

Achievable Rates and Error Exponents for a Class of Mismatched Compound Channels

Priyanka Patel, Francesc Molina[✉], *Member, IEEE*, and Albert Guillén i Fàbregas[✉], *Fellow, IEEE*

Abstract—This paper investigates achievable information rates and error exponents of mismatched decoding when the channel belongs to the class of channels that are close to the decoding metric in terms of relative entropy. For both discrete- and continuous-alphabet channels, we derive approximations of the worst-case achievable information rates and error exponents as a function of the radius of a small relative entropy ball centered at the decoding metric, allowing the characterization of the loss incurred due to imperfect channel estimation. We provide a number of examples including symmetric metrics and modulo-additive noise metrics for discrete systems, and nearest neighbor decoding for continuous-alphabet channels, where we derive the approximation when the channel admits arbitrary statistics and when it is assumed noise-additive with unknown finite second-order moment.

Index Terms—Mismatched decoding, information rates, error exponents, mismatched compound channel.

I. INTRODUCTION

STANDARD results in information theory have been derived assuming that the channel transition probability law is perfectly known to the system designer as well as to the transmitter and receiver [1]. This assumption is too optimistic in practice since there are several environments where the channel is either unknown or difficult to estimate accurately. Therefore, devising the fundamental limits for reliable trans-

mission under channel uncertainty becomes imperative [2], [3], [4].

Most practical receivers aim at estimating the channel law and use this estimate for decoding as if it were perfect. Mismatched decoding (see e.g. [4] and references therein), studies precisely the problem of reliable communication with a fixed decoding metric not necessarily equal to the channel. Beyond channel uncertainty, mismatched decoding encompasses a number of important problems such as bit-interleaved coded modulation, finite-precision arithmetic and zero-error communication [4].

This manuscript delves into channel uncertainty by introducing a parameterization for the mismatched compound channel [4, Ch. 2.2.4]. We model channel uncertainty as a compound channel class comprising all channels that are contained within a small radius relative entropy ball centered at the known decoding metric. Accurate approximations to achievable information rates and error exponents are derived using i.i.d. and constant-composition codes for discrete memoryless channels (DMC), and i.i.d. and cost-constrained codes for continuous-alphabet channels. The key tool used is the approximation of relative entropy adopted by Borade and Zheng in [5]. This, combined with convex optimisation [6] and variational calculus [7] tools, provides closed-form results with a strong conceptual understanding of the sensitivity to small channel uncertainties.

A. Notation

Vectors are indicated in boldface and the corresponding i -th entry is written using a subscript, as \mathbf{x} and x_i . We denote the transpose of \mathbf{x} by \mathbf{x}^T . Random variables are denoted in upper-case. The expectation and the variance of random variable X under probability distribution Q_X are denoted respectively by $\mathbb{E}_{Q_X}[X]$ and $\text{Var}_{Q_X}[X]$, or simply $\mathbb{E}[X]$ and $\text{Var}[X]$; emphasis on Q_X will be made when ambiguity arises. The cardinality of alphabet \mathcal{X} is indicated as $|\mathcal{X}|$. The relative entropy and chi-squared distance between probability distributions P and P' are $D(P||P')$ and $\chi^2(P, P')$, respectively. Throughout the paper, natural logarithms are adopted.

B. Problem Setup

Consider reliable transmission of M messages over a DMC with input X and output Y , taking values from discrete and finite alphabets \mathcal{X} and \mathcal{Y} , respectively. The input distribution is denoted by $Q_X(x) = \Pr[X = x]$ for all $x \in \mathcal{X}$ and the channel

Received 8 March 2024; revised 17 February 2025; accepted 13 May 2025. Date of publication 28 May 2025; date of current version 20 August 2025. This work was supported in part by European Research Council under Grant 725411 and Grant 101142747 and in part by Spanish Ministry of Economy and Competitiveness under Grant PID2020-116683GB-C22. The work of Francesc Molina was supported by the Margarita Salas Fellowship. An earlier version of this paper was presented in part at the 2024 International Zürich Seminar on Information and Communication and in part at the 2024 IEEE International Symposium on Information Theory. (Corresponding author: Albert Guillén i Fàbregas.)

Priyanka Patel was with the Department of Engineering, University of Cambridge, CB2 1PZ Cambridge, U.K., and is now with Google, NIC 4AG London, U.K. (e-mail: pp490@cantab.ac.uk).

Francesc Molina was with the Department of Engineering, University of Cambridge, CB2 1PZ Cambridge, U.K., and also with the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain. He is now with the Department of Information and Communication Technologies, Universitat Pompeu Fabra, 08018 Barcelona, Spain (e-mail: fm585@cam.ac.uk).

Albert Guillén i Fàbregas is with the Department of Engineering, University of Cambridge, CB2 1PZ Cambridge, U.K., and also with the Department of Signal Theory and Communications, Universitat Politècnica de Catalunya, 08034 Barcelona, Spain (e-mail: guillen@ieee.org).

Communicated by S. Yang, Associate Editor for Communications.

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TIT.2025.3574460>.

Digital Object Identifier 10.1109/TIT.2025.3574460

probability distribution is $W(y|x) = \Pr[Y = y|X = x]$ for all pairs $(x, y) \in \mathcal{X} \times \mathcal{Y}$. For transmission, the encoder selects a message $m \in \{1, \dots, M\}$ and transmits the corresponding n -symbol codeword $\mathbf{x}^{(m)} = (x_1^{(m)}, \dots, x_n^{(m)})$ from the codebook $\mathcal{C}_n = \{\mathbf{x}^{(i)}\}_{1 \leq i \leq M}$. Upon observing $\mathbf{y} \in \mathcal{Y}^n$, the decoder estimates the transmitted message, inspired by the maximum-likelihood decoder but using a possibly suboptimal decoding metric $q^n(\mathbf{x}, \mathbf{y})$, as

$$\hat{m} = \operatorname{argmax}_{1 \leq \tilde{m} \leq M} q^n(\mathbf{x}^{(\tilde{m})}, \mathbf{y}). \quad (1)$$

When $q^n(\mathbf{x}, \mathbf{y}) \equiv \prod_{i=1}^n W(y_i|x_i)$, the decoder is the optimal maximum-likelihood (ML) decoder and is commonly referred to as matched. The symbol \equiv denotes equivalence in decoding metric in the sense of [4, Propositions 2.1 and 2.2]. In any other case, the decoder is said to be mismatched. An error is declared when $\hat{m} \neq m$, and the probability of error for \mathcal{C}_n is defined as $p_e(\mathcal{C}_n) = \Pr[\hat{m} \neq m]$. In this work, we will work with product metrics for which we adopt the symbol metric notation $q(x, y)$ and

$$\hat{m} = \operatorname{argmax}_{1 \leq \tilde{m} \leq M} \prod_{i=1}^n q(x_i^{(\tilde{m})}, y_i). \quad (2)$$

After taking logarithms, these become additive metrics.

We consider the relevant case in practice where the decoding metric $q(x, y) = \widehat{W}(y|x) > 0$ is a channel estimate $\widehat{W}(y|x)$ from the output of a channel estimator. We analyze achievable rates and error exponents for an unknown channel $W(y|x)$ and a mismatched decoder that uses $\widehat{W}(y|x)$ as if it were perfect, and impose a constraint on the level of mismatch between the estimated and true channels by defining an appropriate statistical distance. More concretely, let $d(P, P')$ be a statistical distance between the joint probability distributions P and P' over the alphabet $\mathcal{X} \times \mathcal{Y}$. Then we define *small mismatch* as the class of channels [8] such that

$$W \in \{P_{Y|X} : d(Q_X P_{Y|X}, Q_X \widehat{W}) \leq r\}. \quad (3)$$

That is, W is within a ball centered at \widehat{W} of radius r , assumed to be small. This class of channels is useful for modeling small perturbations of the channel around the decoding metric due to good, yet imperfect estimation.

C. Overview of Achievable Rates

A number of achievable rates have been derived in the literature for DMCs and continuous channels with an arbitrary fixed decoding metric $q(x, y) > 0$ (see e.g. [4] for a recent survey). The GMI (generalized mutual information) I_{GMI} [9] is the achievable rate obtained with i.i.d. random codes. Instead, employing constant-composition codes for DMCs or cost-constrained codes with a single per-codeword cost constraint for arbitrary alphabets yields the LM (lower mismatch) rate I_{LM} [10], [11]. In this work, we focus on the dual expressions of the aforementioned rates defined from the *mismatched information density*

$$i_{s,a}(x, y) \triangleq \log \frac{q(x, y)^s e^{a(x)}}{\mathbb{E}_{Q_X}[q(X, y)^s e^{a(X)}]} \quad (4)$$

where $s \geq 0$ and $a(x) \in \mathbb{R}$ for $x \in \mathcal{X}$. Therefore, as [4, Ch. 2.3.1, Ch. 2.3.2]

$$I_{\text{GMI}}(Q_X) \triangleq \sup_{s \geq 0} \mathbb{E}_{Q_X \times W} [i_{s,0}(X, Y)] \quad (5)$$

$$I_{\text{LM}}(Q_X) \triangleq \sup_{s \geq 0, a(\cdot)} \mathbb{E}_{Q_X \times W} [i_{s,a}(X, Y)], \quad (6)$$

where the notation $i_{s,0}(x, y)$ in (5) corresponds to setting $a(x) = 0$ in (4) for every $x \in \mathcal{X}$. Indeed, the GMI rate can be obtained from the LM rate by setting the function $a(x)$ to zero for every input. The GMI and LM rates are tight with respect to their respective ensemble of codes. Thus, the weakness in rate relative to the mismatch capacity C_M , the supremum of all achievable rates, is due to an inherent weakness of employed codes rather than loose mathematical analysis. Hence $I_{\text{GMI}}(Q_X) \leq I_{\text{LM}}(Q_X) \leq C_M$, with equality in specific cases [12], [13].

D. Overview of Error Exponents

An error exponent $E(R)$ is said to be achievable at rate R whenever

$$E_r(R) \leq \lim_{n \rightarrow \infty} -\frac{1}{n} \log \bar{p}_e(n, \lfloor e^{nR} \rfloor), \quad (7)$$

where \bar{p}_e is the average error probability over all randomly generated codes of a specific ensemble. If (7) holds with equality, $E_r(R)$ is said to be the ensemble-tight exponent [4, Ch. 7.1]. For ML decoding,

$$E_r(Q_X, R) \triangleq \max_{\rho \in [0,1]} E_0(Q_X, \rho) - \rho R \quad (8)$$

is the (ensemble-tight) random coding exponent [1] achievable at rate R with i.i.d and constant-composition codes, where the respective Gallager E_0 functions are defined from

$$\varepsilon_{a,\rho}(x, y) \triangleq \left(\frac{\mathbb{E}_{Q_X} [W(y|X)^{\frac{1}{1+\rho}} e^{a(X)}]}{W(y|x)^{\frac{1}{1+\rho}} e^{a(x)}} \right)^\rho \quad (9)$$

as

$$E_0^{\text{iid}}(Q_X, \rho) \triangleq -\log \mathbb{E}_{Q_X \times W} [\varepsilon_{0,\rho}(X, Y)] \quad (10)$$

$$E_0^{\text{cc}}(Q_X, \rho) \triangleq \sup_{a(\cdot)} -\mathbb{E}_{Q_X} [\log \mathbb{E}_W [\varepsilon_{a,\rho}(X, Y)|X]]. \quad (11)$$

It is known that the error exponent $E_r(Q_X, R)$ is positive for rates $R < \frac{\partial E_0(Q_X, \rho)}{\partial \rho} \Big|_{\rho=0} = I(X; Y)$, proving the achievability part of the channel coding theorem. Moreover, $E_0^{\text{iid}}(Q_X, \rho) \geq E_0^{\text{cc}}(Q_X, \rho)$ with equality only for the respective optimal input distributions [14, Eq. (52)].

Analogous to the matched case, the ensemble-tight mismatched random coding error exponents [4, Th. 7.1] of i.i.d. and constant-composition codes are defined using

$$\varepsilon_{s,a,\rho}(x, y) \triangleq \left(\frac{\mathbb{E}_{Q_X} [q(X, y)^s e^{a(X)}]}{q(x, y)^s e^{a(x)}} \right)^\rho \quad (12)$$

to obtain the following dual expressions [4, Ch. 7.2]:

$$E_r^{\text{iid}}(Q_X, R) \triangleq \max_{\rho \in [0,1]} E_0^{\text{iid}}(Q_X, \rho) - \rho R \quad (13)$$

$$E_r^{\text{cc}}(Q_X, R) \triangleq \max_{\rho \in [0,1]} E_0^{\text{cc}}(Q_X, \rho) - \rho R, \quad (14)$$

where the respective mismatched Gallager functions are

$$E_0^{\text{iid}}(Q_X, \rho) \triangleq \sup_{s \geq 0} -\log \mathbb{E}_{Q_X \times W}[\varepsilon_{s,0,\rho}(X, Y)] \quad (15)$$

$$E_0^{\text{cc}}(Q_X, \rho) \triangleq \sup_{s \geq 0, a(\cdot)} -\mathbb{E}_{Q_X}[\log \mathbb{E}_W[\varepsilon_{s,a,\rho}(X, Y)|X]]. \quad (16)$$

For simplicity, we have used the same notation for the mismatched error exponents as for the matched case.

An alternative expression for E_0^{cc} can be formulated by observing that constant-composition and cost-constrained coding achieve the same exponent when two cost functions $c_1(\cdot)$ and $c_2(\cdot)$ are optimized [14, Th. 4]. This is

$$E_0^{\text{cc}}(Q_X, \rho) = \sup_{c_1(\cdot), c_2(\cdot)} E_0^{\text{cost}}(Q_X, \rho, \{c_1, c_2\}) \quad (17)$$

where

$$E_0^{\text{cost}}(Q_X, \rho, \mathbf{c}) \triangleq \sup_{s \geq 0, \lambda, \bar{\lambda}} -\log \mathbb{E}_{Q_X \times W}[\varepsilon_{s,\lambda,\bar{\lambda},\rho}^{\text{cost}}(X, Y)] \quad (18)$$

$$\varepsilon_{s,\lambda,\bar{\lambda},\rho}^{\text{cost}}(x, y) \triangleq \left(\frac{\mathbb{E}_{Q_X}[q(X, y)^s e^{\bar{\lambda}(c(X) - \phi)}]}{q(x, y)^s e^{\lambda(c(x) - \phi)}} \right)^\rho \quad (19)$$

and $\mathbf{c}(x) \triangleq [c_1(x), c_2(x)]^T$, $\lambda \triangleq [\lambda_1, \lambda_2]^T$, $\bar{\lambda} \triangleq [\bar{\lambda}_1, \bar{\lambda}_2]^T$ and $\phi \triangleq [\mathbb{E}[c_1(X)], \mathbb{E}[c_2(X)]]^T$ defined in vector form for ease of notation. We are more interested in the expression (17) for the mismatched E_0 function of constant-composition codes, as it adopts a similar form to that of i.i.d. codes, leading to tighter derivations. In addition, cost-constrained coding remains valid for continuous alphabets. When a matched ML decoder $q(x, y) = W(y|x)$ is used, the matched Gallager E_0 functions (10)–(11), are recovered from (15)–(17) by setting $s = (1 + \rho)^{-1}$, hence recovering the matched error exponents. Using Jensen's inequality [15, Ch. 2], it can be shown that for any pair (W, q)

$$E_r^{\text{iid}}(Q_X, R) \leq E_r^{\text{cc}}(Q_X, R). \quad (20)$$

E. Contributions and Paper Outline

Existing work on mismatched decoding typically only considers channel uncertainty at the receiver for an arbitrary channel metric pair. This paper explores achievable rates and random coding error exponents under channel uncertainty for a class of channels that are close to the decoding metric, modeling the case of accurate yet imperfect channel estimation. This scenario is described as the compound channel [16], [17], where the only knowledge available is the class of channels over which the system will operate. The novelty of this paper lies in analyzing the compound channel under a given mismatched decoding metric. We parameterize the compound channel class by all the channels that lie within a relative entropy ball of radius r . Our primary contribution is the sensitivity analysis of achievable rates and random coding error exponents to small values of r , capturing the effect of accurate but imperfect channel estimation. We also assess the boundary between matched and mismatched decoding, providing novel closed-form results for symmetric decoding metrics and the nearest neighbor decoder. In particular:

- 1) In Section III, we derive worst-case information rates for i.i.d. and constant-composition random codes. We

derive an expansion of the worst-case information rates for small mismatch r and show that the mutual information of the estimated channel cannot be surpassed. Our expansion shows second-order terms proportional to \sqrt{r} , inducing an achievable rate penalty that decays with an infinite slope with respect to the mutual information of the estimated channel achieved at $r = 0$. Examples of symmetric and modulo-additive decoding metrics are presented, demonstrating that the worst-case channel within the ball shares the same structure as the decoding metric.

- 2) In Section IV, we derive worst-case error exponents for i.i.d. and constant-composition random codes. We provide the corresponding expansions with second-order penalty terms again proportional to \sqrt{r} . Examples for the same symmetric and modulo-additive decoding metrics used for achievable rates are provided, yielding similar results.
- 3) In Section V, we extend previous analyses allowing for continuous input and output alphabets. We provide expansions for the worst-case achievable rates and error exponents for i.i.d. and cost-constrained random coding with multiple auxiliary costs. The results are particularized to Gaussian codebooks with nearest neighbor decoding [18]. In particular, we derive a second-order expansion of the worst-case achievable rates for arbitrary channels, not necessarily additive, in a small-radius relative entropy ball around the nearest neighbor decoding metric. We also show a significantly smaller penalty when the channels within the ball are assumed to be additive. The structure of the worst-case channels is also studied, highlighting the discontinuities observed when the estimated channel has unbounded support.

Conclusions are drawn in Section VI. Proofs are found in the Appendices.

II. SMALL MISMATCH: A CLASS OF COMPOUND CHANNELS

We study a class of mismatched compound channels for which DMC $W(y|x)$ is unknown but assumed to lie within a divergence ball \mathcal{B} centered on $\hat{W}(y|x)$, a conditional probability distribution that models an estimate of $W(y|x)$ from the output of a channel estimator. The decoder uses it as the decoding metric, i.e., $q(x, y) = \hat{W}(y|x) > 0$.

Definition 1: For small mismatch (ball radius) between the true and estimated channels W and \hat{W} , we require that

$$W \in \mathcal{B}(\hat{W}, r) \triangleq \{P_{Y|X} : D(\hat{W} \| P_{Y|X} | Q_X) \leq r\} \quad (21)$$

holds for small r , where $D(\hat{W} \| P_{Y|X} | Q_X)$ is the conditional relative entropy

$$D(\hat{W} \| P_{Y|X} | Q_X) \triangleq \sum_{x,y} Q_X(x) \hat{W}(y|x) \log \frac{\hat{W}(y|x)}{P_{Y|X}(y|x)}. \quad (22)$$

Eq. (21) is equivalent to setting $D(Q_X \hat{W} \| Q_X P_{Y|X}) \leq r$.

We have adopted a decoder-centric perspective in which the ball is centered around the known quantity, i.e. the estimated

channel \widehat{W} employed to decode. One of the advantages of this formulation is that for sufficiently small mismatch r we can rely on [5, Eqs. (1)–(4)] and apply a Taylor expansion to the logarithm in the relative entropy and group the $o(\cdot)$ terms. This allows us to express the relative entropy as a function of

$$\theta(y|x) \triangleq P_{Y|X}(y|x) - \widehat{W}(y|x) \quad (23)$$

plus a term of lesser significance, as

$$D(\widehat{W} \| P_{Y|X} | Q_X) = \frac{1}{2} \sum_{x,y} Q_X(x) \frac{\theta(y|x)^2}{\widehat{W}(y|x)} + o\left(\frac{1}{2} \sum_{x,y} Q_X(x) \frac{\theta(y|x)^2}{\widehat{W}(y|x)}\right). \quad (24)$$

Observe that the first term of (24) is exactly the chi-squared distance $\frac{1}{2} \chi^2(Q_X P_{Y|X}, Q_X \widehat{W})$ as defined in [19, Pag. 425]. Vectorizing θ allows us to express (24) in a more convenient form to obtain closed form results. Besides this, there are other statistical distances to consider, such as generic f -divergences [20], [21] or the Rényi divergence of order α [22]. Here, we only consider relative entropy, an f -divergence with $f''(1) = 1$ (f'' stands for the second derivative of f). Our results can be easily extended to a generic f -divergence or Rényi entropy by substituting the $\frac{1}{2}$ by the constant factor $\frac{f''(1)}{2}$ for f -divergences or $\frac{\alpha}{2}$ for the Rényi divergence.

III. ACHIEVABLE RATES

In this section, we derive the worst-case LM and GMI rates. For ease of explanation, we focus on the worst-case LM rate as the corresponding GMI can be obtained straightforwardly by setting $a(x) = 0$ for all $x \in \mathcal{X}$. We rewrite the mismatched information density $i_{s,a}$ in (4) with $q(x, y) = \widehat{W}(y|x)$ as

$$i_{s,a}(x, y) = \log \frac{\widehat{W}(y|x)^s e^{a(x)}}{\mathbb{E}_{Q_X}[\widehat{W}(y|X)^s e^{a(X)}]}. \quad (25)$$

The worst-case rate problem we wish to solve is

$$I_{\text{LM}}(Q_X, \widehat{W}, r) = \min_{W \in \mathcal{B}} \sup_{s \geq 0, a(\cdot)} \mathbb{E}_{Q_X \times W} [i_{s,a}(X, Y)] \quad (26)$$

where the minimization is over all valid conditional probability distributions in the relative entropy ball \mathcal{B} defined in (21). Since the true channel is unknown, problem (26) finds the channel in \mathcal{B} that gives the worst possible LM rate. This gives an indication of the rate loss incurred by accurate but not perfect channel estimation.

Channels $W \in \mathcal{B}$ need to be valid conditional probability distributions, i.e.,

$$W(y|x) \geq 0 \quad \text{for } (x, y) \in \mathcal{X} \times \mathcal{Y} \quad (27)$$

$$\sum_y W(y|x) = 1 \quad \text{for } x \in \mathcal{X}. \quad (28)$$

The constraint $W(y|x) \leq 1$ for $(x, y) \in \mathcal{X} \times \mathcal{Y}$ is guaranteed by construction when (27)–(28) are satisfied. Next, observe that imposing (27) in problem (26) is unnecessary since the relative entropy in (21) implies that positivity already for a strictly positive \widehat{W} .

Theorem 1: Consider a class of DMCs $W(y|x)$, a mismatched decoder based on the channel estimate $\widehat{W}(y|x)$ and a fixed input

distribution $Q_X(x)$ satisfying (21). Then, for a sufficiently small $r \geq 0$, the worst-case LM and GMI rates can be expressed as

$$I_{\text{LM}}(Q_X, \widehat{W}, r) = \sup_{s \geq 0, a(\cdot)} \left\{ I_{s,a}^{\text{ML}}(Q_X, \widehat{W}) - \sqrt{2r V_{s,a}(Q_X, \widehat{W})} + o(\sqrt{r}) \right\} \quad (29)$$

$$I_{\text{GMI}}(Q_X, \widehat{W}, r) = \sup_{s \geq 0} \left\{ I_{s,0}^{\text{ML}}(Q_X, \widehat{W}) - \sqrt{2r V_{s,0}(Q_X, \widehat{W})} + o(\sqrt{r}) \right\} \quad (30)$$

where

$$I_{s,a}^{\text{ML}}(Q_X, \widehat{W}) \triangleq \mathbb{E}_{Q_X \times \widehat{W}} [i_{s,a}(X, Y)] \quad (31)$$

$$V_{s,a}(Q_X, \widehat{W}) \triangleq \mathbb{E}_{Q_X} [\text{Var}_{\widehat{W}} [i_{s,a}(X, Y) | X]]. \quad (32)$$

Proof: The full proof is provided in Appendix A, where we minimize the dual expression for the LM rate (26) including the $o(\cdot)$ term in (24) as a penalty in the objective function. This gives an accurate approximation to relative entropy [5] and an accurate approximation on I_{LM} as $r \rightarrow 0$. The worst-case GMI rate is found by setting $a(x) = 0$ for $x \in \mathcal{X}$.

The following comments are in order from inspection of problem (26), Theorem 1 and its proof in Appendix A:

- 1) *The solution for a chi-squared distance ball.* Let the small mismatch between the true and estimated channels W and \widehat{W} be characterized by a small radius chi-squared distance ball as

$$W \in \tilde{\mathcal{B}}(\widehat{W}, r) \triangleq \left\{ P_{Y|X} : \frac{1}{2} \chi^2(Q_X P_{Y|X}, Q_X \widehat{W}) \leq r \right\} \quad (33)$$

where $\frac{1}{2} \chi^2(Q_X P_{Y|X}, Q_X \widehat{W})$ as defined in [19, Pag. 425] corresponds to the first term in (24). Then, the worst-case LM and GMI rates for sufficiently small $r \geq 0$ are exactly

$$\tilde{I}_{\text{LM}}(Q_X, \widehat{W}, r) = \sup_{s \geq 0, a(\cdot)} \left\{ I_{s,a}^{\text{ML}}(Q_X, \widehat{W}) - \sqrt{2r V_{s,a}(Q_X, \widehat{W})} \right\} \quad (34)$$

$$\tilde{I}_{\text{GMI}}(Q_X, \widehat{W}, r) = \sup_{s \geq 0} \left\{ I_{s,0}^{\text{ML}}(Q_X, \widehat{W}) - \sqrt{2r V_{s,0}(Q_X, \widehat{W})} \right\} \quad (35)$$

achieved for every fixed $s, a(x)$ by the channel distribution

$$\tilde{W}_{s,a}^*(y|x) = \widehat{W}(y|x) \left(1 - \sqrt{2r} \cdot \varphi_{s,a}(x, y) \right) \quad (36)$$

with

$$\varphi_{s,a}(x, y) \triangleq \frac{i_{s,a}(x, y) - \mathbb{E}_{\widehat{W}} [i_{s,a}(x, Y)]}{\sqrt{V_{s,a}(Q_X, \widehat{W})}}. \quad (37)$$

The worst-case channels $\tilde{W}_{\text{LM}}^*, \tilde{W}_{\text{GMI}}^*$ are found after maximizing (34)–(35) over $s, a(x)$. The key observation here is that the difference in the worst-case achievable rate for the mismatched compound channel with relative entropy or chi-squared distance balls is merely a little- o term.

- 2) *Feasibility of worst-case channels.* In our proof, by shifting the $o(\cdot)$ term from the constraint to the objective function as a penalty term, we derive a more convenient

and tractable mathematical analysis that yields valuable insights. However, one consequence of this shift is that the worst-case channels may have negative values, though this effect is absent for sufficiently small r . An additional condition on the ball radius r must be satisfied to guarantee non-negativeness. From $1 - \sqrt{2r} \cdot \varphi_{s,a}(x, y) > 0$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have:

$$r < \min_{(x,y): \varphi_{s,a}(x,y)>0} \frac{1}{2\varphi_{s,a}(x,y)^2}. \quad (38)$$

$\tilde{W}_{s,a}^*(y|x) \leq 1$ for $(x, y) \in \mathcal{X} \times \mathcal{Y}$ is satisfied by construction as (38) guarantees non-negativity and $\sum_y \tilde{W}_{s,a}^*(y|x) = 1$. It is easy to show that the right-hand side of (38) produces a positive value. We simply need to show that $\varphi_{s,a}(x, y)$ is finite everywhere $\hat{W}(y|x) > 0$, which can be easily shown by referring to equation (37). The only situation in which this condition would not be fulfilled is if $V_{s,a} = 0$, which occurs only when $\hat{W}(y|x) = |\mathcal{Y}|^{-1}$, and for which $\tilde{I}_{LM} = 0$. Therefore, it exists a sufficiently small r that fulfills (38). Indeed, in all practical settings we have considered, the bound (38) exceeds the ball radius of interest for the optimal $s, a(x)$ corresponding to the worst-case channels.

- 3) *The solution.* The worst-case channels are always found at the border of the respective ball, i.e., when (21) or (33) holds with equality.
- 4) *Expansion term.* The expansions yield the same penalty term proportional to \sqrt{r} , the square root of the ball radius. These penalties indicate that even a small mismatch causes a severe penalty on the achievable rate.
- 5) *Achievable rate at $r = 0$.* For a fixed \hat{W} , I_{LM} (29) is upper bounded by the mutual information of the estimated channel \hat{W} with input Q_X , $I_{MI}(Q_X, \hat{W})$. This is because $\hat{W} \in \mathcal{B}$ regardless of any $r \geq 0$, so the minimum in (26) can be upper bounded by evaluating at $W = \hat{W}$. Hence

$$I_{LM}(Q_X, \hat{W}, r) \leq \sup_{s \geq 0, a(\cdot)} I_{s,a}^{ML}(Q_X, \hat{W}) \quad (39)$$

$$= I_{MI}(Q_X, \hat{W}), \quad (40)$$

and thus, information rates surpassing $I_{MI}(Q_X, \hat{W})$ cannot be achieved. (39) is tight at $r = 0$, where $\mathcal{B} = \{\hat{W}\}$ and the supremum is achieved for $s = 1$ and $a(x) = 0$ for $x \in \mathcal{X}$ [14, Sec. IV-C]. The same conclusions can be extracted for \tilde{I}_{LM} defined in (34).

- 6) *Input-maximizing distribution.* Analysis of (29) reveals that although the first term is concave in Q_X , the second term is not, though it is comparatively much smaller than the first, especially when $r \rightarrow 0$. In this regime, $s = 1$ and $a(x) = 0$ are the optimal values, and therefore

$$\max_{Q_X} I_{LM} \approx I_{1,0}(Q_X^*, \hat{W}) - \sqrt{2rV_{1,0}(Q_X^*, \hat{W})} \quad (41)$$

where $I_{1,0}(Q_X, \hat{W}) = I_{MI}(Q_X, \hat{W})$ and

$$Q_X^* \triangleq \operatorname{argmax}_{Q_X} I_{MI}(Q_X, \hat{W}) \quad (42)$$

is the capacity-achieving distribution for channel \hat{W} , which can be computed in a complexity efficient way via

the Blahut-Arimoto algorithm [23]. The approximation becomes tighter as $r \rightarrow 0$, and holds with equality only at $r = 0$, in which case

$$\max_{Q_X} I_{LM}(Q_X, \hat{W}, 0) = \max_{Q_X} I_{MI}(Q_X, \hat{W}). \quad (43)$$

The same conclusions can be extracted for \tilde{I}_{LM} in (34).

- 7) *Lower bound on \tilde{I}_{LM} and \tilde{I}_{GMI} .* The worst-case rates (34)–(35) can be lower-bounded by setting $s = 1$ and $a(x) = 0$ for every input as

$$\tilde{I}_{LM}(Q_X, \hat{W}, r) \geq \tilde{I}_{GMI}(Q_X, \hat{W}, r) \quad (44)$$

$$\geq I_{MI}(Q_X, \hat{W}) - \sqrt{2rV_{1,0}(Q_X, \hat{W})}. \quad (45)$$

As $r \rightarrow 0$ the penalty term shrinks until $I_{MI}(Q_X, \hat{W})$ is achieved. The bound is tight at $r = 0$. It allows the evaluation of the gain of introducing $s \geq 0$ before applying Markov's inequality in the achievability proof [4].

A. Example: Symmetric \hat{W} and Equiprobable Q_X

We derive the worst-case GMI for a symmetric \hat{W} and a equiprobable input distribution $Q_X(x) = |\mathcal{X}|^{-1}$ for $x \in \mathcal{X}$. This combination could be used when W is close to being symmetric. Symmetry is defined as Gallager [1, Pag. 94], i.e. when each row or column of \hat{W} is a permutation of each other row or column, respectively. We show the final results here; the complete derivations are found in Appendix C.

The worst-case GMI in (35) can be simplified due to the symmetry of \hat{W} . Specifically, the magnitudes in (35) can be expressed using the first row of \hat{W} , $\hat{W}_{\text{sym}}(y) \triangleq W(y|x_1)$, as

$$I_{s,0}^{ML}(Q_X, \hat{W}) = \log \frac{|\mathcal{X}|}{\sum_y \hat{W}_{\text{sym}}(y)^s} - sH(\hat{W}_{\text{sym}}) \quad (46)$$

$$V_{s,0}(Q_X, \hat{W}) = s^2 \left(\mathbb{E}_{\hat{W}_{\text{sym}}} [\log^2 \hat{W}_{\text{sym}}] - H^2(\hat{W}_{\text{sym}}) \right) \quad (47)$$

where $H(\hat{W}_{\text{sym}})$ is the entropy of the random variable characterized by the probability mass function \hat{W}_{sym} .

Next, we present two additional insightful results. One of the most curious findings of this case study is that the worst-case channel in the chi-squared distance ball $\tilde{\mathcal{B}}$ defined in (33) has the same structure as \hat{W} , i.e. \tilde{W}_{GMI}^* is also a symmetric DMC and is independent of s . This suggests two key points. Firstly, adding $a(x)$ does not offer any advantage, and thus i.i.d. and constant-composition achieve the same rate, as for output-symmetric channels and metrics [4, Ch. 2.4.2]. Secondly, for a scenario where the channel is known to be symmetric but unknown, the worst-case channel in $\tilde{\mathcal{B}}$, when considering only channels with the same structure as \hat{W} , leads to the same solution we obtained.

The next result arises from applying the bounds in point 7) after Theorem 1. Specifically, the worst-case GMI rate can be lower bounded by setting $s = 1$ to yield

$$\tilde{I}_{GMI}(Q_X, \hat{W}, r) \geq C(\hat{W}) - \sqrt{2r\text{Var}_{\hat{W}_{\text{sym}}}[\log \hat{W}_{\text{sym}}]}, \quad (48)$$

where $C(\hat{W}) \triangleq \log |\mathcal{X}| - H(\hat{W}_{\text{sym}})$ is the capacity of DMC \hat{W} . The bound becomes tighter as $r \rightarrow 0$.

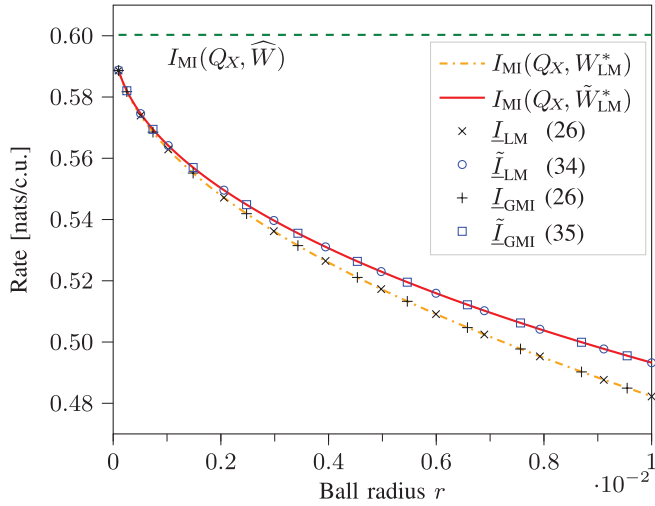


Fig. 1. Information rates (in nats per channel use) computed for Q_X in (53) and estimated channel \hat{W} in (54).

B. Example: Modulo-Additive \hat{W} and Equiprobable Q_X

Further results can be obtained even in a more closed form for the specific case of a modulo-additive \hat{W} by particularizing (46)–(47) for the channel estimate

$$\hat{W} = \begin{bmatrix} \bar{p} & p & \cdots & p \\ p & \bar{p} & \cdots & p \\ \vdots & \vdots & \ddots & \vdots \\ p & p & \cdots & \bar{p} \end{bmatrix} \quad (49)$$

with $p \in [0, |\mathcal{X}|^{-1}]$ and $\bar{p} \triangleq 1 - (|\mathcal{X}| - 1)p \geq p$ defined for compactness of notation.

In this case, we have

$$I_{s,0}^{\text{ML}}(Q_X, \hat{W}) = \log \frac{|\mathcal{X}|}{\bar{p}^s + (1 - \bar{p})p^{s-1}} + s(\bar{p} \log \bar{p} + (1 - \bar{p}) \log p) \quad (50)$$

$$V_{s,0}(Q_X, \hat{W}) = s^2 \cdot \bar{p}(1 - \bar{p}) \log^2(\bar{p}/p) \quad (51)$$

and the worst-case channel in $\tilde{\mathcal{B}}$ is a modulo-additive DMC with crossover probabilities

$$q = p \left(1 + \sqrt{2r} \cdot \sqrt{\frac{\bar{p}}{1 - \bar{p}}} \right). \quad (52)$$

The feasibility condition (38) leads to $r < \frac{1}{2} \frac{\bar{p}}{1 - \bar{p}}$. The worst-case bound $r < \frac{1}{2} \frac{1}{|\mathcal{X}| - 1}$ is achieved when $p = \frac{1}{|\mathcal{X}|}$.

C. Example: Ternary-Input Ternary-Output \hat{W}

We compute the worst-case LM and GMI rates $\tilde{I}_{\text{LM}}, \tilde{I}_{\text{GMI}}$ from (34)–(35), using the following input distribution Q_X and ternary-input ternary-output estimated channel \hat{W} :

$$Q_X = [0.3 \ 0.3 \ 0.4] \quad (53)$$

$$\hat{W} = \begin{bmatrix} 0.85 & 0.05 & 0.10 \\ 0.15 & 0.825 & 0.025 \\ 0.025 & 0.10 & 0.875 \end{bmatrix}. \quad (54)$$

We plot them in Fig. 1 along with \tilde{I}_{LM} and \tilde{I}_{GMI} numerically computed from (26) using an off-the-shelf solver.

$I_{\text{ML}}(Q_X, \hat{W})$ is shown in Fig. 1 for reference. It is achievable as $r \rightarrow 0$. We also plot the mutual information of the worst-case channels obtained from the solver, W_{LM}^* , and from our approximation, \tilde{W}_{LM}^* . The approximation is very accurate for r sufficiently small, thus validating our approach for the regime of interest, i.e. small mismatch r . For $r > 0$ the curves decrease from the reference $I_{\text{ML}}(Q_X, \hat{W})$, as supported by point 5) after Theorem 1. Specifically, both the true and approximate worst-case rates decrease with an infinitely negative gradient at $r = 0$. This behavior arises because the worst-case channel in the relative entropy ball is such that it causes the respective LM and GMI rate to decrease with an infinite slope. This indicates that even a small mismatch can have a significant impact on the achievable rates. The feasibility of the worst-case channels (38) are determined for $r < 2.35$, thus confirming the validity of our approach for these simulations.

The mutual information for the worst-case channels, illustrated by solid and dash-dotted lines, also exhibits the aforementioned behavior as the worst-case channels have a linear dependence on \sqrt{r} . Simulations also show that they are very close to the achievable rates we derived for the mismatched compound channel. This strongly suggests that the achievable rate for the compound channel coincides, within the limits of numerical accuracy, with the maximum rate achievable when communication is over the worst-case channel within the ball and the decoder is matched.

IV. ERROR EXPONENTS

In this section, we study the worst-case mismatched random coding error exponents

$$\underline{E}_r^{\text{iid}}(Q_X, \hat{W}, R, r) = \min_{W \in \mathcal{B}} \max_{\rho \in [0,1]} E_0^{\text{iid}}(Q_X, W, \rho) - \rho R \quad (55)$$

$$\underline{E}_r^{\text{cc}}(Q_X, \hat{W}, R, r) = \min_{W \in \mathcal{B}} \max_{\rho \in [0,1]} E_0^{\text{cc}}(Q_X, W, \rho) - \rho R \quad (56)$$

where

$$E_0^{\text{iid}}(Q_X, W, \rho) \triangleq \sup_{s \geq 0} E_{s,0,\mathbf{0},\rho}(Q_X, W) \quad (57)$$

$$E_0^{\text{cc}}(Q_X, W, \rho) \triangleq \sup_{c_1(\cdot), c_2(\cdot)} \sup_{s \geq 0, \lambda, \bar{\lambda}} E_{s,\lambda,\bar{\lambda},\rho}(Q_X, W) \quad (58)$$

and

$$E_{s,\lambda,\bar{\lambda},\rho}(Q_X, W) \triangleq -\log \mathbb{E}_{Q_X \times W}[\varepsilon_{s,\lambda,\bar{\lambda},\rho}(X, Y)], \quad (59)$$

where we have redefined the *mismatched exponent density* as a function of the channel estimate as

$$\varepsilon_{s,\lambda,\bar{\lambda},\rho}(x, y) \triangleq \left(\frac{\mathbb{E}_{Q_X}[\hat{W}(y|x)^s e^{\lambda \mathcal{I}(c(x)-\phi)}]}{\hat{W}(y|x)^s e^{\bar{\lambda} \mathcal{I}(c(x)-\phi)}} \right)^\rho. \quad (60)$$

We aim to find the worst-case mismatched random coding error exponents $\underline{E}_r^{\text{iid}}$ and $\underline{E}_r^{\text{cc}}$ subject to Definition 1. The following lemma serves as a preliminary to our main theorems.

Lemma 1: Finding the worst-case mismatched random coding error exponent is equivalent to first finding its corresponding worst-case Gallager function, i.e.,

$$\underline{E}_r(Q_X, \hat{W}, R, r) = \max_{\rho \in [0,1]} \min_{W \in \mathcal{B}} E_0(Q_X, W, \rho) - \rho R. \quad (61)$$

Proof: The worst-case error exponent $\underline{E}_r(Q_X, \widehat{W}, R, r)$ is

$$\underline{E}_r(Q_X, \widehat{W}, R, r) = \min_{W \in \mathcal{B}} \max_{\rho \in [0,1]} E_0(Q_X, W, \rho) - \rho R \quad (62)$$

$$= \max_{\rho \in [0,1]} \min_{W \in \mathcal{B}} E_0(Q_X, W, \rho) - \rho R, \quad (63)$$

where, as the problem is concave in $\rho \in [0, 1]$ and convex in $W \in \mathcal{B}$, the minimax theorem [24, Th. 1] has been used to swap the order of the optimizations from (62) to (63). Fixing (Q_X, R) and noticing that ρR is independent of W , the problem is then equivalent to minimizing E_0 prior to maximizing over $\rho \in [0, 1]$. ■

Consequently, we focus on deriving the worst-case mismatched Gallager E_0 functions for i.i.d. and constant-composition codes

$$\underline{E}_0^{\text{iid}}(Q_X, \widehat{W}, \rho, r) = \min_{W \in \mathcal{B}} \sup_{s \geq 0} E_{s,0,0,\rho}(Q_X, W) \quad (64)$$

$$\underline{E}_0^{\text{cc}}(Q_X, \widehat{W}, \rho, r) = \min_{W \in \mathcal{B}} \sup_{c_1(\cdot), c_2(\cdot)} \sup_{s \geq 0, \lambda, \bar{\lambda}} E_{s,\lambda,\bar{\lambda},\rho}(Q_X, W). \quad (65)$$

We describe the results together in the following theorem. The corresponding worst-case error exponents need to be computed by maximizing over $\rho \in [0, 1]$ afterwards, but they are not included herein for ease of readability.

Theorem 2: Consider a class of DMCs $W(y|x)$, a mismatched decoder based on the channel estimate $\widehat{W}(y|x)$ and a fixed input distribution $Q_X(x)$ satisfying (21). Then, for sufficiently small $r \geq 0$, the worst-case mismatched Gallager function for i.i.d. and constant-composition random coding can be expressed as

$$\underline{E}_0^{\text{iid}}(Q_X, \widehat{W}, \rho, r) = \sup_{s \geq 0} \left\{ E_{s,0,0,\rho}^{\text{ML}}(Q_X, \widehat{W}) - \log \left(1 + \sqrt{2r V_{s,0,0,\rho}(Q_X, \widehat{W})} \right) + o(\sqrt{r}) \right\} \quad (66)$$

$$\underline{E}_0^{\text{cc}}(Q_X, \widehat{W}, \rho, r) = \sup_{c_1(\cdot), c_2(\cdot)} \sup_{s \geq 0, \lambda, \bar{\lambda}} \left\{ E_{s,\lambda,\bar{\lambda},\rho}^{\text{ML}}(Q_X, \widehat{W}) - \log \left(1 + \sqrt{2r V_{s,\lambda,\bar{\lambda},\rho}(Q_X, \widehat{W})} \right) + o(\sqrt{r}) \right\} \quad (67)$$

where

$$E_{s,\lambda,\bar{\lambda},\rho}^{\text{ML}}(Q_X, \widehat{W}) \triangleq -\log \mathbb{E}_{Q_X \times \widehat{W}} [\varepsilon_{s,\lambda,\bar{\lambda},\rho}(X, Y)] \quad (68)$$

$$V_{s,\lambda,\bar{\lambda},\rho}(Q_X, \widehat{W}) \triangleq \frac{\mathbb{E}_{Q_X} [\text{Var}_{\widehat{W}} [\varepsilon_{s,\lambda,\bar{\lambda},\rho}(X, Y) | X]]}{\mathbb{E}_{Q_X \times \widehat{W}}^2 [\varepsilon_{s,\lambda,\bar{\lambda},\rho}(X, Y)]}. \quad (69)$$

Proof: The proof is found in Appendix B and follows similar steps as the worst-case LM rate in Appendix A. ■

Corollary 1: For sufficiently small r , the worst-case mismatched Gallager functions can be expanded as

$$\underline{E}_0^{\text{iid}}(Q_X, \widehat{W}, \rho, r) = \sup_{s \geq 0} \left\{ E_{s,0,0,\rho}^{\text{ML}}(Q_X, \widehat{W}) - \sqrt{2r V_{s,0,0,\rho}(Q_X, \widehat{W})} + o(\sqrt{r}) \right\} \quad (70)$$

$$\underline{E}_0^{\text{cc}}(Q_X, \widehat{W}, \rho, r) = \sup_{c_1(\cdot), c_2(\cdot)} \sup_{s \geq 0, \lambda, \bar{\lambda}} \left\{ E_{s,\lambda,\bar{\lambda},\rho}^{\text{ML}}(Q_X, \widehat{W}) - \sqrt{2r V_{s,\lambda,\bar{\lambda},\rho}(Q_X, \widehat{W})} + o(\sqrt{r}) \right\} \quad (71)$$

Proof: The expressions come from a first-order Puiseux series expansion of (66)–(67) around $r = 0$. The expansions incur an error of order $O(r)$ that is absorbed in $o(\sqrt{r})$ as the latter dominates the summation for small $r \geq 0$. ■

The following observations are in order:

- 1) *The solution for a chi-squared distance ball.* The worst-case mismatched E_0 functions for a small mismatch given by a chi-squared distance ball as defined in (33) are exactly

$$\underline{E}_0^{\text{iid}}(Q_X, \widehat{W}, \rho, r) = \sup_{s \geq 0} \left\{ E_{s,0,0,\rho}^{\text{ML}}(Q_X, \widehat{W}) - \log \left(1 + \sqrt{2r V_{s,0,0,\rho}(Q_X, \widehat{W})} \right) \right\} \quad (72)$$

$$\underline{E}_0^{\text{cc}}(Q_X, \widehat{W}, \rho, r) = \sup_{c_1(\cdot), c_2(\cdot)} \sup_{s \geq 0, \lambda, \bar{\lambda}} \left\{ E_{s,\lambda,\bar{\lambda},\rho}^{\text{ML}}(Q_X, \widehat{W}) - \log \left(1 + \sqrt{2r V_{s,\lambda,\bar{\lambda},\rho}(Q_X, \widehat{W})} \right) \right\} \quad (73)$$

achieved for fixed $s, \lambda, \bar{\lambda}, c_1(\cdot), c_2(\cdot)$ by the channel distribution

$$\tilde{W}_{s,\lambda,\bar{\lambda}}^*(y|x) = \widehat{W}(y|x) \cdot \left(1 + \frac{\sqrt{2r} \cdot \varphi_{s,\lambda,\bar{\lambda},\rho}(x, y)}{\mathbb{E}_{Q_X \times \widehat{W}} [\varepsilon_{s,\lambda,\bar{\lambda},\rho}(X, Y)]} \right) \quad (74)$$

with

$$\varphi_{s,\lambda,\bar{\lambda},\rho}(x, y) \triangleq \frac{\varepsilon_{s,\lambda,\bar{\lambda},\rho}(x, y) - \mathbb{E}_{\widehat{W}} [\varepsilon_{s,\lambda,\bar{\lambda},\rho}(x, Y)]}{\sqrt{V_{s,\lambda,\bar{\lambda},\rho}(Q_X, \widehat{W})}}. \quad (75)$$

The worst-case channels $\tilde{W}_{\text{iid}}^*, \tilde{W}_{\text{cc}}^*$ are found after maximizing (72)–(73) over $s, \lambda, \bar{\lambda}, c_1(\cdot), c_2(\cdot)$.

- 2) *Feasibility of worst-case channels.* Similarly to what we did for the worst-case achievable rates, the following condition over the ball radius needs to be satisfied to guarantee non-negativity of the obtained distributions:

$$r < \min_{(x,y): \varphi_{s,\lambda,\bar{\lambda},\rho}(x,y) > 0} \frac{1}{2 \varphi_{s,\lambda,\bar{\lambda},\rho}(x, y)^2}. \quad (76)$$

- 3) *The solution.* The worst-case channels are always found at the border of the respective ball, i.e., when (21) or (33) holds with equality.

- 4) *Expansion term.* The expansion yields a penalty term proportional to \sqrt{r} at most. This indicates that even a small mismatch causes a severe penalty on the E_0 function and thus on the error exponent.

- 5) *Achievable E_0 function at $r = 0$.* For a fixed \widehat{W} ,

$$\underline{E}_0^{\text{cc}}(Q_X, \widehat{W}, \rho, r) \leq \sup_{c_1(\cdot), c_2(\cdot)} \sup_{s \geq 0, \lambda, \bar{\lambda}} E_{s,\lambda,\bar{\lambda},\rho}^{\text{ML}}(Q_X, \widehat{W}) \quad (77)$$

coinciding with the matched Gallager function for channel \widehat{W} . The proof follows because $\widehat{W} \in \mathcal{B}$ for any $r \geq 0$, so the minimum in (65) can be upper bounded by evaluating at $W = \widehat{W}$. The bound is tight at $r = 0$ as $\mathcal{B} = \{\widehat{W}\}$.

- 6) *E_0 function at $\rho = 0$.* The worst-case LM and GMI rates cannot be obtained by differentiating the worst-case E_0 functions and evaluating at $\rho = 0$. This is in contrast to

what occurs for $r = 0$, making our separate analysis of achievable rates and error exponents highly relevant.

A. Example: Symmetric \widehat{W} and Equiprobable Q_X

As in Section III-A, we derive the worst-case i.i.d. Gallager function for symmetric [1, Pag. 94] estimated channels \widehat{W} and equiprobable input $Q_X(x) = |\mathcal{X}|^{-1}$, showing only the final results; the full derivations are found in Appendix C.

The worst-case mismatched Gallager function yields

$$\begin{aligned} \tilde{E}_0^{\text{iid}}(Q_X, \widehat{W}, \rho, r) = \sup_{s \geq 0} \left\{ \log \frac{|\mathcal{X}|^\rho}{\kappa_s^\rho \kappa_{1-sp}} \right. \\ \left. - \log \left(1 + \sqrt{2r} \sqrt{\frac{\kappa_{1-2sp}}{\kappa_{1-sp}^2} - 1} \right) \right\} \quad (78) \end{aligned}$$

where $\kappa_t \triangleq \sum_y \widehat{W}(y|x_1)^t$ is an auxiliary function computed from the first row of \widehat{W} .

The same observations in Section III-A apply here: the worst-case channel in $\tilde{\mathcal{B}}$ exhibits the same structure as \widehat{W} , suggesting that i.i.d. and constant-composition random coding achieve the same error exponent.

B. Example: Modulo-Additive \widehat{W} and Equiprobable Q_X

Following from the previous example, we now analyze the specific case of a modulo-additive \widehat{W} by particularizing (78) for the transition matrix in (49). This gives $\kappa_t = \bar{p}^t + (|\mathcal{X}| - 1)p^t = \bar{p}^t + (1 - \bar{p})p^{t-1}$, and thus

$$\begin{aligned} \tilde{E}_0^{\text{iid}}(Q_X, \widehat{W}, \rho, r) = \\ \sup_{s \geq 0} \left\{ \log \frac{|\mathcal{X}|^\rho}{(\bar{p}^{1-sp} + (1 - \bar{p})p^{-sp}) (\bar{p}^s + (1 - \bar{p})p^{s-1})^\rho} \right. \\ \left. - \log \left(1 + \sqrt{2r} \sqrt{\frac{\bar{p}^{1-2sp} + (1 - \bar{p})p^{-2sp}}{(\bar{p}^{1-sp} + (1 - \bar{p})p^{-sp})^2} - 1} \right) \right\}. \quad (79) \end{aligned}$$

C. Example: Ternary-Input Ternary-Output \widehat{W}

We study the worst-case \tilde{E}_0^{iid} (72) and \tilde{E}_0^{cc} (73) for the following input distribution and channel estimate:

$$Q_X = [0.3 \ 0.3 \ 0.4] \quad (80)$$

$$\widehat{W} = \begin{bmatrix} 0.85 & 0.05 & 0.10 \\ 0.025 & 0.945 & 0.03 \\ 0.025 & 0.10 & 0.875 \end{bmatrix}. \quad (81)$$

An off-the-shelf solver is used to numerically compute the true $\underline{E}_0^{\text{iid}}$ (64) and $\underline{E}_0^{\text{cc}}$ (65), which are plotted in Fig. 2 along with the matched Gallager functions E_0^{iid} (10) and E_0^{cc} (11) under maximum-likelihood decoding using \widehat{W} . The graph shows that $\underline{E}_0^{\text{iid}} \leq \underline{E}_0^{\text{cc}}$ for all r , and the approximations \tilde{E}_0^{iid} and \tilde{E}_0^{cc} show the same behavior. Much like $I_{\text{MI}}(Q_X, \widehat{W})$ in Section III, the estimated channel's matched Gallager functions are achievable for $r = 0$, but as soon as $r > 0$, the mismatched Gallager functions decrease rapidly from their matched counterparts. The worst-case mismatched Gallager functions drop with infinite slope at $r = 0$, as expected from their dependence on \sqrt{r} in our findings.

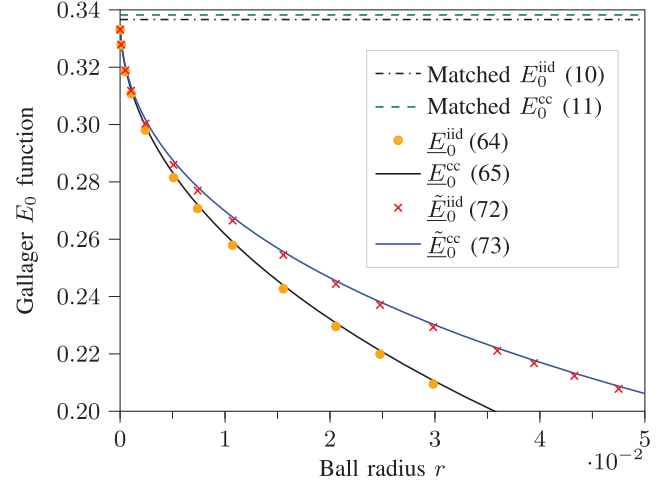


Fig. 2. Matched and mismatched Gallager functions computed for $\rho = 0.7$, input distribution Q_X in (80) and estimated channel \widehat{W} in (81).

The curves have similar graphical shapes, with approximated curves increasingly higher than the true curves. The error terms appear to be negative at the optimal point in our simulation. Apparently, with increasing r , we increasingly overestimate the worst-case Gallager functions, so we also overestimate the error exponents. Additionally, the simulation shows that the approximations deviate from their true values to a greater extent than those for the achievable rates. The reason is that the logarithm in (66)–(67) is more sensitive to small variations in contrast to the linear dependence in (31). Consequently, the validity of our approximation seems to be restricted to a lower range of values of r for the mismatched error exponents, in contrast to the large range of values valid for the mismatched achievable rates.

V. CONTINUOUS ALPHABETS

In this section, we extend the study of worst-case achievable rates and Gallager E_0 functions to continuous-alphabet channels by allowing the alphabets \mathcal{X} and \mathcal{Y} to take infinite values. Specifically, our focus is on $\mathcal{X} = \mathcal{Y} = \mathbb{R}$. We use $Q_X(x)$ and $W(y|x)$ to denote the continuous probability input and channel distributions, respectively, and redefine small mismatch for continuous-alphabet channels as follows.

Definition 2: For small mismatch (ball radius) between the continuous-alphabet channels W and \widehat{W} , we require that

$$W \in \mathcal{B} = \left\{ P_{Y|X} \in \mathcal{S} : D(\widehat{W} \| P_{Y|X} | Q_X) \leq r \right\} \quad (82)$$

holds for small r , where \mathcal{S} is a function space admitting probability density functions $P_{Y|X}(y|x)$ over $(x, y) \in \mathbb{R}^2$, and $D(\widehat{W} \| P_{Y|X} | Q_X)$ defines the conditional relative entropy in integral form as

$$D(\widehat{W} \| P_{Y|X} | Q_X) \triangleq \iint Q_X(x) \widehat{W}(y|x) \log \frac{\widehat{W}(y|x)}{P_{Y|X}(y|x)} dx dy. \quad (83)$$

We analyze both i.i.d. and cost-constrained codes since constant-composition random coding is undefined for continuous alphabets. As we shall see, most of the previous expressions of information rates and E_0 functions remain

invariant, by substituting probability distributions with the corresponding probability densities and summations with integrals. The novelty is that the formulation of the continuous case allows for unprecedented closed-form results for the nearest neighbor decoder that further extend [2], [4], [18].

A. i.i.d. Random Coding

For i.i.d. random coding, the mismatched information density and mismatched exponent density are defined as

$$i_s(x, y) \triangleq \log \frac{\widehat{W}(y|x)^s}{\int Q_X(\bar{x}) \widehat{W}(y|\bar{x})^s d\bar{x}} \quad (84)$$

$$\varepsilon_{s,\rho}(x, y) \triangleq \left(\frac{\int Q_X(\bar{x}) \widehat{W}(y|\bar{x})^s d\bar{x}}{\widehat{W}(y|x)^s} \right)^\rho, \quad (85)$$

where we have expressed expectations in integral form for reasons of emphasis.

We want to solve the following problems to determine the worst-case GMI and E_0 function:

$$I_{\text{GMI}}(Q_X, \widehat{W}, r) = \min_{W \in \mathcal{B}} \sup_{s \geq 0} \mathbb{E}_{Q_X \times W} [i_s(X, Y)] \quad (86)$$

$$\underline{E}_0^{\text{iid}}(Q_X, \widehat{W}, r) = \min_{W \in \mathcal{B}} \sup_{s \geq 0} -\log \mathbb{E}_{Q_X \times W} [\varepsilon_{s,\rho}(X, Y)] \quad (87)$$

Their solutions are particular cases of cost-constrained random coding, and are thus described below.

B. Cost-Constrained Random Coding

We study codebooks in which, unlike i.i.d. random coding, codewords are constrained to satisfy multiple per-codeword cost functions [14]. Specifically, we consider codebooks in which each codeword $\mathbf{x}^{(m)}$ satisfies

$$\frac{1}{n} \sum_{i=1}^n c_0(x_i^{(m)}) \leq P \quad (88)$$

$$\left| \frac{1}{n} \sum_{i=1}^n c_\ell(x_i^{(m)}) - \phi_\ell \right| \leq \frac{\delta}{n}, \quad \ell = 1, \dots, L \quad (89)$$

where $\phi_\ell \triangleq \mathbb{E}[c_\ell(X)]$ for $\ell = 1, \dots, L$ and $\delta > 0$. Observe that these input cost functions might be related to the channel having an input cost constraint $c_0(\cdot)$, or might be design cost constraints that allow us to shape codewords in a certain way as for the cost functions $c_\ell(\cdot)$ for $\ell = 1, \dots, L$.

The corresponding mismatched information and exponent densities now read

$$i_{s,\lambda}(x, y) \triangleq \log \frac{\widehat{W}(y|x)^s e^{\lambda^T c(x)}}{\int Q_X(\bar{x}) \widehat{W}(y|\bar{x})^s e^{\lambda^T c(\bar{x})} d\bar{x}} \quad (90)$$

$$\varepsilon_{s,\lambda,\bar{\lambda},\rho}(x, y) \triangleq \left(\frac{\int Q_X(\bar{x}) \widehat{W}(y|\bar{x})^s e^{\lambda^T (c(\bar{x}) - \phi)} d\bar{x}}{\widehat{W}(y|x)^s e^{\lambda^T (c(x) - \phi)}} \right)^\rho \quad (91)$$

where we have defined the following vectors to denote compactly the cost functions and corresponding weights:

$$\mathbf{c}(x) \triangleq [c_0(x), c_1(x), \dots, c_L(x)]^T \quad (92)$$

$$\boldsymbol{\phi} \triangleq [P, \phi_1, \dots, \phi_L]^T \quad (93)$$

$$\boldsymbol{\lambda} \triangleq [\lambda_0, \lambda_1, \dots, \lambda_L]^T \quad (94)$$

$$\bar{\boldsymbol{\lambda}} \triangleq [\bar{\lambda}_0, \bar{\lambda}_1, \dots, \bar{\lambda}_L]^T \quad (95)$$

with $\lambda_0, \bar{\lambda}_0 \geq 0$ and $\lambda_\ell, \bar{\lambda}_\ell \in \mathbb{R}$ for $\ell = 1, \dots, L$. The problems we want to solve are

$$I_{\text{cost}}(Q_X, \widehat{W}, r) = \min_{W \in \mathcal{B}} \sup_{s \geq 0, \lambda} \mathbb{E}_{Q_X \times W} [i_{s,\lambda}(X, Y)] \quad (96)$$

$$\underline{E}_0^{\text{cost}}(Q_X, \widehat{W}, r) = \min_{W \in \mathcal{B}} \sup_{s \geq 0, \lambda, \bar{\lambda}} -\log \mathbb{E}_{Q_X \times W} [\varepsilon_{s,\lambda,\bar{\lambda},\rho}(X, Y)]. \quad (97)$$

The main difference with respect to the i.i.d. case is the need to further optimize over vectors $\boldsymbol{\lambda}, \bar{\boldsymbol{\lambda}}$. Note that (86)–(87) can be obtained from (96)–(97) by setting $\boldsymbol{\lambda} = \bar{\boldsymbol{\lambda}} = \mathbf{0}$. The following theorem provides the solutions to all problems.

Theorem 3: Consider a family of continuous-alphabet channels $W(y|x)$, a mismatched decoder based on the channel estimate $\widehat{W}(y|x)$ and a fixed input distribution $Q_X(x)$ satisfying (82). Then, for every $r \geq 0$ sufficiently small:

(i) the worst-case achievable rates for cost-constrained and i.i.d. random coding can be expressed as

$$I_{\text{cost}}(Q_X, \widehat{W}, r) = \sup_{s \geq 0, \lambda} \left\{ I_{s,\lambda}^{\text{ML}}(Q_X, \widehat{W}) - \sqrt{2rV_{s,\lambda}(Q_X, \widehat{W})} + o(\sqrt{r}) \right\} \quad (98)$$

$$I_{\text{GMI}}(Q_X, \widehat{W}, r) = \sup_{s \geq 0} \left\{ I_{s,\mathbf{0}}^{\text{ML}}(Q_X, \widehat{W}) - \sqrt{2rV_{s,\mathbf{0}}(Q_X, \widehat{W})} + o(\sqrt{r}) \right\} \quad (99)$$

where

$$I_{s,\lambda}^{\text{ML}}(Q_X, \widehat{W}) \triangleq \mathbb{E}_{Q_X \times \widehat{W}} [i_{s,\lambda}(X, Y)] \quad (100)$$

$$V_{s,\lambda}(Q_X, \widehat{W}) \triangleq \mathbb{E}_{Q_X} [\text{Var}_{\widehat{W}} [i_{s,\lambda}(X, Y)|X]]. \quad (101)$$

(ii) the worst-case mismatched E_0 functions for cost-constrained and i.i.d. random coding can be expressed as

$$\underline{E}_0^{\text{cost}}(Q_X, \widehat{W}, \rho, r) = \sup_{s \geq 0, \lambda, \bar{\lambda}} \left\{ E_{s,\lambda,\bar{\lambda},\rho}^{\text{ML}}(Q_X, \widehat{W}) - \log \left(1 + \sqrt{2rV_{s,\lambda,\bar{\lambda},\rho}(Q_X, \widehat{W})} \right) + o(\sqrt{r}) \right\} \quad (102)$$

$$\underline{E}_0^{\text{iid}}(Q_X, \widehat{W}, \rho, r) = \sup_{s \geq 0} \left\{ E_{s,\mathbf{0},\mathbf{0},\rho}^{\text{ML}}(Q_X, \widehat{W}) - \log \left(1 + \sqrt{2rV_{s,\mathbf{0},\mathbf{0},\rho}(Q_X, \widehat{W})} \right) + o(\sqrt{r}) \right\} \quad (103)$$

where

$$E_{s,\lambda,\bar{\lambda},\rho}^{\text{ML}}(Q_X, \widehat{W}) \triangleq -\log \mathbb{E}_{Q_X \times \widehat{W}} [\varepsilon_{s,\lambda,\bar{\lambda},\rho}(X, Y)] \quad (104)$$

$$V_{s,\lambda,\bar{\lambda},\rho}(Q_X, \widehat{W}) \triangleq \frac{\mathbb{E}_{Q_X} [\text{Var}_{\widehat{W}} [\varepsilon_{s,\lambda,\bar{\lambda},\rho}(X, Y)|X]]}{\mathbb{E}_{Q_X \times \widehat{W}} [\varepsilon_{s,\lambda,\bar{\lambda},\rho}(X, Y)]^2}. \quad (105)$$

Proof: The solution to worst-case achievable rates is found by resorting to the calculus of variations [7]; see Appendix D for the proof. The worst-case mismatched Gallager E_0 function is a reformulation of the worst-case achievable rates, as shown in Appendix B, followed by the steps in Appendix D. ■

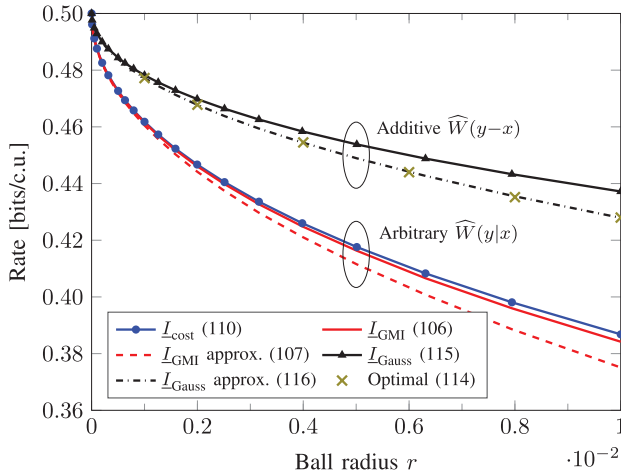


Fig. 3. Worst-case GMI and cost-constraint rates versus ball radius r for Gaussian codewords $Q_X(x) = \mathcal{N}(x; 1)$ and the nearest neighbor decoder $\hat{W}(y|x) = \mathcal{N}(y-x; 1)$. These computations assume that $o(\sqrt{r}) = 0$. The optimal values from (114) are depicted using cross markers.

C. Example: Gaussian Codebooks and Nearest Neighbor Decoding for Arbitrary Continuous-Alphabet Channels

We next study the nearest neighbor decoder when combined with Gaussian inputs by particularizing the above results to the estimated channel $\hat{W}(y|x) = \mathcal{N}(y-x; \hat{\sigma}^2)$ and input distribution $Q_X(x) = \mathcal{N}(x; P)$. We make no additional assumption on the nature of the memoryless channel, other than being in a relative entropy ball of radius r centered at the nearest neighbor decoder. Appendix E details the specific computations; here we only show the final results.

1) *Gaussian i.i.d. Codebooks*: the worst-case GMI rate can be expressed as

$$I_{\text{GMI}}(Q_X, \hat{W}, r) = \sup_{s \geq 0} \left\{ \frac{1}{2} \log(1 + s\Gamma) + \frac{\Gamma(1-s)}{2(s^{-1} + \Gamma)} - \sqrt{r \cdot \frac{\Gamma(2 + s^2\Gamma)}{(s^{-1} + \Gamma)^2}} + o(\sqrt{r}) \right\} \quad (106)$$

where $\Gamma \triangleq P/\hat{\sigma}^2$ denotes the estimated signal-to-noise ratio assuming an additive noise structure even if the channels in \mathcal{B} are not necessarily additive.

Moreover, as $r \rightarrow 0$ (106) is dominated by the first two terms, for which $s = 1$ is the optimal value and thus (106) can be accurately approximated as

$$I_{\text{GMI}}(Q_X, \hat{W}, r) \approx \frac{1}{2} \log(1 + \Gamma) - \sqrt{r \cdot \frac{\Gamma(2 + \Gamma)}{(1 + \Gamma)^2}}. \quad (107)$$

Intuitively, information rates close to second order coding rates can be achieved when the true channel is close to a Gaussian channel in terms of relative entropy. This approximation, aided by (106), helps quantify the gain of the parameter $s \geq 0$. Fig. 3 presents a numerical example of this gain.

A similar analysis can be performed for the worst-case Gallager E_0 function (see Appendix E-B for details). The corresponding terms in the approximation are given by

$$E_{s,0,0,\rho}^{\text{ML}}(Q_X, \hat{W}) = \frac{\rho}{2} \log(1 + \Gamma s)$$

$$+ \frac{1}{2} \log \left(1 + \frac{\rho\Gamma(1-s-\rho s)}{\Gamma + s^{-1}} \right) \quad (108)$$

and

$$V_{s,0,0,\rho}(Q_X, \hat{W}) = \frac{1}{\sqrt{\Gamma + s^{-1}}} \frac{\Gamma + s^{-1} + \rho\Gamma(1-s-\rho s)}{\sqrt{\Gamma + s^{-1} + 2\rho\Gamma(1-s-2\rho s)}} - \frac{\sqrt{2}}{\rho s} \frac{\Gamma + s^{-1} + \rho\Gamma(1-s-\rho s)}{\Gamma + s^{-1} + \rho\Gamma(2-s-2\rho s)}. \quad (109)$$

2) *Gaussian Spherical Codebooks*: when codewords satisfy the per-codeword power constraint (88) with $c_0(x) = x^2$, we have the following achievable rate

$$I_{\text{cost}}(Q_X, \hat{W}, r) = \sup_{s \geq 0, t \geq 0} \left\{ \frac{1}{2} \log(g(\Gamma) + s\Gamma) + \frac{\Gamma(1-s) + s^{-1}g(\Gamma)(1-g(\Gamma))}{2(\Gamma + s^{-1}g(\Gamma))} - \sqrt{r \cdot \frac{\Gamma(2g^2(\Gamma) + s^2\Gamma)}{(s^{-1}g(\Gamma) + \Gamma)^2}} + o(\sqrt{r}) \right\} \quad (110)$$

where $g(\Gamma) \triangleq 1 - 2\lambda\Gamma$. Setting $\lambda = 0$ gives that $g(\Gamma) = 1$ and thus $I_{\text{cost}} = I_{\text{GMI}}$. In general, we have that $I_{\text{cost}} \geq I_{\text{GMI}}$. Numerical evidence of such a gain is shown in Fig. 3.

D. Example: Gaussian Codebooks and Nearest Neighbor Decoding for Additive Noise Channels

Further closed-form results can be obtained when we add the assumption that the channel adds the noise Z , but its probability density function $W(z)$ for $z \in \mathbb{R}$ is unknown. In this case, we redefine \mathcal{B} to contain only additive noise channels $W(y|x) = W(y-x)$. That is,

$$\mathcal{W} \in \mathcal{B} = \left\{ P_{Y|X} \in \mathcal{S} : D(\hat{W} \| P_{Y|X}) \leq r \right\} \quad (111)$$

where thanks to the additive structure of both channel and decoding metric, the relative entropy becomes independent of the input distribution Q_X , as

$$D(\hat{W} \| P_{Y|X}) \triangleq \int_{\mathbb{R}} \hat{W}(y) \log \frac{\hat{W}(y)}{P_{Y|X}(y)} dy. \quad (112)$$

We next study the cases of Gaussian i.i.d. and spherical codebooks in Section V-D.1 and Gaussian i.i.d. codebooks with a fixed cost in Section V-D.2. The same steps in Appendix D are followed to derive all the results below.

1) *Gaussian i.i.d. And Spherical Codebooks*: For additive noise channels with a known second-order moment $\mathbb{E}[Z^2]$, Lapidoth [18] showed that the rate

$$I_{\text{Gauss}}(Q_X, \hat{W}) = \frac{1}{2} \log \left(1 + \frac{P}{\mathbb{E}[Z^2]} \right) \quad (113)$$

is achievable by the nearest neighbor decoder $\hat{W}(y|x) = \mathcal{N}(y-x; 1)$ regardless of the additive noise distribution provided that its second-order moment $\mathbb{E}[Z^2] < \infty$ is known. We consider instead an available estimate for the unknown second-order moment which we decompose into mean $\hat{\mu}$ and variance $\hat{\sigma}^2$, and a uncertainty level given by the ball radius r . Hence, the channel estimate is the nearest neighbor decoder $\hat{W}(y|x) = \mathcal{N}(y-x-\hat{\mu}; \hat{\sigma}^2)$. We wish to solve

$$I_{\text{Gauss}}(Q_X, \hat{W}, r) = \min_{W \in \mathcal{B}} \frac{1}{2} \log \left(1 + \frac{P}{\mathbb{E}[(Z-\hat{\mu})^2]} \right) \quad (114)$$

where the dependence on the noise distribution appears only as a function of $\mathbb{E}[(Z - \hat{\mu})^2]$. In other words, we wish to find the noise distribution in \mathcal{B} that gives the highest second-order moment with respect to $\hat{\mu}$. By approximating the relative entropy ball as before, the problem can be solved analytically, leading to the following rate

$$I_{\text{Gauss}}(Q_X, \hat{W}, r) = \frac{1}{2} \log \left(1 + \frac{\Gamma}{1 + 2\sqrt{r}} \right) + o(\sqrt{r}) \quad (115)$$

which can be expanded to facilitate comparison with (107):

$$I_{\text{Gauss}}(Q_X, \hat{W}, r) = \frac{1}{2} \log(1 + \Gamma) - \frac{\Gamma \sqrt{r}}{1 + \Gamma} + o(\sqrt{r}). \quad (116)$$

In contrast to [4, Th. 3.9], there is no penalty for considering the non-zero mean provided that the true channel is sufficiently close to a Gaussian of mean $\hat{\mu}$ and variance $\hat{\sigma}$. The result also gives a second-order term proportional to \sqrt{r} , but the scaling is advantageous compared to (107), where channel additivity is not assumed. Simulations in Fig. 3 support this statement since knowing the additive structure of the channel operation leads to a significant rate increase, especially at low signal-to-noise ratios. Fig. 3 also shows the true optimal values obtained by solving (114) as detailed in Section V-E. These values are found to be very close to our approximation in (116) when $o(\sqrt{r}) = 0$, a coincidence that appears to be accidental. More importantly, both approximations in (115)–(116) are remarkably close to the true solution.

2) *Fixed-Cost Gaussian Codebooks*: For additive noise channels with a known second-order moment $\mathbb{E}[Z^2]$, Scarlett et al. [4, Th. 3.9] showed that the rate

$$I_{\text{fixed-cost}}(Q_X, \hat{W}) = \frac{1}{2} \log \left(1 + \frac{P}{\text{Var}[Z]} \right) \quad (117)$$

is achievable by cost-constrained random coding with a single auxiliary cost $c_1(x) = x$. In other words, the cost function $c_1(x) = x$ is able to cancel the effect of the unknown channel mean. We wish to solve

$$I_{\text{fixed-cost}}(Q_X, \hat{W}, r) = \min_{W \in \mathcal{B}} \frac{1}{2} \log \left(1 + \frac{P}{\text{Var}[Z]} \right) \quad (118)$$

or equivalently, find the noise distribution in \mathcal{B} with the highest variance. The following rate is achieved

$$I_{\text{fixed-cost}}(Q_X, \hat{W}, r) = \frac{1}{2} \log \left(1 + \frac{\Gamma}{1 + 2\sqrt{r}} \right) + o(\sqrt{r}) \quad (119)$$

coinciding with (115). Interestingly, the mean estimate $\hat{\mu}$ only appears in the constraint $W \in \mathcal{B}$ but the achievable rate does not depend explicitly on it. The picture is slightly different to what occurs in (115), where the channel mean has to be accurately estimated so as to design the nearest neighbor decoder accordingly. This reinforces the usefulness of cost-constrained coding, as a simple auxiliary cost function makes the achievable rate independent of the channel mean.

E. Worst-Case Channels

Throughout the manuscript, we have primarily focused on our approximation of relative entropy by chi-squared distance, which has produced the small terms $o(\sqrt{r})$ in every case we

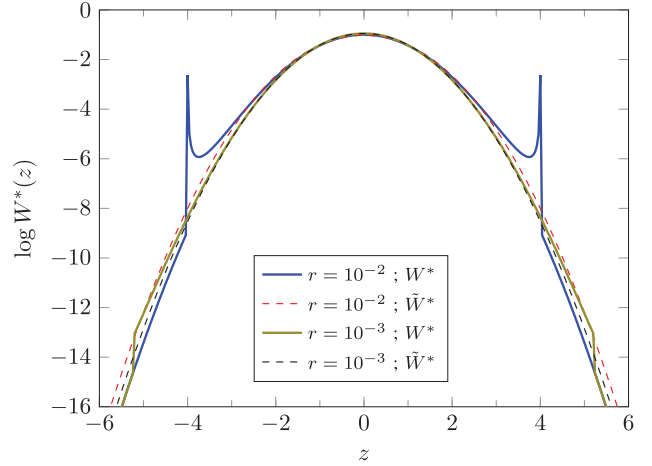


Fig. 4. Worst-case channel (in log scale) in the relative entropy (W^* in (124)) and chi-squared (\tilde{W}^* in (125)) balls. Computations assume Gaussian codebooks $Q_X(x) = \mathcal{N}(x; 1)$ and the nearest neighbor decoder $\hat{W}(y|x) = \mathcal{N}(y-x; 1)$. The worst-case GMI rates for W^* are shown in Fig. 3.

have analyzed. We have not given as much attention to the worst-case channels when examining both relative entropy and chi-squared balls. In this respect, the continuous-alphabet case is likely one of the most intriguing cases to study as the structure of the respective worst-case channels when \hat{W} has unbounded support substantially differs depending on whether we consider a relative entropy or chi-squared distance ball.

We concentrate on the worst-case additive channel when Gaussian codebooks are combined with the nearest neighbor decoder, corresponding to

$$\begin{aligned} W^*(z) &\triangleq \arg\max_{W \in \mathcal{B}} \mathbb{E}[Z^2] \\ &= \arg\max_{W \in \mathcal{B}} \int_{\mathbb{R}} z^2 W(z) dz. \end{aligned} \quad (120)$$

After tedious but straightforward analysis following standard variational tools for broken extremals [7, Sec. 15], the worst-case channel in \mathcal{B} is defined from the piece-wise expression

$$W_{z_0}^*(z) = \begin{cases} \hat{W}(z) & \text{if } |z| \geq z_0 \\ \hat{W}(z) \cdot \frac{\rho}{z^2 - \lambda} & \text{if } |z| < z_0 \end{cases} \quad (121)$$

for $\rho < 0$ and $0 < z_0 < \lambda$ satisfying

$$\rho = \left(\int_{-z_0}^{z_0} \frac{\hat{W}(z)}{z^2 - \lambda} dz \right)^{-1} \quad (122)$$

$$r = \int_{-z_0}^{z_0} \hat{W}(z) \log \left(\frac{z^2 - \lambda}{\rho} \right) dz. \quad (123)$$

The worst-case channel probability density function is found after maximizing over $z_0 > 0$ as

$$W^*(z) = \arg\max_{z_0 > 0} \int_{\mathbb{R}} z^2 W_{z_0}^*(z) dz. \quad (124)$$

The solution contrasts with the worst-case channel in the chi-squared distance ball $\tilde{\mathcal{B}}$ computed as in [25, Appendix C]

$$\begin{aligned} \tilde{W}^*(z) &= \arg\max_{W \in \tilde{\mathcal{B}}} \mathbb{E}[Z^2] \\ &= \hat{W}(z) \left(1 + \sqrt{r} \cdot \frac{z^2 - \hat{\sigma}^2}{\hat{\sigma}^2} \right). \end{aligned} \quad (125)$$

Fig. 4 illustrates these differences, showing that the discontinuities at $z = \pm z_0$ for W^* are not present in the corresponding approximation, \hat{W}^* , computed when the chi-squared distance is instead adopted. The primary reason for this difference is that relative entropy, viewed as a functional of the channel W , is not continuous everywhere when the channel estimate \hat{W} has unbounded support, such as in the case of nearest neighbor decoding. In these situations, the functional admits a broken extremal as the maximizing distribution, in which only a portion of the support is optimized.

VI. CONCLUSION

We have studied a family of mismatched compound channels by assuming that the true channel lies within a small radius relative entropy ball centered at the decoding metric. We have shown that a second order expansion of the relative entropy ball for a small radius allows the derivation of closed-form results to worst-case (minimum) information rates and error exponents for discrete and continuous memoryless channels. Specifically, for both discrete- and continuous-alphabet memoryless channels, we derive penalty terms that are proportional to the square root of the ball radius. We have obtained new explicit expressions for the worst-case rate achievable with Gaussian codebooks and nearest neighbor decoding which complement the analytical findings derived by Lapidoth and Narayan [2], Scarlett et al. [4], Lapidoth [18], and Merhav et al. [26].

APPENDIX A PROOF OF THEOREM 1

Here, we provide the full derivation for the worst-case LM rate in Theorem 1. Let us define

$$I_{s,a}(Q_X, W) \triangleq \sum_{x,y} Q_X(x) W(y|x) i_{s,a}(x, y) \quad (126)$$

$$I_{s,a}^{\text{ML}}(Q_X, \hat{W}) \triangleq \sum_{x,y} Q_X(x) \hat{W}(y|x) i_{s,a}(x, y), \quad (127)$$

and let small mismatch be characterized by Definition 1. Then, we formulate the worst-case LM rate I_{LM} as

$$I_{\text{LM}}(Q_X, \hat{W}, r) = \min_{W \in \mathcal{B}} \sup_{s \geq 0, a(\cdot)} I_{s,a}(Q_X, W). \quad (128)$$

Since $I_{s,a}$ is linear in W and concave in $s \geq 0$ and $a(x)$ [4, Ch. 2.3.4], we can switch the order of the optimizations in application of the minimax theorem [24, Th. 1]. Then, by minimizing $I_{s,a}$ over all channels $W \in \mathcal{B}$ we obtain

$$I_{s,a}(Q_X, \hat{W}, r) \triangleq \min_{W \in \mathcal{B}} I_{s,a}(Q_X, W) \quad (129)$$

$$= I_{s,a}^{\text{ML}}(Q_X, \hat{W}) + \min_{W \in \mathcal{B}} \theta^T \nabla I_{s,a} \quad (130)$$

where:

$$\theta \triangleq \left[\theta(y_1|x_1), \dots, \theta(y_{|\mathcal{Y}|}|x_1), \theta(y_1|x_2), \dots, \theta(y_{|\mathcal{Y}|}|x_{|\mathcal{X}|}) \right]^T \quad (131)$$

with $\theta(y_i|x_j) = W(y_i|x_j) - \hat{W}(y_i|x_j)$ as defined in (23), and

$$\nabla I_{s,a} \triangleq \left[Q_X(x_1) i_{s,a}(x_1, y_1), \dots, Q_X(x_1) i_{s,a}(x_1, y_{|\mathcal{Y}|}), \right. \\ \left. Q_X(x_2) i_{s,a}(x_2, y_1), \dots, Q_X(x_2) i_{s,a}(x_2, y_{|\mathcal{Y}|}), \right]$$

$$\dots, Q_X(x_{|\mathcal{X}|}) i_{s,a}(x_{|\mathcal{X}|}, y_{|\mathcal{Y}|}) \Big]^T \quad (132)$$

is the gradient of $I_{s,a}$ (126) with respect to each $W(y_i|x_j)$.

The constraint $W \in \mathcal{B}$ over all valid conditional probability distributions is also written in vector form as the following set of $1+|\mathcal{X}|$ constraints

$$\begin{cases} D(\hat{W}||W|Q_X) = d(\theta) + o(d(\theta)) \leq r \\ \theta^T \mathbf{1}_j = 0 \quad \text{for } 1 \leq j \leq |\mathcal{X}| \end{cases} \quad (133)$$

under the following definitions:

$$d(\theta) \triangleq \frac{1}{2} \theta^T \mathbf{K} \theta \quad (134)$$

$$\mathbf{K} \triangleq \text{diag} \left(\frac{Q_X(x_1)}{\hat{W}(y_1|x_1)}, \dots, \frac{Q_X(x_{|\mathcal{X}|})}{\hat{W}(y_{|\mathcal{Y}|}|x_{|\mathcal{X}|})} \right) \quad (135)$$

$$\mathbf{1}_j \triangleq [0 \dots 0 \ 1_{(1,j)} \dots 1_{(|\mathcal{Y}|,j)} \ 0 \dots 0]^T \quad (136)$$

where the upper side of (133) guarantees that sought channels are within the ball of radius r , and the constraints in the lower side guarantee that $\sum_y W(y|x_j) = 1$ for every $x_j \in \mathcal{X}$.

Therefore, problem in (130) can be written equivalently as

$$\min_{W \in \mathcal{B}} \theta^T \nabla I_{s,a} = \min_{\substack{d(\theta) + o(d(\theta)) \leq r \\ \theta^T \mathbf{1}_j = 0, 1 \leq j \leq |\mathcal{X}|}} \theta^T \nabla I_{s,a}, \quad (137)$$

a convex problem that can be solved using the Lagrangian method [6]. This leads to a system of implicit equations that must be solved numerically using a trial-and-error algorithm for the unknown variables.

One of the key points to note is that there is no solution to (137) that satisfies $D(\hat{W}||W|Q_X) < r$, with strict inequality. As the objective function and the equality constraints are linear in θ , the following minimum

$$\min_{\theta^T \mathbf{1}_j = 0, 1 \leq j \leq |\mathcal{X}|} \theta^T \nabla I_{s,a} \quad (138)$$

is unbounded unless we impose boundary constraints on W . In this case, the minimum is achieved by the worst-case channel

$$W^*(y|x) = 1\{i_{s,a}(x, y) = \min_{\bar{y}} i_{s,a}(x, \bar{y})\} \quad (139)$$

for which the constraint $D(\hat{W}||W^*|Q_X) = \infty \leq r$ is never satisfied. Therefore, the inequality in (133) needs to be set with equality, and we will henceforth consider it with equality. The geometric interpretation is that, because the objective function is linear, the optimal solution is found at the boundaries of the ellipsoid defined by the constraint. Specifically, the objective function intersects the ellipsoid at two points, one of which corresponds to the minimum we are seeking.

As we have said before, the solution to (137) is not explicit. To achieve results that allow for conceptual understanding, we derive an approximation that is highly accurate as $r \rightarrow 0$:

- 1) We rely on the fact that $o(d(\theta))$ has a negligible impact for sufficiently small r . This is supported by [5], where it is stated that $d(\theta)$ well approximates $D(\hat{W}||W|Q_X)$ when \hat{W} is in a neighborhood of W . Therefore, we treat $o(d(\theta)) = \gamma$ as a small constant term independent of θ , absorbed in the constraint value via $\bar{r} \triangleq r - \gamma$. We then solve the following convex problem [6]

$$J(\bar{r}) = \min_{\substack{d(\theta) = \bar{r} \\ \theta^T \mathbf{1}_j = 0, 1 \leq j \leq |\mathcal{X}|}} \theta^T \nabla I_{s,a} \quad (140)$$

$$= \max_{\rho, \nu_1, \dots, \nu_{|\mathcal{X}|}} \min_{\theta} \mathcal{L}(\bar{r}) \quad (141)$$

where $\mathcal{L}(\bar{r})$ is the following Lagrangian

$$\mathcal{L}(\bar{r}) \triangleq \theta^T \nabla I_{s,a} + \rho(d(\theta) - \bar{r}) + \sum_j \nu_j \theta^T \mathbf{1}_j. \quad (142)$$

2) We expand $J(\bar{r}) = J(r - \gamma)$ for small γ as

$$J(\bar{r}) = J(r) - J'(r)\gamma \quad (143)$$

where the first term coincides with the minimum in (141) when $\bar{r} = r$ given by

$$J(r) = \max_{\rho, \nu_1, \dots, \nu_{|\mathcal{X}|}} \min_{\theta} \mathcal{L}(r), \quad (144)$$

and the derivative in the second term is precisely the optimal Lagrange multiplier associated with the constraint $d(\theta) = r$ in problem (144): $\rho^* = -J'(r)$.

Our approach yields $J(\bar{r}) = J(r) + \rho^* o(d(\theta))$, which, when evaluated at the optimal θ^* , produces $J(\bar{r}) = J(r) + \rho^* o(r)$.

Therefore, our approach establishes the following equivalence between the optimization problems when adopting the relative entropy ball \mathcal{B} in (21) or the chi-squared distance ball $\tilde{\mathcal{B}}$ in (33) for sufficiently small $r \geq 0$

$$\min_{W \in \mathcal{B}} \theta^T \nabla I_{s,a} = \min_{W \in \tilde{\mathcal{B}}} \theta^T \nabla I_{s,a} + \rho^* o(r) \quad (145)$$

where

$$\min_{W \in \tilde{\mathcal{B}}} \theta^T \nabla I_{s,a} \triangleq \max_{\rho, \nu_1, \dots, \nu_{|\mathcal{X}|}} \min_{\theta} \mathcal{L}(r). \quad (146)$$

A. The Solution

Problem (146) is linear with quadratic constraints, for which the KKT conditions are necessary and sufficient for optimality [6]. The stationary point equations are obtained by differentiating with respect to θ^T , ρ , and each ν_j , and equating to zero. This gives the following system of equations:

$$\theta^* = -\frac{\mathbf{K}^{-1}}{\rho^*} \left(\nabla I_{s,a} + \sum_j \nu_j^* \mathbf{1}_j \right) \quad (147)$$

$$\frac{1}{2} \theta^{*T} \mathbf{K} \theta^* = r \quad (148)$$

$$\theta^{*T} \mathbf{1}_j = 0 \quad \text{for } j = 1, \dots, |\mathcal{X}|. \quad (149)$$

We obtain

$$\nu_j^* = -Q_X(x_j) \mathbb{E}_{\hat{W}}[i_{s,a}(x_j, Y)] \quad (150)$$

$$\rho^* = \sqrt{V_{s,a}(Q_X, \hat{W})} / (2r) \quad (151)$$

$$V_{s,a}(Q_X, \hat{W}) \triangleq \mathbb{E}_{Q_X}[\text{Var}_{\hat{W}}[i_{s,a}(X, Y)|X]] \quad (152)$$

and

$$\theta^* = \frac{\mathbf{K}^{-1}}{\rho^*} \left(\sum_j Q_X(x_j) \mathbb{E}_{\hat{W}}[i_{s,a}(x_j, Y)] \mathbf{1}_j - \nabla I_{s,a} \right) \quad (153)$$

where the element $\theta^*(y|x)$ is extracted from θ^* (153) as

$$\theta^*(y|x) = -\sqrt{2r} \cdot \hat{W}(y|x) \varphi_{s,a}(x, y), \quad (154)$$

$$\varphi_{s,a}(x, y) \triangleq \frac{i_{s,a}(x, y) - \mathbb{E}_{\hat{W}}[i_{s,a}(x, Y)]}{\sqrt{V_{s,a}(Q_X, \hat{W})}}. \quad (155)$$

The worst-case channel distribution obtained by isolating $\tilde{W}_{s,a}^*$ from $\theta^*(y|x) = \tilde{W}_{s,a}^*(y|x) - \hat{W}(y|x)$ is

$$\tilde{W}_{s,a}^*(y|x) = \hat{W}(y|x) \left(1 - \sqrt{2r} \cdot \varphi_{s,a}(x, y) \right). \quad (156)$$

We first substitute (154) into (145), and account for the term $\rho^* o(r) = o(\sqrt{r})$. We then substitute the result into (130), yielding the following expression:

$$\begin{aligned} \underline{I}_{s,a}(Q_X, \hat{W}, r) &= \underline{I}_{s,a}^{\text{ML}}(Q_X, \hat{W}) \\ &\quad - \sqrt{2r V_{s,a}(Q_X, \hat{W})} + o(\sqrt{r}). \end{aligned} \quad (157)$$

It is straightforward to check that the solution corresponds to a minimum as the cost function is linear with quadratic constraints and $\rho^* > 0$.

B. Discussion

The only remaining issue to be proven is whether the worst-case channels are non-negative. This arises because the constraint $d(\theta) = r$ in (146) does not necessarily guarantee non-negativity, unlike the case with $D(\hat{W}||W|Q_X) = d(\theta) + o(d(\theta)) = r$ for $\hat{W} > 0$ in the original problem (137). Further analysis reveals that this issue is fully resolved when r is small, without the need to impose additional constraints.

To evidence the above, we have solved the full optimization problem including constraints to ensure that all valid conditional channel distributions are non-negative, as $W(y|x) \geq 0$ for $(x, y) \in \mathcal{X} \times \mathcal{Y}$, through the slack variables $\mu(x, y) \geq 0$ satisfying $\mu(x, y)W(y|x) = 0$. The solution can be expressed in terms of the following variables

$$i_{s,a}^\mu(x, y) \triangleq i_{s,a}(x, y) - \mu(x, y) \quad (158)$$

$$\varphi_{s,a}^\mu(y|x) \triangleq \frac{i_{s,a}^\mu(x, y) - \mathbb{E}_{\hat{W}}[i_{s,a}^\mu(x, Y)]}{\sqrt{V_{s,a}^\mu(Q_X, \hat{W})}} \quad (159)$$

$$V_{s,a}^\mu(Q_X, \hat{W}) \triangleq \mathbb{E}_{Q_X}[\text{Var}_{\hat{W}}[i_{s,a}^\mu(X, Y)|X]], \quad (160)$$

achieving the following minimum

$$\begin{aligned} \underline{I}_{s,a}(Q_X, \hat{W}, r) &= \max_{\mu(\cdot) \geq 0} \left\{ \mathbb{E}_{Q_X \times \hat{W}}[i_{s,a}^\mu(X, Y)] \right. \\ &\quad \left. - \sqrt{2r V_{s,a}^\mu(Q_X, \hat{W})} + o(\sqrt{r}) \right\} \end{aligned} \quad (161)$$

for the worst-case channel distribution

$$\tilde{W}_{s,a}^*(y|x) = \hat{W}(y|x) \left(1 - \sqrt{2r} \cdot \varphi_{s,a}^\mu(x, y) \right). \quad (162)$$

From the above expression, it is evident that when r is sufficiently small, the second term has a minimal impact on the structure of $\hat{W}(y|x) > 0$. This confirms the validity of our approach in the absence of positivity constraints for the regime of r values that make $\tilde{W}_{s,a}^*(y|x) \geq 0$ when $\mu(x, y) = 0$.

APPENDIX B PROOF OF THEOREM 2

We determine the worst-case Gallager E_0 functions

$$\underline{E}_0^{\text{iid}}(Q_X, \hat{W}, \rho, r) = \min_{W \in \mathcal{B}} \sup_{s \geq 0} E_{s,0,0,\rho}(Q_X, W) \quad (163)$$

$$\underline{E}_0^{\text{cc}}(Q_X, \hat{W}, \rho, r) = \min_{W \in \mathcal{B}} \sup_{c_1(\cdot), c_2(\cdot)} \sup_{s \geq 0, \lambda, \bar{\lambda}} E_{s, \lambda, \bar{\lambda}, \rho}(Q_X, W) \quad (164)$$

using analogous arguments to those for the proof of the worst-case LM rate in Appendix A. In particular, as $\underline{E}_0^{\text{iid}}$ can be obtained from $\underline{E}_0^{\text{cc}}$ by setting $\lambda = \bar{\lambda} = \mathbf{0}$ and by deleting the optimization over $c_1(\cdot), c_2(\cdot)$, we will only solve (164).

Firstly, as the problem is concave in $s \geq 0, c_1(\cdot), c_2(\cdot)$ and convex in $W \in \mathcal{B}$, we make use of Fan's minimax Theorem [24] to move the minimization as the inner problem. Hence, our approach reduces to solving

$$\underline{E}_{s, \lambda, \bar{\lambda}, \rho}(Q_X, \hat{W}, r) \triangleq \min_{W \in \mathcal{B}} -\log \mathbb{E}_{Q_X \times W}[\varepsilon_{s, \lambda, \bar{\lambda}, \rho}(X, Y)]. \quad (165)$$

Now, by adding and subtracting the term $E_{s, \lambda, \bar{\lambda}, \rho}^{\text{ML}}(Q_X, \hat{W}) \triangleq -\log \mathbb{E}_{Q_X \times \hat{W}}[\varepsilon_{s, \lambda, \bar{\lambda}, \rho}(X, Y)]$ and using $\theta(y|x) = W(y|x) - \hat{W}(y|x)$, problem (165) can be rewritten as

$$\begin{aligned} \underline{E}_{s, \lambda, \bar{\lambda}, \rho}(Q_X, \hat{W}, r) &= E_{s, \lambda, \bar{\lambda}, \rho}^{\text{ML}}(Q_X, \hat{W}) \\ &\quad - \log \left(1 + \max_{W \in \mathcal{B}} \sum_{x, y} \frac{Q_X(x) \theta(y|x) \varepsilon_{s, \lambda, \bar{\lambda}, \rho}(x, y)}{\mathbb{E}_{Q_X \times \hat{W}}[\varepsilon_{s, \lambda, \bar{\lambda}, \rho}(X, Y)]} \right). \end{aligned} \quad (166)$$

Note that the inner maximization problem,

$$\max_{W \in \mathcal{B}} \sum_{x, y} \frac{Q_X(x) \theta(y|x) \varepsilon_{s, \lambda, \bar{\lambda}, \rho}(x, y)}{\mathbb{E}_{Q_X \times \hat{W}}[\varepsilon_{s, \lambda, \bar{\lambda}, \rho}(X, Y)]}, \quad (167)$$

has the same form as problem (137) in Appendix A, but with a substitution of $\varepsilon_{s, \lambda, \bar{\lambda}, \rho}(x, y) / \mathbb{E}_{Q_X \times \hat{W}}[\varepsilon_{s, \lambda, \bar{\lambda}, \rho}(X, Y)]$ by $i_{s, a}(x, y)$ and a change from minimization to maximization.

The solution is obtained straightforwardly by mimicking the derivation in Appendix A. This gives

$$\theta^*(y|x) = \sqrt{2r} \cdot \frac{\hat{W}(y|x) \varphi_{s, \lambda, \bar{\lambda}, \rho}(x, y)}{\mathbb{E}_{Q_X \times \hat{W}}[\varepsilon_{s, \lambda, \bar{\lambda}, \rho}(X, Y)]} \quad (168)$$

with

$$\varphi_{s, \lambda, \bar{\lambda}, \rho}(x, y) \triangleq \frac{\varepsilon_{s, \lambda, \bar{\lambda}, \rho}(x, y) - \mathbb{E}_{\hat{W}}[\varepsilon_{s, \lambda, \bar{\lambda}, \rho}(x, Y)]}{\sqrt{V_{s, \lambda, \bar{\lambda}, \rho}(Q_X, \hat{W})}} \quad (169)$$

$$V_{s, \lambda, \bar{\lambda}, \rho}(Q_X, \hat{W}) \triangleq \frac{\mathbb{E}_{Q_X}[\text{Var}_{\hat{W}}[\varepsilon_{s, \lambda, \bar{\lambda}, \rho}(X, Y)|X]]}{\mathbb{E}_{Q_X \times \hat{W}}^2[\varepsilon_{s, \lambda, \bar{\lambda}, \rho}(X, Y)]} \quad (170)$$

achieving the following objective

$$\begin{aligned} \underline{E}_{s, \lambda, \bar{\lambda}, \rho}(Q_X, \hat{W}, r) &= E_{s, \lambda, \bar{\lambda}, \rho}^{\text{ML}}(Q_X, \hat{W}) \\ &\quad - \log \left(1 + \sqrt{2r V_{s, \lambda, \bar{\lambda}, \rho}(Q_X, \hat{W})} + o(\sqrt{r}) \right) \end{aligned} \quad (171)$$

$$= E_{s, \lambda, \bar{\lambda}, \rho}^{\text{ML}}(Q_X, \hat{W})$$

$$- \log \left(1 + \sqrt{2r V_{s, \lambda, \bar{\lambda}, \rho}(Q_X, \hat{W})} \right) + o(\sqrt{r}) \quad (172)$$

where for the second inequality we have used that $\log(x + o(\sqrt{r})) = \log(x) + o(\sqrt{r})$.

APPENDIX C THE CASE OF A SYMMETRIC \hat{W}

For a symmetric (as Gallager [1, Pag. 94]) estimated channel \hat{W} and an equiprobable input distribution $Q_X(x) = |\mathcal{X}|^{-1}$, expressions can be obtained in terms of

$$\kappa_t \triangleq \sum_x \hat{W}(y|x)^t = \sum_y \hat{W}_{\text{sym}}(y)^t \quad (173)$$

where $\hat{W}_{\text{sym}}(y) \equiv \hat{W}_{\text{sym}}(y|x_1)$ denotes the first row of \hat{W} . The second equality in (173) holds due to the symmetry of \hat{W} .

A. Computations Relative to the Achievable Rate

The mismatched information density reads

$$\begin{aligned} i_{s, 0}(x, y) &= \log \frac{\hat{W}(y|x)^s}{\frac{1}{|\mathcal{X}|} \sum_{\bar{x}} \hat{W}(y|\bar{x})^s} \\ &= \log \frac{|\mathcal{X}|}{\kappa_s} + \log \hat{W}(y|x)^s. \end{aligned} \quad (174)$$

Therefore,

$$\begin{aligned} I_{s, 0}^{\text{ML}}(Q_X, \hat{W}) &= \mathbb{E}_{Q_X \times \hat{W}}[i_{s, 0}(X, Y)] \\ &= \log \frac{|\mathcal{X}|}{\kappa_s} + s \cdot \mathbb{E}_{\hat{W}_{\text{sym}}}[\log \hat{W}_{\text{sym}}] \end{aligned} \quad (175)$$

and the expansion terms proportional to \sqrt{r} are

$$\begin{aligned} V_{s, 0}(Q_X, \hat{W}) &= \mathbb{E}_{Q_X}[\text{Var}_{\hat{W}}[i_{s, 0}(X, Y)|X]] \\ &= \mathbb{E}_{Q_X}[\text{Var}_{\hat{W}}[s \cdot \log \hat{W}(Y|X)|X]] \\ &= s^2 \left(\mathbb{E}_{\hat{W}_{\text{sym}}}[\log^2 \hat{W}_{\text{sym}}] - \mathbb{E}_{\hat{W}_{\text{sym}}}^2[\log \hat{W}_{\text{sym}}] \right). \end{aligned} \quad (176)$$

B. Computations Relative to the Gallager E_0 Function

The mismatched exponent density reads

$$\begin{aligned} \varepsilon_{s, \mathbf{0}, \mathbf{0}, \rho}(x, y) &= \left(\frac{\frac{1}{|\mathcal{X}|} \sum_{\bar{x}} \hat{W}(y|\bar{x})^s}{\hat{W}(y|x)^s} \right)^\rho \\ &= \frac{\kappa_s^\rho}{|\mathcal{X}|^\rho} \hat{W}(y|x)^{-s\rho}. \end{aligned} \quad (177)$$

Therefore

$$\begin{aligned} E_{s, \mathbf{0}, \mathbf{0}, \rho}^{\text{ML}}(Q_X, \hat{W}) &= -\log \mathbb{E}_{Q_X \times \hat{W}}[\varepsilon_{s, \mathbf{0}, \mathbf{0}, \rho}(X, Y)] \\ &= \log \frac{|\mathcal{X}|^\rho}{\kappa_s^\rho \kappa_{1-s\rho}} \end{aligned} \quad (178)$$

and the expansion terms proportional to \sqrt{r} are

$$\begin{aligned} V_{s, \mathbf{0}, \mathbf{0}, \rho}(Q_X, \hat{W}) &= \frac{\mathbb{E}_{Q_X}[\text{Var}_{\hat{W}}[\varepsilon_{s, \mathbf{0}, \mathbf{0}, \rho}(X, Y)|X]]}{\mathbb{E}_{Q_X \times \hat{W}}^2[\varepsilon_{s, \mathbf{0}, \mathbf{0}, \rho}(X, Y)]} \\ &= \frac{\mathbb{E}_{Q_X}[\text{Var}_{\hat{W}}[\hat{W}(Y|X)^{-s\rho}|X]]}{\mathbb{E}_{Q_X \times \hat{W}}^2[\hat{W}(Y|X)^{-s\rho}]} \\ &= \frac{\kappa_{1-2s\rho}^2 - \kappa_{1-s\rho}^2}{\kappa_{1-s\rho}^2}. \end{aligned} \quad (179)$$

APPENDIX D PROOF OF THEOREM 3

The continuous-alphabet case is similar to the discrete-alphabet case, but now we work with probability density functions instead of probability mass functions, and integrals instead of summations. Therefore, we resort to the calculus of variations to address the minimization in the function space.

We want to solve the following variational problem

$$\min_{W \in \mathcal{B}} \int \int_{\mathbb{R}^2} Q_X(x) W(y|x) i_{s,\lambda}(x, y) dx dy \quad (180)$$

where the constraint $W \in \mathcal{B}$ considers conditional probability density functions $W(y|x) \geq 0$ that satisfy

$$\begin{cases} d(\widehat{W} \| W | Q_X) + o(d(\widehat{W} \| W | Q_X)) & \leq r \\ \int_{\mathbb{R}} W(y|x) dy & = 1 \text{ for } x \in \mathbb{R} \end{cases} \quad (181)$$

with

$$d(\widehat{W} \| W | Q_X) \triangleq \int \int_{\mathbb{R}^2} Q_X(x) \frac{(W(y|x) - \widehat{W}(y|x))^2}{2\widehat{W}(y|x)} dx dy. \quad (182)$$

For sufficiently small r , the same trick as in (141) can be applied to move the $o(\cdot)$ term in the constraint as a penalty term in the objective function. We will omit the technical derivations, as it follows the same rules outlined in Appendix A, and focus on solving the following variational problem

$$\min_{\substack{d(\widehat{W} \| W | Q_X) = r \\ \int_{\mathbb{R}} W(y|x) dy = 1}} \int \int_{\mathbb{R}^2} Q_X(x) W(y|x) i_{s,\lambda}(x, y) dx dy + \rho^* o(r) \quad (183)$$

where the differentiability of the objective functional and constraints is assumed under an appropriately chosen norm for channels with continuous derivatives, as functionals are either linear or quadratic in W .

Now, we proceed by minimizing the Lagrangian

$$\mathcal{L}[W] = \int \int_{\mathbb{R}^2} F(x, y, W(y|x)) dx dy \quad (184)$$

with

$$\begin{aligned} F(x, y, W) &\triangleq Q_X(x) W i_{s,\lambda}(x, y) + v(x) W \\ &\quad + \rho Q_X(x) \frac{(W - \widehat{W}(y|x))^2}{2\widehat{W}(y|x)} \end{aligned} \quad (185)$$

and where we have obviated W -independent constant terms. We follow the rules in [7, Sec. V], where the stationary point $W^*(y|x)$ is found by solving the Euler-Lagrange equation

$$\frac{\partial F}{\partial W} - \frac{\partial}{\partial x} \frac{\partial F}{\partial W_x} - \frac{\partial}{\partial y} \frac{\partial F}{\partial W_y} = 0 \quad \text{in } (x, y) \in \mathbb{R}^2. \quad (186)$$

In our case, since F in (185) does not depend on the partial derivatives of W , denoted W_x and W_y , the above reduces to solving $\frac{\partial F}{\partial W} = 0$ in $(x, y) \in \mathbb{R}^2$ for $\{W^*(y|x), v^*(x), \rho^*\}$ jointly with the problem constraints. Specifically, we have

$$W^*(y|x) = \widehat{W}(y|x) \left(1 - \frac{Q_X(x) i_{s,\lambda}(x, y) + v^*(x)}{\rho^* Q_X(x)} \right) \quad (187)$$

with

$$v^*(x) \triangleq -Q_X(x) \mathbb{E}_{\widehat{W}}[i_{s,\lambda}(x, Y)] \quad (188)$$

$$\rho^* \triangleq \sqrt{V_{s,\lambda}(Q_X, \widehat{W}) / (2r)} \quad (189)$$

$$V_{s,\lambda}(Q_X, \widehat{W}) \triangleq \mathbb{E}_{Q_X}[\text{Var}_{\widehat{W}}[i_{s,\lambda}(X, Y)|X]]. \quad (190)$$

The stationary point $W^*(y|x)$ corresponds to the global minimum as the problem is linear in W subject to quadratic constraints in W and $\rho^* > 0$.

APPENDIX E GAUSSIAN CODEBOOKS AND NEAREST NEIGHBOR DECODING

For Gaussian i.i.d. codebooks $Q_X(x) = \mathcal{N}(x; P)$ and the nearest neighbor decoder $\widehat{W}(y|x) = \mathcal{N}(y-x; \hat{\sigma}^2)$, we have

$$\widehat{W}(y|x)^s = \alpha_s \cdot \mathcal{N}(y-x; \hat{\sigma}^2 s^{-1}) \quad (191)$$

$$\mathbb{E}_{Q_X}[\widehat{W}(y|X)^s] = \alpha_s \cdot \mathcal{N}(y; \hat{\sigma}^2 s^{-1} + P) \quad (192)$$

with $\alpha_s \triangleq \sqrt{s^{-1}(2\pi\hat{\sigma}^2)^{1-s}}$.

A. Achievable Rate With Gaussian i.i.d. Codebooks

The mismatched information density is (cf. [27, (28)-(29)])

$$i_{s,0}(x, y) = C_1 - \frac{s(y-x)^2}{2\hat{\sigma}^2} + \frac{1}{2} \frac{y^2}{\hat{\sigma}^2 s^{-1} + P} \quad (193)$$

with

$$C_1 \triangleq \frac{1}{2} \log \left(1 + \frac{sP}{\hat{\sigma}^2} \right), \quad (194)$$

and the respective expectations are

$$\mathbb{E}_{\widehat{W}}[i_{s,0}(x, Y)] = C_1 + \frac{1}{2} \frac{x^2 - sP}{\hat{\sigma}^2 s^{-1} + P} \quad (195)$$

$$\mathbb{E}_{Q_X \times \widehat{W}}[i_{s,0}(X, Y)] = C_1 + \frac{P}{2} \frac{1-s}{\hat{\sigma}^2 s^{-1} + P}. \quad (196)$$

Note that $I_{s,0}^{\text{ML}}(Q_X, \widehat{W}) = \mathbb{E}_{Q_X \times \widehat{W}}[i_{s,0}(X, Y)]$.

Now we turn our attention to the term $V_{s,0}(Q_X, \widehat{W}) = \mathbb{E}_{Q_X \times \widehat{W}}[i_{s,0}^2(X, Y)] - \mathbb{E}_{Q_X}[\mathbb{E}_{\widehat{W}}^2[i_{s,0}(X, Y)|X]]$. After tedious but otherwise straightforward calculations, we find its two terms as follows

$$\begin{aligned} \mathbb{E}_{Q_X \times \widehat{W}}[i_{s,0}^2(X, Y)] &= K_1 + \frac{3s^2}{4} + \frac{3(\hat{\sigma}^2 + P)^2}{4(\hat{\sigma}^2 s^{-1} + P)^2} \\ &\quad - \frac{1}{2} \frac{s(3\hat{\sigma}^2 + P)}{(\hat{\sigma}^2 s^{-1} + P)^2} \end{aligned} \quad (197)$$

$$\mathbb{E}_{Q_X}[\mathbb{E}_{\widehat{W}}^2[i_{s,0}(X, Y)|X]] = K_1 + \frac{P^2(3-2s+s^2)}{4(\hat{\sigma}^2 s^{-1} + P)^2} \quad (198)$$

where $K_1 \triangleq C_1^2 + C_1 P(1-s)(\hat{\sigma}^2 s^{-1} + P)^{-2}$, leading to

$$V_{s,0}(Q_X, \widehat{W}) = \frac{P}{2} \frac{2\hat{\sigma}^2 + s^2 P}{(\hat{\sigma}^2 s^{-1} + P)^2}. \quad (199)$$

B. E_0 Function With Gaussian i.i.d. Codebooks

The expressions for the worst-case mismatched Gallager E_0 functions can be obtained from

$$\mathbb{E}'_{Q_X}[\widehat{W}(y|X)^s] \triangleq \alpha'_s \frac{\sqrt{2\pi P_s t^{-1}}}{\sqrt{(2\pi P_s)^t}} \cdot \mathcal{N}(y; P_s t^{-1}) \quad (200)$$

with $P_s \triangleq \hat{\sigma}^2 s^{-1} + P$, and

$$A(s, t) \triangleq \int_{\mathbb{R}} \mathbb{E}'_{Q_X}[\widehat{W}(y|X)^s] \mathbb{E}_{Q_X}[\widehat{W}(y|X)^{1-st}] dy$$

$$\begin{aligned}
&= \int_{\mathbb{R}} \frac{\alpha_s^t \alpha_{1-st}}{\sqrt{(2\pi P_s)^t}} \sqrt{\frac{P_s}{P_s + tP_{1-st}}} \mathcal{N}\left(y; \frac{P_s P_{1-st}}{P_s + tP_{1-st}}\right) dy \\
&= \frac{\alpha_s^t \alpha_{1-st}}{\sqrt{(2\pi P_s)^t}} \sqrt{\frac{P_s}{P_s + tP_{1-st}}} \\
&= \sqrt{\left(1 + \frac{P_s}{\hat{\sigma}^2}\right)^{-t} \left(1 + \frac{tP(1-s-ts)}{P + \hat{\sigma}^2 s^{-1}}\right)^{-1}}. \quad (201)
\end{aligned}$$

Therefore

$$\begin{aligned}
E_{s,0,0,\rho}^{\text{ML}}(Q_X, \hat{W}) &= -\log A(s, \rho) \\
&= \frac{\rho}{2} \log\left(1 + \frac{P_s}{\hat{\sigma}^2}\right) \\
&\quad + \frac{1}{2} \log\left(1 + \frac{\rho P(1-s-\rho s)}{P + \hat{\sigma}^2 s^{-1}}\right). \quad (202)
\end{aligned}$$

The dispersion term $V_{s,0,0,\rho}(Q_X, \hat{W})$ is computed as the following sum:

$$V_{s,0,0,\rho}(Q_X, \hat{W}) = V_1(Q_X, \hat{W}) - V_2(Q_X, \hat{W}). \quad (203)$$

The first term is

$$\begin{aligned}
V_1(Q_X, \hat{W}) &\triangleq \frac{\mathbb{E}_{Q_X \times \hat{W}}[\varepsilon_{s,0,0,\rho}^2(X, Y)]}{\mathbb{E}_{Q_X \times \hat{W}}^2[\varepsilon_{s,0,0,\rho}(X, Y)]} \\
&= \frac{A(s, 2\rho)}{A(s, \rho)^2} \\
&= \frac{1}{\sqrt{P_s}} \frac{P + \hat{\sigma}^2 s^{-1} + \rho P(1-s-\rho s)}{\sqrt{P + \hat{\sigma}^2 s^{-1} + 2\rho P(1-s-2\rho s)}}. \quad (204)
\end{aligned}$$

The second term is

$$V_2(Q_X, \hat{W}) = \frac{\mathbb{E}_{Q_X}[\mathbb{E}_{\hat{W}}^2[\varepsilon_{s,0,0,\rho}(X, Y)|X]]}{\mathbb{E}_{Q_X \times \hat{W}}^2[\varepsilon_{s,0,0,\rho}(X, Y)]}. \quad (205)$$

We compute the numerator through the steps

$$\mathbb{E}_{\hat{W}}^2[\varepsilon_{s,0,0,\rho}(x, Y)] = \frac{\alpha_{1-\rho s}^2 \alpha_s^{2\rho}}{\rho(2\pi P_s)^{\rho-1}} \mathcal{N}^2\left(x; \frac{P_s - \rho s P}{\rho(1-s\rho)}\right) \quad (206)$$

$$\mathbb{E}_{Q_X} \mathbb{E}_{\hat{W}}^2[\varepsilon_{s,0,0,\rho}(X, Y)|X] = \frac{1}{s\rho} \frac{\sqrt{2}(\hat{\sigma}^2 s^{-1})^\rho P_s^{1-\rho}}{P_s + \rho P(2-s-2s\rho)} \quad (207)$$

and thus

$$V_2(Q_X, \hat{W}) = \frac{\sqrt{2}}{\rho s} \frac{P + \hat{\sigma}^2 s^{-1} + \rho P(1-s-\rho s)}{P + \hat{\sigma}^2 s^{-1} + \rho P(2-s-2\rho s)}. \quad (208)$$

C. Gaussian Spherical Codebooks

For spherical codebooks, we have

$$\mathbb{E}_{Q_X}[\hat{W}^s(y|X)e^{\lambda X^2}] = \frac{\alpha}{\sqrt{g(P)}} \cdot \mathcal{N}(y; \hat{\sigma}^2 s^{-1} + P') \quad (209)$$

and the auxiliary variables

$$C_2 \triangleq \frac{1}{2} \log\left(g(P) + \frac{sP}{\hat{\sigma}^2}\right) \quad (210)$$

$$P' \triangleq P/g(P) \quad (211)$$

$$g(P) \triangleq 1 - 2\lambda P. \quad (212)$$

The mismatched information density is

$$i_{s,\lambda}(x, y) = \lambda x^2 + C_2 - \frac{s(y-x)^2}{2\hat{\sigma}^2} + \frac{1}{2} \frac{y^2}{\hat{\sigma}^2 s^{-1} + P'} \quad (213)$$

and

$$\mathbb{E}_{\hat{W}}[i_{s,\lambda}(x, Y)] = \lambda x^2 + C_2 + \frac{1}{2} \frac{x^2 - sP'}{\hat{\sigma}^2 s^{-1} + P'} \quad (214)$$

$$\mathbb{E}_{Q_X \times \hat{W}}[i_{s,\lambda}(X, Y)] = \lambda P + C_2 + \frac{1}{2} \frac{P - sP'}{\hat{\sigma}^2 s^{-1} + P'}. \quad (215)$$

After straightforward calculations, we can write

$$\begin{aligned}
I_{s,\lambda}^{\text{ML}}(Q_X, \hat{W}) &= \frac{1}{2} \log\left(g(P) + \frac{sP}{\hat{\sigma}^2}\right) \\
&\quad + \frac{P(1-s) + \hat{\sigma}^2 s^{-1} g(P)(1-g(P))}{2(P + \hat{\sigma}^2 s^{-1} g(P))} \quad (216)
\end{aligned}$$

$$V_{s,\lambda}(Q_X, \hat{W}) = \frac{P}{2} \frac{2\hat{\sigma}^2 g^2(P) + s^2 P}{(\hat{\sigma}^2 s^{-1} g(P) + P)^2}. \quad (217)$$

REFERENCES

- [1] R. G. Gallager, *Information Theory and Reliable Communication*. Hoboken, NJ, USA: Wiley, 1968.
- [2] A. Lapidoth and P. Narayan, "Reliable communication under channel uncertainty," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2148–2177, Oct. 1998.
- [3] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 35–43, Jan. 1995.
- [4] J. Scarlett, A. Guillén i Fàbregas, A. Somekh-Baruch, and A. Martinez, "Information-theoretic foundations of mismatched decoding," *Found. Trends Commun. Inf. Theory*, vol. 17, no. 2–3, pp. 149–401, 2020.
- [5] S. Borade and L. Zheng, "Euclidean information theory," in *Proc. IEEE Int. Zürich Seminar Commun.*, Mar. 2008, pp. 14–17.
- [6] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [7] I. M. Gelfand et al., *Calculus of Variations*. Chelmsford, U.K.: Courier Corporation, 2000.
- [8] P. Boroumand and A. Guillén i Fàbregas, "Mismatched binary hypothesis testing: Error exponent sensitivity," *IEEE Trans. Inf. Theory*, vol. 68, no. 10, pp. 6738–6761, Oct. 2022.
- [9] A. Kaplan and S. Shamai (Shitz), "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *Arch. Elektronik Übertragungstechnik*, vol. 47, no. 4, pp. 228–239, Nov. 1993.
- [10] J. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., MIT, Cambridge, MA, USA, 1983.
- [11] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. IT-27, no. 1, pp. 5–12, Jan. 1981.
- [12] E. Asadi Kangarshahi and A. Guillén i Fàbregas, "A single-letter upper bound to the mismatch capacity," *IEEE Trans. Inf. Theory*, vol. 67, no. 4, pp. 2013–2033, Apr. 2021.
- [13] A. Somekh-Baruch, "A single-letter upper bound on the mismatch capacity via multicast transmission," *IEEE Trans. Inf. Theory*, vol. 68, no. 5, pp. 2801–2812, May 2022.
- [14] J. Scarlett, A. Martinez, 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, May 2014.
- [15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 2006.
- [16] A. Lapidoth and I. E. Telatar, "The compound channel capacity of a class of finite-state channels," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, pp. 973–983, May 1998.
- [17] S. Loyka and C. D. Charalambous, "A general formula for compound channel capacity," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, pp. 3971–3991, Jul. 2016.
- [18] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, May 1996.
- [19] A. L. Gibbs and F. E. Su, "On choosing and bounding probability metrics," *Int. Stat. Rev.*, vol. 70, no. 3, pp. 419–435, 2002.
- [20] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *J. Roy. Stat. Soc., Ser. B Methodol.*, vol. 28, no. 1, pp. 131–142, Jan. 1966.

- [21] I. Csiszár, "Information-type measures of difference of probability distributions and indirect observation," *Studia Scientiarum Mathematicarum Hungarica*, vol. 2, pp. 229–318, Jan. 1967.
- [22] P. Harremoës and T. van Erven, "Rényi divergence and Kullback–Leibler divergence," *IEEE Trans. Inf. Theory*, vol. 60, pp. 3797–3820, 2014.
- [23] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. IT-18, no. 1, pp. 14–20, Jan. 1972.
- [24] K. Fan, "Minimax Theorems," *Proc. Nat. Acad. Sci. USA*, vol. 39, no. 1, pp. 42–47, 1953.
- [25] F. Molina, P. Patel, and A. Guillén i Fàbregas, "Nearest neighbor decoding for a class of compound channels," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Jul. 2024, pp. 1510–1513.
- [26] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.
- [27] J. Scarlett, V. Y. F. Tan, and G. Durisi, "The dispersion of nearest-neighbor decoding for additive non-Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, pp. 81–92, Jan. 2017.

Priyanka Patel received the B.A. and Master of Engineering degrees in information, computer and electrical engineering from the University of Cambridge in 2023. Since 2023, she has been a Software Engineer at Google, London. Her broad research interests include information theory and communication theory. She received the 2023 IET Student Prize for her Masters work.

Francesc Molina (Member, IEEE) was born in Barcelona, Spain, in 1993. He received the M.Sc. degree in telecommunications engineering and the Ph.D. degree in signal theory and communications from the Universitat Politècnica de Catalunya (UPC), Spain, in 2017 and 2021, respectively. From 2018 to 2022, he was with the Signal Processing and Communications Group, UPC, first as a Ph.D. Researcher and later as a Post-Doctoral Researcher. In 2023, he joined the Signal Processing and Communications Laboratory, University of Cambridge, U.K., as a Post-Doctoral Research Fellow. Since 2024, he has been a Post-Doctoral Researcher with the Information Theory and Coding Group, Universitat Pompeu Fabra, Spain. His main research interests span the fields of wireless communications and information theory. He was a recipient of the Best Ph.D. Thesis Award from Spanish Chapter of the IEEE Aerospace and Electronic Systems Society.

Albert Guillén i Fàbregas (Fellow, IEEE) received the degree in telecommunications engineering from the Universitat Politècnica de Catalunya, the degree in electronics engineering from the Politecnico di Torino, in 1999, and the Ph.D. degree in communication systems from the École Polytechnique Fédérale de Lausanne (EPFL) in 2004.

He is a Professor at the Department of Engineering, University of Cambridge and a Fellow of the Emmanuel College and a part-time Researcher at the Universitat Politècnica de Catalunya. He has held appointments at New Jersey Institute of Technology, Telecom Italia, European Space Agency (ESA), Institut Eurécom, University of South Australia, Universitat Pompeu Fabra (ICREA Research Professor), as well as visiting appointments at EPFL, École Nationale des Télécommunications (Paris), Universitat Pompeu Fabra, University of South Australia, Centrum Wiskunde and Informatica, and Texas A&M University at Qatar. His research interests include information theory, communication theory, coding theory, and statistical inference.

Dr. Guillén i Fàbregas is a fellow of the Institute for Mathematics and its Applications (IMA) and a member of the Young Academy of Europe. He received the Starting, Consolidator, Advanced, and Proof of Concept Grants from European Research Council, the Young Authors Award of the 2004 European Signal Processing Conference, the 2004 Best Doctoral Thesis Award from Spanish Institution of Telecommunications Engineers, and a Research Fellowship from Spanish Government to join ESA. Since 2013, he has been an Editor of the Foundations and Trends in Communications and Information Theory, Now Publishers and was an Associate Editor of IEEE TRANSACTIONS ON INFORMATION THEORY (2013–2020) and IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS (2007–2011).