [1] revealed a rich phase structure and characterized the soft-covering phenomenon in Neyman–Pearson terms.

A natural follow-up question that may arise concerns the characterization of *quenched* error exponents, namely, the long-block limits of the normalized expected logarithms (as opposed to logarithms of expectations) of the FA and MD probabilities, and a further question is whether they also reflect the almost-sure limits, in which case, they can be referred to as *typical-code* error exponents, and we say that the *self-averaging* property (which is actually an ergodic property) holds true. Obviously, by Jensen’s inequality, the quenched FA and MD error exponents cannot be smaller than their respective annealed counterparts, but can we characterize them beyond that?

We show in this work that the quenched FA and MD error exponents are equal to the respective annealed error exponents of [1] for all  $R$  and  $\tau$ , and they also admit self-averaging. This equivalence between quenched and annealed exponents is quite surprising in view of the fact that in other related problem areas in information theory, there are gaps between annealed and quenched quantities. Random coding error exponents of channel coding (see, e.g., [12, 13]), which are traditionally defined in the annealed sense, are smaller in general than their typical-code counterparts (see, e.g., [14], [15] and references therein) at a certain range of low coding rates. In statistical physics, the analogous distinction between quenched and annealed free energies of disordered systems with random parameters is fundamental [16], [17], and the gaps are visible, especially at low temperatures, and they convey different meanings. In Section 5, an attempt is made to provide some insight and intuition behind the equivalence between quenched and

annealed exponents in our problem, as opposed to the inequivalence in the other problem areas.

We then proceed to study the effect of mismatched detection in the context of soft covering. Suppose that the detector implements the likelihood ratio test (LRT) w.r.t. the wrong channel model, say  $\tilde{W}$ , instead of the true channel  $W$ . As in mismatched decoding (see, e.g., [18] and many references therein), this mismatch may arise from either a wrong assumption on the channel statistics, or due to limitations on the implementation of the exact LRT. We show that self-averaging persists under mismatch, characterize the degradation of the exponents, and identify a precise connection between the shift in the soft-covering critical rate and the generalized mutual information (GMI) of mismatched decoding.

The outline of this article is as follows. Section 2 recalls the setup and notation. Sections 3 and 4 prove the main self-averaging theorems. Section 5 discusses why the Jensen gap vanishes. Section 6 develops the mismatched detection theory, including the annealed mismatched exponents, self-averaging under mismatch, the phase structure, and a connection to the GMI. Finally, in Section 7 we summarize and conclude.

## 2 Setup, Notation Conventions and Objectives

We adopt the same notation conventions as in [1]. Let  $W : \mathcal{X} \rightarrow \mathcal{Y}$  be a discrete memoryless channel (DMC) with finite input and output alphabets,  $\mathcal{X}$  and  $\mathcal{Y}$ , and let  $P_X$  be a fixed input distribution. We denote by  $P_Y$  the output marginal induced by  $P_X$  and  $W$ , that is,

$$P_Y(y) = \sum_{x \in \mathcal{X}} P_X(x)W(y|x), \quad \forall y \in \mathcal{Y}. \quad (1)$$

For a given positive integer  $n$  and a sequence  $x^n = (x_1, \dots, x_n) \in \mathcal{X}^n$ , the channel output distribution is given by

$$W^n(y^n|x^n) = \prod_{i=1}^n W(y_i|x_i), \quad y^n \in \mathcal{Y}^n, \quad (2)$$

where  $\mathcal{X}^n$  and  $\mathcal{Y}^n$  are the  $n$ -th Cartesian powers of  $\mathcal{X}$  and  $\mathcal{Y}$ , respectively. We also denote

$$P_Y^{\otimes n}(y^n) = \prod_{i=1}^n P_Y(y_i). \quad (3)$$

The *type* (empirical distribution) of  $x^n$  is defined as  $\hat{P}_{x^n}(a) = \frac{1}{n} \#\{i : x_i = a\}$ ,  $a \in \mathcal{X}$ . The *type class*  $\mathcal{T}(Q_X) \subseteq \mathcal{X}^n$  is the set of all sequences of type  $\hat{P}_{x^n}(a) = Q_X$ . For a pair  $(x^n, y^n)$ , the *joint type* is  $Q_{XY}(a, b) = \frac{1}{n} \#\{i : x_i = a, y_i = b\}$ ,  $(a, b) \in \mathcal{X} \times \mathcal{Y}$ , with marginals  $Q_X$ ,  $Q_Y$  and conditional type  $Q_{Y|X}$ . The *conditional type class* is defined as  $\mathcal{T}(Q_{X|Y}|y^n) = \{x^n : (x^n, y^n) \in \mathcal{T}(Q_{XY})\}$ . Likewise,  $\mathcal{T}(Q_{Y|X}|x^n) \triangleq \{y^n : (x^n, y^n) \in \mathcal{T}(Q_{XY})\}$ . We always restrict to types with  $Q_X = P_X$ .

Information functionals derived from a given probability distribution will be subscripted by the notation of this distribution. When this is an empirical distribution  $Q_{XY}$ , the subscript will be abbreviated by  $Q$ , in order to avoid cumbersome notation. Thus,  $H_Q(Y) = -\sum_y Q_Y(y) \log Q_Y(y)$  is the marginal empirical entropy of an auxiliary random vector  $Y$  governed by  $Q_Y$ ,  $H_Q(Y|X) = -\sum_{x,y} Q_{XY}(x,y) \log Q_{Y|X}(y|x)$  is the conditional empirical entropy of  $Y$  given  $X$ , where  $(X,Y)$  are jointly governed by  $Q_{XY}$ , and  $I_Q(X;Y) = H_Q(Y) - H_Q(Y|X)$  is the mutual information under  $Q_{XY}$ .

The channel codebook  $\mathcal{C} = \{x^n(1), \dots, x^n(M)\}$ ,  $M = e^{nR}$ ,  $R$  being the coding rate in nats per channel use, has its codewords drawn independently and uniformly from the type class  $\mathcal{T}(P_X)$ . For a given type  $Q_{XY}$ ,  $y^n \in \mathcal{Y}^n$ , let  $N(Q_{XY}|y^n)$  denote the number of codewords  $\{x^n(m)\}$  in  $\mathcal{C}$  whose joint type with  $y^n$  is  $Q_{XY}$ . Since the codewords are randomly selected independently,  $N(Q_{XY}|y^n)$  is a binomial random variable with  $M = e^{nR}$  trials and success rate of the exponential order of  $e^{-nI_Q(X;Y)}$ . When  $I_Q(X;Y) \leq R$ , there are typically about  $e^{n[R-I_Q(X;Y)]} \rightarrow \infty$  codewords in  $\mathcal{T}(Q_{XY}|y^n)$ . We therefore refer to  $Q_{XY}$  as a *bulk type* whenever  $I_Q(X;Y) \leq R$ . On the other hand, when  $R \leq I_Q(X;Y)$  the probability that  $N(Q_{XY}|y^n) \geq 1$  tends to zero at the exponential rate of  $e^{n[I_Q(X;Y)-R]}$ , and so, there are typically no codewords at all in  $\mathcal{T}(Q_{XY}|y^n)$ , and then  $Q_{XY}$  is referred to as a *sparse type*.

The hypothesis testing problem pertaining to soft covering is in distinguishing between the following two hypotheses concerning the probability distribution that governs a vector of observations,  $y^n \in \mathcal{Y}^n$ :

$\mathcal{H}_0$ :  $y^n$  is governed by the i.i.d. distribution,  $P_Y^{\otimes n}$ .

$\mathcal{H}_1$ :  $y^n$  is governed by the mixture distribution  $\frac{1}{M} \sum_{m=1}^M W^n(y^n|x^n(m))$ .

Defining

$$S(y^n, \mathcal{C}) = \sum_{m=1}^M W^n(y^n|x^n(m)), \quad (4)$$

the LLR statistic is defined as

$$\Lambda(y^n, \mathcal{C}) = \frac{1}{n} \log \frac{\frac{1}{M} \sum_{m=1}^M W^n(y^n|x^n(m))}{P_Y^{\otimes n}(y^n)} = \frac{1}{n} \log \frac{S(y^n, \mathcal{C})}{M \cdot P_Y^{\otimes n}(y^n)}. \quad (5)$$

For a given threshold parameter  $\tau$ , if  $\Lambda(y^n, \mathcal{C}) \geq \tau$ , hypothesis  $\mathcal{H}_1$  is accepted, otherwise, hypothesis  $\mathcal{H}_0$  is accepted. The FA and MD probabilities (for a given  $\mathcal{C}$  known to the detector) are then defined as

$$\alpha_n(\tau, R, \mathcal{C}) \triangleq \sum_{\{y^n: \Lambda(y^n, \mathcal{C}) \geq \tau\}} P_Y^{\otimes n}(y^n) \quad (6)$$

$$\beta_n(\tau, R, \mathcal{C}) \triangleq \frac{1}{M} \sum_{m=1}^M \sum_{\{y^n: \Lambda(y^n, \mathcal{C}) < \tau\}} W^n(y^n|x^n(m)). \quad (7)$$

The corresponding annealed (average-code) exponents are defined as

$$E_{\text{FA}}^{\text{a}}(\tau, R) \triangleq - \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \{ \alpha_n(\tau, R, \mathcal{C}) \}, \quad (8)$$

$$E_{\text{MD}}^{\text{a}}(\tau, R) \triangleq - \lim_{n \rightarrow \infty} \frac{1}{n} \log \mathbb{E} \{ \beta_n(\tau, R, \mathcal{C}) \} \quad (9)$$

where the expectations are w.r.t. the random selection of  $\mathcal{C}$ , and where the existence of these limits was shown in [1]. For a joint type  $Q_{XY}$  with  $Q_X = P_X$ , define:

$$D_{\text{m}}(Q_Y) \triangleq D(Q_Y \| P_Y), \quad (10)$$

$$D_{\text{c}}(Q_{XY}) \triangleq D(Q_{Y|X} \| W | P_X) = \sum_{x,y} P_X(x) Q_{Y|X}(y|x) \log \frac{Q_{Y|X}(y|x)}{W(y|x)}, \quad (11)$$

and observe that  $D_{\text{m}}(Q_Y) \leq D_{\text{c}}(Q_{XY})$  by the data processing inequality of the KL divergence (see also [1]). We also define

$$\lambda(Q_{XY}, R) \triangleq D_{\text{m}}(Q_Y) - D_{\text{c}}(Q_{XY}) + [I_Q(X; Y) - R]_+, \quad (12)$$

and

$$\begin{aligned} \Delta(Q_Y, R) &\triangleq \max_{\substack{Q'_{XY}: Q'_X = P_X, Q'_Y = Q_Y \\ I_{Q'}(X; Y) \leq R}} [D_{\text{m}}(Q_Y) - D_{\text{c}}(Q'_{XY})] \\ &= \max_{\substack{Q'_{XY}: Q'_X = P_X, Q'_Y = Q_Y \\ I_{Q'}(X; Y) \leq R}} \lambda(Q'_{XY}, R). \end{aligned} \quad (13)$$

In [1] the following single-letter expressions for the annealed exponents  $E_{\text{FA}}^{\text{a}}(\tau, R)$  and  $E_{\text{MD}}^{\text{a}}(\tau, R)$  were derived:

$$E_{\text{FA}}^{\text{a}}(\tau, R) = \min_{\substack{Q_{XY}: Q_X = P_X \\ \lambda(Q_{XY}, R) \geq \tau}} \{ D_{\text{m}}(Q_Y) + [I_Q(X; Y) - R]_+ \}, \quad (14)$$

$$E_{\text{MD}}^{\text{a}}(\tau, R) = \min_{\substack{Q_{XY}: Q_X = P_X \\ \lambda(Q_{XY}, R) < \tau \\ \Delta(Q_Y, R) < \tau}} D_{\text{c}}(Q_{XY}). \quad (15)$$

Some more notation conventions that will be used throughout the sequel include the following. The logarithmic function  $\log(\cdot)$  will be understood to be defined to the base e unless specified otherwise. The notation  $[\cdot]_+$  designates the positive clipping operator, namely, given a real number  $t$ ,  $[t]_+$  is defined as  $\max\{t, 0\}$ . For two positive sequences,  $\{a_n\}_{n \geq 1}$  and  $\{b_n\}_{n \geq 1}$ , the notation  $a_n \doteq b_n$  means equivalence in the exponential scale, i.e.,  $\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$ .

Our first objective in this work is to derive the quenched (typical-code) FA and MD exponents, defined as

$$E_{\text{FA}}^{\text{q}}(\tau, R) \triangleq - \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \{ \log \alpha_n(\tau, R, \mathcal{C}) \}, \quad (16)$$

$$E_{\text{MD}}^{\text{q}}(\tau, R) \triangleq - \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \{ \log \beta_n(\tau, R, \mathcal{C}) \} \quad (17)$$

where it should be noted that here the exponents are defined in terms of expected logarithms, as opposed to eqs. (8) and (9), that are defined in terms of logarithms of expectations. We will prove that  $E_{\text{FA}}^{\text{a}}(\tau, R) = E_{\text{FA}}^{\text{a}}(\tau, R)$  and  $E_{\text{MD}}^{\text{a}}(\tau, R) = E_{\text{MD}}^{\text{a}}(\tau, R)$ , and moreover that these common values are also the almost-sure limits of  $\{-\frac{1}{n} \log \alpha_n(\tau, R, \mathcal{C})\}$  and  $\{-\frac{1}{n} \log \beta_n(\tau, R, \mathcal{C})\}$ , respectively — i.e., the exponents are self-averaging.

Our second objective will be to expand the scope to mismatched hypothesis testing, where the detector assumes the wrong channel  $\tilde{W}$  instead of the true one,  $W$ . It turns out that the expressions of the error exponents extend relatively straightforwardly to the mismatched case. Moreover, both the quenched-annealed equivalence and the self-averaging properties are preserved. The more interesting and less trivial part is the effect of mismatch on the critical rate  $R$  beyond which there is no value of the threshold  $\tau$  where both error exponents are strictly positive, which in the matched case is  $I(X; Y)$  — the soft-covering limit. As mentioned before, it will turn out that there is a non-trivial relationship between the mismatched critical rate and the GMI of mismatched decoding.

### 3 The FA Exponent

Our first theorem in this work asserts that  $\{-\frac{1}{n} \log \alpha_n(\tau, R, \mathcal{C})\}_{n \geq 1}$  is self-averaging and its almost-sure limit is  $E_{\text{FA}}^{\text{a}}(\tau, R)$ . It will then follow that  $\{-\frac{1}{n} \mathbb{E}\{\log \alpha_n(\tau, R, \mathcal{C})\}\}_{n \geq 1}$  converges to the same limit, which then must coincide with  $E_{\text{FA}}^{\text{a}}(\tau, R)$ .

**Theorem 1.** *For all  $\tau \in \mathbb{R}$  and  $R > 0$ :*

$$\frac{1}{n} \log \alpha_n(\tau, R, \mathcal{C}) \xrightarrow{n \rightarrow \infty} -E_{\text{FA}}^{\text{a}}(\tau, R) \quad \text{a.s. over } \mathcal{C}. \quad (18)$$

*Proof.* Let  $Q_{XY}^*$  be the minimizer in (14), so  $\lambda(Q_{XY}^*, R) \geq \tau$  and  $E_{\text{FA}}^{\text{a}}(\tau, R) = D_{\text{m}}(Q_Y^*) + [I_{Q^*}(X; Y) - R]_+$ . Fix  $\epsilon > 0$ . The proof has two parts: the lower bound, asserting that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \log \alpha_n(\tau, R, \mathcal{C}) \geq -E_{\text{FA}}^{\text{a}}(\tau, R) \quad \text{a.s.},$$

and the upper bound, namely,

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \log \alpha_n(\tau, R, \mathcal{C}) \leq -E_{\text{FA}}^{\text{a}}(\tau, R) \quad \text{a.s.}$$

Beginning from the lower bound, we distinguish between two cases, according to whether  $Q_{XY}^*$  is sparse or bulk. Suppose first that  $Q_{XY}^*$  is sparse ( $I_{Q^*}(X; Y) > R$ ). Then, in view of eq. (14), we have

$$E_{\text{FA}}^{\text{a}}(\tau, R) = D_{\text{m}}(Q_Y^*) + I_{Q^*}(X; Y) - R. \quad (19)$$

Define

$$\ell(Q_{XY}) \triangleq H_Q(Y|X) + D_c(Q_{XY}), \quad (20)$$

so that  $W^n(y^n|x^n) = e^{-n\ell(Q_{XY})}$  for  $(x^n, y^n) \in \mathcal{T}(Q_{XY})$ . Every codeword  $x^n(m)$  together with every  $y^n \in \mathcal{A}_m \triangleq \mathcal{T}(Q_{Y|X}^*|x^n(m))$  contributes  $W^n(y^n|x^n(m)) = e^{-n\ell(Q_{XY}^*)}$  to  $S(y^n, \mathcal{C})$ . Since  $P_Y^{\otimes n}(y^n) = e^{-n[H_{Q^*}(Y)+D_m(Q_Y^*)]}$  and  $M = e^{nR}$ , we get

$$\begin{aligned} \Lambda(y^n, \mathcal{C}) &\geq \frac{1}{n} \log \frac{e^{-n\ell(Q_{XY}^*)}}{e^{nR} \cdot e^{-n(H_{Q^*}(Y)+D_m(Q_Y^*))}} \\ &= D_m(Q_Y^*) - D_c(Q_{XY}^*) + I_{Q^*}(X; Y) - R \\ &= \lambda(Q_{XY}^*, R) \geq \tau. \end{aligned} \quad (21)$$

Here we used the fact that for sparse  $Q_{XY}^*$ ,  $I_{Q^*}(X; Y) - R = [I_{Q^*}(X; Y) - R]_+$ , and so,  $D_m(Q_Y^*) - D_c(Q_{XY}^*) + I_{Q^*}(X; Y) - R = \lambda(Q_{XY}^*, R)$ . It follows that every  $y^n \in \mathcal{A}_m$  satisfies the FA condition, for every  $m$  and every codebook  $\mathcal{C}$ . The probability of  $\mathcal{A}_m$  under  $P_Y^{\otimes n}$  satisfies, at the exponential scale,

$$P_Y^{\otimes n}(\mathcal{A}_m) \doteq e^{nH_{Q^*}(Y|X)} \cdot e^{-n[H_{Q^*}(Y)+D_m(Q_Y^*)]} = e^{-n[I_{Q^*}(X; Y)+D_m(Q_Y^*)]}. \quad (22)$$

Clearly,

$$\alpha_n(\tau, R, \mathcal{C}) \geq P_Y^{\otimes n} \left\{ \bigcup_{m=1}^M \mathcal{A}_m \right\}, \quad (23)$$

which we lower-bound via inclusion-exclusion. This is the key structural observation:  $P_Y^{\otimes n}(\mathcal{A}_m) \doteq e^{-n[I_{Q^*}(X; Y)+D_m(Q_Y^*)]}$  is the *same deterministic quantity*  $Z$  for every  $m$  and every codebook  $\mathcal{C}$ , since it depends only on the fixed type  $Q_{XY}^*$  and the fixed distribution  $P_Y$ , not on which specific codeword  $x^n(m)$  was drawn. Accordingly,

$$\sum_{m=1}^M P_Y^{\otimes n}(\mathcal{A}_m) = M \cdot Z \doteq e^{-nE_{\text{FA}}^a(\tau, R)} \quad (24)$$

deterministically. For the pairwise overlaps, since  $x^n(m)$  and  $x^n(m')$  are independent for  $m \neq m'$ , for each fixed  $y^n \in \mathcal{T}(Q_Y^*)$  the events  $\{y^n \in \mathcal{A}_m\}$  and  $\{y^n \in \mathcal{A}_{m'}\}$  are independent. Hence:

$$\begin{aligned} \mathbb{E} \{P_Y^{\otimes n}(\mathcal{A}_m \cap \mathcal{A}_{m'})\} &= \sum_{y^n} P_Y^{\otimes n}(y^n) \cdot \Pr\{\mathcal{A}_m \text{ includes } y^n\} \cdot \Pr\{\mathcal{A}_{m'} \text{ includes } y^n\} \\ &\doteq e^{-n[D_m(Q_Y^*)+2I_{Q^*}(X; Y)]}, \end{aligned} \quad (25)$$

and so:

$$\mathbb{E} \left\{ \sum_{m \neq m'} P_Y^{\otimes n}(\mathcal{A}_m \cap \mathcal{A}_{m'}) \right\} \doteq e^{-n[D_m(Q_Y^*)+2I_{Q^*}(X; Y)-2R]}. \quad (26)$$

In the sparse case  $I_{Q^*}(X; Y) > R$ , the exponent in (26) exceeds  $E_{\text{FA}}^a(\tau, R) = D_m(Q_Y^*) + I_{Q^*}(X; Y) - R$  by the amount  $I_{Q^*}(X; Y) - R > 0$ . Hence by Markov's inequality and the

Borel–Cantelli lemma, for almost every  $\mathcal{C}$  and all large  $n$ :

$$\sum_{m \neq m'} P_Y^{\otimes n}(\mathcal{A}_m \cap \mathcal{A}_{m'}) \leq \frac{1}{2} e^{-n[E_{\text{FA}}^{\text{a}}(\tau, R) + \epsilon/2]}. \quad (27)$$

By the inclusion-exclusion principle, using (24) and the a.s. event (27):

$$P_Y^{\otimes n} \left\{ \bigcup_{m=1}^M \mathcal{A}_m \right\} \geq \sum_{m=1}^M P_Y^{\otimes n}(\mathcal{A}_m) - \sum_{m \neq m'} P_Y^{\otimes n}(\mathcal{A}_m \cap \mathcal{A}_{m'}) \geq \frac{1}{2} M \cdot Z \doteq \frac{1}{2} e^{-nE_{\text{FA}}^{\text{a}}(\tau, R)}, \quad (28)$$

so  $\alpha_n(\tau, R, \mathcal{C}) \geq e^{-n[E_{\text{FA}}^{\text{a}}(\tau, R) + \epsilon]}$  a.s.

Consider next the case where  $Q_{XY}^*$  is bulk ( $I_{Q^*}(X; Y) \leq R$ ). In this case,

$$E_{\text{FA}}^{\text{a}}(\tau, R) = D_m(Q_Y^*). \quad (29)$$

Since the feasible set  $\{\lambda(Q_{XY}, R) \geq \tau\}$  is closed and the infimum of  $D_m(Q_Y) + [I_Q(X; Y) - R]_+$  over it equals  $E_{\text{FA}}^{\text{a}}(\tau, R)$ , for any  $\epsilon > 0$  we may choose  $Q_{XY}^*$  with  $\lambda(Q_{XY}^*, R) > \tau$  (strictly inside the feasible set) and  $D_m(Q_Y^*) \leq E_{\text{FA}}^{\text{a}}(\tau, R) + \epsilon/2$ . As mentioned before,  $N(Q_{XY}^* | y^n)$  is a binomial random variable with  $M = e^{nR}$  trials and success rate of the exponential order of  $e^{-nI_{Q^*}(X; Y)}$ . Since  $\mu_n \triangleq \mathbb{E}\{N(Q_{XY}^* | y^n)\} \doteq e^{n[R - I_{Q^*}(X; Y)]} \rightarrow \infty$ , the Chernoff bound yields, for any  $\delta \in (0, 1)$ :

$$\Pr\{N(Q_{XY}^* | y^n) < (1 - \delta)\mu_n\} \leq e^{-\delta^2 \mu_n / 2} = \exp\{-\delta^2 e^{n[R - I_{Q^*}(X; Y)]} / 2\}, \quad (30)$$

which is doubly exponentially small as a function of  $n$ . By the union bound over  $|\mathcal{T}(Q_Y^*)| \leq e^{nH_{Q^*}(Y)}$   $Q_Y^*$ -typical sequences and the Borel-Cantelli lemma, almost surely, for all large  $n$ ,  $N(Q_{XY}^* | y^n) \geq (1 - \delta)\mu_n$  for all  $y^n \in \mathcal{T}(Q_Y^*)$  at the same time. On this a.s. event, for every  $y^n \in \mathcal{T}(Q_Y^*)$ :

$$\begin{aligned} \Lambda(y^n, \mathcal{C}) &\geq \frac{1}{n} \log \frac{(1 - \delta) e^{n[R - I_{Q^*}(X; Y)]} \cdot e^{-n\ell(Q_{XY}^*)}}{e^{nR} \cdot e^{-n[H_{Q^*}(Y) + D_m(Q_Y^*)]}} \\ &= D_m(Q_Y^*) - D_c(Q_{XY}^*) + \frac{\log(1 - \delta)}{n} \\ &= \lambda(Q_{XY}^*, R) + \frac{\log(1 - \delta)}{n} > \tau + \frac{\log(1 - \delta)}{n} \end{aligned} \quad (31)$$

which exceeds  $\tau$  for all large  $n$  (since  $\lambda(Q_{XY}^*, R) > \tau$  and  $\frac{\log(1 - \delta)}{n} \rightarrow 0$ ), where we have used  $\lambda(Q_{XY}^*, R) = D_m(Q_Y^*) - D_c(Q_{XY}^*)$  (as  $[I_{Q^*}(X; Y) - R]_+ = 0$  for bulk  $Q^*$ ). Therefore,  $\alpha_n(\tau, R, \mathcal{C}) \geq P_Y^{\otimes n}(\mathcal{T}(Q_Y^*)) \doteq e^{-nD_m(Q_Y^*)} \geq e^{-n[E_{\text{FA}}^{\text{a}}(\tau, R) + \epsilon/2]}$ , so  $\alpha_n(\tau, R, \mathcal{C}) \geq e^{-n[E_{\text{FA}}^{\text{a}}(\tau, R) + \epsilon]}$  a.s. for all large  $n$ .

Moving on to the upper bound, we use the fact that  $\alpha_n(\tau, R, \mathcal{C})$  is a non-negative random variable whose expectation over the randomness of  $\mathcal{C}$  is the *annealed* FA probability:

$$\mathbb{E}\{\alpha_n(\tau, R, \mathcal{C})\} = \sum_{y^n} P_Y^{\otimes n}(y^n) \cdot \Pr\{\Lambda(y^n, \mathcal{C}) \geq \tau\} = \alpha_n^{\text{a}}(\tau, R). \quad (32)$$

By the annealed analysis of [1, Theorem 1],

$$\alpha_n^a(\tau, R) \doteq e^{-nE_{\text{FA}}^a(\tau, R)}, \quad (33)$$

so for any  $\epsilon > 0$  and all large  $n$ , we have by Markov's inequality,

$$\Pr \left\{ \mathcal{C} : \alpha_n(\tau, R, \mathcal{C}) \geq e^{-n[E_{\text{FA}}^a(\tau, R) - \epsilon]} \right\} \leq \frac{e^{-nE_{\text{FA}}^a(\tau, R)}}{e^{-n[E_{\text{FA}}^a(\tau, R) - \epsilon]}} = e^{-n\epsilon}, \quad (34)$$

which is summable across all  $n \geq 1$ , and so, by the Borel-Cantelli lemma,

$$\alpha_n(\tau, R, \mathcal{C}) \leq e^{-n[E_{\text{FA}}^a(\tau, R) - \epsilon]} \quad (35)$$

eventually almost surely for all large  $n$ . Combining with the a.s. lower bound, we observe that  $\frac{1}{n} \log \alpha_n(\tau, R, \mathcal{C}) \rightarrow -E_{\text{FA}}^a(\tau, R)$  a.s.  $\square$

The result for the expected logarithm of the FA probability is asserted in the following corollary:

**Corollary 1.**

$$-\frac{1}{n} \mathbb{E} \{ \log \alpha_n(\tau, R, \mathcal{C}) \} \xrightarrow{n \rightarrow \infty} E_{\text{FA}}^a(\tau, R), \quad (36)$$

or, equivalently,

$$E_{\text{FA}}^q(\tau, R) = E_{\text{FA}}^a(\tau, R). \quad (37)$$

*Proof.* Let  $U_n \triangleq -\frac{1}{n} \log \alpha_n(\tau, R, \mathcal{C}) \geq 0$ . By Theorem 1, we know that  $U_n \rightarrow E_{\text{FA}}^a(\tau, R)$  a.s. By Fatou's lemma,

$$\liminf_{n \rightarrow \infty} \mathbb{E} \{ U_n \} \geq \mathbb{E} \{ \liminf_{n \rightarrow \infty} U_n \} = E_{\text{FA}}^a(\tau, R). \quad (38)$$

On the other hand, by Jensen's inequality applied to the logarithmic function,

$$\mathbb{E} \{ U_n \} = -\frac{1}{n} \mathbb{E} \{ \log \alpha_n(\tau, R, \mathcal{C}) \} \leq -\frac{1}{n} \log \mathbb{E} \{ \alpha_n(\tau, R, \mathcal{C}) \} = -\frac{1}{n} \log \alpha_n^a(\tau, R) \rightarrow E_{\text{FA}}^a(\tau, R). \quad (39)$$

Combining both bounds, we conclude that  $\lim_{n \rightarrow \infty} \mathbb{E} \{ U_n \} = E_{\text{FA}}^a(\tau, R)$ .  $\square$

## 4 The MD Exponent

We now present parallel results for the MD exponent.

**Theorem 2.** For all  $\tau \in \mathbb{R}$  and  $R > 0$ :

$$\frac{1}{n} \log \beta_n(\tau, R, \mathcal{C}) \xrightarrow{n \rightarrow \infty} -E_{\text{MD}}^a(\tau, R) \quad \text{a.s. over } \mathcal{C}. \quad (40)$$

*Proof.* The upper bound follows by Markov's inequality and the Borel–Cantelli lemma, exactly as for the FA exponent. We henceforth focus on the lower bound.

By the symmetry of the random coding mechanism, it may be assumed without loss of generality that  $x^n(1)$  is the transmitted codeword. Let  $x^n \triangleq x^n(1)$  be drawn uniformly from  $\mathcal{T}(P_X)$  independently of  $\mathcal{C}' \triangleq \{x^n(2), \dots, x^n(M)\}$ , and let  $y^n \sim W^n(\cdot|x^n)$ . Define

$$S'(y^n, \mathcal{C}') \triangleq \sum_{m=2}^M W^n(y^n|x^n(m)), \quad (41)$$

so that

$$S(y^n, \mathcal{C}) = W^n(y^n|x^n) + S'(y^n, \mathcal{C}'). \quad (42)$$

All almost-sure statements below are over the random selection of  $\mathcal{C}$ . Pick a feasible type  $Q_{XY}^*$  with  $Q_X^* = P_X$ ,  $\lambda(Q_{XY}^*, R) < \tau$ ,  $\Delta(Q_Y^*, R) < \tau$ , and  $D_c(Q_{XY}^*) \leq E_{\text{MD}}^a(\tau, R) + \epsilon/2$ . Such a type exists since the feasible set is non-empty with infimum  $E_{\text{MD}}^a(\tau, R)$ . Set  $\delta \triangleq \tau - \Delta(Q_Y^*, R) > 0$ .

For any  $y^n \in \mathcal{T}(Q_Y^*)$ , the LLR  $\Lambda(y^n, \mathcal{C})$  simplifies to:

$$\Lambda(y^n, \mathcal{C}) = \frac{1}{n} \log S(y^n, \mathcal{C}) - R + H_{Q^*}(Y) + D_m(Q_Y^*). \quad (43)$$

Define

$$\theta(Q_Y^*) \triangleq \tau + R - H_{Q^*}(Y) - D_m(Q_Y^*), \quad (44)$$

so that  $\Lambda(y^n, \mathcal{C}) < \tau$  iff  $S(y^n, \mathcal{C}) < e^{n\theta(Q_Y^*)}$ . For  $(x^n, y^n) \in \mathcal{T}(Q_{XY}^*)$ ,  $W^n(y^n|x^n) = e^{-n\ell(Q_{XY}^*)}$  where

$$\ell(Q_{XY}^*) \triangleq H_{Q^*}(Y|X) + D_c(Q_{XY}^*). \quad (45)$$

Using the identity  $H_{Q^*}(Y) = H_{Q^*}(Y|X) + I_{Q^*}(X; Y)$ , we have

$$\ell(Q_{XY}^*) + \theta(Q_Y^*) = \tau - \lambda(Q_{XY}^*, R) + [R - I_{Q^*}(X; Y)]_+ > 0 \quad (46)$$

since  $\lambda(Q_{XY}^*, R) < \tau$  and  $[R - I_{Q^*}(X; Y)]_+ \geq 0$ . Hence  $W^n(y^n|x^n) < e^{n\theta(Q_Y^*)}$ , and for all large  $n$ :

$$W^n(y^n|x^n) < \frac{1}{2} e^{n\theta(Q_Y^*)}. \quad (47)$$

Next, let us decompose  $S'(y^n, \mathcal{C}')$  by joint types:

$$S'(y^n, \mathcal{C}') = \sum_{\substack{Q'_{XY}: Q'_X = P_X \\ Q'_Y = Q_Y^*}} N'(Q'_{XY}|y^n) \cdot e^{-n\ell(Q'_{XY})}, \quad (48)$$

where  $N'(Q'_{XY}|y^n) \triangleq \#\{m \geq 2 : (x^n(m), y^n) \in \mathcal{T}(Q'_{XY})\}$ . Since  $y^n \in \mathcal{T}(Q_Y^*)$ , only types  $Q'_{XY}$  with  $Q'_Y = Q_Y^*$  contribute to  $S'(y^n, \mathcal{C}')$ .

We next distinguish between sparse and bulk types  $Q'_{XY}$ .

For sparse types,  $\mathbb{E}\{N'(Q'_{XY}|y^n)\} \doteq e^{-n[I_{Q'}(X;Y)-R]} \rightarrow 0$ , so by Markov's inequality and the Borel–Cantelli lemma,  $N'(Q'_{XY}|y^n) = 0$  eventually a.s. for all large  $n$ , simultaneously over all  $y^n \in \mathcal{T}(Q_Y^*)$  and all (polynomially many) sparse  $Q'_{XY}$ .

For bulk types,  $N'(Q'_{XY}|y^n)$  is a binomial random variable with  $M-1$  trials and success rate of the exponential order of  $e^{-nI_{Q'}(X;Y)}$ . Denoting  $\mu'_n \triangleq \mathbb{E}\{N'(Q'_{XY}|y^n)\} \doteq e^{n[R-I_{Q'}(X;Y)]}$ , the Chernoff bound yields  $\Pr\{N'(Q'_{XY}|y^n) > 2\mu'_n\} \leq e^{-\mu'_n/3}$ , which is doubly exponentially small. Applying the union bound over all  $y^n \in \mathcal{T}(Q_Y^*)$  and all (polynomially many) bulk types  $Q'_{XY}$ , the total failure probability is still doubly exponentially small. By the Borel–Cantelli lemma, eventually almost surely for all large  $n$ :

$$N'(Q'_{XY}|y^n) \leq 2\mu'_n \quad \text{for all } y^n \in \mathcal{T}(Q_Y^*) \text{ and all bulk } Q'. \quad (49)$$

Call this a.s. event  $\Omega$ . On  $\Omega$ , for any  $y^n \in \mathcal{T}(Q_Y^*)$ , we bound  $S'(y^n, \mathcal{C}')$  by summing only over bulk types (sparse types contribute nothing on  $\Omega$ ):

$$\begin{aligned} S'(y^n, \mathcal{C}') &\leq \sum_{\substack{Q'_{XY}: Q'_Y=Q_Y^* \\ I_{Q'}(X;Y) \leq R}} 2\mu'_n \cdot e^{-n\ell(Q'_{XY})} \\ &= 2 \sum_{Q'_{XY}} e^{n[R-I_{Q'}(X;Y)]} \cdot e^{-n[H_{Q'}(Y|X)+D_c(Q'_{XY})]} \\ &= 2 \sum_{Q'_{XY}} e^{n[R-H_{Q^*}(Y)-D_c(Q'_{XY})]} \end{aligned} \quad (50)$$

$$= 2 \sum_{Q'_{XY}} e^{n[\theta(Q_Y^*)+\lambda(Q'_{XY}, R)-\tau]} \quad (51)$$

$$\begin{aligned} &\leq 2 \text{poly}(n) \cdot e^{n\theta(Q_Y^*)} \cdot e^{-n\delta} \\ &< \frac{1}{2} e^{n\theta(Q_Y^*)} \end{aligned} \quad (52)$$

for all large  $n$ , where  $\text{poly}(n)$  denotes the (polynomial) number of types. Step (50) uses  $I_{Q'}(X;Y) + H_{Q'}(Y|X) = H_{Q'}(Y) = H_{Q^*}(Y)$  since  $Q'_Y = Q_Y^*$ . Step (51) uses the identity, valid for all bulk  $Q'_{XY}$  with  $Q'_Y = Q_Y^*$  (so  $[I_{Q'}(X;Y) - R]_+ = 0$ ):

$$R - H_{Q^*}(Y) - D_c(Q'_{XY}) = \theta(Q_Y^*) + \lambda(Q'_{XY}, R) - \tau, \quad (53)$$

which follows by substituting  $\theta(Q_Y^*) = \tau + R - H_{Q^*}(Y) - D_m(Q_Y^*)$  and  $\lambda(Q'_{XY}, R) = D_m(Q_Y^*) - D_c(Q'_{XY})$  (the latter is because  $[I_{Q'}(X;Y) - R]_+ = 0$  for bulk  $Q'_{XY}$ ). The penultimate step uses  $\lambda(Q'_{XY}, R) \leq \Delta(Q_Y^*, R) = \tau - \delta$  for every bulk  $Q'_{XY}$  with  $Q'_Y = Q_Y^*$ , by definition of  $\Delta(Q_Y^*, R)$ .

It follows that on  $\Omega$ , for all large  $n$  and all  $y^n \in \mathcal{T}(Q_{Y|X}^*|x^n)$ :

$$S(y^n, \mathcal{C}) = W^n(y^n|x^n) + S'(y^n, \mathcal{C}') < \frac{1}{2} e^{n\theta(Q_Y^*)} + \frac{1}{2} e^{n\theta(Q_Y^*)} = e^{n\theta(Q_Y^*)}, \quad (54)$$

and so,  $\Lambda(y^n, \mathcal{C}) < \tau$  for every such  $y^n$ . Therefore:

$$\begin{aligned} \beta_n(\tau, R, \mathcal{C}) &\geq W^n(\mathcal{T}(Q_{Y|X}^* | x^n) | x^n) \\ &\doteq e^{-nD_c(Q_{XY}^*)} \\ &\geq e^{-n[E_{\text{MD}}^a(\tau, R) + \epsilon]} \quad \text{a.s.}, \end{aligned} \tag{55}$$

where the second line uses the method of types and the third uses  $D_c(Q_{XY}^*) \leq E_{\text{MD}}^a(\tau, R) + \epsilon/2$ . Combining both bounds and letting  $\epsilon \rightarrow 0$ :  $\frac{1}{n} \log \beta_n(\tau, R, \mathcal{C}) \rightarrow -E_{\text{MD}}^a(\tau, R)$  a.s.  $\square$

We have the following analogous corollary for the quenched MD exponent.

**Corollary 2.**

$$-\frac{1}{n} \mathbb{E}\{\log \beta_n(\tau, R, \mathcal{C})\} \xrightarrow{n \rightarrow \infty} E_{\text{MD}}^a(\tau, R), \tag{56}$$

or, equivalently,

$$E_{\text{MD}}^a(\tau, R) = E_{\text{MD}}^a(\tau, R). \tag{57}$$

*Proof.* The proof is the same as for Corollary 1, with  $U_n \triangleq -\frac{1}{n} \log \beta_n(\tau, R, \mathcal{C})$  and  $E_{\text{FA}}^a(\tau, R)$  replaced by  $E_{\text{MD}}^a(\tau, R)$ .  $\square$

**Remark 1.** Since Corollaries 1 and 2 establish that the annealed and quenched exponents coincide, there is no longer any need to distinguish between them. Henceforth, we write  $E_{\text{FA}}(\tau, R)$  and  $E_{\text{MD}}(\tau, R)$  for the common value, dropping the superscript <sup>a</sup>.

## 5 Discussion on the Annealed-Quenched Equivalence

The self-averaging results of Sections 3 and 4 tell us that for large  $n$ , both  $\frac{1}{n} \log \alpha_n(\tau, R, \mathcal{C})$  and  $\frac{1}{n} \log \beta_n(\tau, R, \mathcal{C})$  are essentially constants across the vast majority of codebooks  $\mathcal{C}$ : the *Jensen gap* of both exponents vanishes. Such a behavior should not be taken for granted, because it is not common to other exponents associated with random coding, for example, the classical (annealed) random coding error exponent (see, e.g., [12], [13, Section 5.6]) as opposed to the typical-code (quenched) error exponent, which differ in general at low coding rates — see, e.g., [14], [15] and references therein. Moreover, the Jensen gap between the annealed and quenched free energies of disordered systems with random parameters is positive in general at low temperatures. It is then natural to wonder what it is that makes the difference between the problem setting at hand and the other problem settings in that respect.

The key structural reason is that  $S(y^n, \mathcal{C}) = \sum_{m=1}^M W^n(y^n | x^n(m))$  is a sum of  $M = e^{nR}$  *independent* terms for each fixed  $y^n$  (independence follows from the independent generation of codewords). The coefficient of variation therefore decays as  $e^{-nR/2} \rightarrow 0$ , so  $S(y^n, \mathcal{C})/\mathbb{E}\{S(y^n, \mathcal{C})\} \rightarrow 1$

for each  $y^n$ , and the self-averaging of  $\frac{1}{n} \log \alpha_n(\tau, R, \mathcal{C})$  follows. This is in contrast to spin glasses, where the terms of the partition function are *correlated* through the shared coupling matrix, causing large relative fluctuations and a positive Jensen gap; and to channel coding, where the error probability depends on a *maximum* rather than a sum, making it sensitive to rare bad codewords. For  $\beta_n(\tau, R, \mathcal{C})$ , the mechanism is slightly different: the dominant contribution is the deterministic transmitted-type weight  $W^n(y^n|x^n(1)) = e^{-n\ell(Q_{XY}^*)}$ , and the interferer sum concentrating *below* the threshold  $e^{n\theta(Q_Y^*)}$  is what gives self-averaging.

## 6 Mismatched Detection

So far we have assumed that the detector knows the true channel  $W$  and computes the optimal Neyman–Pearson LLR. We now study what happens when the detector is *mismatched*: it assumes a wrong channel model  $\tilde{W} : \mathcal{X} \rightarrow \mathcal{Y}$  in place of  $W$ . This is the practically relevant scenario where the channel is estimated imperfectly or chosen for computational convenience rather than accuracy. The mismatched setting also opens a genuinely new theoretical question: does self-averaging survive, and how do the exponents degrade?

Let  $\tilde{W} : \mathcal{X} \rightarrow \mathcal{Y}$  be the detector’s assumed channel, with output marginal  $\tilde{P}_Y(y) \triangleq \sum_x P_X(x)\tilde{W}(y|x)$ . The codebook  $\mathcal{C}$  is still drawn from  $\mathcal{T}(P_X)$  under the true channel  $W$ , and the true hypotheses are as before:

$$\begin{aligned} \mathcal{H}_0 : y^n &\sim P_Y^{\otimes n}, \\ \mathcal{H}_1 : y^n &\sim P_{Y^n|\mathcal{C}} \quad (\text{true mixture output under } W). \end{aligned}$$

The detector constructs its LLR assuming the channel  $\tilde{W}$ :

$$\tilde{\Lambda}(y^n, \mathcal{C}) \triangleq \frac{1}{n} \log \frac{\sum_{m=1}^M \tilde{W}^n(y^n|x^n(m))}{M \cdot \tilde{P}_Y^{\otimes n}(y^n)}. \quad (58)$$

The mismatched detector decides in favor of  $\mathcal{H}_1$  iff  $\tilde{\Lambda}(y^n, \mathcal{C}) \geq \tau$ , otherwise it decides in favor of  $\mathcal{H}_0$ . The error probabilities are:

$$\tilde{\alpha}_n(\tau, R, \mathcal{C}) \triangleq \sum_{\{y^n: \tilde{\Lambda}(y^n, \mathcal{C}) \geq \tau\}} P_Y^{\otimes n}(y^n) \quad (59)$$

$$\tilde{\beta}_n(\tau, R, \mathcal{C}) \triangleq \frac{1}{M} \sum_{m=1}^M \sum_{\{y^n: \tilde{\Lambda}(y^n, \mathcal{C}) < \tau\}} W^n(y^n|x^n(m)). \quad (60)$$

Note that  $\tilde{\alpha}_n$  and  $\tilde{\beta}_n$  are always evaluated under the true measures  $P_Y^{\otimes n}$  and  $W^n$ , as error probabilities are defined with respect to the true distribution regardless of the detector’s beliefs. What differs between modeling choices is the hypotheses the detector formulates. In the model above, the true hypotheses are  $\mathcal{H}_0$  and  $\mathcal{H}_1$  as stated, and the mismatch lies entirely in the test

statistic  $\tilde{\Lambda}$  (e.g. due to a wrong channel estimate or a computationally convenient approximation). A detector who *genuinely believes* the channel is  $\tilde{W}$  would instead formulate hypotheses  $\tilde{\mathcal{H}}_0 : y^n \sim \tilde{P}_Y^{\otimes n}$  and  $\tilde{\mathcal{H}}_1 : y^n \sim \frac{1}{M} \sum_m \tilde{W}^n(\cdot | x^n(m))$ , giving the LRT statistic  $\tilde{\Lambda}(y^n, \mathcal{C})$  a different operational meaning even though its mathematical form is identical.

For a joint type  $Q_{XY}$  with  $Q_X = P_X$ , define the mismatched analogues of the quantities in Section 2:

$$\tilde{D}_m(Q_Y) \triangleq D(Q_Y \| \tilde{P}_Y), \quad (61)$$

$$\tilde{D}_c(Q_{XY}) \triangleq D(Q_{Y|X} \| \tilde{W} | P_X), \quad (62)$$

$$\tilde{\ell}(Q_{XY}) \triangleq H_Q(Y|X) + \tilde{D}_c(Q_{XY}), \quad (63)$$

$$\tilde{\lambda}(Q_{XY}, R) \triangleq \tilde{D}_m(Q_Y) - \tilde{D}_c(Q_{XY}) + [I_Q(X; Y) - R]_+, \quad (64)$$

$$\tilde{\theta}(Q_Y) \triangleq \tau + R - H_Q(Y) - \tilde{D}_m(Q_Y), \quad (65)$$

$$\tilde{\Delta}(Q_Y, R) \triangleq \max_{\substack{Q'_{XY}: Q'_X = P_X, Q'_Y = Q_Y \\ I_{Q'}(X; Y) \leq R}} \tilde{\lambda}(Q'_{XY}, R). \quad (66)$$

Clearly, for  $(x^n, y^n) \in \mathcal{T}(Q_{XY})$ ,  $\tilde{W}^n(y^n | x^n) = e^{-n\tilde{\ell}(Q_{XY})}$  and  $\tilde{P}_Y^{\otimes n}(y^n) = e^{-n[H_Q(Y) + \tilde{D}_m(Q_Y)]}$ , so the condition  $\tilde{\Lambda}(y^n, \mathcal{C}) < \tau$  is equivalent to  $\tilde{S}(y^n, \mathcal{C}) < e^{n\tilde{\theta}(Q_Y)}$ , where  $\tilde{S}(y^n, \mathcal{C}) \triangleq \sum_m \tilde{W}^n(y^n | x^n(m))$ .

The annealed exponents follow from the same derivations as in [1], replacing  $W$  with  $\tilde{W}$  in the test statistic while keeping the true underlying probability measures for the error probabilities.

### Theorem 3.

$$\tilde{E}_{\text{FA}}(\tau, R) = \min_{\substack{Q_{XY}: Q_X = P_X \\ \tilde{\lambda}(Q_{XY}, R) \geq \tau}} [D_m(Q_Y) + [I_Q(X; Y) - R]_+], \quad (67)$$

$$\tilde{E}_{\text{MD}}(\tau, R) = \min_{\substack{Q_{XY}: Q_X = P_X \\ \tilde{\lambda}(Q_{XY}, R) < \tau \\ \tilde{\Delta}(Q_Y, R) < \tau}} D_c(Q_{XY}). \quad (68)$$

*Proof.* The proof follows [1, Theorem 1] with the following substitutions throughout:  $W \rightarrow \tilde{W}$ ,  $P_Y \rightarrow \tilde{P}_Y$ ,  $\ell(Q_{XY}) \rightarrow \tilde{\ell}(Q_{XY})$ ,  $\lambda(Q_{XY}, R) \rightarrow \tilde{\lambda}(Q_{XY}, R)$ ,  $\theta(Q_Y) \rightarrow \tilde{\theta}(Q_Y)$ ,  $\Delta(Q_Y, R) \rightarrow \tilde{\Delta}(Q_Y, R)$ , and  $S(y^n, \mathcal{C}) \rightarrow \tilde{S}(y^n, \mathcal{C}) \triangleq \sum_m \tilde{W}^n(y^n | x^n(m))$ . The key steps are as follows.

The annealed FA probability is:

$$\mathbb{E}\{\tilde{\alpha}_n(\tau, R, \mathcal{C})\} = \sum_{y^n} P_Y^{\otimes n}(y^n) \cdot \Pr\{\tilde{S}(y^n, \mathcal{C}) \geq e^{n\tilde{\theta}(\tilde{P}_{y^n})}\}.$$

For  $y^n \in \mathcal{T}(Q_Y)$ :  $P_Y^{\otimes n}(y^n) = e^{-n[H_Q(Y) + D_m(Q_Y)]}$ ;  $\tilde{W}^n(y^n | x^n(m)) = e^{-n\tilde{\ell}(Q_{XY})}$  for  $(x^n(m), y^n) \in \mathcal{T}(Q_{XY})$ . Crucially, the type-class enumerator  $N(Q_{XY} | y^n) \sim \text{Bin}(M, e^{-nI_Q(X; Y)})$  has the same distribution as in the matched case, as it counts codewords of a given joint type with  $y^n$ , which depends only on the codebook, not on  $\tilde{W}$ . Applying Theorem 4.1 of [19] to bound  $\Pr\{\tilde{S}(y^n, \mathcal{C}) \geq$

$e^{n\tilde{\theta}(\hat{P}_{y^n})}$  exactly as in [1], the dominant type minimizes the true cost  $D_m(Q_Y) + [I_Q(X; Y) - R]_+$  subject to the mismatched constraint  $\tilde{\lambda}(Q_{XY}, R) \geq \tau$ , yielding (67).

The annealed MD probability is:

$$\mathbb{E}\{\tilde{\beta}_n(\tau, R, \mathcal{C})\} = \frac{1}{M} \sum_{m=1}^M \sum_{y^n} W^n(y^n | x^n(m)) \cdot \Pr\{\tilde{S}'(y^n, \mathcal{C}') < e^{n\tilde{\theta}(\hat{P}_{y^n})}\}.$$

The probability that  $(x^n, y^n)$  has joint type  $Q_{XY}$  is  $\doteq e^{-nD_c(Q_{XY})}$ . The event  $\{\tilde{S}(y^n, \mathcal{C}) < e^{n\tilde{\theta}(\hat{P}_{y^n})}\}$  is analyzed via Theorem 4.3 of [19] applied to the counts  $N'(Q'_{XY} | y^n)$ , with thresholds computed using  $\tilde{\ell}(Q_{XY})$  and  $\tilde{\theta}(Q_Y)$ . The feasibility condition reduces to  $\tilde{\lambda}(Q_{XY}, R) < \tau$  and  $\tilde{\Delta}(Q_Y, R) < \tau$ , and the dominant type minimizes the true  $D_c(Q_{XY})$  over this mismatched feasible set, yielding (68).  $\square$

In the FA exponent (67): the *constraint*  $\tilde{\lambda}(Q_{XY}, R) \geq \tau$  uses the mismatched  $\tilde{\lambda}(Q_{XY}, R)$  (the test statistic decides when to fire), while the *cost*  $D_m(Q_Y) + [I_Q(X; Y) - R]_+$  uses the true divergence  $D_m(Q_Y)$  (the true FA probability weight). In the MD exponent (68): the constraints use  $\tilde{\lambda}(Q_{XY}, R)$  and  $\tilde{\Delta}(Q_Y, R)$ , while the cost  $D_c(Q_{XY})$  uses the true channel. The detector's beliefs ( $\tilde{W}$ ) control *when* the test fires, but the true channel ( $W$ ) controls *how costly* each error is.

The next theorem establishes self-averaging under mismatch; the mismatched counterparts of Corollaries 1 and 2 then follow by the same Fatou–Jensen argument.

**Theorem 4.** *For all  $\tau \in \mathbb{R}$ ,  $R > 0$ , and almost every  $\mathcal{C}$ :*

$$\frac{1}{n} \log \tilde{\alpha}_n(\tau, R, \mathcal{C}) \xrightarrow{n \rightarrow \infty} -\tilde{E}_{\text{FA}}(\tau, R), \quad (69)$$

$$\frac{1}{n} \log \tilde{\beta}_n(\tau, R, \mathcal{C}) \xrightarrow{n \rightarrow \infty} -\tilde{E}_{\text{MD}}(\tau, R). \quad (70)$$

*Proof.* The proofs of Theorems 1 and 2 depend on the codebook only through: (a) the independence of codewords, (b) the type-class structure ( $W^n(y^n | x^n) = e^{-n\ell(Q_{XY})}$  for  $(x^n, y^n) \in \mathcal{T}(Q_{XY})$ ), and (c) linearity of expectation over the codebook. None of these depend on whether the test uses  $W$  or  $\tilde{W}$ . Replacing  $W$  by  $\tilde{W}$  throughout (i.e.  $S(y^n, \mathcal{C}) \rightarrow \tilde{S}(y^n, \mathcal{C})$ ,  $\ell(Q_{XY}) \rightarrow \tilde{\ell}(Q_{XY})$ ,  $\lambda(Q_{XY}, R) \rightarrow \tilde{\lambda}(Q_{XY}, R)$ ,  $\theta(Q_Y) \rightarrow \tilde{\theta}(Q_Y)$ ,  $\Delta(Q_Y, R) \rightarrow \tilde{\Delta}(Q_Y, R)$ ) yields the mismatched annealed exponents as the limits, while all concentration arguments go through unchanged.  $\square$

Self-averaging is therefore a structural property of the random coding mechanism, robust to detector mismatch. However, while the Jensen gap remains zero, the exponent *values* change with  $\tilde{W}$  and can degrade severely.

The more interesting part of this section concerns the phase structure and the mismatched critical rate. The central question is the following: for which rates  $R$  does a tradeoff interval between the two positive exponents exist under the mismatched detector? Define:

$$\tau^*(R) \triangleq \tilde{\lambda}(P_X \otimes W, R) = D(P_Y \| \tilde{P}_Y) - D(W \| \tilde{W} | P_X) + [I(X; Y) - R]_+, \quad (71)$$

the value of  $\tilde{\lambda}(Q_{XY}, R)$  at the true channel type  $Q_{XY}^* = P_X \otimes W$ . In particular, at  $R = I(X; Y)$ :

$$\tau^* \triangleq \tau^*(I(X; Y)) = D(P_Y \| \tilde{P}_Y) - D(W \| \tilde{W} | P_X) \leq 0, \quad (72)$$

with  $\tau^* = 0$  iff  $\tilde{W} = W$  (by the data processing inequality). By the same inequality,  $\tau^*(R) \leq [I(X; Y) - R]_+$  for all  $R$ , with equality in particular when  $\tilde{W} = W$ .

**Proposition 1.** (i) *The mismatched critical rate, the infimum of rates above which no tradeoff interval exists, is:*

$$R_c = \inf_{\substack{Q_{XY}: Q_X=P_X \\ Q_Y=P_Y}} \max\{I_{Q'}(X; Y), I(X; Y) - D(W \| \tilde{W} | P_X) + \tilde{D}_c(Q_{XY})\}. \quad (73)$$

*In the matched case,  $R_c = I(X; Y)$ , which is attained by  $Q_{XY} = P_X \otimes W$ .*

(ii) *For  $R < R_c$ , the tradeoff interval  $(\tilde{\Delta}(P_Y, R), \tau^*(R))$  is non-empty (this is precisely the content of part (i)), and both exponents are simultaneously positive for  $\tau \in (\tilde{\Delta}(P_Y, R), \tau^*(R))$ .*

*Proof.* Let  $Q_{XY}^* \triangleq P_X \otimes W$ , so that  $D_m(Q_Y^*) = 0$ ,  $D_c(Q_{XY}^*) = 0$ , and  $\tilde{\lambda}(Q_{XY}^*, R) = \tau^*(R)$ . The tradeoff interval  $(\tilde{\Delta}(P_Y, R), \tau^*(R))$  is non-empty iff  $\tilde{\Delta}(P_Y, R) < \tau^*(R)$ . For bulk  $Q'_{XY}$  with  $Q'_Y = P_Y$  and  $I_{Q'}(X; Y) \leq R$ :

$$\tilde{\lambda}(Q'_{XY}, R) = D(P_Y \| \tilde{P}_Y) - \tilde{D}_c(Q'_{XY}), \quad (74)$$

so  $\tilde{\Delta}(P_Y, R) = D(P_Y \| \tilde{P}_Y) - \min_{\{Q'_{XY}: Q'_Y=P_Y, I_{Q'}(X; Y) \leq R\}} \tilde{D}_c(Q'_{XY})$ . And  $\tau^*(R) = D(P_Y \| \tilde{P}_Y) - D(W \| \tilde{W} | P_X) + (I(X; Y) - R)$  (using  $[I(X; Y) - R]_+ = I(X; Y) - R$  since  $R < I(X; Y)$  in this case). The condition  $\tilde{\Delta}(P_Y, R) < \tau^*(R)$  becomes the following: for all  $Q'_{XY}$  with  $Q'_Y = P_Y$  and  $I_{Q'}(X; Y) \leq R$ ,  $R < I(X; Y) - D(W \| \tilde{W} | P_X) + \tilde{D}_c(Q'_{XY})$ . Combined with  $R \geq I_{Q'}(X; Y)$ , the tradeoff interval first disappears when there exists  $Q'_{XY}$  with  $Q'_Y = P_Y$  such that

$$R \geq \max\{I_{Q'}(X; Y), I(X; Y) - D(W \| \tilde{W} | P_X) + \tilde{D}_c(Q'_{XY})\}. \quad (75)$$

Taking the infimum over all such  $Q'_{XY}$  gives (73). For  $\tilde{W} = W$ :  $Q'_{XY} = Q_{XY}^*$  gives  $I(X; Y)$ , and any other  $Q'_{XY}$  with  $Q'_Y = P_Y$  has  $D_c(Q'_{XY}) > 0$ , giving a larger value. Part (ii) follows from the zero-region characterizations of  $\tilde{E}_{\text{FA}}(\tau, R)$  and  $\tilde{E}_{\text{MD}}(\tau, R)$ .  $\square$

## 6.1 Discussion

Taking  $Q'_{XY} = Q^*_{XY}$  in (73) gives  $R_c \leq I(X; Y)$ . The data processing inequality gives  $\tilde{D}_c(Q_{XY}) \geq D(P_Y \| \tilde{P}_Y)$  for all  $Q_{XY}$  with  $Q_Y = P_Y$ , so from (73):

$$R_c \geq I(X; Y) - D(W \| \tilde{W} | P_X) + D(P_Y \| \tilde{P}_Y) \triangleq I_{\text{GMI}}^{(1)}, \quad (76)$$

where  $I_{\text{GMI}}^{(1)}$  is the generalized mutual information (GMI) of mismatched decoding evaluated at  $s = 1$  [20, 21], and its connection to the full GMI is discussed below. Thus,  $I_{\text{GMI}}^{(1)} \leq R_c \leq I(X; Y)$  (the same interval that contains the LM capacity in mismatched channel decoding), and equivalently, (73) can be written as:

$$R_c = \inf_{\substack{Q_{XY}: Q_X = P_X \\ Q_Y = P_Y}} \max \{ I_Q(X; Y), I_{\text{GMI}}^{(1)} + \tilde{D}_c(Q_{XY}) - D(P_Y \| \tilde{P}_Y) \}. \quad (77)$$

The matched soft-covering point ( $\tau = 0, R = I(X; Y)$ ) shifts to  $(\tau^*, R_c)$  under mismatch: the threshold shifts from 0 to  $\tau^* \leq 0$ , and the critical rate shifts from  $I(X; Y)$  to  $R_c \leq I(X; Y)$ . The threshold shift  $|\tau^*| = I(X; Y) - I_{\text{GMI}}^{(1)}$  equals the GMI capacity loss (see below). Mismatch can only lower  $R_c$ , never raise it above  $I(X; Y)$ , because for  $R \geq I(X; Y)$  the true-type codewords ( $N(Q^*_{XY} | y^n) \doteq e^{n(R - I(X; Y))} \geq 1$ , where  $Q^*_{XY} = P_X \otimes W$ ) push  $\tilde{S}(y^n, \mathcal{C})$  past the threshold on their own, regardless of the weights  $e^{-n\tilde{\ell}(Q_{XY})}$  of other types.

It is interesting to examine two extreme examples of mismatch. The first is a degenerate mismatched channel  $\tilde{W}(y|x) = V(y)$ , which gives  $\tilde{\Lambda}(y^n, \mathcal{C}) \equiv 0$  identically (the mismatched sum  $\tilde{S}(y^n, \mathcal{C}) = M \cdot V^{\otimes n}(y^n)$  cancels the denominator exactly). For  $\tau > 0$ :  $\tilde{\alpha}_n = 0$  and  $\tilde{\beta}_n = 1$ ; for  $\tau < 0$ :  $\tilde{\alpha}_n = 1$  and  $\tilde{\beta}_n = 0$ . In neither case are both exponents simultaneously positive, so  $R_c = 0$ : the collapse occurs at every  $R > 0$ , the worst possible case. The second example is a BSC with the wrong polarity, where  $W = \text{BSC}(p)$ ,  $\tilde{W} = \text{BSC}(1 - p)$ , and  $P_X = (\frac{1}{2}, \frac{1}{2})$ . Since  $\tilde{W}(y|x) = W(y|\bar{x})$ , flipping every codeword bit maps the mismatched sum to the matched sum:  $\tilde{S}(y^n, \mathcal{C}) \stackrel{d}{=} S(y^n, \mathcal{C})$  over balanced binary codebooks, because  $\bar{\mathcal{C}} \stackrel{d}{=} \mathcal{C}$  under the uniform input distribution. Consequently the mismatched and matched exponents are identical, and  $R_c = I(X; Y)$ , despite  $D(W \| \tilde{W} | P_X)$  being large and  $I_{\text{GMI}}^{(1)} \ll I(X; Y)$ . The test statistic carries no less information about  $y^n$  under  $\tilde{W}$  than under  $W$ ; the wrong-polarity assumption is invisible to the LRT. This shows that the lower bound  $R_c \geq I_{\text{GMI}}^{(1)}$  can be far from tight: codebook symmetry can make a severely mismatched detector perform exactly as well as the matched one.

The quantity  $I_{\text{GMI}}^{(1)}$  defined in (76) can be written as:

$$I_{\text{GMI}}^{(1)} = \sum_{x,y} P_X(x) W(y|x) \log \frac{\tilde{W}(y|x)}{\tilde{P}_Y(y)}, \quad (78)$$

i.e. the mutual information computed with  $\tilde{W}$  in the numerator and  $\tilde{P}_Y$  in the denominator, averaged under the true  $P_X \otimes W$ . This is the  $s = 1$  special case of the full GMI [20]:

$$I_{\text{GMI}} = \max_{s \geq 0} \sum_{x,y} P_X(x) W(y|x) \log \frac{\tilde{W}(y|x)^s}{\sum_{x'} P_X(x') \tilde{W}(y|x')^s}, \quad (79)$$

which satisfies  $I_{\text{GMI}} \geq I_{\text{GMI}}^{(1)}$  and  $I_{\text{GMI}} \geq 0$  always, while  $I_{\text{GMI}}^{(1)}$  can be negative for severely mismatched  $\tilde{W}$ . Note that  $I_{\text{GMI}}^{(1)} \geq 0$  iff  $D(W \|\tilde{W}|P_X) \leq I(X;Y) + D(P_Y \|\tilde{P}_Y)$ , i.e. iff the mismatch is not too severe. The lower bound  $R_c \geq I_{\text{GMI}}^{(1)}$  in (76) is therefore informative only when  $I_{\text{GMI}}^{(1)} > 0$ ; when  $I_{\text{GMI}}^{(1)} \leq 0$  the bound is trivial since  $R_c \geq 0$  always. The identity  $I_{\text{GMI}}^{(1)} = I(X;Y) - D(W \|\tilde{W}|P_X) + D(P_Y \|\tilde{P}_Y)$  follows from (78) by writing

$$\log \frac{\tilde{W}(y|x)}{\tilde{P}_Y(y)} = \log \frac{W(y|x)}{P_Y(y)} - \log \frac{W(y|x)}{\tilde{W}(y|x)} + \log \frac{P_Y(y)}{\tilde{P}_Y(y)}$$

and taking expectations under  $P_X \otimes W$ . Therefore  $|\tau^*| = D(W \|\tilde{W}|P_X) - D(P_Y \|\tilde{P}_Y) = I(X;Y) - I_{\text{GMI}}^{(1)}$ : the soft-covering threshold shift equals the  $s = 1$  GMI capacity loss.

## 7 Conclusion

We have proved that both the FA and MD error exponents for the soft-covering hypothesis testing problem are *self-averaging*: for almost every random codebook  $\mathcal{C}$ , both  $\frac{1}{n} \log \alpha_n(\mathcal{C}) \rightarrow -E_{\text{FA}}(\tau, R)$  and  $\frac{1}{n} \log \beta_n(\mathcal{C}) \rightarrow -E_{\text{MD}}(\tau, R)$ . The proofs use elementary tools: Markov's inequality (for both upper bounds), the Chernoff bound and Borel–Cantelli (for the lower bounds). The results confirm that the annealed exponents of [1] describe the typical codebook behavior, not just the ensemble average.

We then extended the analysis to mismatched detection, where the detector uses a wrong channel  $\tilde{W}$ . Self-averaging persists under mismatch, and the mismatched exponents are characterized by the same minimization formulas as the matched case, with  $\tilde{\lambda}(Q_{XY}, R)$ ,  $\tilde{\Delta}(Q_Y, R)$  replacing their matched counterparts in the constraints while the true divergences  $D_m(Q_Y)$ ,  $D_c(Q_{XY})$  appear in the costs. The key structural finding is that for  $R \geq I(X;Y)$ , no tradeoff interval exists regardless of  $\tilde{W}$ , but for  $R < R_c$  the tradeoff interval exists with width depending on the severity of the mismatch. The mismatched critical rate satisfies  $I_{\text{GMI}}^{(1)} \leq R_c \leq I(X;Y)$ , and the natural operating threshold shifts from  $\tau = 0$  to  $\tau^* = D(P_Y \|\tilde{P}_Y) - D(W \|\tilde{W}|P_X) \leq 0$ . The magnitude of this shift equals the mismatched decoding capacity loss:  $|\tau^*| = I(X;Y) - I_{\text{GMI}}^{(1)}$ .

## References

- [1] N. Merhav, “Soft covering through the lens of hypothesis testing,” submitted for publication, 2026. Available on-line at: <https://arxiv.org/pdf/2605.19573>

- [2] A. D. Wyner, “The common information of two dependent random variables,” *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 163–179, 1975.
- [3] T. Han and S. Verdú, “Approximation theory of output statistics,” *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, 1993.
- [4] P. Cuff, “Distributed channel synthesis,” *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.
- [5] P. Cuff, “A stronger soft-covering lemma and applications,” *Proc. 2nd Workshop on Physical-Layer Methods for Wireless Security*, Florence, Italy, pp. 40–43, September 2015.
- [6] P. Cuff, “Soft covering with high probability,” *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, pp. 2963–2967, 2016.
- [7] M. Hayashi, “General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1562–1575, 2006.
- [8] M. Bastani Parizi, E. Telatar, and N. Merhav, “Exact random coding secrecy exponents for the wiretap channel,” *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 509–531, 2017.
- [9] L. Yu and V. Y. F. Tan, “Rényi resolvability and its applications to the wiretap channel,” *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1862–1897, 2019.
- [10] S. Yagli and P. Cuff, “Exact exponent for soft covering,” *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 7635–7654, 2019.
- [11] S.-B. Li, K. Li, and L. Yu, “Two-parameter Rényi information quantities with applications to privacy amplification and soft covering,” submitted, 2026. [arXiv:2511.02297](https://arxiv.org/abs/2511.02297).
- [12] R. G. Gallager, “A simple derivation of the coding theorem and some applications,” *IEEE Trans. Inf. Theory*, vol. IT-11, pp. 3–18, 1965.
- [13] R. G. Gallager, *Information Theory and Reliable Communication*, John Wiley & Sons, New York 1968.
- [14] A. Barg and G. D. Forney, “Random codes: minimum distances and error exponents,” *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2568–2573, 2002.
- [15] N. Merhav, “Error exponents of typical random codes,” *IEEE Trans. Inform. Theory*, vol. 64, no. 9, pp. 6223–6235, September 2018.

- [16] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond*. World Scientific, 1987.
- [17] M. Mézard and A. Montanari, *Information, Physics, and Computation*, Oxford University Press, New York 2009.
- [18] J. Scarlett, A. Guillén i Fàbregas, A. Somekh-Baruch and A. Martinez, “Information-Theoretic Foundations of Mismatched Decoding,” *Foundations and Trends in Communications and Information Theory*, vol. 17, no. 23, pp. 149–401, 2020.
- [19] N. Merhav and N. Weinberger, “A Toolbox for Refined Information-Theoretic Analyses with Applications,” *Foundations and Trends in Comm. and Inf. Theory*, vol. 22, no. 1, pp. 1–184, 2025.
- [20] S. Kaplan and S. Shamai (Shitz), “Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment,” *AEU Arch. Elektron. Übertrag.*, vol. 47, pp. 228–239, 1993.
- [21] J. Scarlett, A. Martínez, and A. Guillén i Fàbregas, “Mismatched decoding: Error exponents, second-order rates and saddlepoint approximations,” *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2647–2666, 2014.