A Toolbox for Refined Information-Theoretic Analyses with Applications

Other titles in Foundations and Trends® in Communications and Information Theory

Channel Simulation: Theory and Applications to Lossy Compression and Differential Privacy
Cheuk Ting Li
ISBN: 978-1-63828-486-4

Maximizing Entropy with an Expectation Constraint and One-Parameter Exponential Families of Distributions: A Reexamination
David L. Neuhoff
ISBN: 978-1-63828-480-2

Codes for Adversaries: Between Worst-Case and Average-Case Jamming
Bikash Kumar Dey, Sidharth Jaggi, Michael Langberg, Anand D. Sarwate and Yihan Zhang
ISBN: 978-1-63828-460-4

Universal Features for High-Dimensional Learning and Inference
Shao-Lun Huang, Anuran Makur, Gregory W. Wornell and Lizhong Zheng
ISBN: 978-1-63828-176-4

Ultra-Reliable Low-Latency Communications: Foundations, Enablers, System Design, and Evolution Towards 6G
Nurul Huda Mahmood, Italo Atzeni, Eduard Axel Jorswieck and Onel Luis Alcaraz López
ISBN: 978-1-63828-180-1

Probabilistic Amplitude Shaping
Georg Böcherer
ISBN: 978-1-63828-178-8

A Toolbox for Refined Information-Theoretic Analyses with Applications

Neri Merhav
Technion – Israel Institute of Technology
merhav@ee.technion.ac.il

Nir Weinberger
Technion – Israel Institute of Technology
nirwein@technion.ac.il

the essence of knowledge
Boston — Delft

Foundations and Trends® in Communications and Information Theory

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

N. Merhav and N. Weinberger. *A Toolbox for Refined Information-Theoretic Analyses with Applications*. Foundations and Trends® in Communications and Information Theory, vol. 22, no. 1, pp. 1–184, 2025.

ISBN: 978-1-63828-501-4

Full text available at: http://dx.doi.org/10.1561/0100000142

Foundations and Trends® in Communications and Information Theory

Volume 22, Issue 1, 2025

Editorial Board

Editorial Scope

Foundations and Trends® in Communications and Information Theory publishes survey and tutorial articles in the following topics:

- Coded modulation
- Coding theory and practice
- Communication complexity
- Communication system design
- Cryptology and data security
- Data compression
- Data networks
- Demodulation and Equalization
- Denoising
- Detection and estimation
- Information theory and statistics
- Information theory and computer science
- Joint source/channel coding
- Modulation and signal design
- Multiuser detection
- Multiuser information theory
- Optical communication channels
- Pattern recognition and learning
- Quantization
- Quantum information processing
- Rate-distortion theory
- Shannon theory
- Signal processing for communications
- Source coding
- Storage and recording codes
- Speech and Image Compression
- Wireless Communications

Information for Librarians

Foundations and Trends® in Communications and Information Theory, 2025, Volume 22, 4 issues. ISSN paper version 1567-2190. ISSN online version 1567-2328 . Also available as a combined paper and online subscription.

Contents

A Toolbox for Refined Information-Theoretic Analyses with Applications

Neri Merhav and Nir Weinberger

Technion – Israel Institute of Technology, Israel;
merhav@ee.technion.ac.il, nirwein@technion.ac.il

ABSTRACT

This monograph offers a toolbox of mathematical techniques that have been effective and widely applicable in information-theoretic analyses. The first tool is a generalization of the method of types to Gaussian settings, and then to general exponential families. The second tool is Laplace and saddle-point integration, which allow to refine the results of the method of types, and can obtain various precise asymptotic results. The third is the type class enumeration method, a principled method to evaluate the exact random-coding exponent of coded systems, which results in the best known exponent in various problems. The fourth is a subset of tools aimed at evaluating the expectation of non-linear functions of random variables, either via integral representations, by a refinement of Jensen's inequality via change-of-measure, by complementing Jensen's inequality with a reversed inequality, or by a class of generalized Jensen's inequalities that are applicable for functions beyond convex/concave. Various examples of all these tools are provided throughout the monograph.

Neri Merhav and Nir Weinberger (2025), "A Toolbox for Refined Information-Theoretic Analyses with Applications", Foundations and Trends® in Communications and Information Theory: Vol. 22, No. 1, pp 1–184. DOI: 10.1561/0100000142.

1

Introduction

This monograph is concerned with a set of analytical tools for information-theoretic analyses. The use of analytical methods to address challenging combinatorial problems is a classical method in mathematics, and includes various widely used techniques such as Stirling's approximation, Chernoff's bound, transform methods (with interchanging summation or integration order), among others. Analytical techniques also formed the basis of the inception of information-theory by Shannon [182]: On the face of it, and even at a deeper look, efficient coding for noisy channels is a formidable combinatorial problem, in a high dimensional space. Shannon addressed that challenge using analytical techniques:

1. The asymptotic equipartition property, and the estimation of volumes in high dimensional spaces, which allows to evaluate the size of high-probability sets. In the proof of the *noisy channel coding theorem* for discrete memoryless channels (DMCs), this allows to show that when an n-dimensional codeword is transmitted, the set of likely outputs has size roughly given by $e^{nH(Y|X)}$, where $H(Y|X)$ is the conditional entropy of the channel output Y conditioned on the input X, and the total set of likely outputs has roughly size of $e^{nH(Y)}$ (where $H(Y)$ is the entropy of Y).

2. The random-coding argument, which establishes the existence of optimal codes by evaluating the ensemble-average of randomly chosen code, and forms the basis for achievability (direct) results.

3. Convexity of information-measures, which is used to establish data-processing theorems, and consequently forms the basis for impossibility (converse) results.

Combining these ideas directly led, among other results, to the analytical formula for the capacity of DMCs, given by $C = \max_{P_X} I(X;Y)$ (where $I(X;Y) = H(Y) - H(Y|X)$ is the mutual information). Since Shannon's work, these ideas have been continuously extended and refined in numerous ways.

The goal of this monograph is to follow this path and propose a set of advanced analytical tools that are affirmed to be efficient and widely applicable for information-theoretic problems, allowing to obtain accurate and refined performance measure characterizations. Sections 2 and 3 to follow address the problem of estimating volumes in high dimensions, first, via a generalized method of types and, second, via the more advanced saddle-point method; Section 4 describes the *type class enumeration method* (TCEM), a tight analysis method of the performance of random-coding ensembles, and Section 5 considers various aspects of convexity and Jensen's inequality, mostly related to the computation of the expected values of non-linear functions. We next describe each one of these with more detail.

In Section 2, we describe a generalization of the method of types [38], [41], which was originally developed for finite alphabets, to Gaussian distributions, which are distributions over a continuous alphabet, and more generally, to distributions from exponential families. We introduce the notion of a typical set with respect to (WRT) a given parametric family of probability distributions. Such typical sets are defined in a way that the probability of each vector in the set is roughly the same for all possible distributions in the defined parametric family. This generalizes both the notion of weak typicality (a family consisting of a single distribution), and the notion of strong typicality for finite alphabets (the family is the set of all possible PMFs). Moreover, it allows to define, *e.g.*, typical sets for the Gaussian distribution. A key property of typical sets is their *volume*, because if an event of interest can be represented as the union over typical sets, then its probability can be accurately determined on the exponential scale using the volumes of these sets, and the probability of a single representative element from each of these sets. We thus develop a general method to evaluate the volumes of typical sets, and demonstrate its use on memoryless Gaussian sources, on Gaussian sources conditioned on other vectors, and on Gaussian sources with memory. We then generalize this method to distributions from an exponential family.

While the method of types is a general and widely applicable approach that leads to useful exponential bounds, there are settings which require more delicate analysis, and thus, more advanced tools. In Section 3, we begin by describing the Laplace method of integration, and exemplify its use in the problems of universal coding and extreme-value statistics. We then discuss the closely-related saddle-point method of integration in the complex plane, and show how it allows to accurately evaluate the size of type classes, volumes of hyper-spheres, and large-deviations probabilities, not only in the exact exponential rate, but also with the exact pre-exponential factor. We show that this method is applicable beyond parametric models. We further demonstrate its use for the evaluation of the number of lattice points in an L_1 ball, and the evaluation of the volume of an intersection of a hyper-sphere and hyperplane, refining the analysis of Section 2.

In Section 4, we consider coded settings and ensembles of random codes. We introduce the TCEM, which is a principled method for deriving the error exponent of random codes. We first describe the standard techniques commonly used to derive bounds on the error exponent, such as Jensen's inequality and its implications, and various types of union bounds. While these methods indeed turned out to be effective in the error-exponent analysis of basic settings, such as point-to-point channels and standard decoding rules, there is no general guarantee that they are accurate in more advanced scenarios. Indeed, we survey various settings in which these methods are sub-optimal, and do not provide the exact random-coding error exponent. As an alternative, we show that ensemble-average error probabilities (and other related performance measures) may be expressed via *type class* *enumerators* (TCEs), and specifically, via their (non-integer) moments and tail probabilities. We demonstrate this both on basic settings as well as more involved ones. We explore the probabilistic and statistical properties of TCEs, and then discuss a number of settings in multi-user information theory, in distributed compression and in hypothesis testing, for generalized decoding rules such as those allowing erasures and list outputs, and for the analysis of the typical random code. We outline how the TCEM is used in each of these settings, and how it allows to obtain, among other things, exact error-exponents for optimal decoding rules. In Appendix B we show that the exponents obtained by the TCEM can also be computed effectively.

In Section 5, we address the problem of evaluating the expectation of a non-linear function $f(\cdot)$ of a random variable (RV) X. In many cases, this function is either convex or concave, and so a natural course of action is to bound it using Jensen's inequality. However, there is no guarantee that the resulting bound is tight enough for the intended application. We present two general and useful strategies that can be employed in such cases. The first one is based on finding an *integral representation* of the function. Then, we interchange the expectation and integral order, and obtain an alternative expression for $\mathbb{E}\{f(X)\}$. The technique is useful if computing the inner expectation is simpler than the original expectation, or if it can be evaluated more accurately. After evaluating the inner expectation, the expectation $\mathbb{E}\{f(X)\}$ of interest can be computed by solving a one-dimensional integral. For example, when $f(t) = \ln(t)$, this allows to replace the evaluation of the expected logarithm with the evaluation of its moment-generating function (MGF). This is especially appealing since if $X = \sum_{i=1}^{n} X_i$ is the sum of n independent and identically distributed (IID) RVs, then its MGF is the n-th power of the MGF of just one of them. In accordance, this transforms the original expectation, which is an integral in $\mathbb{R}^n$, to a one-dimensional integral. We focus on the logarithmic function $f(t) = \ln(t)$ (and its integer powers), as well as the power function $f(t) = t^\rho$ for some $\rho > 0$ (even non-integer), and exemplify the use of this technique in a multitude of problems such as differential entropy for generalized multivariate Cauchy densities, ergodic capacity of the Rayleigh single-input multiple-output (SIMO) channel, and moments of guesswork.

The second strategy exploits convexity or concavity properties, but goes beyond the standard Jensen's inequality. This strategy may come in various flavors. First, a change of measure can be performed before using Jensen's inequality, and then the alternative measure can be optimized over a given class to improve the bound. As a notable example, when $f(t) = \ln(t)$, this reproduces the Donsker–Varadhan variational characterization of the Kullback–Leibler (KL) divergence. Second, one may use Jensen's inequality, but accompany it with an inequality in the opposite direction, *i.e.*, a reverse Jensen's inequality (RJI), in order to evaluate its tightness. We provide a few techniques, all of which rely on a general form of such a RJI. Third, the "supporting-line" approach used to prove Jensen's inequality may be generalized to cases in which the function whose expected value is sought of is not convex/concave, but takes a more complicated form, such as the composition or a multiplication of a different function with a convex/concave function. A generalized version of Jensen's inequality can still be derived, by properly optimizing the supporting line. We exemplify the use of this technique in various problems involving the evaluation of data compression performance and channel capacity.

In summary, we present a diverse toolbox of analytical techniques, indispensable to every information-theorist aiming to obtain tight and accurate results. We mention in passing other analytical techniques widely used in information theory, such as central-limit theorems extensively used in non-vanishing error regimes [198], concentration of measure bounds [169], statistical-physics methods such as the cavity and the replica method [151], and various methods described in the recent book [56]. These complement the tools outlined in this monograph.

This monograph was invited and written following a plenary talk by the first author, at the 2023 IEEE International Symposium on Information Theory (ISIT 2023), Taipei, Taiwan, June 25-30, 2023. It should be pointed out that some of the proposed techniques (like in Sections 2, 4, and many parts of Section 5) are original, while others are not new (like in Section 3).

Appendices

A

On the Tightness of Chernoff's Bound via the Method of Types

Let P be a memoryless source over an alphabet $\mathcal{X}$. For simplicity, we focus on finite-alphabet sources, though a similar derivation can be carried out using the extended method of types developed in Section 2 for more general sources. Let f be a real function of probability distributions over $\mathcal{X}$, and $\alpha \in \mathbb{R}$. Then,

$$\Pr\left[f(\hat{P}_{\boldsymbol{x}}) \geq \alpha\right]$$

$$= \sum_{\boldsymbol{x}\in\mathcal{X}^n} P(\boldsymbol{x}) \cdot \mathbb{1}\left[f(\hat{P}_{\boldsymbol{x}}) \geq \alpha\right] \tag{A.1}$$

$$\stackrel{(a)}{=} \sum_{\boldsymbol{x}\in\mathcal{X}^n} P(\boldsymbol{x}) \cdot \inf_{s\geq 0} e^{ns[f(\hat{P}_{\boldsymbol{x}})-\alpha]} \tag{A.2}$$

$$\stackrel{(b)}{=} \sum_{Q} e^{-n\cdot D(Q||P)} \cdot \inf_{s\geq 0} e^{ns[f(\hat{P}_{\boldsymbol{x}})-\alpha]} \tag{A.3}$$

$$\stackrel{(c)}{\doteq} \exp\left[-n \cdot \min_{Q}\left\{D(Q||P) - \inf_{s\geq 0} s\left[f(\hat{P}_{\boldsymbol{x}}) - \alpha\right]\right\}\right] \tag{A.4}$$

$$= \exp\left[-n \cdot \min_{Q}\sup_{s\geq 0}\left\{D(Q||P) - s\left[f(\hat{P}_{\boldsymbol{x}}) - \alpha\right]\right\}\right] \tag{A.5}$$

$$\stackrel{(d)}{\leq} \exp\left[-n \cdot \sup_{s\geq 0}\min_{Q}\left\{D(Q||P) - s\left[f(\hat{P}_{\boldsymbol{x}}) - \alpha\right]\right\}\right] \tag{A.6}$$

$$= \inf_{s \geq 0} \exp \left[-n \cdot \min_{Q} \left\{ D(Q||P) - s \left[f(\hat{P}_{\boldsymbol{x}}) - \alpha \right] \right\} \right] \tag{A.7}$$

$$\stackrel{(e)}{\doteq} \inf_{s \geq 0} \sum_{\boldsymbol{x} \in \mathcal{X}^n} P(\boldsymbol{x}) \cdot e^{ns[f(\hat{P}_{\boldsymbol{x}}) - \alpha]} \tag{A.8}$$

$$= \inf_{s \geq 0} \mathbb{E} \left[e^{ns[f(\hat{P}_{\boldsymbol{x}}) - \alpha]} \right], \tag{A.9}$$

where (a) follows from the elementary bound $\mathbb{1}\{t \geq \alpha\} \leq e^{ns(t-\alpha)}$ that holds for any $s \geq 0$, (b) follows from the probability of a type class [(2.12) in Section 2.2.1], and where the summation is over all possible types, (c) follows since the number of possible types is polynomial in n [(2.2) in Section 2.2.1], and so the sum is exponentially on the same scale as the maximum element, (d) follows since maximin is always less or equal than the minimax, and (e) follows again from the method of types, reversing the reasoning above.

The final term in (A.9) is exactly Chernoff's bound for the event $\{f(\hat{P}_{\boldsymbol{x}}) \geq \alpha\}$. Importantly, if f is concave then the *minimax theorem* [188] implies the inequality in (d) above is, in fact, an equality, and so the chain of passages is exponentially tight. In many applications, f is affine (*e.g.*, the empirical mean of some cost) and thus concave, and so Chernoff's bound is assured to be *tight*. See [49] for a thorough discussion.

B

Computation of Exponents

In this appendix, we describe two possible approaches to efficiently compute or bound the exponents obtained using the TCEM. This aspect is an indispensable part of the TCEM, since it is possible for an error exponent to take a rather intricate formula. Indeed, recall that the TCEM exponents are given by Csiszár–Körner-style formulas, *e.g.*, as in (4.10). Thus, they involve a constrained optimization problem over joint distributions, and the dimensionality of the optimized joint distributions increases with the alphabet sizes of the problem (*e.g.*, input and output alphabets of the channel). Thus, a direct optimization, using an exhaustive search or "general-purpose" global optimization over the probability simplex may be prohibitively complex.

The first approach we consider is based on *Lagrange duality* [21] (see also [180, Appendix]), in which the original exponent optimization problem is considered to be the *primal* optimization problem. When deriving instead the *dual* optimization problem of the exponent, the result is a Gallager-style bound [71, Chapter 5], which is often rather easy to compute and plot for an entire range of rates, rather than for a specific rate; see (B.19) in what follows for a typical formula. This is especially useful in multiuser problems [59], for which even problem instances with binary alphabets lead to optimization problems in non-trivial dimensions. For example, for a broadcast channel problem with input alphabet $\mathcal{X}$ and two receivers, each with an alphabet $\mathcal{Y}$, a joint distribution of the input and the two outputs has dimensionality $|\mathcal{X}| \cdot |\mathcal{Y}|^2 - 1$, which is at least 7. In some of the problems, the number of optimization variables for the Gallager-style bound does not increase with the alphabet size of the source or channel. The downside is that, as we shall see, the derivation might include the utilization of bounds that may sacrifice tightness. Indeed, in minimization optimization problems, the value of the dual problem is a *lower bound* on the value of the primal problem, and if the primal optimization problem is convex then *strong duality* holds (under typically mild conditions) [21, Chapter 5], and both values are equal. However, there is no guarantee that the primal optimization problem of the exponent is convex, and sometimes obtaining reasonably simple dual problems requires additional steps, which may also sacrifice tightness.

The second approach is based on utilization of convex optimization solvers. While the optimization problem involved in the computation of the exponent may not be convex as is, in many cases it is possible to develop a procedure that allows to compute it by only solving convex optimization problems.

Moreover, typically, the primal problem involves mostly *minimization* operators (over joint types), while the dual problem involves *maximization* operators (over scalar parameters). From this aspect, the dual exponent is preferable, because even a sub-optimal choice of the dual variables leads to a valid bound on the exponent. Thus, *e.g.*, a coarse exhaustive search on the dual variables may be performed and still lead to a tight bound. In contrast, the minimization in the primal problem must be performed accurately in order to obtain a valid numerical value of the exponent. Nonetheless, it also possible for the primal problem to include a maximization operator (possibly intertwined between minimization operators), and the same holds for such maximization problems — any sub-optimal choice leads to a valid bound. In fact, in some cases, an educated guess for the maximizing primal variable may be proposed, and in some settings it is possible to show that this choice is actually optimal.

B.1 Exponent Computation by Lagrange Duality

Lagrange duality is based on the *minimax theorem* [188], stating the minimax value of a functional convex in the minimization variable and concave in the maximization variable equals to the maximin value. We will next exemplify this technique on the random-coding error exponent $E_{\mathrm{rc},\alpha}(R, P_X)$ from (4.27), and derive a Lagrange dual lower bound on its value. As we have seen, if we consider the MMI rule, then the random-coding error exponent is greatly simplified to the standard random-coding error exponent in (4.10), which only contains a minimization over $Q_{Y|X}$ (with the minimization over $\tilde{Q}_{Y|X}$ removed). In accordance, it is not very difficult to obtain a dual Lagrange form of this exponent. In order to demonstrate a few other techniques that are generally useful for the TCE-based exponents, we will next let $\alpha(\cdot)$ be general, yet restricted to be a linear function of Q_{XY}, given by $\alpha(Q_{XY}) \triangleq \sum_{x\in\mathcal{X},y\in\mathcal{Y}} \alpha(x,y)\cdot Q(x,y)$ (this includes, *e.g.*, the ML decoder).

Let us start by writing the objective function of $E_{\mathrm{rc},\alpha}(R, P_X)$ using a dual variable $\rho \in \mathbb{R}$ as

$$\begin{aligned}
&E_{\mathrm{rc},\alpha}(R, P_X)\\
&= \min_{Q_{Y|X},\tilde{Q}_{Y|X}} D(Q_{Y|X}||W|P_X) + \left[I(P_X\times\tilde{Q}_{Y|X}) - R\right]_+ && \text{(B.1)}\\
&= \min_{Q_{Y|X},\tilde{Q}_{Y|X}} D(Q_{Y|X}||W|P_X) + \max\left\{I(P_X\times\tilde{Q}_{Y|X}) - R, 0\right\} && \text{(B.2)}\\
&\overset{(*)}{=} \min_{Q_{Y|X},\tilde{Q}_{Y|X}} D(Q_{Y|X}||W|P_X) + \max_{\rho\in[0,1]} \rho\cdot\left[I(P_X\times\tilde{Q}_{Y|X}) - R\right] && \text{(B.3)}\\
&= \min_{Q_{Y|X},\tilde{Q}_{Y|X}} \max_{\rho\in[0,1]} D(Q_{Y|X}||W|P_X) + \rho\cdot\left[I(P_X\times\tilde{Q}_{Y|X}) - R\right], && \text{(B.4)}
\end{aligned}$$

where $(*)$ follows from the identity $\max\{t,0\} = \max_{\rho\in[0,1]} \rho t$. Now, the objective function is linear, and hence concave, in the maximizing variable ρ, and the interval $[0,1]$ is convex. Moreover, $D(Q_{Y|X}||W|P_X)$ is convex in $Q_{Y|X}$ and $\rho\cdot I(P_X\times\tilde{Q}_{Y|X})$ is convex in $\tilde{Q}_{Y|X}$ (for $\rho\geq 0$), hence the objective functional is jointly convex in $(Q_{Y|X},\tilde{Q}_{Y|X})$. The constraint set for $(Q_{Y|X},\tilde{Q}_{Y|X})$, given by

$$\Big\{Q_{Y|X}, \tilde{Q}_{Y|X} \colon (P_X \times Q_{Y|X})_Y = (P_X \times \tilde{Q}_{Y|X})_Y, \\ \alpha(P_X \times \tilde{Q}_{Y|X}) \geq \alpha(P_X \times Q_{Y|X})\Big\}, \tag{B.5}$$

is the intersection of a hyperplane and a half space. We also note the implicit constraint that $Q_{Y|X}$ and $\tilde{Q}_{Y|X}$ are conditional probabilities, *i.e.*, $\sum_{y\in\mathcal{Y}} Q_{Y|X}(y|x) = \sum_{y\in\mathcal{Y}} \tilde{Q}_{Y|X}(y|x) = 1$ for all $x \in \mathcal{X}$ and $Q_{Y|X}(y|x), \tilde{Q}_{Y|X}(y|x) \geq 0$ for all $x \in \mathcal{X}, y \in \mathcal{Y}$. These are also convex constraints, and since the intersection of convex sets is convex, the constraint set for $(Q_{Y|X}, \tilde{Q}_{Y|X})$ is convex. So, the minimax theorem [188] implies that

$$E_{\mathrm{rc},\alpha}(R, P_X) = \\ \max_{\rho\in[0,1]} \min_{Q_{Y|X}, \tilde{Q}_{Y|X}} D(Q_{Y|X}||W|P_X) + \rho \cdot \Big[I(P_X \times \tilde{Q}_{Y|X}) - R\Big] \tag{B.6}$$

over the constraint set. We next focus on the inner minimization for a given $\rho \in [0,1]$. Following Lagrange duality [21, Chapter 5], we introduce dual variables $\lambda \geq 0$ and $\{\nu(y)\}_{y\in\mathcal{Y}} \subset \mathbb{R}$. The variable λ is for the inequality constraint $\alpha(P_X \times \tilde{Q}_{Y|X}) \geq \alpha(P_X \times Q_{Y|X})$, whereas the variables $\{\nu(y)\}_{y\in\mathcal{Y}}$ are for the constraint of equal output marginals, that is, the $|\mathcal{Y}|$ constraints $(P_X \times Q_{Y|X})_Y = (P_X \times \tilde{Q}_{Y|X})_Y$. Note that the constraint that $Q_{Y|X}$ and $\tilde{Q}_{Y|X}$ are conditional probability distributions is kept implicit. Hence, the minimization of interest is

$$\min_{Q_{Y|X}, \tilde{Q}_{Y|X}} \max_{\lambda\geq 0} \max_{\{\nu(y)\}_{y\in\mathcal{Y}}} D(Q_{Y|X}||W|P_X) + \rho \cdot \Big[I(P_X \times \tilde{Q}_{Y|X}) - R\Big] \\ + \sum_{y\in\mathcal{Y}} \nu(y) \cdot \left[\sum_{x\in\mathcal{X}} P_X(x)\left(\tilde{Q}_{Y|X}(y|x) - Q_{Y|X}(y|x)\right)\right] \\ + \lambda \cdot \left[\sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} \alpha(x,y) \cdot P_X(x)\left(Q_{Y|X}(y|x) - \tilde{Q}_{Y|X}(y|x)\right)\right]. \tag{B.7}$$

The minimax theorem now implies that we may interchange the minimization and maximization order. We next focus on the minimization, and begin by expressing the mutual information term via the *golden formula* using an arbitrary probability distribution S_Y on $\mathcal{Y}$, as

$$I(P_X \times \tilde{Q}_{Y|X}) = D(\tilde{Q}_{Y|X}||\tilde{Q}_Y|P_X) - D(\tilde{Q}_Y||S_Y) \tag{B.8}$$
$$= \min_{S_Y} D(\tilde{Q}_{Y|X}||S_Y|P_X). \tag{B.9}$$

Using this relation and slightly re-organizing the objective function, we are left with the minimization of the functional

$$\begin{aligned}\min_{S_Y} D(Q_{Y|X}||W|P_X)+ \\ &\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_X(x)Q_{Y|X}(y|x)\cdot[-\nu(y)+\lambda\cdot\alpha(x,y)] \\ &+\rho D(\tilde{Q}_{Y|X}||S_Y|P_X) \\ &+\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_X(x)\tilde{Q}_{Y|X}(y|x)\cdot[\nu(y)-\lambda\cdot\alpha(x,y)]\end{aligned} \tag{B.10}$$

over $(Q_{Y|X}, \tilde{Q}_{Y|X})$. It can be noticed that the minimization over $Q_{Y|X}$ is decoupled from the minimization over $\tilde{Q}_{Y|X}$, and each of them can be solved directly. Alternatively, we may use the *Donsker–Varadhan* variational formula [20, Corollary 4.15], [53], stating that for any two probability measures P_1 and P_2 on $\mathcal{Z}$ and a function $f\colon \mathcal{Z}\to\mathbb{R}$ that does not depend on P_1

$$\min_{P_2}\{D(P_2||P_1)+\mathbb{E}_{P_2}[f(Z)]\} = -\ln\mathbb{E}_{P_1}\left[e^{-f(Z)}\right]. \tag{B.11}$$

Let $W(\cdot|x)$ denote the conditional output of the channel given $x\in\mathcal{X}$. By employing (B.11) separately for each $x\in\mathcal{X}$ we get

$$\begin{aligned}&\min_{Q_{Y|X}} D(Q_{Y|X}||W|P_X)+\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_X(x)Q_{Y|X}(y|x)\cdot[-\nu(y)+\lambda\cdot\alpha(x,y)] \\ &=\sum_{x\in\mathcal{X}} P_X(x)\cdot\Bigg\{\min_{Q_{Y|X=x}} D(Q_{Y|X=x}||W(\cdot|x)) \\ &\qquad +\sum_{y\in\mathcal{Y}} Q_{Y|X}(y|x)\cdot[-\nu(y)+\lambda\cdot\alpha(x,y)]\Bigg\} \qquad (\text{B.12}) \\ &=-\sum_{x\in\mathcal{X}} P_X(x)\cdot\ln\left(\sum_{y\in\mathcal{Y}} W(y|x)\cdot e^{\nu(y)-\lambda\cdot\alpha(x,y)}\right). \qquad (\text{B.13})\end{aligned}$$

Similarly, the minimization over $\tilde{Q}_{Y|X}$ leads to

$$\sum_{x\in\mathcal{X}} P_X(x)\cdot\left\{\min_{\tilde{Q}_{Y|X=x}} \rho D(\tilde{Q}_{Y|X=x}||S_Y) + \sum_{y\in\mathcal{Y}} \tilde{Q}_{Y|X}(y|x)\cdot[\nu(y)-\lambda\cdot\alpha(x,y)]\right\}$$

$$= \min_{S_Y} -\rho \sum_{x\in\mathcal{X}} P_X(x)\cdot\ln\left(\sum_{y\in\mathcal{Y}} S_Y(y)\cdot e^{-[\nu(y)+\lambda\cdot\alpha(x,y)]/\rho}\right) \tag{B.14}$$

$$\geq \min_{S_Y} -\rho\ln\left(\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_X(x)S_Y(y)\cdot e^{-[\nu(y)+\lambda\cdot\alpha(x,y)]/\rho}\right), \tag{B.15}$$

where the inequality follows from convexity and Jensen inequality, yet is *not* guaranteed to be tight. Since $\rho\in[0,1]$, minimizing this last term over S_Y corresponds to maximizing

$$\sum_{y\in\mathcal{Y}} S_Y(y)\sum_{x\in\mathcal{X}} P_X(x)\cdot e^{-[\nu(y)+\lambda\cdot\alpha(x,y)]/\rho}, \tag{B.16}$$

which, due to Schwarz–Cauchy inequality, occurs when

$$S_Y(y) = \frac{\sum_{x\in\mathcal{X}} P_X(x)\cdot e^{-[\nu(y)+\lambda\cdot\alpha(x,y)]/\rho}}{\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} P_X(x)\cdot e^{-[\nu(y)+\lambda\cdot\alpha(x,y)]/\rho}}. \tag{B.17}$$

The minimal value over S_Y is then

$$\min_{S_Y} -\rho\ln\left(\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} P_X(x)S_Y(y)\cdot e^{-[\nu(y)+\lambda\cdot\alpha(x,y)]/\rho}\right)$$

$$= -\rho\ln\left(\frac{\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_X(x)e^{-[\nu(y)+\lambda\cdot\alpha(x,y)]/\rho}\right)^2}{\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} P_X(x)\cdot e^{-[\nu(y)+\lambda\cdot\alpha(x,y)]/\rho}}\right). \tag{B.18}$$

We thus conclude the dual lower bound

$$E_{\mathrm{rc},\alpha}(R,P_X)$$
$$\geq -\sum_{x\in\mathcal{X}} P_X(x)\cdot\ln\left(\sum_{y\in\mathcal{Y}} W(y|x)\cdot e^{\nu(y)-\lambda\cdot\alpha(x,y)}\right)$$

$$-\rho \ln\left(\frac{\sum_{y\in\mathcal{Y}}\left(\sum_{x\in\mathcal{X}} P_X(x)e^{-[\nu(y)+\lambda\cdot\alpha(x,y)]/\rho}\right)^2}{\sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} P_X(x)\cdot e^{-[\nu(y)+\lambda\cdot\alpha(x,y)]/\rho}}\right), \qquad \text{(B.19)}$$

for any choice of $\rho \in [0,1]$, $\lambda \geq 0$ and $\{\nu(y)\}_{y\in\mathcal{Y}} \subset \mathbb{R}$.

Let us compare the primal optimization in (B.1), with the dual lower bound (B.19). The primal problem is a minimization problem of dimension $2|\mathcal{X}|(|\mathcal{Y}|-1)$ over a constrained set $(Q_{Y|X}, \tilde{Q}_{Y|X})$ (the constraints further reduce the dimension by $|\mathcal{Y}|+1$). For the exact exponent, this minimization must be accurately solved. By comparison, the dual exponent is a lower bound on the exact exponent [recall (B.15)], and can be maximized over dimension $|\mathcal{Y}|+2$. Nonetheless, this maximization can be performed in a crude manner, since any choice of the dual parameters leads to a valid lower bound on the exponent.

For additional derivations of dual Lagrange exponents formulations and Gallager-style bounds, see [41, Exercise 10.24] and [165] (in Russian), and in the context of the TCEM, see [11], [137], [177].

B.2 Exponent Computation Procedures with Convex Optimization Solvers

As we have seen, we may write

$$E_{\mathrm{rc},\alpha}(R,P_X) = \max_{\rho\in[0,1]} \min_{Q_{Y|X},\tilde{Q}_{Y|X}} D(Q_{Y|X}||W|P_X) + \rho\cdot\left[I(P_X\times\tilde{Q}_{Y|X}) - R\right], \qquad \text{(B.20)}$$

and when $\alpha(Q_{XY})$ is a linear function of Q_{XY}, then the feasible set of $(Q_{Y|X}, \tilde{Q}_{Y|X})$ is convex. Hence, the inner minimization problem is a convex optimization problem that can be efficiently solved. However, in principle, it should be solved for the continuous set of values $\rho \in [0,1]$. We next describe an alternative method to evaluate $E_{\mathrm{rc},\alpha}(R,P_X)$.

Let us write $E_{\mathrm{rc},\alpha}(R,P_X) = \min\{E_-(R), E_+(R)\}$ where[1]

$$E_-(R) = \min_{Q_{Y|X},\tilde{Q}_{Y|X}} D(Q_{Y|X}||W|P_X), \qquad \text{(B.21)}$$

[1]For brevity, we omit the explicit dependence on the score α and the input distribution P_X.

where the minimization is over the set

$$\Big\{Q_{Y|X}, \tilde{Q}_{Y|X} \colon (P_X \times Q_{Y|X})_Y = (P_X \times \tilde{Q}_{Y|X})_Y,$$
$$\alpha(P_X \times \tilde{Q}_{Y|X}) \geq \alpha(P_X \times Q_{Y|X}),\ I(P_X \times \tilde{Q}_{Y|X}) \leq R\Big\}, \quad \text{(B.22)}$$

and where

$$E_+(R) = \min_{Q_{Y|X}, \tilde{Q}_{Y|X}} D(Q_{Y|X}||W|P_X) + I(P_X \times \tilde{Q}_{Y|X}) - R, \quad \text{(B.23)}$$

where the minimization over the set

$$\Big\{Q_{Y|X}, \tilde{Q}_{Y|X} \colon (P_X \times Q_{Y|X})_Y = (P_X \times \tilde{Q}_{Y|X})_Y,$$
$$\alpha(P_X \times \tilde{Q}_{Y|X}) \geq \alpha(P_X \times Q_{Y|X}),\ I(P_X \times \tilde{Q}_{Y|X}) \geq R\Big\}. \quad \text{(B.24)}$$

Note that the only difference between $E_-(R)$ and $E_+(R)$ is the constraint $I(P_X \times \tilde{Q}_{Y|X}) \gtreqless R$, and due to the continuity of the objective function, we have included the points $\{I(P_X \times \tilde{Q}_{Y|X}) = R\}$ in both problems. Now, since the KL divergence is also a convex function of $Q_{Y|X}$, it can be seen that the objective function is jointly convex in $\{Q_{Y|X}, \tilde{Q}_{Y|X}\}$ for both optimization problems. Since $\alpha(Q_{XY})$ is a linear function of Q_{XY}, the set $\{Q_Y = \tilde{Q}_Y,\ \alpha(P_X \times \tilde{Q}_{Y|X}) \geq \alpha(P_X \times Q_{Y|X})\}$ is a convex set. Furthermore, the set $\{I(P_X \times \tilde{Q}_{Y|X}) \leq R\}$ is also a convex set, and thus so is its intersection with the previous set. Consequently, the minimization problem of $E_-(R)$ is a convex optimization problem [21] (of dimension $2|\mathcal{X}| \times (|\mathcal{Y}| - 1)$), which can be efficiently solved, *e.g.*, using software packages such as `CVX` [78]. In contrast, the minimization problem of $E_+(R)$ involves the set $\{I(P_X \times \tilde{Q}_{Y|X}) \geq R\}$, which is *not* a convex set.

We thus proceed as follows. First, let us solve $E_+(R)$ for $R = 0$. In this case, the constraint $I(P_X \times Q_{Y|X}) \geq R$ is idle, and so

$$E_+(0) = \min_{Q_{Y|X}, \tilde{Q}_{Y|X}:\, \alpha(P_X \times \tilde{Q}_{Y|X}) \geq \alpha(P_X \times Q_{Y|X})} D(Q_{Y|X}||W|P_X) + I(P_X \times \tilde{Q}_{Y|X}). \tag{B.25}$$

This is a convex optimization problem, which can be efficiently solved. Let us denote the solution of this problem as $(Q^{(0)}_{Y|X}, \tilde{Q}^{(0)}_{Y|X})$. Now, as long as $R \leq R_{\text{cr}} \triangleq I(\tilde{Q}^{(0)}_{Y|X})$, then the objective function in $E_+(R)$ is minimized by the unconstrained solution $(Q^{(0)}_{Y|X}, \tilde{Q}^{(0)}_{Y|X})$, even if the constraint $I(P_X \times Q_{Y|X}) \geq R$ is imposed. For these rates it thus holds that $E_+(R) = E_+(0) - R$. Now, if $R \geq R_{\text{cr}}$ then the unconstrained solution $(Q^{(0)}_{Y|X}, \tilde{Q}^{(0)}_{Y|X})$ does not solve $E_+(R)$, and so the solution must be obtained on the boundary $\{I(P_X \times \tilde{Q}_{Y|X}) = R\}$. However, for such rates

$$\begin{aligned} E_+&(R) \\ &= \min_{Q_{Y|X}, \tilde{Q}_{Y|X}:\, I(P_X \times \tilde{Q}_{Y|X}) = R} D(Q_{Y|X}||W|P_X) + I(P_X \times \tilde{Q}_{Y|X}) - R && \text{(B.26)} \\ &= \min_{Q_{Y|X}, \tilde{Q}_{Y|X}:\, I(P_X \times \tilde{Q}_{Y|X}) = R} D(Q_{Y|X}||W|P_X) && \text{(B.27)} \\ &\geq \min_{Q_{Y|X}, \tilde{Q}_{Y|X}:\, I(P_X \times \tilde{Q}_{Y|X}) \leq R} D(Q_{Y|X}||W|P_X) && \text{(B.28)} \\ &= E_-(R), && \text{(B.29)} \end{aligned}$$

where all the above minimization operators are under the constraint $\alpha(P_X \times \tilde{Q}_{Y|X}) \geq \alpha(P_X \times Q_{Y|X})$, and the inequality holds since the feasible set is larger for $E_-(R)$. Consequently, for rates $R \geq R_{\text{cr}}$, the exponent is given by $\min\{E_-(R), E_+(R)\} = E_-(R)$.

To conclude, despite the fact that the minimization problem of $E_+(R)$ is not a convex optimization problem, the exponent can be computed for all rates by only solving convex optimization problems. To summarize, this is done by the following procedure: (1) Solve the optimization problem for $E_+(0)$, and compute the critical rate R_{cr}. (2) Solve the optimization problem $E_-(R)$ for any $R > R_{\text{cr}}$. The exponent is

$$\begin{cases} E_+(0) - R, & 0 \leq R \leq R_{\text{cr}} \\ E_-(R), & R > R_{\text{cr}} \end{cases}. \tag{B.30}$$

Note that this method requires solving two convex optimization problems at most for each rate, and the first one for finding $E_+(0)$ one is common to all rates.

For additional computational algorithms, see, for example, [64, Section V] for the computation of the exponent of the interference channel, [216, Appendix A] for the exponents of joint detection and decoding, and [215, Section VI] for exponents of distributed hypothesis testing.

C

The Derivation of the Expurgated Exponent

In this appendix, we outline the expurgation argument that follows the TCEM method. The proof follows [128, Appendix]. Let us focus on a specific codeword index m. We showed in Section 4.3 that, effectively, $\overline{N}_m(Q_{X\tilde{X}}) \sim \text{Binomial}(e^{nR}, e^{-nI(Q_{X\tilde{X}})})$. Thus, we separate between *typically populated* joint types ($I(Q_{X\tilde{X}}) \leq R$) and *typically empty* joint types ($I(Q_{X\tilde{X}}) > R$). First, for the populated types, for any $\epsilon > 0$, it holds by (4.66) that

$$\Pr\left[\overline{N}_m(Q_{X\tilde{X}}) \geq e^{n(R-I(Q_{X\tilde{X}})+\epsilon)}\right] \doteq e^{-n\infty}. \tag{C.1}$$

Taking the union over an exponentially number of codewords e^{nR} and a polynomial number of joint types, it follows from the union bound that

$$\mathcal{F} \triangleq \bigcup_{m=1}^{e^{nR}} \bigcup_{Q_{X\tilde{X}}:\, Q_X = Q_{\tilde{X}} = P_X,\, I(Q_{X\tilde{X}}) \geq R} \left\{\overline{N}_m(Q_{X\tilde{X}}) \geq e^{n(R-I(Q_{X\tilde{X}})+\epsilon)}\right\} \tag{C.2}$$

satisfies $\Pr[\mathcal{F}] \doteq e^{-n\infty}$. Since by (4.67) the lower tail also similarly decays double-exponentially, for the sake of exponent analysis, the TCE are *effectively* deterministic, for all codewords in the codebook and all joint types with $I(Q_{X\tilde{X}}) \leq R$, and is given by

$$\overline{N}_m(Q_{X\tilde{X}}) \doteq e^{n[R-I(Q_{X\tilde{X}})]}. \tag{C.3}$$

Second, for the empty types for which $I(Q_{X\tilde{X}}) > R$, it holds by (4.66) that

$$\Pr\left[\overline{N}_m(Q_{X\tilde{X}}) \geq 1\right] \doteq e^{-n[I(Q_{X\tilde{X}})-R]}, \tag{C.4}$$

which is exponentially small. Thus, we do not expect to observe other codewords $\tilde{m} \neq m$ which have joint type $Q_{X\tilde{X}}$ with $\boldsymbol{X}_m$. Indeed, the event

$$\mathcal{E}_m \triangleq \left\{ \bigcup_{Q_{X\tilde{X}}\colon Q_X=Q_{\tilde{X}}=P_X,\, I(Q_{X\tilde{X}})>R} \left\{\overline{N}_m(Q_{X\tilde{X}}) \geq 1\right\} \right\} \tag{C.5}$$

is the event that the mth codeword is a a-typical neighboring codeword, in the sense that there exists a $Q_{X\tilde{X}}$ with $I(Q_{X\tilde{X}}) > R$ and at least one neighboring codeword $\boldsymbol{X}_{\tilde{m}}$ so that $\hat{Q}_{\boldsymbol{X}_m\boldsymbol{X}_{\tilde{m}}} = Q_{X\tilde{X}}$. By the union bound, since the number of joint types increases polynomially with n, $p_n \triangleq \Pr[\mathcal{E}_m] \doteq e^{-n(I(Q_{X\tilde{X}})-R)}$. Thus, on the average, we expect that $p_n e^{nR}$ codewords will have such a-typical neighboring codewords. So, the event

$$\mathcal{E}^* \triangleq \left\{ \frac{1}{e^{nR}} \sum_{m=1}^{e^{nR}} \mathbb{1}\{\mathcal{E}_m\} \geq 2p_n \right\}, \tag{C.6}$$

in which more than $2p_n e^{nR}$ have such a-typical neighboring codeword has low probability. Indeed, Markov's inequality, which does not require independence of the events $\{\mathcal{E}_m\}$, implies that $\Pr[\mathcal{E}^*] \leq \frac{1}{2}$. Hence, with probability larger than $1/2 - \Pr[\mathcal{F}] \geq 1/3$, both $\mathcal{F}^c$ and $[\mathcal{E}^*]^c$ hold. We thus may choose a codebook $\mathcal{C}_n$ that belongs to the event $\mathcal{F}^c \cap [\mathcal{E}^*]^c$. The number of codewords in this codebook for which $\mathbb{1}\{\mathcal{E}_m\} = 1$ is less than $3p_n e^{nR}$. Thus, we can *expurgate* those codewords from the codebook, and obtain a new codebook $\mathcal{C}_n^*$ which satisfies: (1) Its size is larger than $|\mathcal{C}_n^*| \geq e^{nR}(1-3p_n) \doteq e^{nR}$. (2) Its TCEs $\overline{N}_m^*(Q_{X\tilde{X}})$ are only smaller than those of the original codebook, and specifically, $\overline{N}_m^*(Q_{X\tilde{X}}) = 0$ for all $Q_{X\tilde{X}}$ with $I(Q_{X\tilde{X}}) > R$. (3) $\overline{N}_m^*(Q_{X\tilde{X}}) \leq e^{n(R-I(Q_{X\tilde{X}})+\epsilon)}$ for all $Q_{X\tilde{X}}$ with $I(Q_{X\tilde{X}}) \leq R$.

For such a codebook, and after taking $\epsilon \downarrow 0$, the error probability bound in (4.38) is given by

$$P_{\mathsf{e}} \leq \exp\left[-n \cdot E_{\text{ex}}(R, P_X)\right], \tag{C.7}$$

where $E_{\text{ex}}(R, P_X)$ is as defined in (4.14).

Compared to the TCEM, the properties of codebook $\mathcal{C}_n^*$ traditionally follow from the *packing lemma* [41, Exercise 10.2], [42] (which is somewhat similar) or from a *graph decomposition lemma* [40, Corollary to Lemma 2]. In the latter case, equipped with the existence of such a codebook, [40] derived a bound for decoders with general score $\alpha(\cdot)$, and when $\alpha(\cdot)$ is set to be the ML decoder, then this exponent is shown to be at least as high as both the random-coding error exponent and the expurgated exponent.

D

Proofs for Section 4.3

Before proving Theorems 4.1, 4.2 and 4.3, we recall the following Chernoff tail bounds of a binomial RV $X \sim \text{Binomial}(m, p)$. If $r > p$ then $rm > \mathbb{E}[X] = pm$ and so the probability of the upper tail is

$$e^{-m \cdot D(r||p) - o(m)} \leq \Pr\left[X > rm\right] \leq e^{-m \cdot D(r||p)}, \tag{D.1}$$

where $D(r||p) \triangleq r \ln \frac{r}{p} + (1-r) \ln \frac{(1-r)}{(1-p)}$ is the binary KL divergence. If $r < p$ then this probability $\Pr[X > rm] \geq \Pr[X > \lfloor \mathbb{E}[X] \rfloor] \geq 1/2$, and the so the exponent is zero. Similarly, if $r < p$ then the probability of the lower tail is

$$e^{-m \cdot D(r||p) - o(m)} \leq \Pr\left[X < rm\right] \leq e^{-m \cdot D(r||p)}, \tag{D.2}$$

and if $r > p$ then the exponent is zero.

We will also need the following simple lemma regarding the KL divergence.

Lemma D.1. *Let $\{a_n, b_n\}$ be sequences in $(0,1)$ such that $a_n = o(1)$ and $b_n = o(1)$. Then,*

$$D(a_n||b_n) \sim \begin{cases} b_n & \frac{a_n}{b_n} = o(1) \\ a_n \ln \frac{a_n}{b_n}, & \frac{a_n}{b_n} = \omega(1) \end{cases}, \tag{D.3}$$

where for a sequence $\{c_n\}$, *the notation* $c_n = o(1)$ *means that* $\lim_{n\to\infty} c_n = 0$ *and the notation* $c_n = \omega(1)$ *means that* $\lim_{n\to\infty} c_n = \infty$.

Proof. We use the expansion $\ln(1+x) = x + \Theta(x^2)$ throughout. If $\frac{a_n}{b_n} = o(1)$ then it holds that

$$
\begin{aligned}
&(1-a_n)\ln\left[\frac{1-a_n}{1-b_n}\right] \\
&= (1-a_n)\ln(1-a_n) - (1-a_n)\ln(1-b_n) &\text{(D.4)}\\
&= -a_n(1-a_n) + \Theta(a_n^2) + b_n(1-a_n) + \Theta(b_n^2) &\text{(D.5)}\\
&= (b_n - a_n)(1-a_n) + \Theta(b_n^2) &\text{(D.6)}\\
&= b_n \cdot \left[\left(1 - \frac{a_n}{b_n}\right) - a_n(1-a_n) + \Theta(b_n^2)\right] &\text{(D.7)}\\
&\sim b_n, &\text{(D.8)}
\end{aligned}
$$

and so for all n large enough

$$\left|a_n \ln\frac{a_n}{b_n}\right| = a_n \ln\frac{b_n}{a_n} = -b_n \cdot \frac{a_n}{b_n}\ln\frac{a_n}{b_n} = -o(b_n) \tag{D.9}$$

since $\lim_{t\downarrow 0} t\ln t = 0$. This is negligible compared to the first term.

If $\frac{a_n}{b_n} = \omega(1)$ then

$$
\begin{aligned}
&\left|(1-a_n)\ln\left(\frac{1-a_n}{1-b_n}\right)\right| \\
&= |(1-a_n)\ln(1-a_n) - (1-a_n)\ln(1-b_n)| &\text{(D.10)}\\
&= \left|(1-a_n)\left[-a_n + \Theta(a_n^2) + b_n + \Theta(b_n^2)\right]\right| &\text{(D.11)}\\
&= \Theta(a_n), &\text{(D.12)}
\end{aligned}
$$

which is negligible compared to $a_n \ln\frac{a_n}{b_n} = \omega(a_n)$. □

We are now ready to prove Theorem 4.1, which provides exact exponents of the tail probabilities of the TCE N.

Proof of Theorem 4.1. In the case of a TCE, we are dealing with both an exponential number of trials and an exponentially decaying success probability, and so we consider the events $\{N > e^{n\lambda}\}$ and $\{N < e^{n\lambda}\}$

for some $\lambda \in \mathbb{R}$. Throughout, we will use the asymptotic expansion of the binary KL divergence in Lemma D.1.

We distinguish between two cases:

1. If $A > B$ then the mean value $\mathbb{E}[N] = e^{n(A-B)}$ is exponentially large. For the upper tail, we assume $\lambda > A - B$, for which
$$\Pr\left[N > e^{n\lambda}\right] \leq \exp\left[-e^{nA} \cdot D(e^{-n(A-\lambda)}||e^{-nB})\right]. \tag{D.13}$$
Since $A - B < \lambda$ then ${}^{e^{-n(A-\lambda)}}/_{e^{-nB}} = \omega(1)$ and the exponent is
$$e^{nA} \cdot D(e^{-n(A-\lambda)}||e^{-nB}) \sim e^{nA}e^{-n(A-\lambda)} \ln \frac{e^{-n(A-\lambda)}}{e^{-nB}} \tag{D.14}$$
$$= n(\lambda - (A - B))e^{n\lambda}. \tag{D.15}$$
Thus, the right-tail probability decays double-exponentially. Similarly, for the lower tail, we assume $\lambda < A - B$, for which
$$\Pr\left[N < e^{n\lambda}\right] \leq \exp\left[-e^{nA} \cdot D(e^{-n(A-\lambda)}||e^{-nB})\right]. \tag{D.16}$$
Since $A - B > \lambda$ then ${}^{e^{-n(A-\lambda)}}/_{e^{-nB}} = o(1)$ and the exponent is
$$e^{nA} \cdot D(e^{-n(A-\lambda)}||e^{-nB}) \sim e^{n(A-B)}. \tag{D.17}$$
Thus, the lower-tail probability also decays double-exponentially.

2. If $B > A$ then the mean value $\mathbb{E}[N] = e^{-n(B-A)} \leq 1$ is exponentially small. For the upper tail, we set $\lambda > 0 > A-B$ and obtain a double-exponentially decay, exactly as in the previous case. Next, as N is integer, for $\lambda \leq 0$, Markov's inequality implies that
$$\Pr\left[N > e^{n\lambda}\right] = \Pr\left[N \geq 1\right] \leq \mathbb{E}[N] = \exp\left[-n(B - A)\right]. \tag{D.18}$$
On the other hand,
$$\Pr\left[N > e^{n\lambda}\right] \geq \Pr\left[N = 1\right] = \binom{e^{nA}}{1} \cdot e^{-nB} \cdot (1 - e^{-nB})^{e^{nA}-1} \tag{D.19}$$
$$= e^{-n(B-A)} \cdot (1 - e^{-nB})^{e^{nA}-1} \tag{D.20}$$
$$\sim \exp\left[-n(B - A)\right], \tag{D.21}$$
which shows that Markov's inequality is exponentially tight in this case, and hence $\Pr[N > e^{n\lambda}] \doteq e^{-n(B-A)}$. The variable N has no lower tail since the above implies that $\Pr[N = 0] \geq 1 - e^{-n(B-A)}$.

Combining the two cases leads to the claimed result. □

We next prove Theorem 4.2, which states the exponent of $\mathbb{E}[N^s]$.

Proof of Theorem 4.2. We separate again between two cases, depending on the sign of $A - B$.

1. If $A > B$ then we know that any exponential deviation from the mean leads to a double-exponentially decay. Hence, for any $\lambda > A - B$

$$\begin{aligned}\mathbb{E}\left[N^s\right] = \Pr[N \le e^{n\lambda}] \cdot \mathbb{E}\left[N^s | N \le e^{n\lambda}\right] \\ + \Pr[N > e^{n\lambda}] \cdot \mathbb{E}\left[N^s | N \ge e^{n\lambda}\right] \quad &\text{(D.22)} \\ \dot{\le} e^{n\lambda s} + e^{-n\infty} \cdot e^{nsA} \quad &\text{(D.23)} \\ \doteq e^{n\lambda s}, \quad &\text{(D.24)}\end{aligned}$$

where we have used the fact that $N \le e^{nA}$ with probability 1, and write $e^{-n\infty}$ for a probability that decays super-exponentially. Taking the limit $\lambda \downarrow A - B$ shows that

$$\mathbb{E}\left[N^s\right] \dot{\le} e^{n(A-B)s}. \quad \text{(D.25)}$$

A matching lower bound can be derived in an analogous way: For any $\lambda < A - B$

$$\begin{aligned}\mathbb{E}\left[N^s\right] = \Pr[N \ge e^{n\lambda}] \cdot \mathbb{E}\left[N^s | N \ge e^{n\lambda}\right] \\ + \Pr[N < e^{n\lambda}] \cdot \mathbb{E}\left[N^s | N < e^{n\lambda}\right] \quad &\text{(D.26)} \\ \ge \left[1 - \Pr[N < e^{n\lambda}]\right] \cdot e^{n\lambda s} \quad &\text{(D.27)} \\ \sim e^{n\lambda s}, \quad &\text{(D.28)}\end{aligned}$$

after taking the limit $\lambda \uparrow A - B$. Hence,

$$\mathbb{E}\left[N^s\right] \doteq e^{n(A-B)s}. \quad \text{(D.29)}$$

2. If $A < B$ then we take $\lambda > 0$ to obtain

$$\mathbb{E}\left[N^s\right] = \Pr[1 \le N \le e^{n\lambda}] \cdot \mathbb{E}\left[N^s | 1 \le N \le e^{n\lambda}\right]$$

$$+\Pr[N > e^{n\lambda}] \cdot \mathbb{E}\left[N^s | N \geq e^{n\lambda}\right] \tag{D.30}$$

$$\dot{\leq} \Pr[N \geq 1] \cdot e^{n\lambda} + e^{-n\infty} \cdot e^{nsA} \tag{D.31}$$

$$\dot{\leq} e^{-n(B-A)} \cdot e^{n\lambda}. \tag{D.32}$$

Taking the limit $\lambda \downarrow 0$ shows that

$$\mathbb{E}\left[N^s\right] \dot{\leq} e^{-n(B-A)}. \tag{D.33}$$

A lower bound is obtained by

$$\mathbb{E}\left[N^s\right] \geq \Pr[N = 1] \cdot 1^s \geq [1 + o(1)] \cdot e^{-n(B-A)}, \tag{D.34}$$

which shows that the upper bound is tight.

Combining the two cases leads to the claimed result. □

We finally prove Theorem 4.3, which states that the probability of an intersection of lower tail events of a set of TCEs is exponentially equivalent to either 0 or 1.

Proof of Theorem 4.3. If there is a $j^* \in [k_n]$ so that $B_{j^*} < A_{j^*}$ and $\lambda < A_{j^*} - B_{j^*}$ then $\Pr[N_{j^*} < e^{n\lambda}] \doteq e^{-n\infty}$. So,

$$\Pr\left[\bigcap_{j=1}^{k_n} \left\{N_j < e^{n\lambda}\right\}\right] \leq \min_{1 \leq j \leq k_n} \Pr\left[N_j < e^{n\lambda}\right] \doteq e^{-n\infty}. \tag{D.35}$$

Otherwise, if all $j = 1, \ldots, k_n$ it holds that either $B_j > A_j$ or $\lambda > A_j - B_j$ then (4.66) implies that $\Pr[N_j > e^{n\lambda}] \dot{\leq} e^{-n\infty}$ for all $j = 1, \ldots, k_n$. Thus, from the union bound, as $n \to \infty$

$$\Pr\left[\bigcap_{j=1}^{k_n} \left\{N_j \leq e^{n\lambda}\right\}\right] = 1 - \Pr\left[\bigcup_{j=1}^{k_n} \left\{N_j > e^{n\lambda}\right\}\right] \tag{D.36}$$

$$\geq 1 - \sum_{j=1}^{k_n} \Pr\left[N_j > e^{n\lambda}\right] \tag{D.37}$$

$$\geq 1 - k_n \cdot \max_{1 \leq j \leq k_n} \Pr\left[N_j > e^{n\lambda}\right] \tag{D.38}$$

$$\geq 1 - k_n \cdot e^{-\min_{1 \leq j \leq k_n} E_j} \tag{D.39}$$

$$\to 1. \tag{D.40}$$

Combining (D.35) and (D.40) leads to the stated claim. □

References

[1] R. Ahlswede and I. Csiszár, "Hypothesis testing with communication constraints," *IEEE Trans. Inf. Theory*, vol. 32, no. 4, 1986, pp. 533–542.

[2] R. Ahlswede and G. Dueck, "Good codes can be produced by a few permutations," *IEEE Trans. Inf. Theory*, vol. 28, no. 3, 1982, pp. 430–443.

[3] M. A. Ali, H. Budak, and Z. Zhang, "A new extension of quantum Simpson's and quantum Newton's inequalities for quantum differentiable convex functions," *Mathematical Methods in the Applied Sciences*, 2021. DOI: 10.1002/mma.7889.

[4] S. M. Ali and S. D. Silvey, "A general class of coefficients of divergence of one distribution from another," *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 28, no. 1, 1966, pp. 131–142.

[5] Y. Altuğ and A. B. Wagner, "Moderate deviations in channel coding," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, 2014, pp. 4417–4426.

[6] Y. Altuğ and A. B. Wagner, "Refinement of the random coding bound," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, 2014, pp. 6005–6023.

[7] D. Anade, J.-M. Gorce, P. Mary, and S. M. Perlaza, "An upper bound on the error induced by saddlepoint approximations – applications to information theory," *Entropy*, vol. 22, no. 6, 2020, p. 690. DOI: 10.3390/e22060690.

[8] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 55, no. 7, 2009, pp. 3051–3073.

[9] E. Arikan and N. Merhav, "Guessing subject to distortion," *IEEE Trans. Inf. Theory*, vol. 44, no. 3, 1998, pp. 1041–1056.

[10] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels (corresp.)," *IEEE Trans. Inf. Theory*, vol. 19, no. 3, 1973, pp. 357–359.

[11] R. Averbuch and N. Merhav, "Exact random coding exponents and universal decoders for the asymmetric broadcast channel," *IEEE Trans. Inf. Theory*, vol. 64, no. 7, 2018, pp. 5070–5086.

[12] R. Averbuch, N. Weinberger, and N. Merhav, "Expurgated bounds for the asymmetric broadcast channel," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, 2019, pp. 3412–3435.

[13] R. R. Bahadur and R. R. Rao, "On deviations of the sample mean," *Ann. Math. Statist.*, vol. 31, no. 4, 1960, pp. 1015–1027.

[14] A. Barg and G. D. Forney, "Random codes: Minimum distances and error exponents," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, 2002, pp. 2568–2573.

[15] L. A. Bassalygo, S. I. Gel'fand, and M. S. Pinsker, "Simple methods for deriving lower bounds in the theory of codes," *Probl. Pered. Inform*, vol. 27, no. 4, 1991, pp. 3–8.

[16] P. Bergmans, "Random coding theorem for broadcast channels with degraded components," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, 1973, pp. 197–207.

[17] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo-codes. 1," in *Proceedings of ICC'93-IEEE International Conference on Communications*, IEEE, vol. 2, pp. 1064–1070, 1993.

[18] P. Billingsley, "Statistical methods in Markov chains," *Ann. Math. Statist.*, vol. 32, 1961, pp. 12–40.

[19] R. E. Blahut, "Hypothesis testing and information theory," *IEEE Trans. Inf. Theory*, vol. 20, no. 4, 1974, pp. 405–417.

[20] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.

[21] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[22] L. B. Boza, "Asymptotically optimal tests for finite Markov chains," *Ann. Math. Stat.*, vol. 42, 1971, pp. 1992–2007.

[23] N. G. de Bruijn, *Asymptotic Methods in Analysis*. Dover Publications, 1981.

[24] H. Budak, M. A. Ali, and M. Tarhanaci, "Some new quantum Hermite–Hadamard like inequalities for coordinated convex functions," *Journal of Optimization Theory and Applications*, vol. 186, no. 3, 2020, pp. 899–910.

[25] I. Budimir, S. S. Dragomir, and J. Pečari, "Further reverse results for Jensen's discrete inequality and applications in information theory," *Journal of Inequalities in Pure and Applied Mathematics*, vol. 2, 2001, pp. 1–14.

[26] M. V. Burnashev, "Data transmission over a discrete channel with feedback. random transmission time," *Problemy peredachi informatsii*, vol. 12, no. 4, 1976, pp. 10–30.

[27] D. Cao and V. Y. F. Tan, "Exact error and erasure exponents for the asymmetric broadcast channel," *IEEE Trans. Inf. Theory*, vol. 66, no. 2, 2019, pp. 865–885.

[28] V. Chandar, A. Tchamkerten, and D. Tse, "Asynchronous capacity per unit cost," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, 2012, pp. 1213–1226.

[29] V. Chandar, A. Tchamkerten, and G. W. Wornell, "Optimal sequential frame synchronization," *IEEE Trans. Inf. Theory*, vol. 54, no. 8, 2008, pp. 3725–3728.

[30] J. Chen, D.-k. He, A. Jagmohan, and L. A. Lastras-Montaño, "On universal variable-rate Slepian-Wolf coding," in *Proc. of IEEE International Conference on Communications*, IEEE, pp. 1426–1430, 2008.

[31] S.-Y. Chung, G. D. Forney, T. J. Richardson, and R. Urbanke, "On the design of low-density parity-check codes within 0.0045 dB of the Shannon limit," *IEEE Communications letters*, vol. 5, no. 2, 2001, pp. 58–60.

[32] M. Costa, "Writing on dirty paper (corresp.)," *IEEE Trans. Inf. Theory*, vol. 29, no. 3, 1983, pp. 439–441.

[33] T. Cover, "Broadcast channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, 1972, pp. 2–14.

[34] T. Cover and A. E. Gamal, "Capacity theorems for the relay channel," *IEEE Trans. Inf. Theory*, vol. 25, no. 5, 1979, pp. 572–584.

[35] T. Cover and C. Leung, "An achievable rate region for the multiple-access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 27, no. 3, 1981, pp. 292–298.

[36] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. John Wiley & Sons, 2006.

[37] I. Csiszár, "Eine informationstheoretische ungleichung und ihre anwendung auf den beweis der ergodizität von markoffschen ketten," *A Magyar Tudományos Akadémia Matematikai Kutató Intézetének Közleményei*, vol. 8, no. 1-2, 1963, pp. 85–108.

[38] I. Csiszár, "The method of types," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, 1998, pp. 2505–2523.

[39] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Trans. Inf. Theory*, vol. 24, no. 3, 1978, pp. 339–348.

[40] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, 1981, pp. 5–12.

[41] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*. Cambridge University Press, 2011.

[42] I. Csiszár, J. Körner, and K. Marton, "A new look at the error exponent of discrete memoryless channels," in *Proc. of International Symposium on Information Theory*, 107 (abstract), 1977.

[43] A. G. D'yachkov, "Bounds on the average error probability for a code ensemble with fixed composition," *Problemy Peredachi Informatsii*, vol. 16, no. 4, 1980, pp. 3–8.

[44] G. Dasarathy and S. C. Draper, "On reliability of content identification from databases based on noisy queries," in *2011 IEEE International Symposium on Information Theory Proceedings*, IEEE, pp. 1066–1070, 2011.

[45] G. Dasarathy and S. C. Draper, "Upper and lower bounds on the reliability of content identification," in *23th International Zurich Seminar on Communications (IZS 2014)*, ETH-Zürich, 2014.

[46] L. D. Davisson, "Universal noiseless coding," *IEEE Trans. Inf. Theory*, vol. 29, no. 6, 1973, pp. 783–795.

[47] L. D. Davisson, G. Longo, and A. Sgarro, "The error exponent for noiseless encoding of finite ergodic Markov sources," *IEEE Trans. Inf. Theory*, vol. 27, no. 4, 1981, pp. 431–438.

[48] D. De Caen, "A lower bound on the probability of a union," *Discrete mathematics*, vol. 169, no. 1-3, 1997, pp. 217–220.

[49] A. Dembo and O. Zeitouni, *Large Deviations and Applications*. Jones and Bartlett Publishers, 1993.

[50] B. Derrida, "Random-energy model: Limit of a family of disordered models," *Physical Review Letters*, vol. 45, no. 2, 1980, p. 79.

[51] B. Derrida, "The random energy model," CEA Centre d'Etudes Nucleaires de Saclay, Tech. Rep., 1980.

[52] B. Derrida, "Random-energy model: An exactly solvable model of disordered systems," *Physical Review B*, vol. 24, no. 5, 1981, p. 2613.

[53] M. D. Donsker and S. R. S. Varadhan, "Asymptotic evaluation of certain Markov process expectations for large time. iv," *Communications on pure and applied mathematics*, vol. 36, no. 2, 1983, pp. 183–212.

[54] S. S. Dragomir, "Some reverses of the Jensen inequality for functions of selfadjoint operators in Hilbert spaces," *Journal of Inequalities and Applications*, 2010. DOI: https://doi.org/10.1155/2010/496821.

[55] S. S. Dragomir, "Some reverses of the Jensen inequality with applications," *Bulletin of the Australian Mathematical Society*, vol. 87, 2013, pp. 177–194.

[56] M. Drmota and W. Szpankowski, *Analytic Information Theory: From Compression to Learning*. Cambridge University Press, 2023.

[57] G. Dueck and J. Körner, "Reliability function of a discrete memoryless channel at rates above capacity (corresp.)," *IEEE Trans. Inf. Theory*, vol. 25, no. 1, 1979, pp. 82–85.

[58] P. Dupuis and R. S. Ellis, *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, 1997.

[59] A. El Gamal and Y.-H. Kim, *Network Information Theory*. Cambridge University Press, 2011.

[60] P. Elias, "Coding for noisy channels," in *IRE Conv. Rec.*, vol. 3, pp. 37–46, 1955.

[61] P. Elias, "List decoding for noisy channels," in *IRE WESCON Conf. Rec.*, vol. 2, pp. 94–104, 1957.

[62] T. Erseghe, "Coding in the finite-blocklength regime: Bounds based on Laplace integrals and their asymptotic approximations," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, 2016, pp. 6854–6883.

[63] S. E. Esipov and T. J. Newman, "Interface growth and Burgers turbulence: The problem of random initial conditions," *Physical Review E*, vol. 48, no. 2, 1993, pp. 1046–1050.

[64] R. H. Etkin, N. Merhav, and E. Ordentlich, "Error exponents of optimum decoding for the interference channel," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, 2009, pp. 40–56.

[65] R. M. Fano, *Transmission of Information: A Statistical Theory of Communications*. M.I.T. Press, 1961.

[66] M. Feder and N. Merhav, "Universal composite hypothesis testing: A competitive minimax approach," *IEEE Trans. Inf. Theory*, vol. 48, no. 6, 2002, pp. 1504–1517.

[67] J. Font-Segura, G. Vázquez-Vilar, A. Martinez, A. Guillén i Fàbregas, and A. Lancho, "Saddlepoint approximations of lower and upper bounds to the error probability in channel coding," in *Proc. 2018 52nd Annual Conference on Information Sciences and Systems (CISS 2018)*, Princeton, NJ, USA, 2018.

[68] G. D. Forney, "Exponential error bounds for erasure, list, and decision feedback schemes," *IEEE Trans. Inf. Theory*, vol. 14, no. 2, 1968, pp. 206–220.

[69] R. G. Gallager, "Low-density parity-check codes," *IRE Transactions on information theory*, vol. 8, no. 1, 1962, pp. 21–28.

[70] R. G. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inf. Theory*, vol. 11, no. 1, 1965, pp. 3–18.

[71] R. G. Gallager, *Information Theory and Reliable Communication*, vol. 588. Springer, 1968.

[72] R. G. Gallager, "The random coding bound is tight for the average code (corresp.)," *IEEE Trans. Inf. Theory*, vol. 19, no. 2, 1973, pp. 244–246.

[73] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, no. 3, 1974, pp. 3–14.

[74] R. G. Gallager, "Source coding with side information and universal coding," M.I.T., LIDS- P-937, 1976.

[75] S. I. Gel'fand and M. S. Pinsker, "Coding for channels with random parameters," *Probl. Contr. Inform. Theory*, vol. 9, no. 1, 1980, pp. 19–31.

[76] S. Ginzach, N. Merhav, and I. Sason, "Random-coding error exponent of variable-length codes with a single-bit noiseless feedback," in *2017 IEEE Information Theory Workshop (ITW)*, IEEE, pp. 584–588, 2017.

[77] V. D. Goppa, "Nonprobabilitistic mutual information without memory," *Problems of Control and Inform., Theory*, vol. 4, 1975, pp. 97–102.

[78] M. Grant, S. P. Boyd, and Y. Ye, "CVX users' guide," 2009. URL: http://www.%20stanford.%20edu/boyd/software.%20html.

[79] T. S. Han and S. Amari, "Statistical inference under multiterminal data compression," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, 1998, pp. 2300–2324.

[80] T. S. Han and K. Kobayashi, "A new achievable rate region for the interference channel," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, 1981, pp. 49–60.

[81] T. S. Han and K. Kobayashi, "Exponential-type error probabilities for multiterminal hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, 2006, pp. 2–14.

[82] E. A. Haroutunian, "Bounds for the exponent of the probability of error for a semicontinuous memoryless channel," *Problemy Peredachi Informatsii*, vol. 4, no. 4, 1968, pp. 37–48.

[83] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, 2006, pp. 1562–1575.

[84] M. Hayashi, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Trans. Inf. Theory*, vol. 57, no. 6, 2011, pp. 3989–4001.

[85] M. Hayashi and V. Y. F. Tan, "Asymmetric evaluations of erasure and undetected error probabilities," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, 2015, pp. 6560–6577.

[86] E. Hille, *Analytic Function Theory*, vol. 2. American Mathematical Soc., 1959.

[87] J. Honda, "Exact asymptotics of random coding error probability for general memoryless channels," in *Proc. 2018 IEEE International Symposium on Information Theory (ISIT 2018)*, pp. 1844–1848, Vail, CO, U.S.A, 2018.

[88] W. Huleihel and N. Merhav, "Universal decoding for Gaussian intersymbol interference channels," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, 2015, pp. 1606–1618.

[89] W. Huleihel and N. Merhav, "Random coding error exponents for the two-user interference channel," *IEEE Trans. Inf. Theory*, vol. 63, no. 2, 2016, pp. 1019–1042.

[90] W. Huleihel, S. Salamatian, N. Merhav, and M. Médard, "Gaussian intersymbol interference channels with mismatch," *IEEE Trans. Inf. Theory*, vol. 65, no. 7, 2019, pp. 4499–4517.

[91] W. Huleihel, N. Weinberger, and N. Merhav, "Erasure/list random coding error exponents are not universally achievable," *IEEE Trans. Inf. Theory*, vol. 62, no. 10, 2016, pp. 5403–5421.

[92] P. A. Humblet, "Generalization of Huffman coding to minimize the probability of buffer overflow," *IEEE Trans. Inf. Theory*, vol. 27, 1981, pp. 230–232.

[93] T. Ignatenko and F. M. J. Willems, "Biometric security from an information-theoretical perspective," *Foundations and Trends® in Communications and Information Theory*, vol. 7, no. 2–3, 2012, pp. 135–316.

[94] A. Ingber, T. Courtade, and T. Weissman, "Compression for quadratic similarity queries," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, 2015, pp. 2729–2747.

[95] S. Janson, "New versions of Suen's correlation inequality," *Random Structures and Algorithms*, vol. 13, no. 3-4, 1998, pp. 467–483.

[96] T. Jebara and A. Pentland, "On reversing Jensen's inequality," in *Proc. 13th International Conference on Neural Information Processing Systems (NIPS 2000)*, pp. 213–219, Denver, CO, U.S.A., 2000.

[97] F. Jelinek, "Buffer overflow in variable length coding of fixed rate sources," *IEEE Trans. Inf. Theory*, vol. 14, no. 3, 1968, pp. 490–501.

[98] F. Jelinek, "Evaluation of expurgated bound exponents," *IEEE Trans. Inf. Theory*, vol. 14, no. 3, 1968, pp. 501–505.

[99] F. Jelinek, *Probabilistic Information Theory*. McGraw-Hill, 1968.

[100] R. Johannesson and K. S. Zigangirov, *Fundamentals of Convolutional Coding*. John Wiley & Sons, 2015.

[101] Y. Kaspi and N. Merhav, "Error exponents for broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 57, no. 1, 2010, pp. 101–123.

[102] B. G. Kelly and A. B. Wagner, "Reliability in source coding with side information," *IEEE Trans. Inf. Theory*, vol. 58, no. 8, 2012, pp. 5086–5111.

[103] G. Keshet, Y. Steinberg, and N. Merhav, "Channel coding in the presence of side information," *Foundations and Trends® in Communications and Information Theory*, vol. 4, no. 6, 2008, pp. 445–586.

[104] S. Khan, M. A. Khan, and Y.-M. Chu, "Converses of Jensen inequality derived from the Green functions with applications in information theory," *Mathematical Methods in Applied Sciences*, vol. 43, 2020, pp. 2577–2587.

[105] S. Khan, M. A. Khan, and Y.-M. Chu, "New converses of Jensen inequality via Green functions with applications," *Revista de la Real Academia de Ciencias Exactas, Fisicas y Naturales. Serie A, Matematicas*, vol. 114, no. 3, 2020.

[106] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, 1977, pp. 60–64.

[107] J. Körner and A. Sgarro, "Universally attainable error exponents for broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, 1980, pp. 670–679.

[108] R. E. Krichevsky and V. K. Trofimov, "The performance of universal encoding," *IEEE Trans. Inf. Theory*, vol. 27, no. 2, 1981, pp. 199–207.

[109] A. Lancho, J. Östman, G. Durisi, T. Koch, and G. Vázquez-Vilar, "Saddlepoint approximations for short-packet wireless communications," *IEEE Trans. on Wireless Communications*, vol. 19, no. 7, 2020, pp. 4831–4846.

[110] E. L. Lehmann, *Theory of Point Estimation*. New York: John Wiley & Sons, 1983.

[111] Y. Liang, H. V. Poor, and S. Shamai (Shitz), "Information theoretic security," *Foundations and Trends® in Communications and Information Theory*, vol. 5, no. 4–5, 2009, pp. 355–580.

[112] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error-Correcting Codes*, vol. 16. Elsevier, 1977.

[113] A. Martinez and A. Guillén i Fàbregas, "Random-coding bounds for threshold decoders: Error exponent and saddlepoint approximation," in *Proc. 2011 IEEE International Symposium on Information Theory (ISIT 2011)*, pp. 2905–2909, St. Petersburg, Russia, 2011.

[114] A. Martinez and A. Guillén i Fàbregas, "Saddlepoint approximation of random-coding bounds," in *The 2011 Information Theory and Applications Workshop (ITA 2011)*, La Jolla, CA, USA, 2011.
[115] N. Merhav, "On the estimation of the model order in exponential families," *IEEE Trans. Inf. Theory*, vol. 35, no. 5, 1989, pp. 1109–1114.
[116] N. Merhav, "Universal decoding for memoryless Gaussian channels with a deterministic interference," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, 1993, pp. 1261–1269.
[117] N. Merhav, "A large-deviations notion of perfect secrecy," *IEEE Trans. Inf. Theory*, vol. 49, no. 2, 2003, pp. 506–508.
[118] N. Merhav, "Error exponents of erasure/list decoding revisited via moments of distance enumerators," *IEEE Trans. Inf. Theory*, vol. 54, no. 10, 2008, pp. 4439–4447.
[119] N. Merhav, "Relations between random coding exponents and the statistical physics of random codes," *IEEE Trans. Inf. Theory*, vol. 55, no. 1, 2008, pp. 83–92.
[120] N. Merhav, "Statistical physics and information theory," *Foundations and Trends in Communications and Information Theory*, vol. 6, no. 1-2, 2009.
[121] N. Merhav, "On optimum strategies for minimizing exponential moments of a given cost function," *Communications in Information and Systems*, vol. 11, no. 4, 2011, pp. 343–368.
[122] N. Merhav, "Subset-sum phase transitions and data compression," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2011, 2011, P09017.
[123] N. Merhav, "Another look at expurgated bounds and their statistical-mechanical interpretation," *arXiv preprint arXiv:1301.4117*, 2013.
[124] N. Merhav, "Universal decoding for arbitrary channels relative to a given class of decoding metrics," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, 2013, pp. 5566–5576.
[125] N. Merhav, "Erasure/list exponents for Slepian–Wolf decoding," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, 2014, pp. 4463–4471.

[126] N. Merhav, "Exact correct-decoding exponent of the wiretap channel decoder," *IEEE Trans. Inf. Theory*, vol. 60, no. 12, 2014, pp. 7606–7615.

[127] N. Merhav, "Exact random coding error exponents of optimal bin index decoding," *IEEE Trans. Inf. Theory*, vol. 60, no. 10, 2014, pp. 6024–6031.

[128] N. Merhav, "List decoding – random coding exponents and expurgated exponents," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, 2014, pp. 6749–6759.

[129] N. Merhav, "Statistical physics of random binning," *IEEE Trans. Inf. Theory*, vol. 61, no. 5, 2015, pp. 2454–2464.

[130] N. Merhav, "Universal decoding for source–channel coding with side information," *Communications in Information and Systems*, vol. 16, no. 1, 2016, pp. 17–58.

[131] N. Merhav, "Correction to "the generalized stochastic likelihood decoder: Random coding and expurgated bounds"," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, 2017, pp. 6827–6829.

[132] N. Merhav, "Reliability of universal decoding based on vector-quantized codewords," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, 2017, pp. 2696–2709.

[133] N. Merhav, "The generalized stochastic likelihood decoder: Random coding and expurgated bounds," *IEEE Trans. Inf. Theory*, vol. 63, no. 8, 2017, pp. 5039–5051.

[134] N. Merhav, "Ensemble performance of biometric authentication systems based on secret key generation," *IEEE Trans. Inf. Theory*, vol. 65, no. 4, 2018, pp. 2477–2491.

[135] N. Merhav, "Error exponents of typical random codes," *IEEE Trans. Inf. Theory*, vol. 64, no. 9, 2018, pp. 6223–6235.

[136] N. Merhav, "Lower bounds on exponential moments of the quadratic error in parameter estimation," *IEEE Trans. Inf. Theory*, vol. 64, no. 12, 2018, pp. 7636–7648.

[137] N. Merhav, "A Lagrange-dual lower bound to the error exponent of the typical random code," *IEEE Trans. Inf. Theory*, vol. 66, no. 6, 2019, pp. 3456–3464.

[138] N. Merhav, "Error exponents of typical random codes for the colored Gaussian channel," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, 2019, pp. 8164–8179.

[139] N. Merhav, "Error exponents of typical random trellis codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 4, 2019, pp. 2067–2077.

[140] N. Merhav, "Universal decoding for asynchronous Slepian–Wolf encoding," *IEEE Trans. Inf. Theory*, vol. 67, no. 5, 2020, pp. 2652–2662.

[141] N. Merhav, "On more general distributions of random binning for Slepian–Wolf encoding," *IEEE Trans. Inf. Theory*, vol. 68, no. 2, 2021, pp. 737–751.

[142] N. Merhav, "Reversing Jensen's inequality for information-theoretic analyses," *Information*, vol. 13, no. 1, 2022, p. 39. DOI: 10.3390/info13010039.

[143] N. Merhav, "D-semifaithful codes that are universal over both memoryless sources and distortion measures," *IEEE Trans. Inf. Theory*, vol. 69, no. 7, 2023, pp. 4746–4757.

[144] N. Merhav, "Some families of Jensen-like inequalities with application to information theory," *Entropy*, vol. 25, no. 5, 2023, p. 752. DOI: 10.3390/e25050752.

[145] N. Merhav and M. Feder, "Minimax universal decoding with an erasure option," *IEEE Trans. Inf. Theory*, vol. 53, no. 5, 2007, pp. 1664–1675.

[146] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai (Shitz), "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, 1994, pp. 1953–1967.

[147] N. Merhav and E. Sabbag, "Optimal watermark embedding and detection strategies under limited detection resources," *IEEE Trans. Inf. Theory*, vol. 54, no. 1, 2008, pp. 255–274.

[148] N. Merhav and I. Sason, "An integral representation of the logarithmic function with applications in information theory," *Entropy*, vol. 22, no. 1, 2020, p. 51. DOI: 10.3390/e22010051.

[149] N. Merhav and I. Sason, "Some useful integral representations for information-theoretic analyses," *Entropy*, vol. 22, no. 6, 2020, p. 707. DOI: 10.3390/e22060707.

[150] N. Merhav and M. J. Weinberger, "On universal simulation of information sources using training data," *IEEE Trans. Inf. Theory*, vol. 50, no. 1, 2004, pp. 5–20.

[151] M. Mézard and A. Montanari, *Information, Physics and Computation*. Oxford University Press, 2009.

[152] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge University Press, 2017.

[153] P. Moulin, "The log-volume of optimal codes for memoryless channels, asymptotically within a few nats," *IEEE Trans. Inf. Theory*, vol. 63, no. 4, 2017, pp. 2278–2313.

[154] P. Moulin and Y. Wang, "Capacity and random-coding exponents for channel coding with side information," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, 2007, pp. 1326–1347.

[155] S. Natarajan, "Large deviations, hypotheses testing, and source coding for finite Markov chains," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, 1985, pp. 360–365.

[156] A. Nazari, A. Anastasopoulos, and S. S. Pradhan, "Error exponent for multiple-access channels: Lower bounds," *IEEE Trans. Inf. Theory*, vol. 60, no. 9, 2014, pp. 5095–5115.

[157] A. Nazari, R. Venkataramanan, D. Krithivasan, S. S. Pradhan, and A. Anastasopoulos, "Typicality graphs: Large deviation analysis," *arXiv preprint arXiv:1010.1317*, 2010.

[158] T. Neuschel, "Apéry polynomials and the multivariate saddle point method," *Contr. Approx.*, vol. 40, 2014, pp. 487–507.

[159] R. Nevanlinna and V. Paatero, *Introduction to Complex Analysis*, vol. 310. American Mathematical Society, 2007.

[160] Y. Oohama and T. S. Han, "Universal coding for the Slepian-Wolf data compression system and the strong converse theorem," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, 1994, pp. 1908–1919.

[161] L. Ozarow, "The capacity of the white Gaussian multiple access channel with feedback," *IEEE Trans. Inf. Theory*, vol. 30, no. 4, 1984, pp. 623–629.

[162] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, 1991.

[163] M. B. Parizi and E. Telatar, "On the secrecy exponent of the wire-tap channel," in *Proc. 2015-Fall IEEE Information Theory Workshop (ITW 2015-Fall)*, IEEE, pp. 287–291, 2015.
[164] M. B. Parizi, E. Telatar, and N. Merhav, "Exact random coding secrecy exponents for the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 63, no. 1, 2016, pp. 509–531.
[165] G. S. Poltyrev, "Random coding bounds for discrete memoryless channels," *Problemy Peredachi Informatsii*, vol. 18, no. 1, 1982, pp. 12–26.
[166] Y. Polyanskiy, "Channel coding: Non-asymptotic fundamental limits," Ph.D. dissertation, Department of Electrical Engineering, Princeton University, 2010.
[167] Y. Polyanskiy and S. Verdú, "Channel dispersion and moderate deviations limits for memoryless channels," in *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, IEEE, pp. 1334–1339, 2010.
[168] S. S. Pradhan, J. Chou, and K. Ramchandran, "Duality between source coding and channel coding and its extension to the side information case," *IEEE Trans. Inf. Theory*, vol. 49, no. 5, 2003, pp. 1181–1203.
[169] M. Raginsky and I. Sason, "Concentration of measure inequalities in information theory, communications, and coding," *Foundations and Trends® in Communications and Information Theory*, vol. 10, no. 1-2, 2013, pp. 1–246.
[170] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, 2012, pp. 6282–6303.
[171] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.
[172] P. Ruján, "Finite temperature error-correcting codes," *Physical review letters*, vol. 70, no. 19, 1993, p. 2968.
[173] J. Scarlett, "On the dispersions of the Gel'fand–Pinsker channel and dirty paper coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 9, 2015, pp. 4569–4586.

[174] J. Scarlett, A. 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, 2014, pp. 2647–2666.

[175] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Multiuser random coding techniques for mismatched decoding," *IEEE Trans. Inf. Theory*, vol. 62, no. 7, 2016, pp. 3950–3970.

[176] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Mismatched multi-letter successive decoding for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 64, no. 4, 2017, pp. 2253–2266.

[177] J. Scarlett, L. Peng, N. Merhav, A. Martinez, and A. Guillén i Fàbregas, "Expurgated random-coding ensembles: Exponents, refinements, and connections," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, 2014, pp. 4449–4462.

[178] J. Scarlett and V. Y. F. Tan, "Second-order asymptotics for the Gaussian MAC with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 61, no. 12, 2015, pp. 6700–6718.

[179] J. Scarlett, A. G. i Fàbregas, A. Somekh-Baruch, and A. Martinez, "Information-theoretic foundations of mismatched decoding," *Foundations and Trends® in Communications and Information Theory*, vol. 17, no. 2–3, 2020, pp. 149–401.

[180] J. M. Scarlett, "Reliable communication under mismatched decoding," Ph.D. dissertation, University of Cambridge, 2014.

[181] G. Schwartz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, 1978, pp. 461–464.

[182] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, 1948, pp. 379–423.

[183] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. i," *Information and Control*, vol. 10, no. 1, 1967, pp. 65–103.

[184] C. E. Shannon, R. G. Gallager, and E. R. Berlekamp, "Lower bounds to error probability for coding on discrete memoryless channels. ii," *Information and Control*, vol. 10, no. 5, 1967, pp. 522–552.

[185] H. Shimokawa, T. S. Han, and S. Amari, "Error bound of hypothesis testing with data compression," in *Proceedings of 1994 IEEE International Symposium on Information Theory*, IEEE, p. 114, 1994.

[186] N. Shulman, "Communication over an unknown channel via common broadcasting," Ph.D. dissertation, Tel-Aviv Univ., Tel-Aviv, Israel, 2003.

[187] S. Simić, "On a new converse of Jensen's inequality," *Publications de l'Institut Mathématique*, vol. 85, no. 99, 2009, pp. 107–110.

[188] M. Sion, "On general minimax theorems," *Pacific Journal of Mathematics*, vol. 8, no. 1, 1958, pp. 171–176.

[189] D. Slepian and J. K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. Inf. Theory*, vol. 19, no. 4, 1973, pp. 471–480.

[190] A. Somekh-Baruch and N. Merhav, "Exact random coding exponents for erasure decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, 2011, pp. 6444–6454.

[191] A. Somekh-Baruch, J. Scarlett, and A. Guillén i Fàbregas, "Generalized random Gilbert–Varshamov codes," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, 2019, pp. 3452–3469.

[192] J. Song, S. Still, R. D. H. Rojas, I. P. Castillo, and M. Marsili, "Optimal work extraction and mutual information in a generalized Szilárd engine," *Physical Review E.*, vol. 103, 2021, p. 052 121.

[193] R. Tamir and N. Merhav, "Error exponents in the bee identification problem," *IEEE Trans. Inf. Theory*, vol. 67, no. 10, 2021, pp. 6564–6582.

[194] R. Tamir and N. Merhav, "Trade-offs between error exponents and excess-rate exponents of typical Slepian–Wolf codes," *Entropy*, vol. 23, no. 3, 2021, p. 265.

[195] R. Tamir and N. Merhav, "Universal decoding for the typical random code and for the expurgated code," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, 2021, pp. 2156–2168.

[196] R. Tamir and N. Merhav, "Error exponents of the dirty-paper and Gel'fand–Pinsker channels," *IEEE Trans. Inf. Theory*, vol. 69, no. 12, 2023, pp. 7479–7498.

[197] R. Tamir, N. Merhav, N. Weinberger, and A. Guillén i Fàbregas, "Large deviations behavior of the logarithmic error probability of random codes," *IEEE Trans. Inf. Theory*, vol. 66, no. 11, 2020, pp. 6635–6659.

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

[199] V. Y. F. Tan and M. Tomamichel, "The third-order term in the normal approximation for the AWGN channel," in *Proc. 2014 IEEE International Symposium on Information Theory (ISIT 2014)*, pp. 2077–2081, Honolulu, HI, U.S.A, 2014.

[200] A. Tandon, V. Y. F. Tan, and L. R. Varshney, "The bee-identification problem: Bounds on the error exponent," *IEEE Transactions on Communications*, vol. 67, no. 11, 2019, pp. 7405–7416.

[201] A. E. Taylor, *General Theory of Functions and Integration.* Courier Corporation, 1985.

[202] A. Tchamkerten, V. Chandar, and G. W. Wornell, "Communication under strong asynchronism," *IEEE Trans. Inf. Theory*, vol. 55, no. 10, 2009, pp. 4508–4528.

[203] A. Tchamkerten, V. Chandar, and G. W. Wornell, "Asynchronous communication: Capacity bounds and suboptimality of training," *IEEE Trans. Inf. Theory*, vol. 59, no. 3, 2012, pp. 1227–1255.

[204] S. Tridenski and A. Somekh-Baruch, "The method of types for the AWGN channel," *arXiv preprint arXiv:2307.13322*, 2023.

[205] L. V. Truong, G. Cocco, J. Font-Segura, and A. Guillén i Fàbregas, "Concentration properties of random codes," *IEEE Trans. Inf. Theory*, vol. 69, no. 12, 2023, pp. 7499–7537.

[206] L. V. Truong and A. Guillén i Fàbregas, "Generalized random Gilbert-Varshamov codes: Typical error exponent and concentration properties," *IEEE Trans. Inf. Theory*, 2023.

[207] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication.* Cambridge university press, 2005.

[208] E. Tuncel, “Capacity/storage tradeoff in high-dimensional identification systems,” *IEEE Trans. Inf. Theory*, vol. 55, no. 5, 2009, pp. 2097–2106.

[209] G. Ungerboeck, “Channel coding with multilevel/phase signals,” *IEEE Trans. Inf. Theory*, vol. 28, no. 1, 1982, pp. 55–67.

[210] A. Vamvatsikos, *On the Wagner–Anantharam outer bound and achievable Gaussian source coding exponents*, 2007.

[211] G. Vázquez-Vilar, A. Guillén i Fàbregas, T. Koch, and A. Lancho, “Saddlepoint approximation of the error probability of binary hypothesis testing,” in *Proc. 2018 IEEE International Symposium on Information Theory (ISIT 2018)*, pp. 2306–2310, Vail, CO, USA, 2018.

[212] A. J. Viterbi and J. K. Omura, *Principles of Digital Communication and Coding*. Dover Publications, 2009.

[213] D. Wang, V. Chandar, S.-Y. Chung, and G. W. Wornell, “Error exponents in asynchronous communication,” in *2011 IEEE International Symposium on Information Theory Proceedings*, IEEE, pp. 1071–1075, 2011.

[214] M. J. Weinberger, N. Merhav, and M. Feder, “Optimal sequential probability assignment for individual sequences,” *IEEE Trans. Inf. Theory*, vol. 40, no. 2, 1994, pp. 384–396.

[215] N. Weinberger and Y. Kochman, “On the reliability function of distributed hypothesis testing under optimal detection,” *IEEE Trans. Inf. Theory*, vol. 65, no. 8, 2019, pp. 4940–4965.

[216] N. Weinberger and N. Merhav, “Codeword or noise? Exact random coding exponents for joint detection and decoding,” *IEEE Trans. Inf. Theory*, vol. 60, no. 9, 2014, pp. 5077–5094.

[217] N. Weinberger and N. Merhav, “Optimum tradeoffs between the error exponent and the excess-rate exponent of variable-rate Slepian–Wolf coding,” *IEEE Trans. Inf. Theory*, vol. 61, no. 4, 2015, pp. 2165–2190.

[218] N. Weinberger and N. Merhav, “A large deviations approach to secure lossy compression,” *IEEE Trans. Inf. Theory*, vol. 63, no. 4, 2017, pp. 2533–2559.

[219] N. Weinberger and N. Merhav, "Channel detection in coded communication," *IEEE Trans. Inf. Theory*, vol. 63, no. 10, 2017, pp. 6364–6392.

[220] N. Weinberger and N. Merhav, "Simplified erasure/list decoding," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, 2017, pp. 4218–4239.

[221] P. Whittle, "Some distributions and moment formulae for the Markov chain," *J. Roy. Stat. Soc.*, B, vol. 17, 1955, pp. 235–242.

[222] F. M. J. Willems, T. Kalker, J. Goseling, and J.-P. Linnartz, "On the capacity of a biometrical identification system," in *IEEE International Symposium on Information Theory*, pp. 82–82, 2003.

[223] J. M. Wozencraft, "List decoding," *Quarterly Progress Report*, vol. 48, 1958, pp. 90–95.

[224] G. Wunder, B. Groß, R. Fritschek, and R. F. Schaefer, "A reverse Jensen inequality result with application to mutual information estimation," in *Proc. 2021 IEEE Information Theory Workshop (ITW 2021)*, Kanazawa, Japan, 2021.

[225] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Trans. Inf. Theory*, vol. 22, no. 1, 1976, pp. 1–10.

[226] A. D. Wyner, "The wire-tap channel," *Bell System Technical Journal*, vol. 54, no. 8, 1975, pp. 1355–1387.

[227] M. H. Yassaee, M. R. Aref, and A. Gohari, "A technique for deriving one-shot achievability results in network information theory," in *2013 IEEE International Symposium on Information Theory*, IEEE, pp. 1287–1291, 2013.

[228] R. C. Yavas, V. Kostina, and M. Effros, "Third-order analysis of channel coding in the small-to-moderate deviations regime," in *Proc. 2022 IEEE International Symposium on Inform. Theory (ISIT 2022)*, pp. 2309–2314, Espoo, Finland, 2022.

[229] R. Zamir, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*. Cambridge University Press, 2014.

[230] Y. Zhong, F. Alajaji, and L. L. Campbell, "A type covering lemma and the excess distortion exponent for coding memoryless Laplacian sources," in *Proc. 23rd Biennial Symposium on Communications*, pp. 100–103, Kingston, ON, Canada, 2006.

[231] Y. Zhong, F. Alajaji, and L. L. Campbell, "Joint source-channel coding excess distortion exponent for some memoryless continuous-alphabet systems," *IEEE Trans. Inf. Theory*, vol. 55, no. 3, 2009, pp. 1296–1319.

[232] L. Zhou, V. Y. F. Tan, and M. Motani, "Second-order and moderate deviations asymptotics for successive refinement," *IEEE Trans. Inf. Theory*, vol. 63, no. 5, 2017, p. 2921.