

Reliable Communication Under Mismatched Decoding



Jonathan Scarlett

Department of Engineering

University of Cambridge

Supervisor: Albert Guillén i Fàbregas

This dissertation is submitted for the degree of
Doctor of Philosophy

Trinity Hall

August 2014

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other University. This dissertation is the result of my own work, and it includes nothing which is the outcome of work done in collaboration, except where specifically indicated in the text.

This dissertation contains less than 65,000 words including appendices, bibliography, footnotes, tables and equations, and has less than 150 figures.

Jonathan Scarlett

August 2014

Acknowledgements

I am deeply grateful to my supervisor Albert Guillén i Fàbregas for all that he has provided throughout the course of my PhD. He has been of constant help and support in every way possible, and it has truly been a pleasure to work with him. He allowed me complete freedom in choosing which research areas to pursue, and he was always encouraging of my collaborations with others. I have learnt a great amount from him on numerous aspects of academic research, and I am sure that it will reflect positively on my entire career.

This work benefited greatly from the additional supervision of Alfonso Martinez. He provided unique and valuable insight into every problem that we worked on, as well as countless intriguing technical and non-technical discussions. Several of the results that I am most proud of arose from him providing intelligent comments and questions, and persuading me to pursue certain topics further.

I am thankful to all of my past and present colleagues for their ongoing friendship, and for providing a highly enjoyable working environment. A collaboration with Li Peng, as well as several technical discussions with Adrià Tauste Campo, Gonzalo Vazquez-Vilar, and Tobias Koch, are gratefully acknowledged. It was also a great pleasure to work and interact with Jing Guo, Jossy Sayir, Ramji Venkataramanan, Taufiq Asyhari, and Seçkin Yildirim.

There are numerous members of the information theory community that positively influenced my research. Ongoing interactions with Vincent Tan have been highly valuable, and I am particularly thankful for a collaboration that arose as a result. I am similarly grateful to Neri Merhav for several interesting discussions and a collaboration. My knowledge and understanding of information theory benefited from discussions with several other members of the community, including Anelia Somekh-Baruch, Christoph Bunte, Yücel Altuğ, Pierre Moulin, and Ebrahim MolavianJazi.

I am grateful to Jamie Evans and Subhrakanti Dey for their supervision during my time at the University of Melbourne prior to my PhD, and for many helpful discussions regarding academic research. I thank Ian Wassell for his role as advisor during my PhD, and the administrative staff at the University of Cambridge and Universitat Pompeu Fabra for their assistance. I gratefully acknowledge the financial support from the Cambridge Trusts, without which this PhD would not have been possible.

Finally, I express my deepest gratitude to my parents, my brothers, and my late grandparents for their constant care and support throughout all stages of my life, and I sincerely thank all of my friends around the world for making my experiences truly memorable.

Abstract

Information-theoretic studies of channel coding typically seek to characterize the performance of coded communication systems when the encoder and decoder can be optimized. In practice, however, optimal decoding rules are often ruled out due to channel uncertainty and implementation constraints. This thesis studies the problem of channel coding with mismatched decoding, in which the decoding rule is fixed and possibly suboptimal.

In the point-to-point setting, several asymptotic and non-asymptotic bounds are given characterizing the tradeoff between the transmission rate, error probability, and block length. A cost-constrained random-coding ensemble with multiple auxiliary costs is introduced, and is shown to provide asymptotic performance gains at both fixed rates (error exponents) and varying rates (second-order coding rates and moderate deviations), while being directly applicable to continuous channels. Improved bounds are obtained at low rates using expurgation techniques, in which some of the randomly generated codewords are discarded.

For the i.i.d. ensemble, in which each symbol of each codeword is generated independently, a saddlepoint approximation of the random-coding error probability is shown to unify the fixed-rate and varying-rate asymptotics, while being efficiently computable and providing highly accurate estimates of the performance. For the constant-composition ensemble, in which each codeword is drawn from the set of sequences having a given empirical distribution, refined asymptotics are given in the fixed-rate and fixed-error regimes.

Mismatched maximum-metric decoding is studied for the multiple-access channel, in which two users communicate to a common receiver. In the finite-alphabet setting, an exponentially tight analysis is given for constant-composition random coding, and the exponents are shown to recover a known achievable rate region. Alternative expressions for the rates and exponents are derived that remain valid in the continuous-alphabet setting. A mismatched successive decoding rule is compared to the maximum-metric rule, and is shown to provide improved rates in some cases. The fixed-error asymptotics of maximum-likelihood decoding are studied, and an achievability result is given characterizing the speed at which the rates of the users can approach a given point on the boundary of the capacity region.

The performance of multi-user random-coding ensembles is studied for the point-to-point setting, with a particular focus on superposition coding, in which the codewords are generated conditionally on auxiliary codewords. A refined version of superposition coding is introduced, and is shown to yield rates at least as high as the standard version, with a strict improvement possible. Once again, exponentially tight bounds are given in the finite-alphabet setting, and generalizations to continuous alphabets are given.

Contents

Nomenclature	xi
1 Introduction	1
1.1 Channel Coding and Mismatched Decoding	1
1.1.1 System Setup	2
1.1.2 Applications	3
1.1.3 Extensions	5
1.1.4 Previous Work on Mismatched Decoding	5
1.2 Asymptotic and Non-Asymptotic Performance	10
1.2.1 Error Exponents	11
1.2.2 Second-Order Coding Rates	13
1.2.3 Moderate Deviations	14
1.2.4 Finite-Length Performance	15
1.3 Overview of the Thesis	15
1.4 Notation	17
2 Random Coding Bounds and Asymptotics	19
2.1 Introduction	19
2.2 Non-Asymptotic Bounds	20
2.3 Cost-Constrained Random Coding with Multiple Auxiliary Costs	23
2.4 Random-Coding Error Exponents	25
2.4.1 Cost-Constrained Ensemble	25
2.4.2 i.i.d. and Constant-Composition Ensembles	27
2.4.3 Achievable Rates	29
2.4.4 Number of Auxiliary Costs Required	30
2.5 Second-Order Coding Rates	32
2.5.1 Cost-Constrained Ensemble	32

2.5.2	i.i.d. and Constant-Composition Ensembles	36
2.5.3	Number of Auxiliary Costs Required	36
2.6	Moderate Deviations	37
2.7	Expurgated Error Exponents	40
2.7.1	Cost-Constrained Ensemble	40
2.7.2	i.i.d. and Constant-Composition Ensembles	42
2.7.3	Number of Auxiliary Costs Required	43
2.8	Numerical Example	44
2.9	Extensions to Channels with Input Constraints and More General Alphabets .	46
2.10	Proofs	48
2.10.1	Proof of Theorem 2.3	48
2.10.2	Proof of Theorem 2.4	49
2.10.3	Necessary Conditions for the Optimal Parameters	50
2.10.4	Proof of Theorem 2.8	52
2.10.5	Proof of Theorem 2.9	53
3	Refined Asymptotic Bounds	55
3.1	Introduction	55
3.2	Saddlepoint Approximation for i.i.d. Random Coding	56
3.2.1	Preliminary Definitions and Results	57
3.2.2	Approximation for $\text{rcu}_s(n, M)$	58
3.2.3	Approximation for $\text{rcu}(n, M)$	61
3.2.4	Numerical Examples	65
3.3	Prefactor to the i.i.d. Expurgated Exponent	67
3.3.1	Preliminary Definitions	67
3.3.2	Main Result	69
3.4	Constant-Composition Random Coding	69
3.4.1	Setup and Definitions	69
3.4.2	Main Results	71
3.5	Proofs	75
3.5.1	Properties of $E_0^{\text{iid}}(Q, \rho, s)$	75
3.5.2	Asymptotic Behavior of the Saddlepoint Approximation	75
3.5.3	Proof of Theorem 3.1	77
3.5.4	Proof of Theorem 3.2	81
3.5.5	Proof of Theorem 3.3	86

3.5.6	Proofs of Theorems 3.4 and 3.6	89
3.5.7	Proof of Lemma 3.4	92
3.5.8	Proofs of Theorems 3.5 and 3.7	101
4	Multiple-Access Channels	103
4.1	Introduction	103
4.1.1	System Setup	103
4.1.2	Previous Work	104
4.1.3	Contributions	106
4.2	Random-Coding Bounds, Exponents and Rates	107
4.2.1	Non-Asymptotic Bounds	108
4.2.2	Exponents and Rates for the DM-MAC	109
4.2.3	Exponents and Rates for More General Alphabets	119
4.2.4	Application to the Matched MAC	121
4.2.5	Time-Sharing	125
4.2.6	Application to Single-User Mismatched Channels	130
4.3	Multi-Letter Successive Decoding	132
4.3.1	Achievable Rate Region	133
4.3.2	Numerical Example	134
4.4	Second-Order Asymptotics for ML Decoding	135
4.4.1	Preliminaries	135
4.4.2	Main Results	138
4.4.3	Examples	144
4.5	Proofs	149
4.5.1	Proof of Theorem 4.4	149
4.5.2	Proof of Proposition 4.2	152
4.5.3	Proof of Theorem 4.12	153
4.5.4	Formulation of (4.129) in terms of Convex Optimization Problems	159
4.5.5	Proof of Theorem 4.13	161
4.5.6	Proof of Theorem 4.15 ($\beta = 0, \mathcal{U} = \emptyset$)	163
4.5.7	Proof of Theorem 4.15 (General Case)	171
4.5.8	Proof of Theorem 4.14	173
5	Superposition Coding Techniques	176
5.1	Introduction	176
5.2	Standard Superposition Coding	177

5.2.1	Non-Asymptotic Bounds	178
5.2.2	Exponents and Rates for DMCs	179
5.2.3	Comparison to Expurgated Parallel Coding	182
5.2.4	Rates for More General Alphabets	184
5.3	Refined Superposition Coding	185
5.3.1	Rates for DMCs	187
5.3.2	Comparison to Standard Superposition Coding	190
5.3.3	Numerical Examples	191
5.3.4	Dual Expressions and More General Alphabets	194
5.4	Discussion: Primal and Dual Connections	197
5.5	Proofs	197
5.5.1	Proof of Theorem 5.7	197
5.5.2	Proof of Theorem 5.8	200
6	Conclusions and Future Work	204
Appendix A	The Method of Types	208
A.1	Properties for the Single-User Setting	209
A.2	Properties for the MAC	210
Appendix B	Properties of the LM Rate and GMI	211
B.1	Suprema over s and $a(\cdot)$	211
B.2	Optimization of the LM Rate Over Q	215
Appendix C	Sums of Independent Random Variables	217
C.1	Asymptotic Behavior of the Tail Probability	217
C.2	Variations of the Central Limit Theorem	219
Appendix D	Bounds on the Probability of a Multiply-Indexed Union	221
Appendix E	Convex Optimization and Lagrange Duality	226
E.1	Proof of Lemma E.1	229
E.2	Proof of Lemma E.3	230
	Bibliography	233

Nomenclature

Roman Symbols

a, a_l, a^n, a_l^n	Auxiliary cost functions; additive n -letter extensions
C	Channel capacity
$E_r, E_r^{\text{iid}}, E_r^{\text{cc}}, E_r^{\text{cost}}$	Random-coding error exponents
$E_0, E_0^{\text{iid}}, E_0^{\text{cc}}, E_0^{\text{cost}}$	Functions of ρ such that $E_r = \max_{\rho \in [0,1]} E_0 - \rho R$
$E_{\text{ex}}, E_{\text{ex}}^{\text{iid}}, E_{\text{ex}}^{\text{cc}}, E_{\text{ex}}^{\text{cost}}$	Expurgated error exponents
$E_x, E_x^{\text{iid}}, E_x^{\text{cc}}, E_x^{\text{cost}}$	Functions of ρ such that $E_{\text{ex}} = \sup_{\rho \geq 1} E_x - \rho R$
$I_{\text{GMI}}, I_{\text{LM}}$	Generalized mutual information; LM rate
$i_s, i_{s,a}$	Generalized information densities
$I_s, I_{s,a}$	Means of generalized information densities
L	Number of auxiliary cost functions
M	Number of messages
M^*	Highest M for fixed (n, ϵ)
$N(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
n	Block length
o, O	Order notation for sequences
p_e	Average error probability
p_e^*	Smallest p_e for fixed (n, M)

$\bar{p}_e$	Random-coding error probability
$p_{e,\max}$	Maximal error probability
$P, \tilde{P}$	Primal optimization parameters for rates and exponents
$\hat{P}$	Empirical distribution
$P_{\mathbf{X}}$	Random-coding codeword distribution
q, q^n	Decoding metric; multiplicative n -letter extension
Q, Q^n	Input distribution; multiplicative n -letter extension
Q_n	Type that best approximates Q
R	Transmission rate
$R_{\text{cr}}, R_s^{\text{cr}}, R_{s,a}^{\text{cr}}$	Critical rates
rcu, rcu_s	Random-coding union bounds
$\text{rcux}_\rho, \text{rcux}_{\rho,s}$	Expurgated random-coding union bounds
$s, r_l, \bar{r}_l$	Dual optimization parameters for rates and exponents
T^n	Type class
$U_s, U_{s,a}$	Unconditional variances of generalized information densities
U, u	Auxiliary or time-sharing sequence
V	Channel dispersion
$V_s, V_{s,a}$	Conditional variances of generalized information densities
$\mathbf{V}, \mathbf{V}^{\text{iid}}$	Dispersion matrices
W, W^n	Channel transition law; multiplicative n -letter extension
$\mathbf{X}^{(i)}, \mathbf{x}^{(i)}$	i -th codeword in the codebook
$\mathbf{X}, \bar{\mathbf{X}}, \mathbf{x}, \bar{\mathbf{x}}$	Transmitted and non-transmitted codewords
$\mathbf{Y}, \mathbf{y}$	Output sequence

Greek Symbols

ϵ	Target error probability
μ_n	Normalizing constant
ω, Ω, Θ	Order notation for sequences
ϕ_l	Means of auxiliary cost functions
$\rho; \hat{\rho}$	Parameter for exponent functions; optimal value

Other Symbols

$\mathcal{C}, \mathbf{C}$	Deterministic codebook; random codebook
$\mathbf{m}, \hat{\mathbf{m}}$	Message; message estimate
$\mathcal{P}, \mathcal{P}_n$	Set of probability distributions; set of types
$\mathcal{D}_n$	Cost-constrained random coding constraint set
$\mathbf{Q}, \mathbf{Q}^{-1}$	Standard Gaussian tail probability; functional inverse
$\mathbb{R}$	Set of real numbers
$\mathcal{S}, \mathcal{T}$	Primal-domain constraint sets
$\mathcal{S}_n, \mathcal{T}_n$	Primal-domain type-based constraint sets
$\mathcal{U}$	Auxiliary or time-sharing alphabet
$\mathcal{X}$	Input alphabet
$\mathcal{Y}$	Output alphabet
$\mathbb{Z}$	Set of integers

Abbreviations

AWGN	Additive white Gaussian noise
BEC	Binary erasure channel
BSC	Binary symmetric channel

DMC	Discrete memoryless channel
DM-MAC	Discrete memoryless multiple-access channel
GMI	Generalized mutual information
i.i.d.	Independent and identically distributed
KKT	Karush-Kuhn-Tucker
MAC	Multiple-access channel
ML	Maximum likelihood
PDF	Probability density function
PMF	Probability mass function
SC	Superposition coding

Chapter 1

Introduction

1.1 Channel Coding and Mismatched Decoding

In his groundbreaking 1948 paper, Claude E. Shannon posed the problem of determining the conditions under which reliable communication can be achieved over a noisy channel [1]. The general setup is shown in Figure 1.1; the channel is fixed, and the encoder and decoder are to be designed. Modeling the channel as a probabilistic mapping from the input to the output, Shannon proved the remarkable result that one can design the encoder and decoder to achieve an arbitrarily small probability of error, while still transmitting information at a positive rate per channel use. The channel coding theorem gives the supremum of all rates at which this is possible, referred to as the channel capacity.

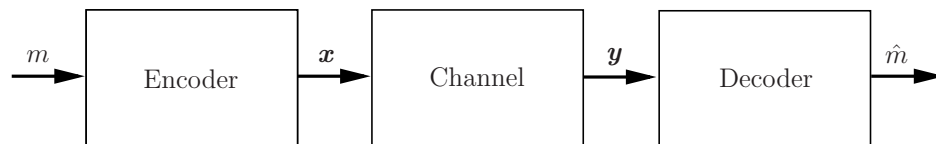


Figure 1.1 Communication channel

The channel coding theorem can be split into two statements: the achievability part, stating that the probability of error can be made arbitrarily small for any rate below the capacity, and the converse part, stating that the probability of error is bounded away from zero for any rate above the capacity. While the converse part must be proved for *any* decoder, there are several decoders that can be used to prove the achievability part, including:

- The maximum-likelihood (ML) decoder [2, Ch. 5], which chooses the most likely code-word under the channel law;

- The joint typicality decoder [1, 3], which looks for a unique codeword such that the empirical mean of a log-probability function is close to its expected value;
- The threshold decoder [4], which looks for a unique codeword whose likelihood exceeds some threshold.

Such decoders are typically studied in the random coding setting [1], where the codewords are independently generated according to some codeword distribution. There exists a code whose performance is at least as good as the average, and one can therefore prove the existence of good codes without explicitly constructing them. This is an instance of the probabilistic method [5].

Practical communication systems are subject to complexity limitations, and many decoders are ruled out. Furthermore, even if the computational complexity is not an issue, each of the above-mentioned decoding rules requires perfect knowledge of the channel, which is rarely the case in practice. Motivated by these issues, we consider the mismatched decoding problem [6–11], in which the decoding rule is fixed and possibly suboptimal. We proceed by formally presenting the system setup, and then discussing its applications and extensions.

1.1.1 System Setup

Here we describe the mismatched decoding setup that will be assumed throughout the thesis (except where stated otherwise). The input and output alphabets are denoted by $\mathcal{X}$ and $\mathcal{Y}$ respectively. The output sequence $\mathbf{y} = (y_1, \dots, y_n)$ is randomly generated from the input sequence $\mathbf{x} = (x_1, \dots, x_n)$ according to

$$W^n(\mathbf{y}|\mathbf{x}) \triangleq \prod_{i=1}^n W(y_i|x_i), \quad (1.1)$$

for some transition law $W(y|x)$. Except where stated otherwise, we assume that $\mathcal{X}$ and $\mathcal{Y}$ are finite, and thus the channel is a discrete memoryless channel (DMC) and W is a conditional probability mass function (PMF). In the case of continuous alphabets, W is instead a conditional probability density function (PDF).

The encoder takes as input a message m equiprobable on $\{1, \dots, M\}$, and transmits the corresponding codeword $\mathbf{x}^{(m)}$ from a codebook $\mathcal{C} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}\}$. The decoder receives the vector $\mathbf{y}$ at the channel output, and forms the estimate

$$\hat{m} = \arg \max_{j \in \{1, \dots, M\}} q^n(\mathbf{x}^{(j)}, \mathbf{y}), \quad (1.2)$$

where $q^n(\mathbf{x}, \mathbf{y}) \triangleq \prod_{i=1}^n q(x_i, y_i)$. The function $q(x, y)$ is a bounded non-negative function called the *decoding metric*, though it need not be a metric in the topological sense. In the case of a tie, a codeword achieving the maximum in (1.2) is selected uniformly at random.¹

An error is said to have occurred if $\hat{m} \neq m$, and we define the average error probability

$$p_e(\mathcal{C}) \triangleq \mathbb{P}[\hat{m} \neq m], \quad (1.3)$$

and the maximal error probability

$$p_{e,\max}(\mathcal{C}) \triangleq \max_{j=1,\dots,M} \mathbb{P}[\hat{m} \neq m \mid m = j]. \quad (1.4)$$

A rate R is said to be achievable if, for all $\delta > 0$, there exists a sequence of codes $\mathcal{C}_n$ of length n with $M \geq e^{n(R-\delta)}$ and vanishing error probability, i.e. $p_e(\mathcal{C}_n) \rightarrow 0$; an equivalent definition would be to use $p_{e,\max}$ in place of p_e . The mismatched capacity is defined to be the supremum of achievable rates.

The case $q(x, y) = W(y|x)$ corresponds to maximum-likelihood (ML) decoding, which is the rule that minimizes the probability of error. It follows that a special case of the mismatched capacity is the (matched) capacity, given by [1]

$$C = \max_Q I(X; Y), \quad (1.5)$$

where

$$I(X; Y) \triangleq \sum_{x,y} Q(x)W(y|x) \log \frac{W(y|x)}{\sum_{\bar{x}} Q(\bar{x})W(y|\bar{x})} \quad (1.6)$$

is the mutual information between the input X and output Y . The maximum in (1.5) is over all probability distributions on $\mathcal{X}$.

1.1.2 Applications

Perhaps the most obvious application of mismatched decoding is that in which the decoder has an incorrect estimate $\widehat{W}(y|x)$ of the channel, but uses the estimate as if it were perfect, i.e. $q(x, y) = \widehat{W}(y|x)$. Such channel uncertainty arises in numerous settings, including wireless communication [13–15].

Communication systems are often designed to combat additive Gaussian noise, and it is therefore of interest to study the performance when a system designed in this way is used on a

¹The subsequent definition of the mismatched capacity does not depend on how ties are resolved, since breaking ties as errors increases the error probability by at most a factor of two [12].

channel with non-Gaussian noise [16]. Mismatch may also occur due to phase offsets, arising from imperfect phase training [9].

In some settings, it may not be possible to implement the ML rule even when the channel is known. For example, its implementation may require infinite-precision arithmetic, and one may therefore be interested in the case that $q(x, y)$ is an approximation of $W(y|x)$ that only requires finite-precision arithmetic to implement [17, 18].

A widely-used technique used in practical systems is bit-interleaved coded modulation (BICM), which allows binary error-correcting codes to be used on non-binary channels by concatenating the binary encoder with an interleaver and a binary labeling [19]. In [20], this setup was studied from a mismatched decoding perspective, and results on the general mismatched decoding problem were applied to obtain performance bounds. Further studies of BICM using mismatched decoding were presented in [21–24].

Interestingly, mismatched decoding has recently been applied to the area of neuroscience [25, 26], with an aim to model and analyze information processing in the brain. For example, [25] suggests that a possible decoding model is one where neural responses exhibit correlations, but the decoder behaves as if they were independent.

The mismatched decoding problem is also interesting from a theoretical perspective. It was shown in [8] that the following quantities are special cases of the mismatched capacity: (i) the *zero undetected error capacity*, defined to be the supremum of rates such that the error probability can be made arbitrarily small and the decoder knows with certainty whether or not an error has occurred; (ii) the *zero-error capacity*, defined to be the supremum of rates such that the error probability is precisely zero. These are both long-standing open problems in information theory, suggesting that finding a computable expression for the mismatched capacity is difficult in general.

It should be noted that, while the definition of the mismatched capacity is seemingly innocuous, it may not be the most suitable performance measure in some of the above-mentioned applications. In particular, the definition involves an optimization over all codebooks, which is questionable unless the codebook designer has knowledge of both the channel and the decoding rule. Fortunately, the results obtained can usually provide insight into the system performance even when the codebook designer does not have such knowledge. For example, under random coding [1], the achievable rate with an optimized input distribution quantifies the performance when the codebook designer has full knowledge of the system, whereas the achievable rate with a fixed input distribution (e.g. uniform) may be a more appropriate performance measure in other cases.

1.1.3 Extensions

The maximum-metric decoding rule in (1.2) is one of many possible decoding rules. Here we outline some alternatives and discuss the reasons for focusing primarily on (1.2).

A simple extension is to consider metrics $q_n(\mathbf{x}, \mathbf{y})$ that need not be written as an n -fold product. Such metrics are highly general, and include the following:

- Joint typicality decoding is recovered by setting $q_n(\mathbf{x}, \mathbf{y}) = \mathbb{1}\{(\mathbf{x}, \mathbf{y}) \in \mathcal{A}\}$, where $\mathcal{A}$ is the typical set (see [1, 3] for details).
- Threshold decoding is recovered by setting $q_n(\mathbf{x}, \mathbf{y}) = \mathbb{1}\{W^n(\mathbf{y}|\mathbf{x}) \geq \gamma\}$ for some γ . Of course, W^n may also be replaced by a different quantity, thus yielding a mismatched version of threshold decoding.
- The maximum mutual information decoder is recovered by setting

$$q_n(\mathbf{x}, \mathbf{y}) = \sum_{x,y} \hat{P}_{\mathbf{x}\mathbf{y}}(x, y) \log \frac{\hat{P}_{\mathbf{x}\mathbf{y}}(x, y)}{\hat{P}_{\mathbf{x}}(x)\hat{P}_{\mathbf{y}}(y)}, \quad (1.7)$$

where $\hat{P}_{\mathbf{x}\mathbf{y}}$ denotes the empirical distribution of $(\mathbf{x}, \mathbf{y})$, and similarly for $\hat{P}_{\mathbf{x}}$ and $\hat{P}_{\mathbf{y}}$. This decoding rule can be used to prove the achievability of (1.5) for any DMC W [27], and it is universal in the sense that the rule does not depend on the channel.

One can also relax the restriction that the decoding rule depends only on the pairs $(\mathbf{x}, \mathbf{y})$; see [28] for a study of a class of decoding rules depending on all pairwise triplets $(\mathbf{x}, \bar{\mathbf{x}}, \mathbf{y})$. The study of iterative coding rules is highly relevant to settings such as bit-interleaved coded modulation [19] and low-density parity check coding [29].

We focus on the single-letter decoding rule in (1.2) since it is more analytically tractable, while still presenting an interesting and challenging problem with many applications. In contrast, under the more general metrics of the form $q_n(\mathbf{x}, \mathbf{y})$, it becomes difficult to perform a unified analysis with computable results. Furthermore, many of the above-mentioned decoding rules are less relevant to practical scenarios, since there are no known efficient methods for implementing or approximating them.

1.1.4 Previous Work on Mismatched Decoding

Achievable Rates

Achievable rates for mismatched decoding have been derived using the following random-coding ensembles (see Section 2.3 for formal definitions):

1. the i.i.d. ensemble, in which each symbol of each codeword is generated independently;
2. the constant-composition ensemble, in which each codeword is drawn uniformly from the set of sequences with a given empirical distribution;
3. the cost-constrained ensemble, in which each codeword is drawn according to an i.i.d. distribution conditioned on an auxiliary cost constraint being satisfied.

While these ensembles all yield the same achievable rate under ML decoding, i.e. the mutual information, this is not true in general under mismatched decoding.

To our knowledge, the first study of the mismatched decoding problem was by Stiglitz [30], who obtained an exponential bound on the error probability. Much of the proceeding work in mismatched decoding focused on the achievable rate called the generalized cutoff rate [17, 31, 32]. Fischer [33] considered the case that $q(x, y)$ is a conditional distribution on $\mathcal{Y}$ given $\mathcal{X}$, and used i.i.d. random coding to show that one can achieve the rate

$$\mathbb{E} \left[\log \frac{q(X, Y)}{\mathbb{E}[q(\bar{X}, Y) | Y]} \right], \quad (1.8)$$

where $(X, Y, \bar{X}) \sim Q(x)W(y|x)Q(\bar{x})$, and Q is an arbitrary input distribution.

The most prominent early works on mismatched decoding are those of Hui [7] and Csiszár-Körner [6], who independently derived the achievable rate commonly referred to as the *LM rate*. This rate can be expressed in two equivalent forms. Merhav *et al.* [9] provided the expression

$$I_{\text{LM}}(Q) \triangleq \sup_{s \geq 0, a(\cdot)} \mathbb{E} \left[\log \frac{q(X, Y)^s e^{a(X)}}{\mathbb{E}[q(\bar{X}, Y)^s e^{a(\bar{X})} | Y]} \right], \quad (1.9)$$

whereas the following expression was given in [6, 7]:

$$I_{\text{LM}}(Q) = \min_{\substack{\tilde{P}_{XY} : \tilde{P}_X = Q, \tilde{P}_Y = P_Y \\ \mathbb{E}_{\tilde{P}}[\log q(X, Y)] \geq \mathbb{E}_{Q \times W}[\log q(X, Y)]}} I_{\tilde{P}}(X; Y), \quad (1.10)$$

where the minimization is over all joint distributions satisfying the specified constraints, $P_Y(y) \triangleq \sum_x P_{XY}(x, y)$, and $I_{\tilde{P}}(X; Y)$ denotes the mutual information in (1.6) under the distribution $\tilde{P}_{XY}$. For binary-input DMCs, a matching converse to the LM rate was reported by Balakirsky [34]. However, in the general case, several examples have been given for which the rate is strictly smaller than the mismatched capacity [8, 9, 35].

The derivation of the LM rate in [7] uses random coding in which the empirical distribution of each codeword is constrained to be close to a given distribution $Q(x)$, whereas [6] uses

constant-composition random coding. Both proofs rely on the input and output alphabets being finite. In the case of general alphabets, Kaplan and Shamai [36] derived the achievable rate known as the *Generalized Mutual Information* (GMI), given by

$$I_{\text{GMI}}(Q) \triangleq \sup_{s \geq 0} \mathbb{E} \left[\log \frac{q(X, Y)^s}{\mathbb{E}[q(\bar{X}, Y)^s | Y]} \right]. \quad (1.11)$$

This rate can equivalently be expressed as follows, using the same notation as (1.10):

$$I_{\text{GMI}}(Q) = \min_{\substack{\tilde{P}_{XY} : \tilde{P}_Y = P_Y \\ \mathbb{E}_{\tilde{P}}[\log q(X, Y)] \geq \mathbb{E}_{Q \times W}[\log q(X, Y)]}} D(\tilde{P}_{XY} \| Q \times P_Y), \quad (1.12)$$

where

$$D(P_Z \| Q_Z) \triangleq \sum_z P_Z(z) \log \frac{P_Z(z)}{Q_Z(z)} \quad (1.13)$$

is the Kullback-Leibler divergence (or divergence for short).

Comparing (1.9) with (1.11), it is clear that $I_{\text{LM}}(Q) \geq I_{\text{GMI}}(Q)$. Motivated by this fact, Ganti *et al.* [11] proved that (1.9) is achievable for more general memoryless channels where the alphabets need not be finite. This was done by generating a number of codewords according to an i.i.d. distribution Q , and then discarding all of the codewords for which $|\frac{1}{n} \sum_{i=1}^n a(x_i) - \mathbb{E}_Q[a(X)]|$ exceeds some threshold, where $a(\cdot)$ can be optimized. An alternative derivation was given by Shamai and Sason [37] using cost-constrained random coding.

In the terminology of [11], (1.10) and (1.12) are primal expressions, and (1.9) and (1.11) are dual expressions. Indeed, the latter can be derived from the former using Lagrange duality techniques [38]. As well as extending readily to general alphabets, the dual expressions have the advantage that an achievable rate can be obtained by substituting any values of s and $a(\cdot)$ into (1.9) and (1.11), whereas (1.10) and (1.12) only give valid bounds after the minimization is performed.

The primal expressions can be generalized slightly by considering decoding rules of the form $q_n(\mathbf{x}, \mathbf{y}) = \alpha(P_{XY})$, where P_{XY} is the empirical distribution of $(\mathbf{x}, \mathbf{y})$. The maximum mutual information rule in (1.7) is an example of such a rule. Subject to minor technical conditions, (1.10) and (1.12) remain valid when each minimum is replaced by an infimum, and each constraint containing q is replaced by $\alpha(\tilde{P}_{XY}) \geq \alpha(Q \times W)$ [6, 8].

Further Results

Since the LM rate and GMI are smaller than the mismatched capacity in general, it is of interest to determine whether the weakness is in the random-coding ensembles or the bounding techniques used in the analysis. Ensemble-tightness results provide an answer to this question by proving that a given rate is the best possible for a given ensemble. For the i.i.d. ensemble with input distribution Q , it is known that the random-coding error probability $\bar{p}_e$ tends to 0 as $n \rightarrow \infty$ when $R < I_{\text{GMI}}(Q)$, whereas $\bar{p}_e \rightarrow 1$ as $n \rightarrow \infty$ when $R > I_{\text{GMI}}(Q)$ [11, 16, 36]. Similarly, for the constant-composition ensemble with input distribution Q , it has been shown that $\bar{p}_e \rightarrow 0$ as $n \rightarrow \infty$ when $R < I_{\text{LM}}(Q)$, whereas $\bar{p}_e \rightarrow 1$ as $n \rightarrow \infty$ when $R > I_{\text{LM}}(Q)$ [9]. This was proved under mild technical conditions in [9], and can be shown to hold more generally similarly to [16, Thm. 1].

Further properties of the LM rate were presented in [7–11]. It was shown by Csiszár and Narayan [8] that the LM rate is positive if and only if the mismatched capacity is positive, which in turn occurs if and only if there exists an input distribution Q such that

$$\mathbb{E}_{Q \times W}[\log q(X, Y)] \geq \mathbb{E}_{Q \times P_Y}[\log q(X, Y)], \quad (1.14)$$

where $P_Y(y) \triangleq \sum_x P_{XY}(x, y)$. It was also shown that I_{LM} is continuous in the pair (Q, W) when restricted to the set of channels satisfying

$$W(y|x) > 0 \implies q(x, y) > 0. \quad (1.15)$$

Channels failing this condition are of little interest, since they also fail the condition in (1.14) whenever the corresponding input x is used.

Another property given in [8] is the following:

$$I_{\text{LM}}(Q) \leq I(X; Y), \quad (1.16)$$

where $I(X; Y)$ is defined in (1.6). Moreover, equality holds if and only if

$$\log q(x, y) = \alpha(x) + \beta(y) + \gamma \log W(y|x) \quad (1.17)$$

for some $\alpha(x)$, $\beta(y)$ and $\gamma > 0$. The sufficiency of (1.17) can be understood by noting that for constant-composition codes, any decoding metric satisfying (1.17) has the same decision regions as the ML decoder. An analogous property to (1.16)–(1.17) holds for the GMI, except that $\alpha(x)$ is removed from (1.17) [20, 36]. It was shown by Lapidath [35] that the mismatched

capacity can equal the matched capacity even when (1.17) fails.

The LM rate (and the GMI) can be improved by considering products of the channel. That is, one can apply the LM rate to the channel $W^{(2)}((y_1, y_2)|(x_1, x_2)) = W(y_1|x_1)W(y_2|x_2)$ with metric $q^{(2)}((x_1, x_2), (y_1, y_2)) = q(x_1, y_1)q(x_2, y_2)$, and similarly for higher orders $W^{(k)}$ and $q^{(k)}$. Denoting the resulting LM rate by $I_{\text{LM}}^{(k)}$, it follows that $\frac{1}{k}I_{\text{LM}}^{(k)}$ is an achievable rate for the original channel. It was shown in [8] that this rate approaches the mismatched capacity as $k \rightarrow \infty$ for erasures-only metrics satisfying $q(x, y) = \max_{x', y'} q(x', y')$ for all (x, y) such that $W(y|x) > 0$. This result was also conjectured to hold for more general decoding metrics.

A multi-letter generalization of the LM rate for channels with memory was presented by Ganti *et al.* [11]. A similar study was performed by Somekh-Baruch [39], who gave a general (multi-letter) formula for the mismatched capacity in the sense of Verdú-Han [40].

A study of additive noise channels with nearest-neighbor decoding was performed by Lapidoth [16]. It was shown that, for both i.i.d. Gaussian coding and coding uniformly on the surface of a sphere, the random-coding rate achieved is the capacity of an additive white Gaussian noise (AWGN) channel with the same signal-to-noise ratio as that of the original (possibly non-Gaussian) channel.

Notions of mismatched capacity beyond that of standard fixed-length block coding were discussed in [9]. In particular, it was shown that higher rates can be achieved if the encoder is allowed to transmit symbols that disagree with the codebook in order to combat the mismatch. For example, if $\mathcal{X} = \mathcal{Y} = \{0, 1\}$, the channel flips each bit deterministically, and the decoder assumes that the channel output is equal to the channel input, then the mismatched capacity (according to the definition given in this chapter) is zero. However, a rate of 1 bit/use can be achieved if the encoder simply flips each bit before transmission.

The notion of mismatched capacity with cost constraints was studied in [11], including an investigation of the mismatched capacity per unit cost. It was observed that $I_{\text{LM}}(Q)$ is not a concave function of Q in general, suggesting that the optimization over Q is difficult. A geometric approach to the mismatched decoding problem was taken in [41] using information geometry techniques, and the problem of choosing the best decoding rule among a given class of decoding rules was studied.

In Appendix B, we complement the above results by providing further properties of the GMI and LM rate, with a particular focus on the optimization of s , $a(\cdot)$ and Q .

Other Settings With Mismatch

The mismatched multiple-access channel (MAC) was studied by Lapidoth [35], who gave an achievable rate region and proved its ensemble tightness. Moreover, it was shown that the

single-user LM rate can be improved by treating the channel as a MAC and generating multiple codebooks in parallel. Building on this approach, Somekh-Baruch [42] derived achievable rates for the mismatched cognitive MAC, where one user knows both messages. The techniques of [35] were applied to the mismatched degraded relay channel in [43].

In [44], Lapidoth considered a rate-distortion setup in which the encoding and decoding are done with respect to different distortion functions. An achievable compression rate analogous to (1.10) was given, and was shown to be asymptotically tight as $k \rightarrow \infty$ when applied to the k -th product of the source. The problem of lossless compression with side information at the decoder and a mismatched decoding rule was studied in [45]. It was shown that this problem is dual to that of mismatched channel coding in the sense that any achievable rate for one setting provides a corresponding achievable rate for the other, and similarly for converse results. The problem of universal decoding relative to a fixed class of mismatched decoding metrics was studied in [46].

The problem of reliable communication with imperfect knowledge of the channel has received significant attention in the context of wireless communication. A mismatched decoding approach was taken in [14] using Gaussian random codes and scaled nearest-neighbor decoding. Extensions to the multiple-antenna setting were presented in [15], and the notion of outage probability with mismatch was studied in [47].

The role of mismatch in practical coding techniques has also received attention. A mismatched version of the Viterbi decoding algorithm was studied in [48], and polar coding with mismatch was studied in [49, 50].

1.2 Asymptotic and Non-Asymptotic Performance

The most common performance measure for information-theoretic studies of channel coding is the capacity. This is a first-order asymptotic performance measure corresponding to an increasingly large block length, and hence its utility may be limited in the presence of constraints on the delay and complexity. In this section, we discuss several asymptotic notions beyond the channel capacity, and then discuss a purely non-asymptotic approach. Our presentation of the existing results is far from exhaustive; we focus on the results that are most relevant to this thesis, and we defer many of the discussions to later sections.

While the definitions below apply to both the matched and mismatched settings, most of the results stated will be for the matched setting, which has been the focus of the majority of the literature. The regimes that we consider can be likened to three regimes concerning the tail probability of a sum of i.i.d. random variables: large deviations, the central limit theorem,

and moderate deviations. In order to elucidate this analogy and familiarize the reader with the relevant terminology, we outline some results on i.i.d. sums in Appendix C.

1.2.1 Error Exponents

Let $p_e^*(n, M)$ be the smallest error probability for a given block length n and number of messages M . The reliability function $E(R)$ at rate $R > 0$ is defined as follows:

$$E(R) \triangleq \liminf_{n \rightarrow \infty} -\frac{1}{n} \log p_e^*(n, \lfloor e^{nR} \rfloor), \quad (1.18)$$

where $\lfloor \cdot \rfloor$ is the floor function. In other words, $E(R)$ is the highest exponent such that one can achieve $p_e \leq e^{-nE(R)+o(n)}$ at rate R . Knowledge of the reliability function provides more insight than the capacity alone, since it not only yields conditions under which the error probability tends to zero, but it also gives the exponential rate of decay of p_e .

Matched Decoding

In the matched setting, lower bounds on $E(R)$ (also known as achievable error exponents) were given by Fano [51], Gallager [2, 52] and Csiszár-Körner [27]. In particular, Gallager used i.i.d. random coding to obtain the exponent [52]

$$E_r(R) \triangleq \max_Q \max_{\rho \in [0,1]} E_0(Q, \rho) - \rho R, \quad (1.19)$$

where the first maximum is over all probability distributions on $\mathcal{X}$, and

$$E_0(Q, \rho) \triangleq -\log \sum_y \left(\sum_x Q(x) W(y|x)^{\frac{1}{1+\rho}} \right)^{1+\rho}. \quad (1.20)$$

This exponent is positive for all rates below the capacity C , and thus proves the achievability part of the channel coding theorem. Furthermore, the exponent coincides with an upper bound on $E(R)$ for all rates in an interval $[R_{\text{cr}}, C]$, where R_{cr} is known as the critical rate. The upper bound, due to Shannon *et al.* [53], is called the sphere packing exponent, and is given by

$$E_{\text{sp}}(R) \triangleq \max_Q \sup_{\rho \geq 0} E_0(Q, \rho) - \rho R. \quad (1.21)$$

The following alternative expressions for E_r and E_{sp} were derived by Csiszár and Körner [27] using the method of types:

$$E_r(R) = \max_Q \min_{\tilde{P}_{XY} : \tilde{P}_X = Q} D(\tilde{P}_{XY} \| Q \times W) + [I_{\tilde{P}}(X; Y) - R]^+ \quad (1.22)$$

$$E_{sp}(R) = \max_Q \min_{\substack{\tilde{P}_{XY} : \tilde{P}_X = Q \\ I_{\tilde{P}}(X; Y) \leq R}} D(\tilde{P}_{XY} \| Q \times W), \quad (1.23)$$

where the minimizations over $\tilde{P}_{XY}$ are over all joint distributions satisfying the specified constraints.

At rates below the critical rate, the reliability function is unknown in general, but improved upper and lower bounds are available. Gallager [2, Ch 5.7] considered an expurgated random-coding ensemble in which the worst half of the codewords are thrown away, yielding the exponent

$$E_{ex}(R) \triangleq \max_Q \sup_{\rho \geq 1} E_x(Q, \rho) - \rho R, \quad (1.24)$$

where

$$E_x(Q, \rho) \triangleq -\rho \log \sum_{x, \bar{x}} Q(x) Q(\bar{x}) \left(\sum_y \sqrt{W(y|x) W(y|\bar{x})} \right)^{1/\rho}. \quad (1.25)$$

This exponent improves on the random-coding exponent at low rates, and is tight in the limit as $R \rightarrow 0$ [54].

A different approach to obtaining expurgated exponents was taken by Csiszár, Körner and Marton [55] (see also [27, Ex. 10.18]), who began by proving the existence of a collection of constant-composition codewords such that any two codewords have a joint empirical distribution satisfying certain properties. By analyzing this collection of codewords using the method of types, an error exponent was obtained that coincides with that of Gallager after the optimization of the input distribution:

$$E_{ex}(R) = \max_Q \min_{\substack{P_{X\bar{X}} : P_X = P_{\bar{X}} = Q \\ I_P(X; \bar{X}) \leq R}} \mathbb{E}_P[d_B(X, \bar{X})] + I_P(X; \bar{X}) - R, \quad (1.26)$$

where $d_B(x, \bar{x}) \triangleq -\log \sum_y \sqrt{W(y|x) W(y|\bar{x})}$ is the Bhattacharyya distance. Yet another characterization was given by Csiszár and Körner [6]:

$$E_{ex}(R) = \max_Q \min_{\substack{P_{X\bar{X}Y} : P_X = P_{\bar{X}} = Q, I_P(X; \bar{X}) \leq R \\ \mathbb{E}_P[\log W(Y|\bar{X})] \geq \mathbb{E}_P[\log W(Y|X)]}} D(P_{X\bar{X}Y} \| Q \times Q \times W) - R. \quad (1.27)$$

For both the random-coding exponent and the expurgated exponent, Gallager's approach yields several advantages including its simplicity and the fact that the analysis is not restricted to finite alphabets. On the other hand, as we will see in Chapter 2 (see also [56]), the constant-composition exponents can be higher for a given input distribution or decoding rule, even though they do not yield an improvement for ML decoding with an optimized input distribution. However, the existing proofs rely heavily on techniques that are valid only when the input and output alphabets are finite; [55] uses the type packing lemma [27, Ch. 10], and [6] uses a combinatorial graph decomposition lemma.

An improved converse exponent at low rates is the straight-line bound [54], which is obtained by combining a vanishing-rate converse with the sphere packing bound and properties of list decoders.

As discussed in Section 1.1.4, in the case that the upper and lower bounds on the error exponent do not coincide, it is of interest to determine whether the weakness in the achievability results is in the random-coding ensemble itself, or the bounding techniques used in the analysis. The ensemble tightness of $E_r(R)$ was proved by Gallager [57] for i.i.d. random coding, and by D'yachkov [58] for constant-composition random coding.

Mismatched Decoding

Existing error exponents for mismatched decoding will be discussed in detail in Chapter 2, so we provide only a brief outline here. The constant-composition ensemble was studied in [6], where both random-coding exponents and expurgated exponents were derived. Random-coding exponents for the i.i.d. ensemble and cost-constrained ensemble were given in [36] and [37] respectively. The exponents in [6, 37] recover the LM rate, while the exponent in [36] recovers the GMI.

In [59], Poltyrev used the idea of mismatched decoding to derive random-coding exponents for maximum-likelihood decoding. Various equivalent forms of (1.19) and (1.22) were obtained by first deriving an exponent for an arbitrary decoding rule, and then performing an optimization of the decoding rule.

1.2.2 Second-Order Coding Rates

The problem of finding the second-order asymptotic expansion of the permissible coding rate for a given error probability was studied by Strassen [60], and later revisited by Polyanskiy *et al.* [61] and Hayashi [62], among others. Such a study is dual to that of error exponents: Instead of studying the behavior of the error probability for a fixed rate, one considers the

behavior of the rate for a fixed target error probability. The capacity gives the first-order characterization, whereas the second-order expansion gives a refined characterization.

For DMCs, the highest number of messages $M^*(n, \epsilon)$ for a given block length n and target error probability $\epsilon \in (0, 1)$ satisfies [60]

$$\log M^*(n, \epsilon) = nC - \sqrt{nV} Q^{-1}(\epsilon) + o(\sqrt{n}), \quad (1.28)$$

where C is the channel capacity, V is a quantity known as the channel dispersion [61], and Q^{-1} is the inverse of the standard Gaussian tail probability $Q(z) \triangleq \int_z^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$. The coefficient to the $\sqrt{n}$ term in (1.28) is often referred to as the second-order coding rate [62], and the right-hand side of (1.28) with the remainder term omitted is often referred to as the normal approximation.

We can interpret C and V as being the mean and variance of the information density

$$i(x, y) \triangleq \log \frac{W(y|x)}{\sum_{\bar{x}} Q(\bar{x})W(y|\bar{x})} \quad (1.29)$$

for some capacity-achieving input distribution Q , with the function $Q^{-1}(\cdot)$ appearing in (1.28) as a result of the central limit theorem. More precisely, letting Π denote the set of capacity-achieving input distributions, we have [60, 61]

$$V = \begin{cases} V_{\min} & \epsilon \in (0, \frac{1}{2}) \\ V_{\max} & \epsilon \in (\frac{1}{2}, 1), \end{cases} \quad (1.30)$$

where

$$V_{\min} \triangleq \min_{Q \in \Pi} \text{Var}[i(X, Y)] \quad (1.31)$$

$$V_{\max} \triangleq \max_{Q \in \Pi} \text{Var}[i(X, Y)] \quad (1.32)$$

with $(X, Y) \sim Q \times W$; note that $i(\cdot, \cdot)$ implicitly depends on Q .

1.2.3 Moderate Deviations

In contrast to error exponents and second-order coding rates, moderate deviations results characterize the asymptotic behavior when $p_e \rightarrow 0$ and $R \rightarrow C$ simultaneously.

Recall that $p_e^*(n, M)$ is the smallest error probability for a given block length n and number of messages M . Let ψ_n be a sequence such that $\psi_n \rightarrow 0$ and $\sqrt{n}\psi_n \rightarrow \infty$, and let V be the

channel dispersion for $\epsilon \in (0, \frac{1}{2})$ (see (1.30)). It was shown in [63, 64] that for any DMC such that $V > 0$, the following holds:

$$\lim_{n \rightarrow \infty} -\frac{1}{n\psi_n^2} \log p_e^*(n, \lfloor e^{n(C-\psi_n)} \rfloor) = \frac{1}{2V}. \quad (1.33)$$

Beyond the discrete memoryless setting, it was shown in [64] that results of the form (1.28) and (1.33) do not directly imply each other, but the values of V must coincide whenever both expansions hold.

1.2.4 Finite-Length Performance

While the above-mentioned asymptotic notions provide more insight into the system performance than the capacity alone, it is often unclear which one (if any) dictates the performance at finite block lengths. Early works providing non-asymptotic bounds include those of Feinstein [4], Gallager [2], and Strassen [60], and a more comprehensive study was recently performed by Polyanskiy *et al.* [61, Sec. II]. The most powerful of the achievability bounds therein, and the one of most interest in this thesis, is the random-coding union (RCU) bound:

$$p_e \leq \text{rcu}(n, M) \triangleq \mathbb{E} \left[\min \left\{ 1, (M-1) \mathbb{P}[W^n(\mathbf{Y}|\bar{\mathbf{X}}) \geq W^n(\mathbf{Y}|\mathbf{X}) | \mathbf{X}, \mathbf{Y}] \right\} \right], \quad (1.34)$$

where $(\mathbf{X}, \mathbf{Y}, \bar{\mathbf{X}}) \sim P_{\mathbf{X}}(\mathbf{x})W^n(\mathbf{y}|\mathbf{x})P_{\mathbf{X}}(\bar{\mathbf{x}})$ for some random-coding distribution $P_{\mathbf{X}}$. Several numerical examples were given in [61] for which, under i.i.d. random coding, the RCU bound gives a tight characterization of the finite-length performance (i.e. it lies close to a non-asymptotic converse bound). Moreover, the bound can be used to prove the achievability parts of all of the asymptotic results outlined in the preceding subsections, with the exception of the expurgated exponent. A notable drawback, however, is that its computation is generally difficult beyond symmetric setups.

1.3 Overview of the Thesis

The structure of the thesis and the main contributions are outlined as follows:

- In Chapter 2, we study random coding bounds and their asymptotic behavior under mismatched decoding. A cost-constrained random-coding ensemble with multiple auxiliary costs is introduced, and is shown to achieve error exponents (both expurgated and non-expurgated), second-order coding rates and moderate deviations results matching

those of constant-composition random coding, while being directly applicable to channels with infinite or continuous alphabets. In each case, the number of auxiliary costs required is shown to be at most two, and sometimes fewer.

- In Chapter 3, we present refined asymptotic results for various random-coding bounds. For i.i.d. random coding, we provide asymptotic estimates of two non-asymptotic bounds (including the mismatched version of the RCU bound) using the saddlepoint approximation [65]. Each expression is shown to characterize the asymptotic behavior of the corresponding random-coding bound at both fixed and varying rates, thus unifying the regimes characterized by error exponents, second-order rates and moderate deviations. Moreover, we obtain a prefactor to the expurgated i.i.d. exponent that behaves as $O(\frac{1}{\sqrt{n}})$, and we explicitly characterize the implied constant. Building on the analysis of the i.i.d. ensemble, we present refined asymptotic bounds for constant-composition random coding in the fixed-rate and fixed-error regimes, providing significant improvements compared to standard methods that introduce polynomial factors in the analysis.
- In Chapter 4, we consider the mismatched multiple-access channel (MAC). In the case of finite alphabets, we obtain error exponents that are tight with respect to the ensemble average, and positive within the interior of Lapidoth's achievable rate region [35]. In the special case of maximum-likelihood decoding, the ensemble tightness of the exponent of Liu-Hughes [66] is proved. We provide alternative expressions for the error exponents and rate regions, including expressions obtained using Lagrange duality which extend immediately to infinite and continuous alphabets. We study a multi-letter successive decoding rule, and we show via a numerical example that it can yield improved rates compared to the corresponding maximum-metric decoder. We also study the second-order asymptotics of the MAC, with a focus on maximum-likelihood decoding and constant-composition random coding. An inner bound on the set of locally achievable second-order coding rates [62, 67] is given for any fixed point on the boundary of the capacity region.
- In Chapter 5, the analysis techniques for the MAC are applied to two types of superposition coding for the point-to-point setting. The standard version is shown to yield an achievable rate which is at least as high as that of Lapidoth's expurgated parallel coding rate [35] after the optimization of the parameters. We introduce a refined version of superposition coding, and show that it achieves rates at least as high as the standard version for any set of random-coding parameters. It is shown that the gap between the two can be significant when the input distribution is fixed. Once again, ensemble-tight

error exponents are given in the case of finite alphabets, and extensions to more general alphabets are given.

- In Chapter 6, we review the main contributions of the thesis, and we present various directions for future research.
- In Appendix A, we give an overview of the method of types [27, Ch. 2], and list the main properties used throughout the thesis. In Appendix B, we provide several properties regarding the optimization of the LM rate and GMI. An introduction to asymptotic results on sums of independent random variables is given in Appendix C, with an aim to familiarize the reader with the relevant concepts and terminology. Upper and lower bounds on the probability of multiply-indexed union are presented in Appendix D; these are used in our analysis of multi-user random-coding techniques. In Appendix E, we provide equivalent forms of various convex optimization problems; these are used to establish equivalences between primal and dual forms of various achievable rates and error exponents.

1.4 Notation

Definitions of the most commonly-used mathematical symbols and abbreviations are listed prior to Chapter 1.

We use bold symbols for vectors (e.g. $\mathbf{x}$), and we denote the corresponding i -th entry using a subscript (e.g. x_i). We denote the Euclidean norm of a vector by $\|\cdot\|$, and the maximum absolute value of the entries of a vector by $\|\cdot\|_\infty$. The set of real numbers is denoted by $\mathbb{R}$, and the set of integers is denoted by $\mathbb{Z}$.

The set of all probability distributions on an alphabet $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$, and the set of all conditional distributions on $\mathcal{Y}$ given $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{Y}|\mathcal{X})$. More precisely, in the case that $\mathcal{X}$ is finite, we write $Q \in \mathcal{P}(\mathcal{X})$ when Q is a PMF on $\mathcal{X}$, and similarly for conditional PMFs. When we consider more general alphabets, we use the same notation to denote more general probability measures; the precise meaning will be clear from the context.

The set of all empirical distributions on a vector in $\mathcal{X}^n$ (i.e. types [27, Sec. 2], [56]) is denoted by $\mathcal{P}_n(\mathcal{X})$, and the set of all conditional empirical distributions (i.e. conditional types) on $\mathcal{Y}^n$ given $\mathcal{X}^n$ is denoted by $\mathcal{P}_n(\mathcal{Y}|\mathcal{X})$. The type of a vector $\mathbf{x}$ is denoted by $\hat{P}_{\mathbf{x}}(\cdot)$, and similarly for conditional types $\hat{P}_{\mathbf{y}|\mathbf{x}}(\cdot|\cdot)$ and joint types $\hat{P}_{\mathbf{xy}}(\cdot, \cdot)$. Given $Q \in \mathcal{P}_n(\mathcal{X})$, the type class $T^n(Q)$ is defined to be the set of all sequences in $\mathcal{X}^n$ with type Q . Given $P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ and $\mathbf{x} \in T^n(P_X)$, the conditional type class $T_{\mathbf{x}}^n(P_{XY})$ (sometimes written as $T_{\mathbf{x}}^n(P_{Y|X})$) is

defined to be the set of all sequences $\mathbf{y} \in \mathcal{Y}^n$ such that $(\mathbf{x}, \mathbf{y}) \in T^n(P_{XY})$. See Appendix A for further definitions and properties relating to types.

The probability of an event is denoted by $\mathbb{P}[\cdot]$, and the symbol $\sim$ means “distributed as”. The marginals of a joint distribution $P_{XY}(x, y)$ are denoted by $P_X(x)$ and $P_Y(y)$. Similarly, $P_{Y|X}(y|x)$ denotes the conditional distribution induced by $P_{XY}(x, y)$. We write $P_X = \tilde{P}_X$ to denote element-wise equality between two probability distributions on the same alphabet. Expectation with respect to a joint distribution $P_{XY}(x, y)$ is denoted by $\mathbb{E}_P[\cdot]$. When the probability distribution is understood from the context, we simply write $\mathbb{E}[\cdot]$. We use similar notation $\text{Var}_P[\cdot]$ and $\text{Var}[\cdot]$ for the variance of a random variable.

Given a distribution $Q(x)$ and a conditional distribution $W(y|x)$, we write $Q \times W$ to denote the joint distribution $Q(x)W(y|x)$, and similarly when there are more than two distributions. For example, given $Q_1(x_1)$, $Q_2(x_2)$ and $W(y|x_1, x_2)$ we have $Q_1 \times Q_2 \times W \sim Q_1(x_1)Q_2(x_2)W(y|x_1, x_2)$. We use standard notation for the entropy $H(X)$, conditional entropy $H(X|Y)$, mutual information $I(X; Y)$, conditional mutual information $I(X; Y|Z)$ and divergence $D(P||Q)$ [27]. In some cases, we explicitly state the underlying distribution of the random variables with a subscript (e.g. $I_P(X; Y) = \sum_{x,y} P_{XY}(x, y) \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)}$).

For two positive sequences f_n and g_n , we write $f_n \doteq g_n$ if $\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{f_n}{g_n} = 0$, and we write $f_n \dot{\leq} g_n$ if $\limsup_{n \rightarrow \infty} \frac{1}{n} \log \frac{f_n}{g_n} \leq 0$ and analogously for $\dot{\geq}$. We write $f_n \asymp g_n$ if $\lim_{n \rightarrow \infty} \frac{f_n}{g_n} = 1$. Furthermore, we write $f_n = O(g_n)$ if $|f_n| \leq c|g_n|$ for some c and sufficiently large n , and $f_n = o(g_n)$ if $\lim_{n \rightarrow \infty} \frac{f_n}{g_n} = 0$. We write $f_n = \Omega(g_n)$ if $g_n = O(f_n)$, $f_n = \omega(g_n)$ if $g_n = o(f_n)$, and $f_n = \Theta(g_n)$ if both $f_n = O(g_n)$ and $f_n = \Omega(g_n)$ hold.

We denote the tail probability of a zero-mean unit-variance Gaussian variable by $Q(\cdot)$, and we denote its functional inverse by $Q^{-1}(\cdot)$. All logarithms have base e , and all rates are in units of nats except in the examples, where bits are used. We define $[c]^+ = \max\{0, c\}$, and we denote the indicator function of an event by $\mathbb{1}\{\cdot\}$. The floor function is denoted by $\lfloor \cdot \rfloor$.

Chapter 2

Random Coding Bounds and Asymptotics

2.1 Introduction

Motivated by the fact that most existing work on mismatched decoding has focused on achievable rates, the main goal of this chapter is to present a more detailed analysis of the error probability, including error exponents, second-order rates, and moderate deviations results. The system setup is described in Section 1.1.1.

Recall that the average error probability for a given codebook is denoted by $p_e(\cdot)$, and the maximal error probability is denoted by $p_{e,\max}(\cdot)$. We use the symbol $\mathcal{C}$ to denote a given (deterministic) codebook, and we use the symbol $\mathbf{C}$ to denote a random codebook with codewords that are independent, each being drawn from some codeword distribution P_X . Our main focus is on bounding the random-coding error probability

$$\bar{p}_e(n, M) \triangleq \mathbb{E}[p_e(\mathbf{C})], \quad (2.1)$$

where $\mathbf{C}$ contains M codewords. There exists at least one codebook achieving an error probability as good as the average: $p_e(\mathcal{C}) \leq \bar{p}_e(n, M)$ for some (deterministic) codebook $\mathcal{C}$ with M codewords of length n .

The notion of ensemble tightness for rates and error exponents can now be formally defined as follows. An achievable rate I is ensemble-tight if

$$\sup \left\{ R : \liminf_{n \rightarrow \infty} \bar{p}_e(n, \lfloor e^{nR} \rfloor) = 0 \right\} = I, \quad (2.2)$$

and an achievable error exponent $E_r(R)$ is ensemble-tight if

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \bar{p}_e(n, \lfloor e^{nR} \rfloor) = E_r(R). \quad (2.3)$$

We will also be interested in analyzing $p_e(\mathcal{C})$ for a random codebook $\mathcal{C}$ containing M' codewords, and using it to bound the error probability for a fixed codebook $\mathcal{C}$ containing $M < M'$ codewords. We will refer to such techniques as expurgated random-coding techniques.

We will use the terminology *random-coding ensemble* when referring to expurgated or non-expurgated random coding with a particular codeword distribution. We will sometimes explicitly refer to an expurgated (or non-expurgated) random-coding ensemble, but the prefix will be omitted when the meaning is clear from the context.

The main contributions of this chapter are as follows:

- In Section 2.2, we present non-asymptotic bounds on the error probability for an arbitrary codeword distribution. These are not only of interest on their own right, but will also provide the starting point for the subsequent asymptotic derivations.
- In Section 2.3, we present a generalization of the cost-constrained ensemble in [2, Ch 7.3], [37] to include multiple auxiliary costs. This ensemble serves as an alternative to constant-composition codes for improving the performance compared to i.i.d. coding, while being applicable to channels with infinite or continuous alphabets.
- Random-coding error exponents, second-order coding rates, moderate deviations results, and expurgated error exponents are presented in Sections 2.4, 2.5, 2.6 and 2.7 respectively. In each case, the asymptotic bound given is seen to be attainable with at most two auxiliary costs, and sometimes fewer. A numerical example is provided in Section 2.8.
- In Section 2.9, we explain how the above-mentioned results can be extended to channels with input constraints and non-finite alphabets. Our techniques are seen to provide, to our knowledge, the most general known derivations of achievable second-order coding rates and moderate deviations results in such cases.

As noted in Section 1.2, the asymptotic regimes considered can be likened to various regimes concerning the tail probability of an i.i.d. sum, an outline of which is given in Appendix C.

Parts of this chapter were presented in [68–72].

2.2 Non-Asymptotic Bounds

In this section, we provide non-asymptotic performance bounds for mismatched decoding with an arbitrary codeword distribution P_X . All of the bounds given extend immediately to general channels and decoding metrics that need not be memoryless nor have finite alphabets.

However, the computation of the bounds is generally difficult beyond memoryless channels exhibiting a high degree of symmetry.

Throughout the section, we assume without loss of generality that the first message is encoded, i.e. $m = 1$. We define the random codewords

$$\{\mathbf{X}^{(i)}\}_{i=1}^M \sim \prod_{i=1}^M P_{\mathbf{X}}(\mathbf{x}^{(i)}). \quad (2.4)$$

We write $\mathbf{X}$ in place of $\mathbf{X}^{(1)}$, and we let $\bar{\mathbf{X}}$ denote an arbitrary codeword $\mathbf{X}^{(i)}$ with $i \neq 1$. The channel output sequence is denoted by $\mathbf{Y}$, yielding

$$(\mathbf{X}, \mathbf{Y}, \bar{\mathbf{X}}) \sim P_{\mathbf{X}}(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) P_{\mathbf{X}}(\bar{\mathbf{x}}). \quad (2.5)$$

The following theorem gives upper and lower bounds on the random-coding error probability defined in (2.1). In particular, we provide a simple extension of the random-coding union (RCU) bound of Polyanskiy *et al.* [61] (see (1.34)) to the mismatched setting, and show that the bound is tight with respect to the ensemble average to within a constant factor.

Theorem 2.1. *For any mismatched memoryless channel (possibly having infinite or continuous alphabets), codeword distribution $P_{\mathbf{X}}(\mathbf{x})$ and constant $s \geq 0$, the random-coding error probability $\bar{p}_e(n, M)$ satisfies*

$$\frac{1}{4} \text{rcu}(n, M) \leq \bar{p}_e(n, M) \leq \text{rcu}(n, M) \leq \text{rcu}_s(n, M), \quad (2.6)$$

where

$$\text{rcu}(n, M) \triangleq \mathbb{E}[\min\{1, (M-1)\mathbb{P}[q^n(\bar{\mathbf{X}}, \mathbf{Y}) \geq q^n(\mathbf{X}, \mathbf{Y}) | \mathbf{X}, \mathbf{Y}]\}] \quad (2.7)$$

$$\text{rcu}_s(n, M) \triangleq \mathbb{E}\left[\min\left\{1, (M-1) \frac{\mathbb{E}[q^n(\bar{\mathbf{X}}, \mathbf{Y})^s | \mathbf{Y}]}{q^n(\mathbf{X}, \mathbf{Y})^s}\right\}\right]. \quad (2.8)$$

Proof. Similarly to [61], we obtain the upper bound rcu by writing

$$\bar{p}_e(n, M) \leq \mathbb{P}\left[\bigcup_{i \neq 1} \{q^n(\mathbf{X}^{(i)}, \mathbf{Y}) \geq q^n(\mathbf{X}, \mathbf{Y})\}\right] \quad (2.9)$$

$$= \mathbb{E}\left[\mathbb{P}\left[\bigcup_{i \neq 1} \{q^n(\mathbf{X}^{(i)}, \mathbf{Y}) \geq q^n(\mathbf{X}, \mathbf{Y})\} \mid \mathbf{X}, \mathbf{Y}\right]\right] \quad (2.10)$$

$$\leq \text{rcu}(n, M), \quad (2.11)$$

where (2.9) follows by upper bounding the random-coding error probability by that of the decoder that breaks ties as errors, and (2.11) follows by applying the truncated union bound. To prove the lower bound in (2.6), it suffices to show that each of the upper bounds in (2.9) and (2.11) is tight to within a factor of two. The matching lower bound to (2.9) follows since whenever a tie occurs it must be between at least two codewords [12], and the matching lower bound to (2.11) follows since the union is over independent events [73, Lemma A.2]. We obtain the upper bound rcu_s by weakening rcu using Markov's inequality. $\square$

As discussed in Section 1.2.1, improved performance bounds can be obtained at low rates by expurgating “bad” codewords from the randomly generated codebook. Non-asymptotic bounds resulting from such techniques are given in the following theorem, which is proved using Gallager's bounding techniques [2, Ch. 5.7]. Following the above terminology, we refer to the resulting bounds as *expurgated random-coding union* (RCUX) bounds.

Theorem 2.2. *For any mismatched memoryless channel (possibly having infinite or continuous alphabets), codeword distribution $P_{\mathbf{X}}$, and parameters $\rho \geq 1$ and $s \geq 0$, there exists a codebook $\mathcal{C}$ with M codewords of length n such that*

$$p_{\text{e,max}}(\mathcal{C}) \leq \text{rcux}_{\rho}(n, M) \leq \text{rcux}_{\rho,s}(n, M), \quad (2.12)$$

where

$$\text{rcux}_{\rho}(n, M) \triangleq \left(4(M-1) \mathbb{E} \left[\mathbb{P} \left[q^n(\bar{\mathbf{X}}, \mathbf{Y}) \geq q^n(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X}, \bar{\mathbf{X}} \right]^{1/\rho} \right] \right)^{\rho} \quad (2.13)$$

$$\text{rcux}_{\rho,s}(n, M) \triangleq \left(4(M-1) \mathbb{E} \left[\mathbb{E} \left[\left(\frac{q^n(\bar{\mathbf{X}}, \mathbf{Y})}{q^n(\mathbf{X}, \mathbf{Y})} \right)^s \mid \mathbf{X}, \bar{\mathbf{X}} \right]^{1/\rho} \right] \right)^{\rho}. \quad (2.14)$$

Proof. It was shown by Gallager [2, Sec. 7.3], using techniques that apply immediately to the mismatched setting, that there exists a codebook $\mathcal{C}$ of size M such that

$$p_{\text{e,max}}(\mathcal{C}) \leq \left(2 \mathbb{E} [p_{\text{e}}(\mathcal{C})^{1/\rho}] \right)^{\rho}, \quad (2.15)$$

where $\mathcal{C}$ contains $M' = 2M - 1$ random codewords. We obtain (2.13) from (2.15) by weakening the expectation as follows:

$$\mathbb{E} [p_{\text{e}}(\mathcal{C})^{1/\rho}] \leq \mathbb{E} \left[\left(\sum_{i \neq 1} \mathbb{P} [q^n(\mathbf{X}^{(i)}, \mathbf{Y}) \geq q^n(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X}, \mathbf{X}^{(i)}] \right)^{1/\rho} \right] \quad (2.16)$$

$$\leq \mathbb{E} \left[2(M-1) \mathbb{P} [q^n(\bar{\mathbf{X}}, \mathbf{Y}) \geq q^n(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X}, \bar{\mathbf{X}}]^{1/\rho} \right], \quad (2.17)$$

where (2.16) follows from the union bound, and (2.17) follows using $M' = 2M - 1$ along with the inequality [2, Ex. 4.15(f)]

$$\left(\sum_i a_i\right)^{1/\rho} \leq \sum_i a_i^{1/\rho}, \quad (2.18)$$

which holds for any $\rho \geq 1$. We obtain (2.14) by applying Markov's inequality to the inner probability in (2.13). $\square$

The RCU bounds will prove useful for obtaining error exponents, second-order coding rates and moderate deviations results, whereas the RCUX bounds will provide improved exponents at low rates.

2.3 Cost-Constrained Random Coding with Multiple Auxiliary Costs

The most commonly-used codeword distributions in the literature are the i.i.d. distribution,

$$P_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n Q(x_i), \quad (2.19)$$

and the constant-composition distribution,

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{|T^n(Q_n)|} \mathbb{1}\{\mathbf{x} \in T^n(Q_n)\}, \quad (2.20)$$

where $Q_n \in \mathcal{P}_n(\mathcal{X})$ is a type such that $\max_x |Q_n(x) - Q(x)| \leq \frac{1}{n}$. The latter is of particular interest in the mismatched setting, since it yields the LM rate, whereas the i.i.d. distribution only yields the GMI (see Section 1.1.4). However, the constant-composition distribution relies on the input being finite.

Here we present a codeword distribution that includes (2.19) and (2.20) as special cases, and that will provide similar performance gains to the constant-composition distribution even for channels with non-finite alphabets. The distribution is given by

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\mu_n} \prod_{i=1}^n Q(x_i) \mathbb{1}\{\mathbf{x} \in \mathcal{D}_n\}, \quad (2.21)$$

where

$$\mathcal{D}_n \triangleq \left\{ \mathbf{x} : \left| \frac{1}{n} \sum_{i=1}^n a_l(x_i) - \phi_l \right| \leq \frac{\delta}{n}, l = 1, \dots, L \right\}, \quad (2.22)$$

and where μ_n is a normalizing constant (i.e. $\mu_n = \mathbb{P}[\mathbf{X}' \in \mathcal{D}_n]$ with $\mathbf{X}' \sim \prod_{i=1}^n Q(x'_i)$), δ is a positive constant, and for each $l = 1, \dots, L$, $a_l(\cdot)$ is a real-valued function on $\mathcal{X}$, and $\phi_l \triangleq \mathbb{E}_Q[a_l(X)]$. We refer to each function $a_l(\cdot)$ as an *auxiliary cost function*, or simply a cost. Observe that each codeword is generated according to an i.i.d. distribution conditioned on the empirical mean of each cost function being close to the true mean. This generalizes the codeword distribution studied in [2, Sec. 7.3], [37] by including multiple costs.

The cost functions $\{a_l(\cdot)\}_{l=1}^L$ in (2.21) should not be viewed as being chosen to meet a system constraint (e.g. power limitations). Rather, they are introduced in order to improve the performance of the random-coding ensemble itself. However, system costs can be handled similarly; see Section 2.9 for details. The constant δ in (2.21) could, in principle, vary with l and n , but a fixed value will suffice for our purposes.

In the case that $L = 0$, it should be understood that $\mathcal{D}_n = \mathcal{X}^n$. In this case, (2.21) reduces to the i.i.d. codeword distribution in (2.19). A less obvious special case of (2.21) is the constant-composition distribution in (2.20), which is recovered from (2.21) by replacing Q by Q_n and choosing the parameters $L = |\mathcal{X}|$, $\delta \in [0, 1)$, and

$$a_l(x) = \mathbb{1}\{x = l\}, \quad l = 1, \dots, |\mathcal{X}|, \quad (2.23)$$

where we assume without loss of generality that $\mathcal{X} = \{1, \dots, |\mathcal{X}|\}$.

The following proposition shows that the normalizing constant μ_n in (2.21) decays at most polynomially fast in n . When $|\mathcal{X}|$ is finite, it follows immediately from the property of types in (A.8) that $\mu_n \geq (n+1)^{-(|\mathcal{X}|-1)}$. For the sake of generality, we present a proof that applies to more general alphabets, subject to minor technical conditions. The case $L = 1$ was handled in [2, Sec. 7.3].

Proposition 2.1. *Fix the input alphabet $\mathcal{X}$ (possibly infinite or continuous), input distribution $Q \in \mathcal{P}(\mathcal{X})$ and auxiliary cost functions $a_1(\cdot), \dots, a_L(\cdot)$. If $\mathbb{E}_Q[a_l(X)^2] < \infty$ for $l = 1, \dots, L$, then there exists a choice of $\delta > 0$ such that the normalizing constant in (2.21) satisfies $\mu_n = \Omega(n^{-L/2})$.*

Proof. This result follows from the multivariate local limit theorem in [74, Cor. 1], which gives asymptotic expressions for probabilities of i.i.d. random vectors taking values in sets of the form (2.22). Let Σ denote the covariance matrix of the vector $[a_1(X), \dots, a_L(X)]^T$. We have by assumption that the entries of Σ are finite. Under the additional assumption $\det(\Sigma) > 0$, [74, Cor. 1] states that $\mu_n = \Theta(n^{-L/2})$ provided that δ is at least as high as the largest span of the $a_l(X)$ ($X \sim Q$) that are lattice variables (see Definition C.1 in Appendix C). If all such variables are non-lattice, then δ can take any positive value.

It only remains to handle the case $\det(\Sigma) = 0$. Suppose that Σ has rank $L' < L$, and assume without loss of generality that $a_1(\cdot), \dots, a_{L'}(\cdot)$ are linearly independent. Up to sets whose probability with respect to Q is zero, the remaining costs $a_{L'+1}(\cdot), \dots, a_L(\cdot)$ can be written as linear combinations of the first L' costs. Letting α denote the largest magnitude of the scalar coefficients in these linear combinations, we see that $x \in \mathcal{D}_n$ provided that

$$\left| \frac{1}{n} \sum_{i=1}^n a_l(x_i) - \phi_l \right| \leq \frac{\delta}{\alpha L' n} \quad (2.24)$$

for $l = 1, \dots, L'$. The proposition follows by choosing δ to be at least as high as $\alpha L'$ times the largest span of the $a_l(X)$ that are lattice variables, and analyzing the first L' costs analogously to the case that $\det(\Sigma) > 0$. $\square$

In accordance with Proposition 2.1, we henceforth assume that the choice of δ is such that $\mu_n = \Omega(n^{-L/2})$.

2.4 Random-Coding Error Exponents

2.4.1 Cost-Constrained Ensemble

Here we present the ensemble-tight error exponent for (non-expurgated) cost-constrained random coding, yielding results for the i.i.d. and constant-composition ensembles as special cases. We define the sets

$$\mathcal{S}(\{a_l\}) \triangleq \{P_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \mathbb{E}_P[a_l(X)] = \phi_l, l = 1, \dots, L\} \quad (2.25)$$

$$\begin{aligned} \mathcal{T}(P_{XY}, \{a_l\}) \triangleq \{ & \tilde{P}_{XY} \in \mathcal{P}(\mathcal{X} \times \mathcal{Y}) : \mathbb{E}_{\tilde{P}}[a_l(X)] = \phi_l, l = 1, \dots, L, \\ & \tilde{P}_Y = P_Y, \mathbb{E}_{\tilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)] \}, \end{aligned} \quad (2.26)$$

where the notation $\{a_l\}$ is used to denote dependence on $a_1(\cdot), \dots, a_L(\cdot)$. The dependence of these sets on Q (via $\phi_l = \mathbb{E}_Q[a_l(X)]$) is kept implicit.

Theorem 2.3. *For any input distribution $Q \in \mathcal{P}(\mathcal{X})$, auxiliary costs $\{a_l\}_{l=1}^L$, and rate $R > 0$, the random-coding error probability for the cost-constrained ensemble in (2.21) satisfies*

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \bar{p}_e(n, \lfloor e^{nR} \rfloor) = E_r^{\text{cost}}(Q, R, \{a_l\}), \quad (2.27)$$

where

$$E_r^{\text{cost}}(Q, R, \{a_l\}) \triangleq \min_{P_{XY} \in \mathcal{S}(\{a_l\})} \min_{\tilde{P}_{XY} \in \mathcal{T}(P_{XY}, \{a_l\})} D(P_{XY} \| Q \times W) + [D(\tilde{P}_{XY} \| Q \times \tilde{P}_Y) - R]^+. \quad (2.28)$$

Proof. See Section 2.10.1. $\square$

The optimization problem in (2.28) is convex when the input distribution and cost functions are fixed. The following theorem gives an alternative expression based on Lagrange duality [38].

Theorem 2.4. *The error exponent in (2.28) can be expressed as*

$$E_r^{\text{cost}}(Q, R, \{a_l\}) = \max_{\rho \in [0, 1]} E_0^{\text{cost}}(Q, \rho, \{a_l\}) - \rho R, \quad (2.29)$$

where

$$E_0^{\text{cost}}(Q, \rho, \{a_l\}) \triangleq \sup_{s \geq 0, \{r_l\}, \{\bar{r}_l\}} -\log \mathbb{E} \left[\left(\frac{\mathbb{E}[q(\bar{X}, Y)^s e^{\sum_{l=1}^L \bar{r}_l(a_l(\bar{X}) - \phi_l)} | Y]}{q(X, Y)^s e^{\sum_{l=1}^L r_l(a_l(X) - \phi_l)}} \right)^\rho \right] \quad (2.30)$$

and $(X, Y, \bar{X}) \sim Q(x)W(y|x)Q(\bar{x})$.

Proof. See Section 2.10.2. $\square$

The derivation of (2.29)–(2.30) via Theorem 2.3 is useful for proving ensemble tightness, but has the disadvantage of being applicable only in the case of finite alphabets. We proceed by giving a direct derivation that does not prove ensemble tightness, but that extends immediately to more general alphabets (see Section 2.9 for further discussion).

Applying (2.21) and $\min\{1, \alpha\} \leq \alpha^\rho$ ($\rho \in [0, 1]$) to rcu_s in (2.8), we obtain

$$\bar{p}_e(n, M) \leq \frac{1}{\mu_n^{1+\rho}} M^\rho \sum_{\mathbf{x} \in \mathcal{D}_n, \mathbf{y}} Q^n(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) \left(\frac{\sum_{\bar{\mathbf{x}} \in \mathcal{D}_n} Q^n(\bar{\mathbf{x}}) q^n(\bar{\mathbf{x}}, \mathbf{y})^s}{q^n(\mathbf{x}, \mathbf{y})^s} \right)^\rho, \quad (2.31)$$

where $Q^n(\mathbf{x}) \triangleq \prod_{i=1}^n Q(x_i)$. From (2.22), each codeword $\mathbf{x} \in \mathcal{D}_n$ satisfies

$$e^{r(a_l^n(\mathbf{x}) - n\phi_l)} e^{|r|\delta} \geq 1 \quad (2.32)$$

for any real number r , where $a_l^n(\mathbf{x}) \triangleq \sum_{i=1}^n a_l(x_i)$. Weakening (2.31) by applying (2.32)

multiple times, we obtain

$$\begin{aligned} \bar{p}_e(n, M) &\leq \frac{e^{\rho \sum_l (|r_l| + |\bar{r}_l|)\delta}}{\mu_n^{1+\rho}} M^\rho \sum_{\mathbf{x} \in \mathcal{D}_n, \mathbf{y}} Q^n(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) \\ &\quad \times \left(\frac{\sum_{\bar{\mathbf{x}} \in \mathcal{D}_n} Q^n(\bar{\mathbf{x}}) q^n(\bar{\mathbf{x}}, \mathbf{y})^s e^{\sum_l \bar{r}_l (a_l^n(\bar{\mathbf{x}}) - n\phi_l)}}{q^n(\mathbf{x}, \mathbf{y})^s e^{\sum_l r_l (a_l^n(\mathbf{x}) - n\phi_l)}} \right)^\rho, \end{aligned} \quad (2.33)$$

where $\{r_l\}$ and $\{\bar{r}_l\}$ are arbitrary. Further weakening (2.33) by replacing the summations over $\mathcal{D}_n$ by summations over all sequences, and expanding each term in the outer summation as product from 1 to n , we obtain

$$\bar{p}_e(n, M) \leq \frac{e^{\rho \sum_l (|r_l| + |\bar{r}_l|)\delta}}{\mu_n^{1+\rho}} M^\rho \left(\sum_{x, y} Q(x) W(y|x) \left(\frac{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{\sum_l \bar{r}_l (a_l(\bar{x}) - \phi_l)}}{q(x, y)^s e^{\sum_l r_l (a_l(x) - \phi_l)}} \right)^\rho \right)^n. \quad (2.34)$$

Since μ_n decays to zero subexponentially in n (cf. Proposition 2.1), we conclude that the prefactor in (2.34) does not affect the exponent. Setting $M = \lfloor e^{nR} \rfloor$, we obtain (2.29).

The preceding analysis can be considered a refinement of that of Shamai and Sason [37], who showed that an achievable error exponent for $L = 1$ is given by

$$E_r^{\text{cost}'}(Q, R, a_1) \triangleq \max_{\rho \in [0, 1]} E_0^{\text{cost}'}(Q, \rho, a_1) - \rho R, \quad (2.35)$$

where

$$E_0^{\text{cost}'}(Q, \rho, a_1) \triangleq \sup_{s \geq 0} -\log \mathbb{E} \left[\left(\frac{\mathbb{E}[q(\bar{X}, Y)^s e^{a_1(\bar{X})} | Y]}{q(X, Y)^s e^{a_1(X)}} \right)^\rho \right]. \quad (2.36)$$

By setting $r_1 = \bar{r}_1 = 1$ in (2.30), we see that E_r^{cost} with $L = 1$ is at least as high as $E_r^{\text{cost}'}$. In Section 2.4.4, we show that the former can be strictly higher.

2.4.2 i.i.d. and Constant-Composition Ensembles

Setting $L = 0$ in (2.30), we recover the exponent of Kaplan and Shamai [36], namely

$$E_r^{\text{iid}}(Q, R) \triangleq \max_{\rho \in [0, 1]} E_0^{\text{iid}}(Q, \rho) - \rho R, \quad (2.37)$$

where

$$E_0^{\text{iid}}(Q, \rho) \triangleq \sup_{s \geq 0} -\log \mathbb{E} \left[\left(\frac{\mathbb{E}[q(\bar{X}, Y)^s | Y]}{q(X, Y)^s} \right)^\rho \right]. \quad (2.38)$$

In the special case of constant-composition random coding (see (2.23)), the constraints $\mathbb{E}_{\tilde{P}}[a_l(X)] = \phi_l$ for $l = 1, \dots, |\mathcal{X}|$ yield $P_X = Q$ and $\tilde{P}_X = Q$ in (2.25) and (2.26) respectively, and thus (2.28) coincides with Csiszár's exponent for constant-composition coding [6]. Hence, we have proved that the exponents of [6, 36] are ensemble-tight.

We henceforth denote the exponent for the constant-composition ensemble by $E_r^{\text{cc}}(Q, R)$. We claim that

$$E_r^{\text{cc}}(Q, R) = \max_{\rho \in [0, 1]} E_0^{\text{cc}}(Q, \rho) - \rho R, \quad (2.39)$$

where

$$E_0^{\text{cc}}(Q, \rho) = \sup_{s \geq 0, a(\cdot)} \mathbb{E} \left[-\log \mathbb{E} \left[\left(\frac{\mathbb{E}[q(\bar{X}, Y)^s e^{a(\bar{X})} | Y]}{q(X, Y)^s e^{a(X)}} \right)^\rho \middle| X \right] \right]. \quad (2.40)$$

To prove this, we first note from (2.23) that

$$\sum_l r_l(a_l(x) - \phi_l) = \sum_{\tilde{x}} r_{\tilde{x}}(\mathbb{1}\{x = \tilde{x}\} - Q(\tilde{x})) \quad (2.41)$$

$$= r(x) - \phi_r, \quad (2.42)$$

where (2.41) follows since $\phi_l = \mathbb{E}_Q[\mathbb{1}\{x = l\}] = Q(l)$, and (2.42) follows by defining $r(x) \triangleq r_x$ and $\phi_r \triangleq \mathbb{E}_Q[r(X)]$. Defining $\bar{r}(x)$ and $\phi_{\bar{r}}$ in a similar fashion, we obtain the following E_0 function from (2.30):

$$E_0^{\text{cc}}(Q, \rho) \triangleq \sup_{s \geq 0, r(\cdot), \bar{r}(\cdot)} -\log \sum_{x, y} Q(x) W(y|x) \left(\frac{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{\bar{r}(\bar{x}) - \phi_{\bar{r}}}}{q(x, y)^s e^{r(x) - \phi_r}} \right)^\rho \quad (2.43)$$

$$\leq \sup_{s \geq 0, r(\cdot), \bar{r}(\cdot)} -\sum_x Q(x) \log \sum_y W(y|x) \left(\frac{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{\bar{r}(\bar{x}) - \phi_{\bar{r}}}}{q(x, y)^s e^{r(x) - \phi_r}} \right)^\rho \quad (2.44)$$

$$= \sup_{s \geq 0, \bar{r}(\cdot)} -\sum_x Q(x) \log \sum_y W(y|x) \left(\frac{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{\bar{r}(\bar{x})}}{q(x, y)^s e^{\bar{r}(x)}} \right)^\rho, \quad (2.45)$$

where (2.44) follows from Jensen's inequality, and (2.45) follows by using the definitions of ϕ_r and $\phi_{\bar{r}}$ to write

$$-\sum_x Q(x) \log \left(\frac{e^{-\phi_{\bar{r}}}}{e^{r(x) - \phi_r}} \right)^\rho = -\sum_x Q(x) \log \left(\frac{1}{e^{\bar{r}(x)}} \right)^\rho = \rho \phi_{\bar{r}}. \quad (2.46)$$

Renaming $\bar{r}(\cdot)$ as $a(\cdot)$, we see that (2.45) coincides with (2.40). It remains to show that equality

holds in (2.44). For $\rho > 0$, this is seen by noting that the choice

$$r(x) = \frac{1}{\rho} \log \sum_y W(y|x) \left(\frac{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{\bar{r}(\bar{x})}}{q(x, y)^s} \right)^\rho \quad (2.47)$$

makes the argument to the logarithm in (2.44) independent of x , thus ensuring that Jensen's inequality holds with equality. The case $\rho = 0$ is trivial.

2.4.3 Achievable Rates

Achievable rates can be obtained from both the primal and dual exponents. The methods for doing this are well-understood [2, 6, 27, 36], so we provide only an outline here. In the primal domain, we observe that the objective in (2.28) is positive whenever the optimizing P_{XY} differs from $Q \times W$. Moreover, similarly to [6, Lem. 1], $\min_{\tilde{P}_{XY} \in \mathcal{T}(P_{XY}, \{a_l\})} D(\tilde{P}_{XY} \| Q \times \tilde{P}_Y)$ is continuous in P_{XY} , and hence taking $P_{XY} \rightarrow Q \times W$ yields the rate

$$\min_{\tilde{P}_{XY} \in \mathcal{T}(Q \times W, \{a_l\})} D(\tilde{P}_{XY} \| Q \times \tilde{P}_Y). \quad (2.48)$$

In particular, this yields the GMI in (1.12) under the i.i.d. ensemble, and the LM rate in (1.10) under the constant-composition ensemble.

In the dual domain, we obtain an achievable rate similarly to [2, Ch. 5] by evaluating the partial derivative of the objective in (2.30) at $\rho = 0$, yielding the rate

$$\sup_{s \geq 0, \{r_l\}, \{\bar{r}_l\}} \mathbb{E} \left[\log \frac{q(X, Y)^s e^{\sum_l r_l(a_l(X) - \phi_l)}}{\mathbb{E}[q(\bar{X}, Y)^s e^{\sum_l \bar{r}_l(a_l(\bar{X}) - \phi_l)} | Y]} \right]. \quad (2.49)$$

In particular, this yields the GMI in (1.11) under the i.i.d. ensemble, and the LM rate in (1.9) under the cost-constrained ensemble with one optimized auxiliary cost (and also under the constant-composition ensemble). Similarly, the E_0 function in (2.36) yields the rate

$$\sup_{s \geq 0} \mathbb{E} \left[\log \frac{q(X, Y)^s e^{a_1(X)}}{\mathbb{E}[q(\bar{X}, Y)^s e^{a_1(\bar{X})} | Y]} \right]. \quad (2.50)$$

Once again, this yields the LM rate when $a_1(\cdot)$ is optimized, though if $a_1(\cdot)$ is chosen suboptimally then it may be strictly smaller than (2.49) with $L = 1$.

2.4.4 Number of Auxiliary Costs Required

We claim that

$$E_r^{\text{iid}}(Q, R) \leq E_r^{\text{cost}}(Q, R, \{a_l\}) \leq E_r^{\text{cc}}(Q, R). \quad (2.51)$$

The first inequality follows by setting $r_l = \bar{r}_l = 0$ in (2.30), and the second inequality follows by setting $r(x) = \sum_l r_l a_l(x)$ and $\bar{r}(x) = \sum_l \bar{r}_l a_l(x)$ in (2.30), and upper bounding the objective by taking the supremum over all $r(\cdot)$ and $\bar{r}(\cdot)$ to recover E_0^{cc} in the form given in (2.43).

Here we study the number of auxiliary costs required for cost-constrained random coding to achieve E_r^{cc} . Such an investigation is of interest in gaining insight into the structure of the codebook, and since the subexponential prefactor in (2.34) grows at a slower rate when L is reduced (see Proposition 2.1). Our results are summarized in the following theorem.

Theorem 2.5. *Fix the input distribution $Q \in \mathcal{P}(\mathcal{X})$ and rate R .*

1. *For any decoding metric, we have*

$$\sup_{a_1(\cdot), a_2(\cdot)} E_r^{\text{cost}}(Q, R, \{a_1, a_2\}) = E_r^{\text{cc}}(Q, R) \quad (2.52)$$

$$\max_Q \sup_{a_1(\cdot)} E_r^{\text{cost}'}(Q, R, a_1) = \max_Q E_r^{\text{cc}}(Q, R). \quad (2.53)$$

2. *If $q(x, y) = W(y|x)$ (ML decoding), then*

$$\sup_{a_1(\cdot)} E_r^{\text{cost}}(Q, R, a_1) = E_r^{\text{cc}}(Q, R) \quad (2.54)$$

$$\sup_{a_1(\cdot)} E_r^{\text{cost}'}(Q, R, a_1) = E_r^{\text{iid}}(Q, R) \quad (2.55)$$

$$\max_Q E_r^{\text{iid}}(Q, R) = \max_Q E_r^{\text{cc}}(Q, R). \quad (2.56)$$

Proof. We have from (2.51) that $E_r^{\text{cost}} \leq E_r^{\text{cc}}$. To obtain the reverse inequality corresponding to (2.52), we set $L = 2$, $\bar{r}_1 = r_2 = 1$ and $r_1 = \bar{r}_2 = 0$ in (2.30). The resulting objective coincides with (2.43) upon identifying $a_1(\cdot) = \bar{r}(\cdot)$ and $a_2(\cdot) = r(\cdot)$.

We now turn to the proof of (2.53). In Section 2.10.3, we show that for any $s > 0$ and $\rho > 0$, the logarithm in (2.40) takes the same value for all x such that $Q(x) > 0$ upon optimizing Q and $a(\cdot)$. It follows that the objective in (2.40) is unchanged when the expectation with respect to X is moved inside the logarithm, thus yielding the objective in (2.36).

We claim that, under ML decoding, we can write E_0^{cc} as

$$E_0^{\text{cc}}(Q, \rho) = \sup_{a(\cdot)} -\log \sum_y \left(\sum_x Q(x) W(y|x)^{\frac{1}{1+\rho}} e^{a(x)-\phi_a} \right)^{1+\rho}, \quad (2.57)$$

where $\phi_a \triangleq \mathbb{E}_Q[a(X)]$. To show this, we make use of the form of E_0^{cc} given in (2.43), and write the summation inside the logarithm as

$$\sum_y \left(\sum_x Q(x) W(y|x)^{1-s\rho} e^{-\rho(r(x)-\phi_r)} \right) \left(\sum_{\bar{x}} Q(\bar{x}) W(y|\bar{x})^s e^{\bar{r}(\bar{x})-\phi_{\bar{r}}} \right)^\rho. \quad (2.58)$$

We now follow [2, Ex. 5.6] and apply Hölder's inequality:

$$\left(\sum_i \alpha_i^{\frac{1}{1+\rho}} \beta_i^{\frac{1}{1+\rho}} \right)^{1+\rho} \leq \left(\sum_i \alpha_i \right) \left(\sum_i \beta_i^\rho \right)^\rho. \quad (2.59)$$

Identifying $\alpha_i = Q(x)W(y|x)^{1-s\rho} e^{-\rho(r(x)-\phi_r)}$ and $\beta_i = (Q(x)W(y|x)^s e^{\bar{r}(x)-\phi_{\bar{r}}})^\rho$, we conclude that (2.58) is lower bounded by

$$\sum_y \left(\sum_x Q(x) W(y|x)^{\frac{1}{1+\rho}} e^{\tilde{r}(x)-\phi_{\tilde{r}}} \right)^{1+\rho}, \quad (2.60)$$

where $\tilde{r}(x) \triangleq \frac{\rho}{1+\rho}(\bar{r}(x) - r(x))$, and $\phi_{\tilde{r}} \triangleq \mathbb{E}_Q[\tilde{r}(X)]$. The necessary and sufficient condition for equality is that $\alpha_i^\rho = \zeta \beta_i$ for all i and some ζ [2, Ex. 4.15]. Some simple manipulations show that this holds provided that $s = \frac{1}{1+\rho}$ and $\bar{r}(\cdot) = -\rho r(\cdot)$. Renaming $\tilde{r}(\cdot)$ as $a(\cdot)$, we obtain (2.57). We can clearly achieve E_r^{cc} using $L = 2$ with the cost functions $r(\cdot)$ and $\bar{r}(\cdot)$. However, since we have chosen one to be a scalar multiple of the other, we conclude that $L = 1$ suffices.

A similar argument using Hölder's inequality reveals that the objective in (2.36) is maximized by $s = \frac{1}{1+\rho}$ and $a_1(\cdot) = 0$, and the objective in (2.38) is maximized by $s = \frac{1}{1+\rho}$, thus yielding (2.55). Finally, combining (2.53) and (2.55), we obtain (2.56). $\square$

Equation (2.52) shows that the cost-constrained ensemble recovers E_r^{cc} using at most two auxiliary costs. In fact, this remains true even in the generalization to more general alphabets; see Section 2.9 for further discussion.

If the input distribution or decoding rule is optimized, then $L = 1$ suffices (see (2.53) and (2.54)), and if both are optimized then $L = 0$ suffices (see (2.56)). The latter result is well-known [56], and is stated for completeness. While (2.53) shows that E_r^{cost} and $E_r^{\text{cost}'}$ coincide when Q is optimized, (2.54)–(2.55) show that the former can be higher for a given Q even when $L = 1$, since E_r^{cc} can exceed E_r^{iid} even under ML decoding [56].

2.5 Second-Order Coding Rates

In this section, we present achievable second-order coding rates (cf. Section 1.2.2) for the ensembles given in Section 2.3, i.e. expansions of the form (1.28) with the equality replaced by $\geq$. To distinguish between the ensembles, we define $M^{\text{iid}}(Q, n, \epsilon)$, $M^{\text{cc}}(Q, n, \epsilon)$ and $M^{\text{cost}}(Q, n, \epsilon)$ to be the maximum number of codewords of length n such that the random-coding error probability does not exceed ϵ for the i.i.d., constant-composition and cost-constrained ensembles respectively, using the input distribution Q .

2.5.1 Cost-Constrained Ensemble

We make use of the following generalization of $i(x, y)$ in (1.29) to the mismatched setting:

$$i_{s,a}(x, y) \triangleq \log \frac{q(x, y)^s e^{a(x)}}{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{a(\bar{x})}}, \quad (2.61)$$

where $s \geq 0$ and $a(\cdot)$ are fixed parameters. We define the quantities

$$I_{s,a}(Q) \triangleq \mathbb{E}[i_{s,a}(X, Y)] \quad (2.62)$$

$$U_{s,a}(Q) \triangleq \text{Var}[i_{s,a}(X, Y)] \quad (2.63)$$

$$V_{s,a}(Q) \triangleq \mathbb{E}[\text{Var}[i_{s,a}(X, Y) | X]], \quad (2.64)$$

where $(X, Y) \sim Q \times W$. From (1.9), we see that the LM rate is equal to $I_{s,a}(Q)$ after optimizing s and $a(\cdot)$.

We can relate (2.62)–(2.64) with the E_0 functions defined in (2.36) and (2.40). Letting $E_0^{\text{cost}'}(Q, \rho, s, a)$ and $E_0^{\text{cc}}(Q, \rho, s, a)$ denote the corresponding objectives with fixed (s, a) in place of the supremum, we have

$$I_{s,a}(Q) = \frac{\partial E_0^{\text{cost}'}(Q, \rho, s, a)}{\partial \rho} \Big|_{\rho=0} = \frac{\partial E_0^{\text{cc}}(Q, \rho, s, a)}{\partial \rho} \Big|_{\rho=0} \quad (2.65)$$

$$U_{s,a}(Q) = -\frac{\partial^2 E_0^{\text{cost}'}(Q, \rho, s, a)}{\partial \rho^2} \Big|_{\rho=0} \quad (2.66)$$

$$V_{s,a}(Q) = -\frac{\partial^2 E_0^{\text{cc}}(Q, \rho, s, a)}{\partial \rho^2} \Big|_{\rho=0}. \quad (2.67)$$

The latter two identities generalize a well-known connection between the exponent and dispersion in the matched case (e.g. see [61, p. 2337]).

The main result of this section is the following theorem, which considers the cost-constrained ensemble. Our proof differs from the usual proof using threshold-based random-

coding bounds [60, 61], but the latter approach can also be used in the present setting [71]. Our analysis can be interpreted as performing a normal approximation of rcu_s in (2.8).

Theorem 2.6. *Fix the input distribution $Q \in \mathcal{P}(\mathcal{X})$, and the parameters $s \geq 0$ and $a(\cdot)$. Using the cost-constrained ensemble in (2.21) with $L = 2$ and*

$$a_1(x) = a(x) \quad (2.68)$$

$$a_2(x) = \mathbb{E}_{W(\cdot|x)}[i_{s,a}(x, Y)], \quad (2.69)$$

the following expansion holds:

$$\log M^{\text{cost}}(Q, n, \epsilon) \geq nI_{s,a}(Q) - \sqrt{nV_{s,a}(Q)} Q^{-1}(\epsilon) + O(\log n). \quad (2.70)$$

Proof. Throughout the proof, we make use of the random variables $(X, Y, \bar{X}) \sim Q(x)W(y|x)Q(\bar{x})$ and $(\mathbf{X}, \mathbf{Y}, \bar{\mathbf{X}}) \sim P_{\mathbf{X}}(\mathbf{x})W^n(\mathbf{y}|\mathbf{x})P_{\mathbf{X}}(\bar{\mathbf{x}})$, and we write $i_{s,a}^n(\mathbf{x}, \mathbf{y}) \triangleq \sum_{i=1}^n i_{s,a}(x_i, y_i)$, and similarly $Q^n(\mathbf{x}) \triangleq \prod_{i=1}^n Q(x_i)$ and $a^n(\mathbf{x}) \triangleq \sum_{i=1}^n a(x_i)$. Probabilities, expectations, etc. containing a realization $\mathbf{x}$ of $\mathbf{X}$ are implicitly conditioned on the event $\mathbf{X} = \mathbf{x}$.

Starting with Theorem 2.1, we weaken rcu_s in (2.8) as follows:

$$\text{rcu}_s(n, M) = \mathbb{E} \left[\min \left\{ 1, (M-1) \frac{\sum_{\bar{\mathbf{x}} \in \mathcal{D}_n} P_{\mathbf{X}}(\bar{\mathbf{x}}) q^n(\bar{\mathbf{x}}, \mathbf{Y})^s}{q^n(\mathbf{X}, \mathbf{Y})^s} \right\} \right] \quad (2.71)$$

$$\leq \mathbb{E} \left[\min \left\{ 1, M e^{2\delta} \frac{\sum_{\bar{\mathbf{x}} \in \mathcal{D}_n} P_{\mathbf{X}}(\bar{\mathbf{x}}) q^n(\bar{\mathbf{x}}, \mathbf{Y})^s e^{a^n(\bar{\mathbf{x}})}}{q^n(\mathbf{X}, \mathbf{Y})^s e^{a^n(\mathbf{X})}} \right\} \right] \quad (2.72)$$

$$\leq \mathbb{E} \left[\min \left\{ 1, \frac{M e^{2\delta}}{\mu_n} \frac{\sum_{\bar{\mathbf{x}}} Q^n(\bar{\mathbf{x}}) q^n(\bar{\mathbf{x}}, \mathbf{Y})^s e^{a^n(\bar{\mathbf{x}})}}{q^n(\mathbf{X}, \mathbf{Y})^s e^{a^n(\mathbf{X})}} \right\} \right] \quad (2.73)$$

$$= \mathbb{P} \left[i_{s,a}^n(\mathbf{X}, \mathbf{Y}) + \log U \leq \log \frac{M e^{2\delta}}{\mu_n} \right] \quad (2.74)$$

$$\leq \mathbb{P} \left[i_{s,a}^n(\mathbf{X}, \mathbf{Y}) + \log U \leq \log \frac{M e^{2\delta}}{\mu_n} \cap \mathbf{X} \in \mathcal{A}_n \right] + \mathbb{P}[\mathbf{X} \notin \mathcal{A}_n] \quad (2.75)$$

$$\leq \max_{\mathbf{x} \in \mathcal{A}_n} \mathbb{P} \left[i_{s,a}^n(\mathbf{x}, \mathbf{Y}) + \log U \leq \log \frac{M e^{2\delta}}{\mu_n} \right] + \mathbb{P}[\mathbf{X} \notin \mathcal{A}_n], \quad (2.76)$$

where (2.72) follows from (2.32), (2.73) follows by substituting the random-coding distribution in (2.21) and summing over all $\bar{\mathbf{x}}$ instead of $\bar{\mathbf{x}} \in \mathcal{D}_n$, and (2.74) follows from the definition of $i_{s,a}^n$ and the identity

$$\mathbb{E}[\min\{1, A\}] = \mathbb{P}[A \geq U], \quad (2.77)$$

where A is an arbitrary non-negative random variable, and U is uniform on $(0, 1)$ and independent of A . Finally, (2.75) holds for any set $\mathcal{A}_n$ by the law of total probability.

We treat the cases $V_{s,a}(Q) > 0$ and $V_{s,a}(Q) = 0$ separately. In the former case, we choose

$$\mathcal{A}_n = \left\{ \mathbf{x} \in \mathcal{D}_n : \left| \frac{1}{n} v_{s,a}^n(\mathbf{x}) - V_{s,a}(Q) \right| \leq \zeta \sqrt{\frac{\log n}{n}} \right\}, \quad (2.78)$$

where ζ is a constant, and $v_{s,a}^n(\mathbf{x}) \triangleq \sum_{i=1}^n v_{s,a}(x_i)$ with

$$v_{s,a}(x) \triangleq \text{Var}[i_{s,a}(x, Y)]. \quad (2.79)$$

Using this definition along with that of $\mathcal{D}_n$ in (2.22) and the choice of the cost function in (2.69), we have for any $\mathbf{x} \in \mathcal{A}_n$ that

$$\left| \mathbb{E}[i_{s,a}^n(\mathbf{x}, \mathbf{Y})] - nI_{s,a}(Q) \right| \leq \delta \quad (2.80)$$

$$\left| \text{Var}[i_{s,a}^n(\mathbf{x}, \mathbf{Y})] - nV_{s,a}(Q) \right| \leq \zeta \sqrt{n \log n}. \quad (2.81)$$

Since the moments of $\log U$ are finite, this implies

$$\left| \mathbb{E}[i_{s,a}^n(\mathbf{x}, \mathbf{Y}) + \log U] - nI_{s,a}(Q) \right| = O(1) \quad (2.82)$$

$$\left| \text{Var}[i_{s,a}^n(\mathbf{x}, \mathbf{Y}) + \log U] - nV_{s,a}(Q) \right| = O(\sqrt{n \log n}). \quad (2.83)$$

Using (2.21) and defining $\mathbf{X}' \sim Q^n(\mathbf{x}')$, we have

$$\mathbb{P}[\mathbf{X} \notin \mathcal{A}_n] \leq \frac{1}{\mu_n} \mathbb{P}[\mathbf{X}' \notin \mathcal{A}_n]. \quad (2.84)$$

We claim that there exists a choice of ζ such that the right-hand side of (2.84) behaves as $O(\frac{1}{\sqrt{n}})$, thus yielding

$$\mathbb{P}[\mathbf{X} \notin \mathcal{A}_n] = O\left(\frac{1}{\sqrt{n}}\right). \quad (2.85)$$

Since $\mu_n = \Omega(n^{-L/2})$ (cf. Proposition 2.1), it suffices to show that $\mathbb{P}[\mathbf{X}' \notin \mathcal{A}_n]$ can be made to behave as $O(n^{-(L+1)/2})$ by a suitable choice of ζ . This follows using $\mathbb{E}_Q[v_{s,a}(X)] = V_{s,a}(Q)$, along with the following moderate deviations result of [75, Thm. 2]: Given an i.i.d. sequence $\{Z_i\}_{i=1}^n$ with $\mathbb{E}[Z_i] = \mu$ and $\text{Var}[Z_i] = \sigma^2 > 0$, we have for any $\eta > 0$ that

$$\mathbb{P}\left[\left|\frac{1}{n} \sum_{i=1}^n Z_i - \mu\right| > \eta \sigma \sqrt{\frac{\log n}{n}}\right] \asymp \frac{2}{\eta \sqrt{2\pi \log n}} n^{-\eta^2/2} \quad (2.86)$$

provided that $\mathbb{E}[Z_i^{\eta^2+2+\delta'}] < \infty$ for some $\delta' > 0$. The latter condition is always satisfied in the

present setting, since we are considering finite alphabets.

We are now in a position to apply the Berry-Esseen theorem for independent and non-identically distributed random variables [76, Sec. XVI.5]. The relevant first and second moments are bounded in (2.82)–(2.83), and the relevant third moment is finite since we are considering finite alphabets. Choosing

$$\log M = nI_{s,a}(Q) - \log \mu_n - 2\delta - \xi_n \quad (2.87)$$

for some ξ_n , and also using (2.76) and (2.85), we obtain from the Berry-Esseen theorem that

$$\bar{p}_e \leq Q\left(\frac{\xi_n + O(1)}{\sqrt{nV_{s,a}(Q) + O(\sqrt{n \log n})}}\right) + O\left(\frac{1}{\sqrt{n}}\right). \quad (2.88)$$

By straightforward rearrangements and a first-order Taylor expansion of the square root function and the function Q^{-1} , we obtain

$$\xi_n \leq \sqrt{nV_{s,a}(Q)} Q^{-1}(\bar{p}_e) + O(\sqrt{\log n}). \quad (2.89)$$

The proof for the case $V_{s,a}(Q) > 0$ is concluded by combining (2.87) and (2.89), and noting from Proposition 2.1 that $\log \mu_n = O(\log n)$.

In the case that $V_{s,a}(Q) = 0$, we again make use of (2.82), but we handle the variance differently. From the definition in (2.64), we in fact have $\text{Var}[i_{s,a}(x, Y)] = 0$ for all x such that $Q(x) > 0$. Thus, for all $x \in \mathcal{D}_n$ we have $\text{Var}[i_{s,a}^n(x, \mathbf{Y})] = 0$ and hence $\text{Var}[i_{s,a}^n(x, \mathbf{Y}) + \log U] = O(1)$. Choosing M as in (2.87) and setting $\mathcal{A}_n = \mathcal{D}_n$, we can write (2.76) as

$$\text{rcu}_s(n, M) \leq \max_{x \in \mathcal{D}_n} \mathbb{P}[i_{s,a}^n(x, \mathbf{Y}) + \log U - nI_{s,a}(Q) \leq -\xi_n] \quad (2.90)$$

$$\leq \frac{O(1)}{(\xi_n - O(1))^2}, \quad (2.91)$$

where (2.91) holds by (2.82) and Chebyshev's inequality provided that ξ_n is sufficiently large so that the $\xi_n - O(1)$ term is positive. Rearranging, we see that we can achieve any target value ϵ of $\bar{p}_e$ with $\xi_n = O(1)$. The proof is concluded using (2.87). $\square$

Observe that the auxiliary cost in (2.68) represents the parameter $a(\cdot)$ appearing in the LM rate in (1.9). The second cost in (2.69) plays a more subtle role, ensuring that the empirical mean of the generalized information density is close to the true mean for all $x \in \mathcal{D}_n$, with a sufficiently small remainder so that the second-order term in (2.70) is not affected.

In Section 2.9, we discuss the extension of the above analysis to input-constrained channels

with continuous alphabets. The proof is nearly unchanged if all of the moments of $i_{s,a}(X, Y)$ are finite, but we also present an alternative approach (with $L > 2$) that only requires finite moments up to the sixth order.

2.5.2 i.i.d. and Constant-Composition Ensembles

The properties of the cost-constrained ensemble used in the proof of Theorem 2.6 are also satisfied by the constant-composition ensemble, so we conclude that (2.70) remains true when M^{cost} is replaced by M^{cc} .

The analogous result for i.i.d. coding does not follow directly from Theorem 2.6, since the proof requires the cost function in (2.69) to be present. However, using similar arguments and the identities $\mathbb{E}[i_s^n(\mathbf{X}, \mathbf{Y})] = nI_s(Q)$ and $\text{Var}[i_s^n(\mathbf{X}, \mathbf{Y})] = nU_s(Q)$ (where $\mathbf{X} \sim Q^n$), we obtain

$$\log M^{\text{iid}}(Q, n, \epsilon) \geq nI_s(Q) - \sqrt{nU_s(Q)} Q^{-1}(\epsilon) + O(1) \quad (2.92)$$

for $s \geq 0$, where $I_s(Q)$ and $U_s(Q)$ are defined as in (2.62)–(2.63) with $a(\cdot) = 0$.

The constant-composition and cost-constrained ensembles provide two advantages over the i.i.d. ensemble: the freedom in choosing the parameter $a(\cdot)$, and the conditional variance in the second-order term instead of the unconditional variance. The latter is advantageous since $V_{s,a}(Q) \leq U_{s,a}(Q)$ (with strict inequality possible), meaning that $V_{s,a}$ corresponds to a smaller backoff from the first-order term in (2.70) when $\epsilon < \frac{1}{2}$.

2.5.3 Number of Auxiliary Costs Required

In accordance with the dispersion analysis for matched decoding [60, 61], the expansion of (2.70) is primarily of interest when the free parameters maximize the first-order term. Here we show that the required number of auxiliary costs can be reduced below the sufficient choice of $L = 2$ in such cases.

We first show that $L = 1$ suffices in the case of ML decoding (i.e. $q(x, y) = W(y|x)$) when the pair (s, a) maximizes $I_{s,a}(Q)$. At a high level, this follows by setting $s = 1$ and $a(\cdot) = 0$ to obtain $I_{s,a}(Q) = I(X; Y)$, and noting that $a_1(\cdot)$ in (2.68) is not needed, since it equals zero. For completeness, we must also show that no other choices yield an improved second-order

term in (2.70). Defining $\tilde{P}(x|y) \triangleq \frac{Q(x)W(y|x)}{\sum_{\bar{x}} Q(\bar{x})W(y|\bar{x})}$ and $\tilde{P}_{s,a}(x|y) \triangleq \frac{Q(x)W(y|x)^s e^{a(x)}}{\sum_{\bar{x}} Q(\bar{x})W(y|\bar{x})^s e^{a(\bar{x})}}$, we have

$$I_{s,a}(Q) = \sum_{x,y} Q(x)W(y|x) \log \frac{\tilde{P}(x|y)}{Q(x)} \frac{\tilde{P}_{s,a}(x|y)}{\tilde{P}(x|y)} \quad (2.93)$$

$$= I(X; Y) - \sum_{x,y} Q(x)W(y|x) \log \frac{\tilde{P}(x|y)}{\tilde{P}_{s,a}(x|y)}, \quad (2.94)$$

where (2.93) follows from simple algebraic manipulations. The second term in (2.94) is a divergence, and is non-negative and equal to zero if and only if $\tilde{P}(x|y) = \tilde{P}_{s,a}(x|y)$ wherever $Q(x)W(y|x) > 0$. Writing the information density in (1.29) as $i(x, y) = \log \frac{\tilde{P}(x|y)}{Q(x)}$, and similarly writing $i_{s,a}(x, y) = \log \frac{\tilde{P}_{s,a}(x|y)}{Q(x)}$, we obtain the equivalent condition that $i(x, y) = i_{s,a}(x, y)$ wherever $Q(x)W(y|x) > 0$. Thus, any choice of (s, a) maximizing $I_{s,a}(Q)$ yields $V_{s,a}(Q) = \mathbb{E}[\text{Var}[i(X, Y) | X]]$, and hence all such choices yield the same second-order term in (2.70).

Next, we claim that the expansion in (2.70) holds for an arbitrary decoding metric even in the absence of the auxiliary cost in (2.69), provided that the triplet (Q, s, a) maximizes $I_{s,a}(Q)$. This follows from the observation that $\mathbb{E}_{W(\cdot|x)}[i_{s,a}(x, Y)]$ takes the same value for all x such that $Q(x) > 0$ under any such parameters; see Section 2.10.3 for the proof. Stated differently, we have $V_{s,a}(Q) = U_{s,a}(Q)$ for any triplet (Q, s, a) maximizing $I_{s,a}(Q)$, thus generalizing the analogous result for ML decoding [61, Lemma 62].

We observe that the number of auxiliary costs required to obtain (2.70) coincides with that of the random-coding exponent (see Theorem 2.5): $L = 2$ suffices in general, $L = 1$ suffices if the metric or input distribution is optimized, and $L = 0$ (i.i.d. coding) suffices when both are optimized.

2.6 Moderate Deviations

In this section, we consider the moderate deviations regime, which was introduced in Section 1.2.3. We again focus on the cost-constrained ensemble. We make use of the definitions of $i_{s,a}(x, y)$, $I_{s,a}(Q)$ and $V_{s,a}(Q)$ in (2.61), (2.62) and (2.64), and we again let

$$E_0^{\text{cc}}(Q, \rho, s, a) \triangleq - \sum_x Q(x) \log \sum_y W(y|x) e^{-\rho i_{s,a}(x,y)} \quad (2.95)$$

denote the objective in (2.40) with fixed parameters in place of the supremum. In contrast to the preceding sections, we consider auxiliary costs that vary with n in (2.21)–(2.22).

Theorem 2.7. Fix the input distribution $Q \in \mathcal{P}(\mathcal{X})$, the parameters $s \geq 0$ and $a(\cdot)$, and a sequence ψ_n such that $\psi_n = o(1)$ and $\psi_n = \omega(\frac{1}{\sqrt{n}})$. Suppose that $V_{s,a}(Q) > 0$, and define $\rho_n \triangleq \frac{\psi_n}{V_{s,a}(Q)}$. Using the cost-constrained ensemble in (2.21) with $L = 2$ and

$$a_1(x) = a(x) \quad (2.96)$$

$$a_2(x) = a_{2,n}(x) = \log \sum_y W(y|x) e^{-\rho_n i_{s,a}(x,y)}, \quad (2.97)$$

we have

$$\liminf_{n \rightarrow \infty} -\frac{1}{n\psi_n^2} \log \bar{p}_e(n, \lfloor e^{n(I_{s,a}(Q) - \psi_n)} \rfloor) \geq \frac{1}{2V_{s,a}(Q)}. \quad (2.98)$$

Proof. Let n be sufficiently large so that $\rho_n \in [0, 1]$. Applying the inequality $\min\{1, \zeta\} \leq \zeta^{\rho_n}$ to (2.73) and using the definition of $i_{s,a}(x, y)$ in (2.61), we obtain

$$\bar{p}_e(n, M_n) \leq \tilde{M}_n^{\rho_n} \sum_{\mathbf{x}, \mathbf{y}} P_{\mathbf{X}}(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) e^{-\rho_n i_{s,a}^n(\mathbf{x}, \mathbf{y})} \quad (2.99)$$

$$= \tilde{M}_n^{\rho_n} \sum_{\mathbf{x}} P_{\mathbf{X}}(\mathbf{x}) \exp \left(\log \sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) e^{-\rho_n i_{s,a}^n(\mathbf{x}, \mathbf{y})} \right), \quad (2.100)$$

where $\tilde{M}_n \triangleq \frac{M_n e^{2\delta}}{\mu_n}$. Combining the choice of $a_2(\cdot)$ in (2.97) (depending on n) with (2.95), (2.100), and the definition of $\mathcal{D}_n$ in (2.22), we obtain

$$\bar{p}_e(n, M_n) \leq \tilde{M}_n^{\rho_n} e^{-nE_0^{\text{cc}}(Q, \rho_n, s, a) + \delta} \quad (2.101)$$

$$= e^{-n(E_0^{\text{cc}}(Q, \rho_n, s, a) - \rho_n \tilde{R}_n) + \delta}, \quad (2.102)$$

where $\tilde{R}_n \triangleq \frac{1}{n} \log \tilde{M}_n$.

We now Taylor expand the function E_0^{cc} similarly to [63]. Using the derivatives given in (2.65) and (2.67), we have the following:

$$E_0^{\text{cc}}(Q, \rho_n, s, a) = \rho_n I_{s,a}(Q) - \frac{\rho_n^2}{2} V_{s,a}(Q) + O(\rho_n^3). \quad (2.103)$$

Substituting (2.103) into (2.102) and recalling that $\rho_n = \frac{\psi_n}{V_{s,a}(Q)}$, we obtain

$$\bar{p}_e(n, M_n) \leq e^{-n \left(\frac{\psi_n}{V_{s,a}(Q)} (I_{s,a}(Q) - \tilde{R}_n) - \frac{\psi_n^2}{2V_{s,a}(Q)} + O(\psi_n^3) \right) + \delta}. \quad (2.104)$$

In accordance with the theorem statement, we set $M_n = e^{n(I_{s,a}(Q) - \psi_n)}$; we assume this to be an integer for notational convenience, and the general case follows similarly. Using the

definitions of $\tilde{M}_n$ and $\tilde{R}_n$ following (2.100) and (2.102) respectively, we obtain

$$\frac{\psi_n}{V_{s,a}(Q)}(I_{s,a}(Q) - \tilde{R}_n) = \frac{\psi_n}{V_{s,a}(Q)}\left(\psi_n - \frac{2\delta}{n} + \frac{1}{n}\log \mu_n\right), \quad (2.105)$$

and hence (2.104) becomes

$$\bar{p}_e(n, M_n) \leq e^{-n\left(\frac{\psi_n^2}{2V_{s,a}(Q)} + \frac{\psi_n}{nV_{s,a}(Q)}(\log \mu_n - 2\delta) + O(\psi_n^3)\right) + \delta}. \quad (2.106)$$

It remains to show that the exponent is dominated by the term containing $\frac{\psi_n^2}{2V_{s,a}(Q)}$. It is easily seen that the terms $O(\psi_n^3)$ and δ are insignificant due to the assumptions $\psi_n = o(1)$ and $\psi_n = \omega\left(\frac{1}{\sqrt{n}}\right)$ respectively, so it remains to handle the term $\frac{\psi_n}{nV_{s,a}(Q)}(\log \mu_n - 2\delta)$.

Proposition 2.1 is stated for fixed auxiliary costs, whereas in the present case, $a_2(\cdot)$ varies with n . However, we can simply make use of the fact that $\mu_n \geq (n+1)^{-(|\mathcal{X}|-1)}$, which follows using types as stated before Proposition 2.1 (A more general approach not relying on types will be given in Section 2.9). We therefore have $\log \mu_n = O(\log n)$, and hence

$$\frac{\psi_n}{nV_{s,a}(Q)}(\log \mu_n - 2\delta) = O\left(\frac{\psi_n \log n}{n}\right) \quad (2.107)$$

$$= O\left(\psi_n^2 \frac{\log n}{\sqrt{n}}\right), \quad (2.108)$$

where (2.108) follows since $\psi_n = \omega\left(\frac{1}{\sqrt{n}}\right)$. Since $\frac{\log n}{\sqrt{n}} \rightarrow 0$, the proof is complete. $\square$

The key step in the preceding proof is choosing the second auxiliary cost in order to obtain the exponential bound in (2.102) without the polynomial prefactor given in (2.34). This is in contrast with [63], where Gallager's bound for i.i.d. coding is used, and the prefactor arises as a constant using standard techniques [2, Ch. 5].

Analogously to the previous two sections, the preceding analysis remains valid for the constant-composition ensemble, and a similar analysis can be performed for the i.i.d. ensemble, with $I_s(Q)$ and $U_s(Q)$ replacing $I_{s,a}(Q)$ and $V_{s,a}(Q)$ respectively. We again proceed by presenting conditions under which the number of auxiliary costs can be reduced.

We claim that, under ML decoding with a fixed input distribution, we can obtain (2.98) with $I_{s,a}(Q) = I(X; Y)$ and $V_{s,a}(Q) = V(Q) \triangleq \mathbb{E}[\text{Var}[i(X, Y) | X]]$ (where $i(\cdot, \cdot)$ is defined in (1.29)) even when $L = 1$, provided that $V(Q) > 0$. To see this, we use (2.54) in Theorem 2.5 to conclude that the following holds when the auxiliary cost $a_{1,n}(\cdot)$ is optimized and allowed to vary with n :

$$\bar{p}_e(n, M_n) \leq e^{-n\left(\sup_{s_n, a_n(\cdot)} E_0^{\text{sc}}(Q, \rho_n, s_n, a_n) - \rho_n \tilde{R}_n\right) + \delta}, \quad (2.109)$$

where we now have $\rho_n = \frac{\psi_n}{V(Q)}$. Note that $\tilde{R}_n$ is bounded independently of (s, a) in the proof of Theorem 2.7, and thus there are no uniformity issues in (2.109). We obtain the desired result by lower bounding the supremum by the specific choices $s_n = 1$ and $a_n(\cdot) = 0$, and repeating the analysis following (2.102).

Consider now the case of a general decoding metric, and suppose that we use cost-constrained random coding ($L = 1$) with an input distribution Q_n and auxiliary cost $a_{1,n}(\cdot)$ that are optimized and allowed to vary with n . We claim that (2.98) holds for any (Q, s, a) maximizing $I_{s,a}(Q)$, provided that $V_{s,a}(Q) > 0$. To see this, we note that similarly to the previous paragraph, (2.53) in Theorem 2.5 implies the following:

$$\bar{p}_e(n, M_n) \leq e^{-n(\sup_{Q_n, s_n, a_n(\cdot)} E_0^{\text{cc}}(Q_n, \rho_n, s_n, a_n) - \rho_n \tilde{R}_n) + \delta}, \quad (2.110)$$

where $\rho_n = \frac{\psi_n}{V_{s,a}(Q)}$. Once again, we obtain the desired result by lower bounding the supremum by the specific choice (Q, s, a) , and repeating the analysis following (2.102).

Finally, we note that $L = 0$ (i.i.d. coding) suffices to obtain (2.98) with $I_{s,a}(Q)$ and $V_{s,a}(Q)$ equal to the channel capacity and dispersion respectively, provided that the latter is positive [63, 64].

2.7 Expurgated Error Exponents

As discussed in Section 1.2.1, expurgation techniques are useful for obtaining improved random-coding bounds and exponents at low rates. In this section, we derive expurgated error exponents using the non-asymptotic bounds given in Theorem 2.2. We follow the general approach of Section 2.4, focusing on the cost-constrained ensemble, and stating the results for the i.i.d. and constant-composition ensembles as special cases.¹

2.7.1 Cost-Constrained Ensemble

Reusing the notation of Section 2.4, we define the sets

$$\mathcal{S}(\{a_l\}) \triangleq \{P_{X\bar{X}} \in \mathcal{P}(\mathcal{X} \times \mathcal{X}) : \mathbb{E}_P[a_l(X)] = \mathbb{E}_P[a_l(\bar{X})] = \phi_l, l = 1, \dots, L\} \quad (2.111)$$

$$\mathcal{T}(\{a_l\}) \triangleq \left\{ P_{X\bar{X}Y} \in \mathcal{P}(\mathcal{X} \times \mathcal{X} \times \mathcal{Y}) : P_{X\bar{X}} \in \mathcal{S}(\{a_l\}), \right. \\ \left. \mathbb{E}_P[\log q(\bar{X}, Y)] \geq \mathbb{E}_P[\log q(X, Y)] \right\}, \quad (2.112)$$

¹Parts of this subsection were done in collaboration with Li Peng, particularly those regarding (2.118).

which implicitly depend on Q through $\{\phi_l\}$. Using the bounds rcux_ρ and $\text{rcux}_{\rho,s}$ given in Theorem 2.2, we have the following.

Theorem 2.8. *For any input distribution $Q \in \mathcal{P}(\mathcal{X})$, auxiliary costs $\{a_l\}_{l=1}^L$, and rate $R > 0$, the following holds under the codeword distribution in (2.21):*

$$\sup_{\rho \geq 1} \lim_{n \rightarrow \infty} -\frac{1}{n} \log \text{rcux}_\rho(n, \lfloor e^{nR} \rfloor) = E_{\text{ex}}^{\text{cost}}(Q, R, \{a_l\}), \quad (2.113)$$

where

$$E_{\text{ex}}^{\text{cost}}(Q, R) \triangleq \min_{\substack{P_{X\bar{X}Y} \in \mathcal{T}(\{a_l\}) \\ D(P_{X\bar{X}} \| Q \times Q) \leq R}} D(P_{X\bar{X}Y} \| Q \times Q \times W) - R, \quad (2.114)$$

and where $Q \times Q \times W$ denotes the joint distribution $Q(x)Q(\bar{x})W(y|x)$.

Proof. See Section 2.10.4. □

For a fixed input distribution and set of cost functions, the optimization problem in (2.114) is convex, thus permitting the following alternative forms based on Lagrange duality.

Theorem 2.9. *For any input distribution $Q \in \mathcal{P}(\mathcal{X})$, auxiliary costs $\{a_l\}$, and rate R , we have*

$$E_{\text{ex}}^{\text{cost}}(Q, R, \{a_l\}) = \sup_{s \geq 0} \min_{\substack{P_{X\bar{X}} \in \mathcal{S}(\{a_l\}) \\ D(P_{X\bar{X}} \| Q \times Q) \leq R}} \mathbb{E}_P[d_s(X, \bar{X})] + D(P_{X\bar{X}} \| Q \times Q) - R \quad (2.115)$$

$$= \sup_{\rho \geq 1} E_{\text{x}}^{\text{cost}}(Q, \rho, \{a_l\}) - \rho R, \quad (2.116)$$

where

$$d_s(x, \bar{x}) \triangleq -\log \mathbb{E} \left[\left(\frac{q(\bar{x}, Y)}{q(x, Y)} \right)^s \middle| X = x \right] \quad (2.117)$$

$$E_{\text{x}}^{\text{cost}}(Q, R, \{a_l\}) \triangleq \sup_{s \geq 0, \{r_l\}, \{\bar{r}_l\}} -\rho \log \mathbb{E} \left[\left(\mathbb{E} \left[\left(\frac{q(\bar{X}, Y)}{q(X, Y)} \right)^s \frac{e^{\sum_{l=1}^L \bar{r}_l(a_l(\bar{X}) - \phi_l)}}{e^{\sum_{l=1}^L r_l(a_l(X) - \phi_l)}} \middle| X, \bar{X} \right] \right)^{1/\rho} \right] \quad (2.118)$$

with $(X, Y, \bar{X}) \sim Q(x)W(y|x)Q(\bar{x})$

Proof. See Section 2.10.5. □

Observe that there are strong similarities between the equivalences in Theorem 2.9 and those in (1.24)–(1.27). However, the former have the notable advantage of holding for any input distribution and decoding metric, whereas (1.24)–(1.27) correspond to ML decoding with an optimized input distribution.

Similarly to Section 2.4, it is of interest to provide a direct derivation of $E_{\text{ex}}^{\text{cc}}$ in the form given in (2.116) without relying on types, thus permitting generalizations beyond the finite alphabet case. We start with (2.14), and write

$$\text{rcux}_{\rho,s}(n, M)^{1/\rho} = 4(M-1) \sum_{\mathbf{x}, \bar{\mathbf{x}}} P_{\mathbf{X}}(\mathbf{x}) P_{\mathbf{X}}(\bar{\mathbf{x}}) \left(\sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \left(\frac{q^n(\bar{\mathbf{x}}, \mathbf{y})}{q^n(\mathbf{x}, \mathbf{y})} \right)^s \right)^{1/\rho} \quad (2.119)$$

$$\leq M \sum_{\mathbf{x}, \bar{\mathbf{x}}} P_{\mathbf{X}}(\mathbf{x}) P_{\mathbf{X}}(\bar{\mathbf{x}}) \left(\sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \left(\frac{q^n(\bar{\mathbf{x}}, \mathbf{y})}{q^n(\mathbf{x}, \mathbf{y})} \right)^s \frac{e^{\sum_{l=1}^L \bar{r}_l(a_l^n(\bar{\mathbf{x}}) - n\phi_l)}}{e^{\sum_{l=1}^L r_l(a_l^n(\mathbf{x}) - n\phi_l)}} \right)^{1/\rho} \quad (2.120)$$

$$\leq M \sum_{\mathbf{x}, \bar{\mathbf{x}}} Q^n(\mathbf{x}) Q^n(\bar{\mathbf{x}}) \left(\sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \left(\frac{q^n(\bar{\mathbf{x}}, \mathbf{y})}{q^n(\mathbf{x}, \mathbf{y})} \right)^s \frac{e^{\sum_{l=1}^L \bar{r}_l(a_l^n(\bar{\mathbf{x}}) - n\phi_l)}}{e^{\sum_{l=1}^L r_l(a_l^n(\mathbf{x}) - n\phi_l)}} \right)^{1/\rho}, \quad (2.121)$$

where (2.120) holds for any $\{r_l\}$ and $\{\bar{r}_l\}$ from (2.32), and (2.121) follows from (2.21) and Proposition 2.1. We obtain (2.118) by setting $M = \lfloor e^{nR} \rfloor$, expanding each term in (2.121) as a product from 1 to n and taking the supremum over $\rho, s, \{r_l\}$ and $\{\bar{r}_l\}$.

2.7.2 i.i.d. and Constant-Composition Ensembles

Setting $L = 0$ in (2.118), we obtain the exponent

$$E_{\text{ex}}^{\text{iid}}(Q, R) \triangleq \sup_{\rho \geq 1} E_{\text{x}}^{\text{iid}}(Q, \rho) - \rho R, \quad (2.122)$$

where

$$E_{\text{x}}^{\text{iid}}(Q, \rho) \triangleq \sup_{s \geq 0} -\rho \log \mathbb{E} \left[\left(\mathbb{E} \left[\left(\frac{q(\bar{X}, Y)}{q(X, Y)} \right)^s \middle| X, \bar{X} \right] \right)^{1/\rho} \right]. \quad (2.123)$$

The objective in (2.123) is concave in s , and under ML decoding (i.e. $q(x, y) = W(y|x)$), it is also unchanged when s is replaced by $1 - s$. From these properties, it follows that $s = \frac{1}{2}$ is optimal for ML decoding, and thus the exponent is the same as that of Gallager (cf. (1.24)).

For the expurgated constant-composition ensemble, the constraints on the auxiliary cost functions in (2.111) reduce to $P_{\mathbf{X}} = P_{\bar{\mathbf{X}}} = Q$, which in turn implies $D(P_{\mathbf{X}\bar{\mathbf{X}}} \| Q \times Q) = I_P(X; \bar{X})$. Thus, (2.114) is precisely the expurgated exponent of Csiszár and Körner [6] (see also (1.27)), which we henceforth denote by $E_{\text{ex}}^{\text{cc}}(Q, R)$.

In the case of ML decoding, the alternative expression for $E_{\text{ex}}^{\text{cc}}$ resulting from (2.115) coincides with the exponent of Csiszár-Körner-Martón [55] (see (1.26)); this can be seen by using

the same reasoning as that following (2.123) to conclude that $s = \frac{1}{2}$ is optimal. We claim that

$$E_{\text{ex}}^{\text{cc}}(Q, R) = \sup_{\rho \geq 1} E_{\text{x}}^{\text{cc}}(Q, \rho) - \rho R, \quad (2.124)$$

where

$$E_{\text{x}}^{\text{cc}}(Q, \rho) = \sup_{s \geq 0, a(\cdot)} \mathbb{E} \left[-\rho \log \mathbb{E} \left[\left(\mathbb{E} \left[\left(\frac{q(\bar{X}, Y)}{q(X, Y)} \right)^s \frac{e^{a(\bar{X})}}{e^{a(X)}} \mid \bar{X} \right] \right)^{1/\rho} \mid X \right] \right]. \quad (2.125)$$

To see this, we follow the steps of (2.43)–(2.46), reusing the notation therein. We obtain the following E_{x} function from (2.118):

$$E_{\text{x}}^{\text{cc}}(Q, \rho) \triangleq \sup_{s \geq 0, r(\cdot), \bar{r}(\cdot)} -\rho \log \sum_{x, \bar{x}} Q(x) Q(\bar{x}) \left(\sum_y W(y|x) \left(\frac{q(\bar{x}, y)}{q(x, y)} \right)^s \frac{e^{\bar{r}(\bar{x}) - \phi_{\bar{r}}}}{e^{r(x) - \phi_r}} \right)^{1/\rho} \quad (2.126)$$

$$\leq \sup_{s \geq 0, r(\cdot), \bar{r}(\cdot)} -\rho \sum_x Q(x) \log \sum_{\bar{x}} Q(\bar{x}) \left(\sum_y W(y|x) \left(\frac{q(\bar{x}, y)}{q(x, y)} \right)^s \frac{e^{\bar{r}(\bar{x}) - \phi_{\bar{r}}}}{e^{r(x) - \phi_r}} \right)^{1/\rho}, \quad (2.127)$$

where (2.127) follows from Jensen's inequality. For any $s \geq 0$, $\rho \geq 1$ and $\bar{r}(\cdot)$, the choice

$$r(x) = \rho \log \sum_{\bar{x}} Q(\bar{x}) \left(\sum_y W(y|x) \left(\frac{q(\bar{x}, y)}{q(x, y)} \right)^s e^{\bar{r}(\bar{x}) - \phi_{\bar{r}}} \right)^{1/\rho} \quad (2.128)$$

yields equality in (2.127). Hence, and by writing

$$-\sum_x Q(x) \log \left(\frac{e^{-\phi_{\bar{r}}}}{e^{r(x) - \phi_r}} \right)^{1/\rho} = -\sum_x Q(x) \log \left(\frac{1}{e^{\bar{r}(x)}} \right)^{1/\rho} = \frac{\phi_{\bar{r}}}{\rho}, \quad (2.129)$$

we see that the supremum over $r(\cdot)$ in (2.126) achieved by (2.128). Renaming $\bar{r}(\cdot)$ as $a(\cdot)$ and using the first equality in (2.129), we obtain (2.125).

2.7.3 Number of Auxiliary Costs Required

Similarly to (2.51), we have for any $\{a_l\}$ that

$$E_{\text{ex}}^{\text{iid}}(Q, R) \leq E_{\text{ex}}^{\text{cost}}(Q, R, \{a_l\}) \leq E_{\text{ex}}^{\text{cc}}(Q, R). \quad (2.130)$$

By setting $r_l = \bar{r}_l = 1$ in (2.118) and comparing the resulting expression with (2.126), we see that the second inequality holds with equality when $L = 2$ and the costs are optimized. We proceed by showing that, analogously to Section 2.4.4, this reduces to $L = 1$ in the case that the input distribution or decoding metric is optimized.

To show that $L = 1$ suffices under ML decoding, we extend the argument following (2.123). Upon setting $q(x, y) = W(y|x)$ and replacing $r(\cdot)$ by $-r(\cdot)$ in (2.126) (which does not affect the supremum), we obtain

$$E_x^{\text{cc}}(Q, \rho) = \sup_{s \geq 0, r(\cdot), \bar{r}(\cdot)} -\rho \log \sum_{x, \bar{x}} Q(x)Q(\bar{x}) \left(\sum_y W(y|x)^{1-s} W(y|\bar{x})^s e^{r(x)-\phi_r} e^{\bar{r}(\bar{x})-\phi_{\bar{r}}} \right)^{1/\rho} \quad (2.131)$$

The objective in (2.131) is jointly concave in s , $r(\cdot)$ and $\bar{r}(\cdot)$, and it is also unchanged when these are replaced by $1-s$, $\bar{r}(\cdot)$ and $r(\cdot)$ respectively. Combining these properties, it follows that the objective cannot decrease upon replacing s by $\frac{1}{2}$ and both $r(\cdot)$ and $\bar{r}(\cdot)$ by $\frac{1}{2}(r(\cdot) + \bar{r}(\cdot))$. Thus, the two auxiliary costs have been replaced by one without decreasing the objective, and we conclude that $L = 1$ suffices.

Consider now the case of a general decoding metric with an optimized input distribution. In Section 2.10.3, we show that for any $s > 0$ and $\rho \geq 1$, $r(x|0)$ in (2.128) takes the same value for all x with $Q(x) > 0$ upon optimizing Q and $a(\cdot)$. Since we identified $a_2(\cdot)$ with $r(\cdot)$ following (2.130), we conclude that $L = 1$ suffices.

Finally, as is seen in (1.24)–(1.27), $L = 0$ (namely, i.i.d. coding) suffices to achieve $E_{\text{ex}}^{\text{cc}}$ under ML decoding with an optimized input distribution Q .

2.8 Numerical Example

Here we provide a numerical example comparing the random-coding and expurgated exponents of Sections 2.4 and 2.7. We consider the channel described by the entries of the $|\mathcal{X}| \times |\mathcal{Y}|$ matrix

$$\begin{bmatrix} 1 - 2\delta_0 & \delta_0 & \delta_0 \\ \delta_1 & 1 - 2\delta_1 & \delta_1 \\ \delta_2 & \delta_2 & 1 - 2\delta_2 \end{bmatrix} \quad (2.132)$$

with $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$. The mismatched decoder chooses the codeword that is closest to $\mathbf{y}$ in terms of Hamming distance. For example, the decoding metric can be taken to be the entries of the matrix in (2.132) with δ_i replaced by $\delta \in (0, \frac{1}{3})$ for $i = 1, 2, 3$. We let $\delta_0 = 0.01$, $\delta_1 = 0.05$, $\delta_2 = 0.25$ and $Q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Under these parameters, we have $I_{\text{GMI}}(Q) = 0.643$ and $I_{\text{LM}}(Q) = 0.728$ bits/use.

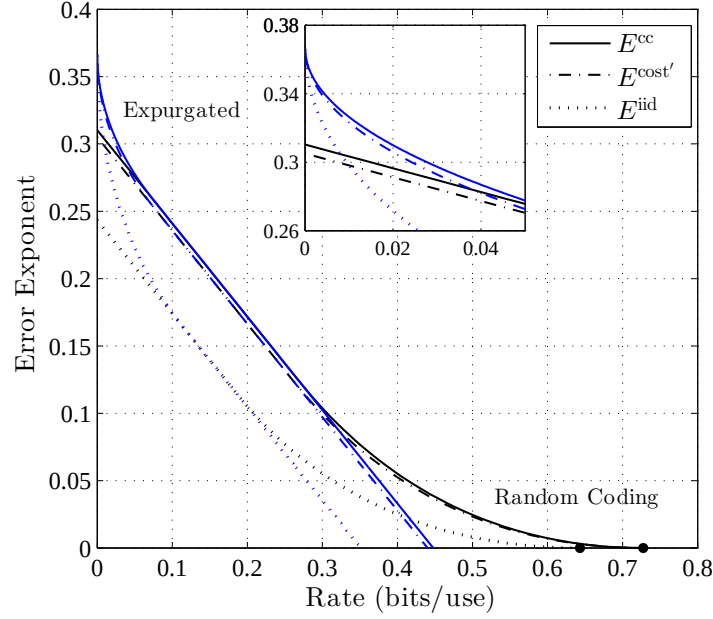


Figure 2.1 Error exponents for the channel defined in (2.132) with $\delta_0 = 0.01$, $\delta_1 = 0.05$, $\delta_2 = 0.25$ and $Q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. The mismatched decoder uses the minimum Hamming distance metric. The corresponding achievable rates $I_{\text{GMI}}(Q)$ and $I_{\text{LM}}(Q)$ are respectively marked on the horizontal axis.

We evaluate the exponents using the software YALMIP [77]. The exponents under consideration are defined in Sections 2.4 and 2.7 with the exception of $E_{\text{ex}}^{\text{cost}'}$, which is obtained from (2.118) by setting $L = 1$ and $r_1 = \bar{r}_1 = 1$. Thus, it is related to $E_{\text{r}}^{\text{cost}'}$ (see (2.35)) in the same way that $E_{\text{ex}}^{\text{cc}}$ is related to E_{r}^{cc} (see (2.39) and (2.124)). For both $E_{\text{r}}^{\text{cost}'}$ and $E_{\text{ex}}^{\text{cost}'}$, we optimize the auxiliary cost $a_1(\cdot)$.

As expected, Figure 2.1 shows that E_{r}^{cc} and $E_{\text{ex}}^{\text{cc}}$ are the highest exponents, with the latter dominating at low rates. The exponents $E_{\text{r}}^{\text{cost}'}$ and $E_{\text{ex}}^{\text{cost}'}$ are only marginally lower, whereas $E_{\text{r}}^{\text{iid}}$ and $E_{\text{ex}}^{\text{iid}}$ are significantly lower. Moreover, $E_{\text{r}}^{\text{iid}}$ achieves the GMI, whereas E_{r}^{cc} and $E_{\text{r}}^{\text{cost}'}$ achieve the LM rate. Interestingly, all of the expurgated exponents approach the same value as $R \rightarrow 0$. This behavior is in fact always observed, as was proved in [69, Thm. 4].

We did not plot $E_{\text{r}}^{\text{cost}}$ and $E_{\text{ex}}^{\text{cost}}$ with $L = 1$ and an optimized auxiliary cost, since in this example the resulting curves were indistinguishable from E_{r}^{cc} and $E_{\text{ex}}^{\text{cc}}$. However, in [68], we show that under other input distributions there may be a visible gap, as well as a larger gap between E_{r}^{cc} and $E_{\text{r}}^{\text{cost}'}$.

2.9 Extensions to Channels with Input Constraints and More General Alphabets

Suppose that each codeword x is constrained to satisfy $\frac{1}{n} \sum_{i=1}^n c(x_i) \leq \Gamma$ for some (system) cost function $c(\cdot)$. The finite-length bounds in Section 2.2 remain valid under any codeword distribution P_X that satisfies the constraint with probability one.

The i.i.d. ensemble is no longer suitable, since in all non-trivial cases it has a positive probability of producing codewords that violate the constraint. On the other hand, the results for the constant-composition ensemble remain unchanged provided that $\sum_x Q(x)c(x) \leq \Gamma$.

For the cost-constrained ensemble, the extensions are less trivial but still straightforward. The main change required is a modification of the definition of $\mathcal{D}_n$ in (2.22) to include a constraint on $\frac{1}{n} \sum_{i=1}^n c(x_i)$. Unlike the auxiliary costs in (2.22), where the sample mean can be above or below the true mean, the system cost of each codeword should be constrained to be less than or equal to its mean. That is, the additional constraint is given by

$$\frac{1}{n} \sum_{i=1}^n c(x_i) \leq \phi_c \triangleq \sum_x Q(x)c(x). \quad (2.133)$$

Using this modified definition of $\mathcal{D}_n$, one can prove the subexponential behavior of μ_n in Proposition 2.1 provided that Q is such that $\phi_c \leq \Gamma$. We proceed by outlining how the analysis and results of Sections 2.4–2.7 are affected when the alphabets need not be finite and the above modification is made. The input distribution Q and channel W may represent discrete transition laws, density functions, or a mixture of the two; where necessary, the summations should be replaced by integrals.

We first observe that the dual-domain error exponent analysis in Sections 2.4 and 2.7 is unchanged provided that $\mathbb{E}[c(X)^2] < \infty$ and $\mathbb{E}[a_l(X)^2] < \infty$ for each $l = 1, \dots, L$, in accordance with Proposition 2.1.

The analysis of the second-order coding rate (cf. Section 2.5) requires more care. In particular, the moderate deviations result [75, Thm. 2], which was used in the proof of Theorem 2.6, requires finite moments up to a certain order depending on ζ in (2.78). In the case that *all* moments of $i_{s,a}(X, Y)$ are finite, the analysis of Section 2.5.1 is nearly unchanged, except that the third absolute moment should be bounded in the set $\mathcal{A}_n$ in (2.78) in the same way as the second moment. An alternative approach is to introduce two further auxiliary costs into the

cost-constrained ensemble, namely

$$a_3(x) = v_{s,a}(x) \quad (2.134)$$

$$a_4(x) = \mathbb{E}[|i_{s,a}(x, Y) - I_x|^3], \quad (2.135)$$

where $v_{s,a}$ is defined in (2.79), $I_x \triangleq \mathbb{E}[i_{s,a}(x, Y)]$, and each expectation is taken with respect to $W(\cdot|x)$. Under these choices, the relevant second and third moments for the Berry-Esseen theorem are bounded within $\mathcal{D}_n$ similarly to (2.82). The only further requirement (along with $\mathbb{E}_Q[c(X)^2] < \infty$ and $\mathbb{E}_Q[a(X)^2] < \infty$) is that $\mathbb{E}_{Q \times W}[i_{s,a}(X, Y)^6] < \infty$, in accordance with Proposition 2.1.

With the preceding modifications, the techniques in Section 2.5 provide, to our knowledge, the most general known second-order achievability proof for memoryless input-constrained channels with infinite or continuous alphabets.² For example, when specialized to the additive white Gaussian noise (AWGN) channel with a maximal power constraint, we obtain the achievability part of the dispersion given by Polyanskiy *et al.* [61], thus providing a simple alternative to the proof therein based on the $\kappa\beta$ bound.

For the moderate deviations analysis (cf. Section 2.6), there remains the issue that Proposition 2.1 is stated only for auxiliary costs not depending on n . Here we present a simple modification showing that Theorem 2.7 remains valid provided that $\mathbb{E}_Q[c(X)^2] < \infty$, $\mathbb{E}_Q[a(X)^2] < \infty$, and the auxiliary cost $a_2(\cdot)$ in (2.97) (depending on ρ) satisfies $\mathbb{E}_Q[a_2(X)^2] < \infty$ uniformly in $\rho \in [0, \eta]$, where η is an arbitrarily small positive constant.

We relax the constraints on $a_l(\cdot)$ in $\mathcal{D}_n$ (cf. (2.22)) to the following:

$$\left| \frac{1}{n} \sum_{i=1}^n a_l(x_i) - \phi_l \right| \leq \delta \psi_n^2, \quad (2.136)$$

where ψ_n appears in the statement of Theorem 2.7. It is readily verified using Chebyshev's inequality and the assumption $\psi_n = \omega(\frac{1}{\sqrt{n}})$ that the probability of an i.i.d. vector $\mathbf{X}' \sim Q^n$ violating (2.136) tends to zero. Thus, recalling that the constraint in (2.133) is present in $\mathcal{D}_n$, the normalizing constant μ_n in (2.21) tends to $\frac{1}{2}$.

The remainder of the analysis is nearly identical to that of Section 2.6, and leads to the following analog of (2.106)

$$\bar{p}_e(n, \lfloor e^{n(I_{s,a}(Q) - \psi_n)} \rfloor) \leq e^{-n \left(\frac{\psi_n^2}{2V_{s,a}(Q)} + \frac{\psi_n}{V_{s,a}(Q)} \left(\frac{1}{n} \log \mu_n - 2\delta \psi_n^2 \right) + o(\psi_n^2) + \delta \psi_n^2 \right)}. \quad (2.137)$$

²Analogous results were stated in [62], but the generality of the proof techniques therein is unclear. In particular, the quantization arguments page 4963 therein require that the rate of convergence from $I(X_m; Y)$ to $I(X; Y)$ is sufficiently fast, where X_m is the quantized input variable with a support of cardinality m .

The desired result follows by taking $\delta \rightarrow 0$.

2.10 Proofs

2.10.1 Proof of Theorem 2.3

The proof is similar to that of Gallager for ML decoding [56], so we omit some details. The codeword distribution in (2.21) can be written as

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\mu_n} \prod_{i=1}^n Q(x_i) \mathbb{1}\{\hat{P}_{\mathbf{x}} \in \mathcal{G}_n\}, \quad (2.138)$$

where $\hat{P}_{\mathbf{x}}$ is the empirical distribution (type) of $\mathbf{x}$, and $\mathcal{G}_n$ is the set of types corresponding to sequences $\mathbf{x} \in \mathcal{D}_n$ (see (2.22)). We define the sets

$$\mathcal{S}_n(\mathcal{G}_n) \triangleq \{P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) : P_X \in \mathcal{G}_n\} \quad (2.139)$$

$$\begin{aligned} \mathcal{T}_n(P_{XY}, \mathcal{G}_n) &\triangleq \{\tilde{P}_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}) : \tilde{P}_X \in \mathcal{G}_n, \\ &\quad \tilde{P}_Y = P_Y, \mathbb{E}_{\tilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)]\}. \end{aligned} \quad (2.140)$$

We have from Theorem 2.1 that $\bar{p}_e \doteq \text{rcu}$. Expanding rcu in terms of types, we obtain

$$\bar{p}_e \doteq \sum_{P_{XY} \in \mathcal{S}_n(\mathcal{G}_n)} \mathbb{P}[(\mathbf{X}, \mathbf{Y}) \in T^n(P_{XY})] \min \left\{ 1, M \sum_{\tilde{P}_{XY} \in \mathcal{T}_n(P_{XY}, \mathcal{G}_n)} \mathbb{P}[(\bar{\mathbf{X}}, \mathbf{y}) \in T^n(\tilde{P}_{XY})] \right\}, \quad (2.141)$$

where $\mathbf{y}$ denotes an arbitrary sequence with type P_Y . From Proposition 2.1, the normalizing constant in (2.138) satisfies $\mu_n \doteq 1$, and hence we can safely proceed from (2.141) as if the codeword distribution were $P_{\mathbf{X}} = Q^n$. Using the properties of types in (A.9)–(A.10), it follows that the two probabilities in (2.141) behave as $e^{-nD(P_{XY} \| Q \times W)}$ and $e^{-nD(\tilde{P}_{XY} \| Q \times P_Y)}$ respectively. Combining this with the fact that the number of joint types grows polynomially in n , and setting $M = \lfloor e^{nR} \rfloor$, we obtain $\bar{p}_e \doteq e^{-nE_{r,n}(Q, R, \mathcal{G}_n)}$, where

$$E_{r,n}(Q, R, \mathcal{G}_n) \triangleq \min_{P_{XY} \in \mathcal{S}_n(\mathcal{G}_n)} \min_{\tilde{P}_{XY} \in \mathcal{T}_n(P_{XY}, \mathcal{G}_n)} D(P_{XY} \| Q \times W) + [D(\tilde{P}_{XY} \| Q \times P_Y) - R]^+. \quad (2.142)$$

Using a simple continuity argument (e.g. see [58, Eq. (30)]), we can replace the minimizations over types by minimizations over joint distributions, and the constraints of the form $|\mathbb{E}_P[a_l(X)] - \phi_l| \leq \frac{\delta}{n}$ can be replaced by $\mathbb{E}_P[a_l(X)] = \phi_l$.

2.10.2 Proof of Theorem 2.4

We make use of Fan's minimax theorem [78], which states that $\min_a \sup_b f(a, b) = \sup_b \min_a f(a, b)$ provided that the minimum is over a compact set, $f(\cdot, b)$ is convex in a for all b , and $f(a, \cdot)$ is concave in b for all a .

Using the identity $[\alpha]^+ = \max_{\rho \in [0,1]} \rho \alpha$ and Fan's minimax theorem, the expression in (2.28) can be written as

$$E_r^{\text{cost}}(Q, R) = \max_{\rho \in [0,1]} \hat{E}_0^{\text{cost}}(Q, \rho) - \rho R, \quad (2.143)$$

where

$$\hat{E}_0^{\text{cost}}(Q, \rho) \triangleq \min_{P_{XY} \in \mathcal{S}(\{a_l\})} \min_{\tilde{P}_{XY} \in \mathcal{T}(P_{XY}, \{a_l\})} D(P_{XY} \| Q \times W) + \rho D(\tilde{P}_{XY} \| Q \times P_Y). \quad (2.144)$$

It remains to show that $\hat{E}_0^{\text{cost}}(Q, \rho) = E_0^{\text{cost}}(Q, \rho)$. We prove this using parts (i) and (ii) of Lemma E.2 in Appendix E in succession. After suitable renaming, part (i) of Lemma E.2 reads as follows:

$$\min_{\tilde{P}_{XY} \in \mathcal{T}(P_{XY}, \{a_l\})} D(\tilde{P}_{XY} \| Q \times P_Y) = \sup_{s \geq 0, \{\bar{r}_l\}} \sum_{x,y} P_{XY}(x, y) \log \frac{q(x, y)^s}{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{\sum_l \bar{r}_l (a_l(\bar{x}) - \phi_l)}}, \quad (2.145)$$

where s and $\{\bar{r}_l\}$ correspond to the Lagrange multipliers for the constraints $\mathbb{E}_{\tilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)]$ and $\mathbb{E}_{\tilde{P}}[a_l(X)] = \phi_l$ respectively. It follows that the minimization in (2.144) is equivalent to

$$\min_{P_{XY} \in \mathcal{S}(\{a_l\})} \sup_{s \geq 0, \{\bar{r}_l\}} \sum_{x,y} P_{XY}(x, y) \left(\log \frac{P_{XY}(x, y)}{Q(x)W(y|x)} + \rho \log \frac{q(x, y)^s}{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{\sum_l \bar{r}_l (a_l(\bar{x}) - \phi_l)}} \right). \quad (2.146)$$

By Fan's minimax theorem [78], the minimum and supremum can be interchanged, and we thus consider the minimization of the objective in (2.146) over $P_{XY} \in \mathcal{S}(\{a_l\})$ with s and $\{\bar{r}_l\}$ fixed. Using part (ii) of Lemma E.2, this minimization has a dual form given by

$$\sup_{\{r_l\}} -\log \sum_{x,y} P_{XY}(x, y) \left(\frac{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{\sum_l \bar{r}_l (a_l(\bar{x}) - \phi_l)}}{q(x, y)^s e^{\sum_l r_l (a_l(x) - \phi_l)}} \right)^\rho, \quad (2.147)$$

where $\{r_l\}$ correspond to Lagrange multipliers in the same way as $\{\bar{r}_l\}$. The proof is concluded by taking the supremum over s and $\{\bar{r}_l\}$.

2.10.3 Necessary Conditions for the Optimal Parameters

Here we consider the optimization of the achievable rates and error exponents for constant-composition random coding over the parameters Q and $a(\cdot)$, considering the Karush-Kuhn-Tucker (KKT) conditions [38, Sec. 5.5.3] corresponding to each. We treat the remaining parameter s as a constant rather than considering its KKT conditions, since these will not be needed for the results we provide.

Optimization of $E_0^{\text{cc}}(Q, \rho)$

If $\rho = 0$ then we trivially have $E_0^{\text{cc}}(Q, \rho) = 0$, so we assume that $\rho > 0$. We write the objective in (2.40) as

$$E_0^{\text{cc}}(Q, \rho, s, a) \triangleq \rho \sum_x Q(x) a(x) - \sum_x Q(x) \log f(x), \quad (2.148)$$

where

$$f(x) \triangleq \sum_y W(y|x) q(x, y)^{-\rho s} \left(\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{a(\bar{x})} \right)^\rho. \quad (2.149)$$

We have the partial derivatives

$$\frac{\partial f(x)}{\partial Q(x')} = \sum_y W(y|x) q(x, y)^{-\rho s} \rho \left(\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{a(\bar{x})} \right)^{\rho-1} q(x', y)^s e^{a(x')} \quad (2.150)$$

$$\triangleq \rho g(x, x') \quad (2.151)$$

$$\frac{\partial f(x)}{\partial a(x')} = \rho Q(x') g(x, x'). \quad (2.152)$$

We proceed by analyzing the necessary KKT conditions for (Q, a) to maximize $E_0^{\text{cc}}(Q, \rho, s, a)$. The KKT condition corresponding to the partial derivative with respect to $a(x')$ is

$$\rho Q(x') - \sum_x Q(x) \frac{\rho Q(x') g(x, x')}{f(x)} = 0, \quad (2.153)$$

which implies that either $Q(x') = 0$ or

$$\sum_x Q(x) \frac{g(x, x')}{f(x)} = 1. \quad (2.154)$$

Similarly, the KKT condition corresponding to $Q(x')$ gives

$$\rho a(x') - \log f(x') - \rho \sum_x Q(x) \frac{g(x, x')}{f(x)} - \lambda = 0 \quad (2.155)$$

for all x' such that $Q(x') > 0$, where λ is the Lagrange multiplier associated with the constraint $\sum_x Q(x) = 1$. Substituting (2.154) into (2.155) gives

$$-\log(f(x')e^{-\rho a(x')}) = \lambda + \rho. \quad (2.156)$$

Using the definition of $f(\cdot)$ in (2.149), we see that (2.156) implies that the logarithm in (2.40) is independent of x .

Optimization of $I_{s,a}(Q)$

We write $I_{s,a}(Q)$ in (2.62) as

$$\begin{aligned} I_{s,a}(Q) = & s \sum_{x,y} Q(x)W(y|x) \log q(x,y) \\ & + \sum_x Q(x)a(x) - \sum_{x,y} Q(x)W(y|x) \log \sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})} \end{aligned} \quad (2.157)$$

and analyze the KKT conditions associated with the maximization over (Q, a) . We omit some details, since the steps are similar to those above. For all x' such that $Q(x') > 0$, the KKT condition for $a(x')$ yields

$$\sum_{x,y} Q(x)W(y|x) \frac{q(x',y)^s e^{a(x')}}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})}} = 1, \quad (2.158)$$

and the KKT condition for $Q(x')$ yields

$$\begin{aligned} s \sum_y W(y|x') \log q(x',y) + a(x') - \sum_y W(y|x') \log \sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})} \\ - \sum_{x,y} Q(x)W(y|x) \frac{q(x',y)^s e^{a(x')}}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})}} - \lambda = 0 \end{aligned} \quad (2.159)$$

where λ is a Lagrange multiplier. Substituting (2.158) into (2.159) and performing some simple rearrangements, we obtain

$$\sum_y W(y|x') \log \frac{q(x',y)^s e^{a(x')}}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})}} = \lambda + 1. \quad (2.160)$$

Optimization of $E_x^{\text{cc}}(Q, \rho)$

We again omit the details to avoid repetition with the preceding analysis. We write the objective in (2.125) in the form

$$E_x^{\text{cc}}(Q, \rho, s, a) = -\rho \sum_x Q(x) \log \sum_{\bar{x}} Q(\bar{x}) e^{(-d_s(x, \bar{x}) + a(\bar{x}))/\rho} + \sum_x Q(x) a(x), \quad (2.161)$$

where d_s is defined in (2.117). For all x' such that $Q(x') > 0$, we obtain the following KKT conditions corresponding to $a(x')$ and $Q(x')$:

$$\sum_x Q(x) \frac{e^{(-d_s(x, x') + a(x'))/\rho}}{\sum_{\bar{x}} Q(\bar{x}) e^{(-d_s(x, \bar{x}) + a(\bar{x}))/\rho}} = 1 \quad (2.162)$$

$$-\rho \log \sum_{\bar{x}} Q(\bar{x}) e^{(-d_s(x', \bar{x}) + a(\bar{x}))/\rho} - \rho \sum_x Q(x) \frac{e^{(-d_s(x, x') + a(x'))/\rho}}{\sum_{\bar{x}} Q(\bar{x}) e^{(-d_s(x, \bar{x}) + a(\bar{x}))/\rho}} + a(x') = \lambda, \quad (2.163)$$

where λ is a Lagrange multiplier. Combining these, it follows that any optimal pair (Q, a) must be such that the logarithm in (2.125) is the same for all x such that $Q(x) > 0$.

2.10.4 Proof of Theorem 2.8

The analysis here is similar to that of Section 2.10.1, and we make use of the definition of the set $\mathcal{G}_n$ therein. Along with the definitions in (2.111)–(2.112), we define

$$\mathcal{T}'(\tilde{P}_{X\bar{X}}) \triangleq \left\{ P_{X\bar{X}Y} \in \mathcal{P}(\mathcal{X} \times \mathcal{X} \times \mathcal{Y}) : P_{X\bar{X}} = \tilde{P}_{X\bar{X}}, \mathbb{E}_P[\log q(\bar{X}, Y)] \geq \mathbb{E}_P[\log q(X, Y)] \right\} \quad (2.164)$$

$$\mathcal{S}_n(\mathcal{G}_n) \triangleq \left\{ P_{X\bar{X}} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{X}) : P_X \in \mathcal{G}_n, P_{\bar{X}} \in \mathcal{G}_n \right\} \quad (2.165)$$

$$\mathcal{T}'_n(\tilde{P}_{X\bar{X}}) \triangleq \mathcal{T}'(\tilde{P}_{X\bar{X}}) \cap \mathcal{P}_n(\mathcal{X} \times \mathcal{X} \times \mathcal{Y}). \quad (2.166)$$

It follows that $P_{X\bar{X}Y} \in \mathcal{T}(\{a_l\})$ if and only if $P_{X\bar{X}Y} \in \mathcal{T}'(\tilde{P}_{X\bar{X}})$ for some $\tilde{P}_{X\bar{X}} \in \mathcal{S}(\{a_l\})$.

Using the codeword distribution in (2.21) and expanding (2.13) in terms of types, we obtain

$$\begin{aligned} & \text{rcux}_\rho(n, M)^{1/\rho} \\ & \doteq M \sum_{\tilde{P}_{X\bar{X}} \in \mathcal{S}_n(\mathcal{G}_n)} \mathbb{P}[(\mathbf{X}, \bar{\mathbf{X}}) \in T^n(P_{X\bar{X}})] \sum_{P_{X\bar{X}Y} \in \mathcal{T}'_n(\tilde{P}_{X\bar{X}})} \mathbb{P}[(\mathbf{x}, \bar{\mathbf{x}}, \mathbf{Y}) \in T^n(P_{X\bar{X}Y}) | \mathbf{X} = \mathbf{x}]^{1/\rho} \end{aligned} \quad (2.167)$$

$$\doteq M \max_{\tilde{P}_{X\bar{X}} \in \mathcal{S}_n(\mathcal{G}_n)} \max_{P_{X\bar{X}Y} \in \mathcal{T}'_n(\tilde{P}_{X\bar{X}})} \exp\left(-nD(P_{X\bar{X}} \| Q \times Q)\right) \exp\left(-\frac{n}{\rho}D(P_{X\bar{X}Y} \| \tilde{P}_{X\bar{X}} \times W)\right) \quad (2.168)$$

$$\doteq M \max_{P_{X\bar{X}Y} \in \mathcal{T}(\{a_l\})} \exp\left(-n\left(\frac{1}{\rho}D(P_{X\bar{X}Y} \| P_{X\bar{X}} \times W) + I_P(X; \bar{X})\right)\right), \quad (2.169)$$

where in (2.167) we define $(\mathbf{x}, \bar{\mathbf{x}})$ to be an arbitrary pair with joint type $\tilde{P}_{X\bar{X}}$, (2.168) follows from Proposition 2.1 and the properties of types in (A.11)–(A.12), and (2.169) follows from the definitions of $\mathcal{S}_n$, $\mathcal{T}'_n$ and $\mathcal{T}$, and by using a continuity argument similar to that of Section 2.10.1. We thus obtain the exponent

$$\sup_{\rho \geq 1} \min_{P_{X\bar{X}Y} \in \mathcal{T}(\{a_l\})} D(P_{X\bar{X}Y} \| P_{X\bar{X}} \times W) + \rho(D(P_{X\bar{X}} \| Q \times Q) - R) \quad (2.170)$$

$$= \min_{P_{X\bar{X}Y} \in \mathcal{T}(\{a_l\})} \sup_{\rho \geq 1} D(P_{X\bar{X}Y} \| P_{X\bar{X}} \times W) + \rho(D(P_{X\bar{X}} \| Q \times Q)), \quad (2.171)$$

where (2.171) follows from Fan's minimax theorem [78], the conditions of which are satisfied here since the objective is linear in ρ and convex in $P_{X\bar{X}Y}$. Finally, we obtain (2.114) from (2.171) using

$$\sup_{\rho \geq 1} \rho\alpha = \begin{cases} \infty & \alpha > 0 \\ \alpha & \alpha \leq 0 \end{cases} \quad (2.172)$$

and the identity

$$D(P_{X\bar{X}Y} \| Q \times Q \times W) = D(P_{X\bar{X}Y} \| P_{X\bar{X}} \times W) + D(P_{X\bar{X}} \| Q \times Q). \quad (2.173)$$

2.10.5 Proof of Theorem 2.9

Recall the definitions of $\mathcal{S}(\{a_l\})$, $\mathcal{T}(\{a_l\})$ and $\mathcal{T}'(\tilde{P}_{X\bar{X}})$ in (2.111), (2.112) and (2.164) respectively. We write (2.114) as

$$E_{\text{ex}}^{\text{cc}}(Q, R) = \min_{\substack{\tilde{P}_{X\bar{X}} \in \mathcal{S}(\{a_l\}) \\ D(P_{X\bar{X}} \| Q \times Q) \leq R}} \min_{P_{X\bar{X}Y} \in \mathcal{T}(\tilde{P}_{X\bar{X}})} D(P_{X\bar{X}Y} \| \tilde{P}_{X\bar{X}} \times W) + D(\tilde{P}_{X\bar{X}} \| Q \times Q) - R, \quad (2.174)$$

where the objective follows from (2.173). For a given $\tilde{P}_{X\bar{X}} \in \mathcal{S}(\{a_l\})$, the quantity $D(P_{X\bar{X}}\|Q \times Q) - R$ is constant, and hence we consider the optimization problem

$$\min_{P_{X\bar{X}Y} \in \mathcal{T}(\tilde{P}_{X\bar{X}})} D(P_{X\bar{X}Y}\|\tilde{P}_{X\bar{X}} \times W). \quad (2.175)$$

By part (iii) of Lemma E.2 in Appendix E, this optimization has the same value as

$$\sup_{s \geq 0} - \sum_{x, \bar{x}} \tilde{P}_{X\bar{X}}(x, \bar{x}) \log \sum_y W(y|x) \left(\frac{q(\bar{x}, y)}{q(x, y)} \right)^s, \quad (2.176)$$

where s is a Lagrange multiplier. Renaming $\tilde{P}_{X\bar{X}}$ as $P_{X\bar{X}}$ and adding $D(P_{X\bar{X}}\|Q \times Q) - R$ (see (2.174)–(2.175)), we obtain the right-hand side of (2.115) with the minimum and supremum in the opposite order. Using Fan's minimax theorem [78], we can safely interchange the two.

The second statement (2.116) follows similarly from part (iv) of Lemma E.2. The parameter ρ corresponds to the Lagrange multiplier for the constraint $D(P_{X\bar{X}}\|Q \times Q) \leq R$, and the parameters $\{r_l\}$ and $\{\bar{r}_l\}$ correspond to Lagrange multipliers for the constraints containing the auxiliary costs.

Chapter 3

Refined Asymptotic Bounds

3.1 Introduction

Error exponents, second-order coding rates and moderate deviations results (see Sections 1.2.1–1.2.3) provide a more refined characterization of the system performance than the capacity alone. However, they all contain higher-order asymptotic terms that can be significant at short to moderate block lengths. In this chapter, we seek refined asymptotic results that characterize some of the higher-order terms.

Early works on characterizing the subexponential prefactor for a given rate under ML decoding include those of Elias [12], Dobrushin [79] and Shannon [80], who studied specific channels exhibiting a high degree of symmetry. More recently, Altuğ and Wagner [81–84] obtained asymptotic prefactors for arbitrary DMCs. The growth rate of the optimal prefactor is not yet fully understood, but matching achievability and converse results have been provided for symmetric channels when the rate is above the critical rate [84].

There has also been substantial effort in studying the third-order term in the regime of fixed error probability, and a complete characterization is known for most DMCs [61, 85–89], as well as the additive white Gaussian noise (AWGN) channel [61, 90]. The achievability techniques of Polyanskiy [85, Sec. 3.4.5] will be of particular interest in this chapter.

Throughout the chapter, we consider the setup described in Section 1.1.1, and we use the random coding bounds and ensembles given in Sections 2.2–2.3. Our main contributions are as follows:

1. In Section 3.2, we provide refined asymptotic results for (non-expurgated) i.i.d. random coding. For the non-asymptotic random-coding bounds rcu and rcu_g introduced in Section 2.2, we give saddlepoint approximations [65] that can be computed efficiently,

and that characterize the asymptotic behavior of the corresponding bounds as $n \rightarrow \infty$ at both fixed and varying rates, this unifying the regimes of error exponents, second-order rates and moderate deviations. In particular, the approximation of rcu is shown to simultaneously capture the behavior of the third-order term of Polyanskiy at a fixed error probability [85, Sec. 3.4.5], and the asymptotic prefactor of Altuğ-Wagner [81] at a fixed rate. Moreover, in the fixed rate setting, a novel characterization of the implied constant is obtained. Using numerical examples, it is shown that the approximations are remarkably accurate even at short block lengths.

2. In Section 3.3, we derive Gallager’s expurgated i.i.d. exponent (and its generalization to mismatched decoding) with a $O(\frac{1}{\sqrt{n}})$ prefactor, thus improving on the original $O(1)$ prefactor. Once again, the implied constant is explicitly characterized.
3. In Section 3.4, we present refined asymptotic bounds for constant-composition random coding. Under some technical conditions, we derive a fixed-error third-order term and a fixed-rate prefactor growth rate analogous to those of i.i.d. random coding. Moreover, we study the reduction in the performance when the technical conditions fail, showing in particular that a loss for sum channels [1] is unavoidable in general.

Our analysis will make use of several techniques and results from the probability literature, including strong large deviations, local limit theorems, Edgeworth expansions, and concentration functions. An introduction to these concepts is given in Appendix C.

Parts of this chapter were presented in [68, 91–93].

3.2 Saddlepoint Approximation for i.i.d. Random Coding

In this section, we obtain estimates of the non-asymptotic bounds rcu and rcu_s (see Theorem 2.1) based on the saddlepoint approximation [65]. We prove the asymptotic tightness in the limit of large n for both fixed and varying rates. In particular, the approximations will be seen to unify the regimes of exponents, second-order coding rates, and moderate deviations, including some higher-order terms. We focus on i.i.d. random coding (i.e. $P_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^n Q(x_i)$), which is particularly amenable to a precise asymptotic analysis.

3.2.1 Preliminary Definitions and Results

As in Section 2.5, we fix $Q \in \mathcal{P}(\mathcal{X})$ and $s > 0$, and define the quantities

$$i_s(x, y) \triangleq \log \frac{q(x, y)^s}{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s} \quad (3.1)$$

$$i_s^n(\mathbf{x}, \mathbf{y}) \triangleq \sum_{i=1}^n i_s(x_i, y_i) \quad (3.2)$$

$$I_s(Q) \triangleq \mathbb{E}[i_s(X, Y)] \quad (3.3)$$

$$U_s(Q) \triangleq \text{Var}[i_s(X, Y)], \quad (3.4)$$

where $(X, Y) \sim Q \times W$. We write rcu_s in (2.8) (with $P_{\mathbf{X}} = Q^n$) as

$$\text{rcu}_s(n, M) = \mathbb{E} \left[\min \left\{ 1, (M-1) e^{-i_s^n(\mathbf{X}, \mathbf{Y})} \right\} \right]. \quad (3.5)$$

Furthermore, we let

$$E_0^{\text{iid}}(Q, \rho, s) \triangleq -\log \mathbb{E} [e^{-\rho i_s(X, Y)}] \quad (3.6)$$

denote the objective in (2.38) with a fixed value of s in place of the supremum. The optimal value of ρ is given by

$$\hat{\rho}(Q, R, s) \triangleq \arg \max_{\rho \in [0, 1]} E_0^{\text{iid}}(Q, \rho, s) - \rho R, \quad (3.7)$$

and the critical rate is defined as

$$R_s^{\text{cr}}(Q) \triangleq \sup \{ R : \hat{\rho}(Q, R, s) = 1 \}. \quad (3.8)$$

We define the following derivatives associated with (3.7):

$$c_1(Q, R, s) \triangleq R - \left. \frac{\partial E_0^{\text{iid}}(Q, \rho, s)}{\partial \rho} \right|_{\rho=\hat{\rho}(Q, R, s)} \quad (3.9)$$

$$c_2(Q, R, s) \triangleq - \left. \frac{\partial^2 E_0^{\text{iid}}(Q, \rho, s)}{\partial \rho^2} \right|_{\rho=\hat{\rho}(Q, R, s)}. \quad (3.10)$$

The following properties of the above quantities are analogous to those of Gallager for ML decoding [2, pp. 141-143], and are proved in a similar fashion (see Section 3.5.1 for details):

1. $\hat{\rho}(Q, R, s)$ is continuous in R .
2. For all $R \geq 0$, we have $c_2(Q, R, s) > 0$ if $U_s(Q) > 0$, and $c_2(Q, R, s) = 0$ if $U_s(Q) = 0$. Furthermore, we have $c_2(Q, I_s(Q), s) = U_s(Q)$.

3. If $U_s(Q) = 0$, then $R_s^{\text{cr}}(Q) = I_s(Q)$.
4. For $R \in [0, R_s^{\text{cr}}(Q))$, we have $\hat{\rho}(Q, R, s) = 1$ and $c_1(Q, R, s) < 0$.
5. For $R \in [R_s^{\text{cr}}(Q), I_s(Q)]$, $\hat{\rho}(Q, R, s)$ is strictly decreasing in R , and $c_1(Q, R, s) = 0$.
6. For $R > I_s(Q)$, we have $\hat{\rho}(Q, R, s) = 0$ and $c_1(Q, R, s) > 0$.

Throughout this section, the arguments to $\hat{\rho}$, c_1 , etc. will be omitted, since their values will be clear from the context.

The density function of a $N(\mu, \sigma^2)$ random variable is denoted by

$$\phi(z; \mu, \sigma^2) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}. \quad (3.11)$$

Our results will distinguish between lattice and non-lattice random variables; see Definition C.1 in Appendix C. In the lattice case, we will make use of

$$\phi_h(z; \mu, \sigma^2) \triangleq \frac{h}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-\mu)^2}{2\sigma^2}}, \quad (3.12)$$

which can be interpreted as an approximation of the integral of $\phi(\cdot; \mu, \sigma^2)$ from z to $z + h$ when h is small.

3.2.2 Approximation for $\text{rcu}_s(n, M)$

In the proof of Theorem 3.1 below, we derive an approximation $\widehat{\text{rcu}}_s$ of rcu_s taking the form

$$\widehat{\text{rcu}}_s(n, M) \triangleq \alpha_n(Q, R, s) e^{-n(E_0^{\text{iid}}(Q, \hat{\rho}, s) - \hat{\rho}R)}, \quad (3.13)$$

where $R = \frac{1}{n} \log M$, and the prefactor α_n varies depending on whether $i_s(X, Y)$ is a lattice variable. In the non-lattice case, the prefactor is given by

$$\alpha_n^{\text{nl}}(Q, R, s) \triangleq \int_0^\infty e^{-\hat{\rho}z} \phi(z; nc_1, nc_2) dz + \int_{-\infty}^0 e^{(1-\hat{\rho})z} \phi(z; nc_1, nc_2) dz. \quad (3.14)$$

In the lattice case, it will prove convenient to consider $R - i_s(X, Y)$ rather than $i_s(X, Y)$. Denoting the offset and span of $R - i_s(X, Y)$ by γ and h respectively, we see that $nR - i_s^n(\mathbf{X}, \mathbf{Y})$ has span h , and its offset can be chosen as

$$\gamma_n \triangleq \min \left\{ n\gamma + ih : i \in \mathbb{Z}, n\gamma + ih \geq 0 \right\}. \quad (3.15)$$

The prefactor for the lattice case is given by

$$\alpha_n^1(Q, R, s) \triangleq \sum_{i=0}^{\infty} e^{-\hat{\rho}(\gamma_n + ih)} \phi_h(\gamma_n + ih; nc_1, nc_2) + \sum_{i=-\infty}^{-1} e^{(1-\hat{\rho})(\gamma_n + ih)} \phi_h(\gamma_n + ih; nc_1, nc_2), \quad (3.16)$$

and the overall prefactor in (3.13) is defined as

$$\alpha_n(Q, R, s) \triangleq \begin{cases} \alpha_n^{\text{nl}}(Q, R, s) & R - i_s(X, Y) \text{ is non-lattice} \\ \alpha_n^1(Q, R, s) & R - i_s(X, Y) \text{ has offset } \gamma \text{ and span } h. \end{cases} \quad (3.17)$$

It is generally easy to verify whether or not $R - i_s(X, Y)$ is a lattice variable by evaluating its values on the support of $Q \times W$. A notable example of the lattice case is that in which there are only two distinct values among these; this occurs in several symmetric settings.

Before proceeding, we give a high-level comparison of the saddlepoint approximation and the normal approximation (the latter corresponding to (2.92) with the remainder term omitted). As stated in Section 3.2.1, we have $c_1(Q, I_s(Q), s) = 0$ and $c_2(Q, I_s(Q), s) = U_s(Q)$; hence, at least at an intuitive level, we expect the two approximations to behave similarly when $R \approx I_s(Q)$. For more general choices of R , the saddlepoint approximation is obtained by first performing an exponential change of measure, and *then* applying a normal-type approximation. This yields an expression having the correct exponential behavior, and yielding significantly better estimates in many cases. The values of c_1 and c_2 are the mean and variance of the exponentially tilted version of $R - i_s(X, Y)$.

While (3.14) and (3.16) are written in terms of integrals and summations, both prefactors are single-letter, and can be computed efficiently to a high degree of accuracy. In the non-lattice case, this is done using the identity

$$\int_a^{\infty} e^{bz} \phi(z; \mu, \sigma^2) dz = e^{\mu b + \frac{1}{2}\sigma^2 b^2} Q\left(\frac{a - \mu - b\sigma^2}{\sigma}\right). \quad (3.18)$$

In the lattice case, we can write each of the summations in (3.16) in the form

$$\sum_i e^{b_0 + b_1 i + b_2 i^2} = e^{-\frac{b_1^2}{4b_2} + b_0} \sum_i e^{b_2 (i + \frac{b_1}{2b_2})^2}, \quad (3.19)$$

where $b_2 < 0$. We can thus obtain an accurate approximation by keeping only the terms in the sum such that i is sufficiently close to $-\frac{b_1}{2b_2}$. Overall, the computational complexity of the saddlepoint approximation is similar to that of the exponent alone.

Theorem 3.1. Fix an input distribution $Q \in \mathcal{P}(\mathcal{X})$ and parameter $s > 0$ such that $U_s(Q) > 0$. For any sequence of integers $\{M_n\}$ such that $M_n \rightarrow \infty$, we have

$$\lim_{n \rightarrow \infty} \frac{\widehat{\text{rcu}}_s(n, M_n)}{\text{rcu}_s(n, M_n)} = 1. \quad (3.20)$$

Proof. See Section 3.5.3. □

A heuristic derivation of the non-lattice version of $\widehat{\text{rcu}}_s$ was provided in [94]; Theorem 3.1 provides a formal derivation, and a treatment of the lattice case. It should be noted that the assumption $U_s(Q) > 0$ is not restrictive, since in the case that $U_s(Q) = 0$ the argument to the expectation in (3.5) is deterministic, and hence rcu_s can easily be computed exactly.

In the case that the rate R is fixed, simpler asymptotic expressions can be obtained. In Section 3.5.2, we prove the following (here $f_n \asymp g_n$ denotes the relation $\lim_{n \rightarrow \infty} \frac{f_n}{g_n} = 1$):

- If $R \in [0, R_s^{\text{cr}}(Q))$ or $R > I_s(Q)$, then

$$\alpha_n(Q, R, s) \asymp 1. \quad (3.21)$$

- If $R = R_s^{\text{cr}}(Q)$ or $R = I_s(Q)$, then

$$\alpha_n(Q, R, s) \asymp \frac{1}{2}. \quad (3.22)$$

- If $R \in (R_s^{\text{cr}}(Q), I_s(Q))$, then

$$\alpha_n^{\text{nl}}(Q, R, s) \asymp \frac{1}{\sqrt{2\pi n c_2} \hat{\rho}(1 - \hat{\rho})} \quad (3.23)$$

$$\alpha_n^1(Q, R, s) \asymp \frac{h}{\sqrt{2\pi n c_2}} \left(e^{-\hat{\rho}\gamma_n} \left(\frac{1}{1 - e^{-\hat{\rho}h}} \right) + e^{(1-\hat{\rho})\gamma_n} \left(\frac{e^{-(1-\hat{\rho})h}}{1 - e^{-(1-\hat{\rho})h}} \right) \right). \quad (3.24)$$

The asymptotic prefactors in (3.21)–(3.24) are related to *exact asymptotics* studies in the statistics literature, which seek to characterize the subexponential prefactor for probabilities that decay at an exponential rate (e.g. see [95]). These prefactors are useful in gaining further insight into the behavior of the error probability compared to the error exponent alone. However, there is a notable limitation which is best demonstrated here using (3.23). The right-hand side of (3.23) characterizes the prefactor to within a multiplicative $1 + o(1)$ term for a given rate, but it diverges as $\hat{\rho} \rightarrow 0$ or $\hat{\rho} \rightarrow 1$. Thus, unless n is large, the estimate obtained by omitting the higher-order terms is inaccurate for rates slightly above $R_s^{\text{cr}}(Q)$ or slightly below $I_s(Q)$.

In contrast, the right-hand side of (3.14) (and similarly (3.16)) remains bounded for all $\hat{\rho} \in [0, 1]$. Furthermore, as Theorem 3.1 shows, this expression characterizes the true behavior of rcu_s to within a multiplicative $1 + o(1)$ term not only for fixed rates, but also when the rate varies with the block length. Thus, it remains suitable for characterizing the behavior of rcu_s even when the rate approaches $I_s(Q)$ (or $R_s^{\text{cr}}(Q)$), as is the case in the regimes of second-order coding rates and moderate deviations. Moreover, in the former regime, the proof of Theorem 3.1 reveals that $\widehat{\text{rcu}}_s = \text{rcu}_s + O(\frac{1}{\sqrt{n}})$, which implies (via a Taylor expansion of Q^{-1}) that $\widehat{\text{rcu}}_s$ captures the behavior in (2.92) up to the $O(1)$ term.

3.2.3 Approximation for $\text{rcu}(n, M)$

In the proof of Theorem 2.1, we obtained rcu_s from rcu using Markov's inequality. In this subsection, we will see that, under some technical assumptions, a refined analysis yields a bound which is tighter than rcu_s , but still amenable to the techniques of the previous subsection.

Technical Assumptions

We define the set

$$\mathcal{Y}_1(Q) \triangleq \left\{ y : q(x, y) \neq q(\bar{x}, y) \text{ for some } x, \bar{x} \text{ such that } Q(x)Q(\bar{x})W(y|x)W(y|\bar{x}) > 0 \right\}, \quad (3.25)$$

and make the following technical assumptions on (W, q, Q) :

$$q(x, y) > 0 \iff W(y|x) > 0 \quad (3.26)$$

$$\mathcal{Y}_1(Q) \neq \emptyset. \quad (3.27)$$

When $q(x, y) = W(y|x)$, (3.26) is trivial, and (3.27) is the *non-singularity or feasibility decoding is suboptimal* condition of [81, 96]. A notable example where this condition fails is the binary erasure channel (BEC). It should be noted that if (3.26) holds but (3.27) fails then we in fact have $\text{rcu} = \text{rcu}_s$ for any $s > 0$, and hence $\widehat{\text{rcu}}_s$ also approximates rcu . This can be seen by noting that rcu_s is obtained from rcu using the inequality $\mathbb{1}\{\bar{q} \geq q\} \leq (\frac{\bar{q}}{q})^s$, which holds with equality when $\frac{\bar{q}}{q} \in \{0, 1\}$.

Definitions

Along with the definitions given in Section 3.2.1, we make use of the reverse conditional distribution

$$\tilde{P}_s(x|y) \triangleq \frac{Q(x)q(x,y)^s}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s}, \quad (3.28)$$

the joint tilted distribution

$$P_{\hat{\rho},s}^*(x,y) = \frac{Q(x)W(y|x)e^{-\hat{\rho}i_s(x,y)}}{\sum_{x',y'} Q(x')W(y'|x')e^{-\hat{\rho}i_s(x',y')}} \quad (3.29)$$

and its Y -marginal $P_{\hat{\rho},s}^*(y)$, and the conditional variance

$$c_3(Q, R, s) \triangleq \mathbb{E} \left[\text{Var} [i_s(X^*, Y^*) | Y^*] \right], \quad (3.30)$$

where $(X^*, Y^*) \sim P_{\hat{\rho},s}^*(y)\tilde{P}_s(x|y)$. Furthermore, we define the set

$$\mathcal{I}_s \triangleq \bigcup_{(x,y): Q(x)W(y|x) > 0, y \in \mathcal{Y}_1(Q)} \{i_s(x,y)\}, \quad (3.31)$$

and let

$$\psi_s \triangleq \begin{cases} 1 & \mathcal{I}_s \text{ does not lie on a lattice} \\ \frac{\bar{h}}{1-e^{-\bar{h}}} & \mathcal{I}_s \text{ lies on a lattice with span } \bar{h}. \end{cases} \quad (3.32)$$

The set $\mathcal{I}_s$ is the support of a random variable that will appear in the analysis of the inner probability in (2.7). It should be noted that $\bar{h}$ in (3.32) can differ from h (the span of $i_s(X, Y)$) in general, and it may even be the case that $\mathcal{I}_s$ lies on a lattice even though $i_s(X, Y)$ is non-lattice. A straightforward example of this phenomenon is seen when $i_s(X, Y)$ takes values on $\{-\sqrt{2}, -1, 1, \sqrt{2}\}$, with only the values $\{-1, 1\}$ corresponding to $y \in \mathcal{Y}_1(Q)$.

We claim that the assumptions in (3.26)–(3.27) imply that $c_3 > 0$ for any R and $s > 0$. To see this, we write

$$\text{Var}_{\tilde{P}_s(\cdot|y)}[i_s(X, y)] = 0 \iff \log \frac{\tilde{P}_s(x|y)}{Q(x)} \text{ is independent of } x \text{ wherever } \tilde{P}_s(x|y) > 0 \quad (3.33)$$

$$\iff q(x, y) \text{ is independent of } x \text{ wherever } Q(x)q(x, y) > 0 \quad (3.34)$$

$$\iff y \notin \mathcal{Y}_1(Q), \quad (3.35)$$

where (3.33) and (3.34) follow from the definition of $\tilde{P}_s$ in (3.28) and the assumption $s > 0$,

and (3.35) follows from (3.26) and the definition of $\mathcal{Y}_1(Q)$ in (3.25). Using (3.1), (3.26) and (3.29), we have

$$P_{\hat{\rho},s}^*(y) > 0 \iff \sum_x Q(x)W(y|x) > 0. \quad (3.36)$$

Thus, from (3.27), we have $P_{\hat{\rho},s}^*(y) > 0$ for some $y \in \mathcal{Y}_1(Q)$, which (along with (3.35)) proves that $c_3 > 0$.

Main Result

The main result of this subsection is written in terms of an approximation of the form

$$\widehat{\text{rcu}}_s^*(n, M) \triangleq \beta_n(Q, R, s) e^{-n(E_0^{\text{iid}}(Q, \hat{\rho}, s) - \hat{\rho}R)}, \quad (3.37)$$

where we again have $R = \frac{1}{n} \log M$. Analogously to the previous subsection, we treat the lattice and non-lattice cases separately:

$$\beta_n(Q, R, s) \triangleq \begin{cases} \beta_n^{\text{nl}}(Q, R, s) & R - i_s(X, Y) \text{ is non-lattice} \\ \beta_n^{\text{l}}(Q, R, s) & R - i_s(X, Y) \text{ has offset } \gamma \text{ and span } h, \end{cases} \quad (3.38)$$

where

$$\beta_n^{\text{nl}}(Q, R, s) \triangleq \int_{\log \frac{\sqrt{2\pi n c_3}}{\psi_s}}^{\infty} e^{-\hat{\rho}z} \phi(z; n c_1, n c_2) dz + \frac{\psi_s}{\sqrt{2\pi n c_3}} \int_{-\infty}^{\log \frac{\sqrt{2\pi n c_3}}{\psi_s}} e^{(1-\hat{\rho})z} \phi(z; n c_1, n c_2) dz \quad (3.39)$$

$$\begin{aligned} \beta_n^{\text{l}}(Q, R, s) &\triangleq \sum_{i=i^*}^{\infty} e^{-\hat{\rho}(\gamma_n + ih)} \phi_h(\gamma_n + ih; n c_1, n c_2) \\ &\quad + \frac{\psi_s}{\sqrt{2\pi n c_3}} \sum_{i=-\infty}^{i^*-1} e^{(1-\hat{\rho})(\gamma_n + ih)} \phi_h(\gamma_n + ih; n c_1, n c_2), \end{aligned} \quad (3.40)$$

and where in (3.40) we use γ_n in (3.15) along with

$$i^* \triangleq \min \left\{ i \in \mathbb{Z} : \gamma_n + ih \geq \log \frac{\sqrt{2\pi n c_3}}{\psi_s} \right\}. \quad (3.41)$$

Theorem 3.2. *Under the setup of Theorem 3.1 and the assumptions in (3.26)–(3.27), we have*

$$\text{rcu}(n, M_n) \leq \widehat{\text{rcu}}_s^*(n, M_n)(1 + o(1)), \quad (3.42)$$

where

$$\text{rcu}_s^*(n, M) \triangleq \mathbb{E} \left[\min \left\{ 1, \frac{M\psi_s}{\sqrt{2\pi n c_3}} e^{-i_s^n(\mathbf{X}, \mathbf{Y})} \right\} \right]. \quad (3.43)$$

Furthermore, we have

$$\lim_{n \rightarrow \infty} \frac{\widehat{\text{rcu}}_s^*(n, M_n)}{\text{rcu}_s^*(n, M_n)} = 1. \quad (3.44)$$

Proof. See Section 3.5.4. $\square$

When the rate does not vary with n , we can apply the same arguments as those given in Section 3.5.2 to obtain the following analogues of (3.21)–(3.24):

- If $R \in [0, R_s^{\text{cr}}(Q))$, then

$$\beta_n(Q, R, s) \asymp \frac{\psi_s}{\sqrt{2\pi n c_3}}, \quad (3.45)$$

and similarly for $R = R_s^{\text{cr}}(Q)$ after multiplying the right-hand side by $\frac{1}{2}$.

- If $R \in (R_s^{\text{cr}}(Q), I_s(Q))$, then

$$\beta_n^{\text{nl}}(Q, R, s) \asymp \left(\frac{\psi_s}{\sqrt{2\pi n c_3}} \right)^{\hat{\rho}} \frac{1}{\sqrt{2\pi n c_2} \hat{\rho} (1 - \hat{\rho})} \quad (3.46)$$

$$\beta_n^{\text{l}}(Q, R, s) \asymp \left(\frac{\psi_s}{\sqrt{2\pi n c_3}} \right)^{\hat{\rho}} \frac{h}{\sqrt{2\pi n c_2}} \left(e^{-\hat{\rho} \gamma'_n} \left(\frac{1}{1 - e^{-\hat{\rho} h}} \right) + e^{(1-\hat{\rho}) \gamma'_n} \left(\frac{e^{-(1-\hat{\rho})h}}{1 - e^{-(1-\hat{\rho})h}} \right) \right), \quad (3.47)$$

where $\gamma'_n \triangleq \gamma_n + i^* h - \log \frac{\sqrt{2\pi n c_3}}{\psi_s} \in [0, h)$ (see (3.41)).

- For $R \geq I_s(Q)$, the asymptotics of β_n coincide with those of α_n (see (3.21)–(3.22)).

When combined with Theorem 3.2, these expansions provide an alternative proof of the main result of Altuğ-Wagner [81], along with a characterization of the multiplicative $\Theta(1)$ terms that were left unspecified in [81]. A simpler version of our analysis can also be used to obtain the prefactors with unspecified constants [91].

Analogously to the previous subsection, in the regime of fixed target error probability we can write (3.44) more precisely as $\widehat{\text{rcu}}_s^* = \text{rcu}_s^* + O(\frac{1}{\sqrt{n}})$, implying that the asymptotic expansions of the rates corresponding to rcu_s^* and $\widehat{\text{rcu}}_s^*$ coincide up to the $O(1)$ term. From the analysis given in [85, Sec. 3.4.5], rcu_s^* yields an expansion of the form (2.92) with the $O(1)$ term replaced by $\frac{1}{2} \log n + O(1)$, and it follows that the same is true of $\widehat{\text{rcu}}_s^*$.

3.2.4 Numerical Examples

Here we provide numerical examples to demonstrate the utility of the saddlepoint approximations given above. Along with $\widehat{\text{rcu}}_s$ and $\widehat{\text{rcu}}_s^*$, we consider (i) the normal approximation, obtained by omitting the remainder term in (2.92), (ii) the error exponent approximation $p_e \approx e^{-nE_r^{\text{iid}}(Q,R)}$, and (iii) the exact asymptotics approximations on the right-hand sides of (3.24) and (3.47). We use the lattice-type versions of the approximations, since we consider examples in which $R - i_s(X, Y)$ is a lattice variable. We observed no significant difference in the accuracy of each approximation in similar non-lattice examples.

We consider the example given in Section 2.8, using the parameters $\delta_0 = 0.01$, $\delta_1 = 0.05$, $\delta_2 = 0.25$, and $Q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. We approximate the summations of the form (3.19) by keeping the 1000 terms¹ whose indices are closest to $-\frac{b_1}{2b_2}$. We choose the free parameter s to be the value that maximizes the error exponent at each rate. For the normal approximation, we choose s to achieve the GMI in (1.11). Defining $R_{\text{cr}}(Q)$ to be the supremum of all rates such that $\hat{\rho} = 1$ when s is optimized, we have $I_{\text{GMI}}(Q) = 0.643$ and $R_{\text{cr}}(Q) = 0.185$ bits/use.

In Figure 3.1, we plot the error probability as a function of the rate with $n = 60$. Despite the fact that the block length is short, rcu_s and $\widehat{\text{rcu}}_s^*$ are indistinguishable at all rates. Similarly, the gap from rcu to $\widehat{\text{rcu}}_s^*$ is small. Consistent with the fact that Theorem 3.2 gives an asymptotic upper bound on rcu rather than an asymptotic equality, $\widehat{\text{rcu}}_s^*$ lies slightly above rcu at low rates. The error exponent approximation is close to rcu_s at low rates, but it is pessimistic at high rates. The normal approximation behaves somewhat similarly to rcu_s , but it is less precise than the saddlepoint approximation, particularly at low rates.

To facilitate the computation of rcu and rcu_s at larger block lengths, we consider the symmetric setup of $\delta_0 = \delta_1 = \delta_2 = \delta = 0.1$ and $Q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Under these parameters, we have $I(X; Y) = 0.633$ and $R_{\text{cr}}(Q) = 0.192$ bits/use. In Figure 3.2, we plot the rate required for each random-coding bound and approximation to achieve a given error probability $\epsilon = 10^{-8}$, as a function of n . Once again, $\widehat{\text{rcu}}_s$ is indistinguishable from rcu_s , and similarly for $\widehat{\text{rcu}}_s^*$ and rcu . The error exponent approximation yields similar behavior to rcu_s at short block lengths, but the gap widens at larger block lengths. The exact asymptotics approximations are accurate other than a divergence near the critical rate, which is to be expected from the discussion following (3.24). In contrast to similar plots with larger target error probabilities (e.g. [61, Fig. 8]), the normal approximation is inaccurate over a wide range of rates.

Further examples demonstrating the accuracy of the saddlepoint approximation can be found in [92, 94].

¹The plots remained the same when this value was increased or decreased by an order of magnitude.

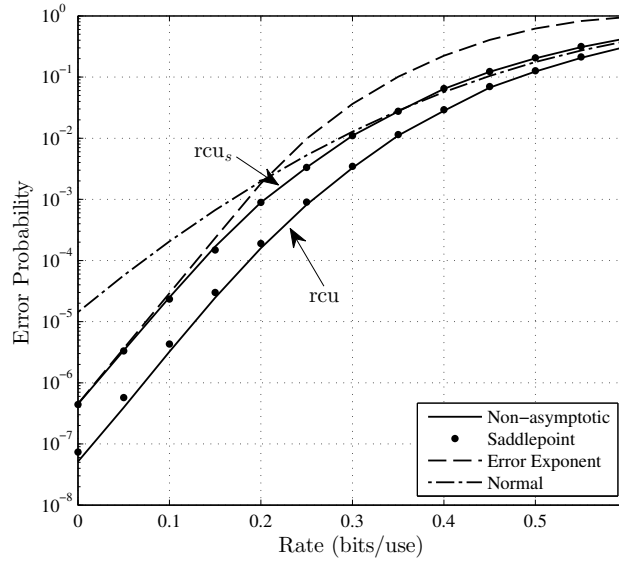


Figure 3.1 i.i.d. random-coding bounds for the channel defined in (2.132) with minimum Hamming distance decoding. The parameters are $n = 60$, $\delta_0 = 0.01$, $\delta_1 = 0.05$, $\delta_2 = 0.25$ and $Q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

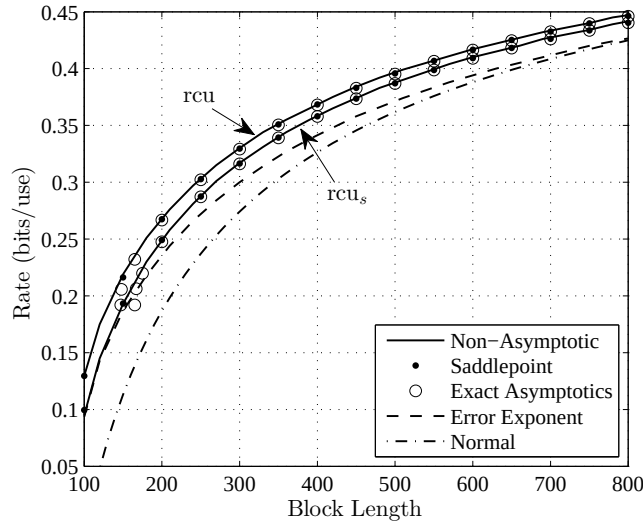


Figure 3.2 Rate required to achieve a target error probability ϵ for the channel defined in (2.132) with ML decoding. The parameters are $\epsilon = 10^{-8}$, $\delta_0 = \delta_1 = \delta_2 = \delta = 0.1$ and $Q = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

3.3 Prefactor to the i.i.d. Expurgated Exponent

In this section, we obtain a refined asymptotic bound of the form $p_e(\mathcal{C}_n) \leq \alpha(n, R)e^{-nE_x^{\text{iid}}(R)}(1 + o(1))$, where $E_x^{\text{iid}}(R)$ is the expurgated exponent given in (2.122). We will see that, under some technical conditions, the prefactor to rcux_ρ in (2.13) behaves as $O(\frac{1}{\sqrt{n}})$, thus improving on Gallager's $O(1)$ prefactor. The analysis builds heavily upon that of Section 3.2, and we thus provide a more concise treatment.

3.3.1 Preliminary Definitions

We define the sets

$$\mathcal{Y}'_1(x, \bar{x}) \triangleq \left\{ y : W(y|x)W(y|\bar{x}) > 0 \right\} \quad (3.48)$$

$$\mathcal{A}(Q) \triangleq \left\{ (x, \bar{x}) : Q(x)Q(\bar{x}) > 0, \frac{q(\bar{x}, y)}{q(x, y)} \neq \frac{q(\bar{x}, y')}{q(x, y')} \text{ for some } y, y' \in \mathcal{Y}'_1(x, \bar{x}) \right\}, \quad (3.49)$$

and make the following technical assumptions:

$$q(x, y) = 0 \iff W(y|x) = 0 \quad (3.50)$$

$$\mathcal{A}(Q) \neq \emptyset. \quad (3.51)$$

It is easily shown that (3.27) holds whenever (3.51) holds by considering the contrapositive statement. However, (3.27) does not imply (3.51) in general.

We proceed by defining several quantities that are analogous to those appearing in Section 3.2.3. We let

$$E_x^{\text{iid}}(Q, \rho, s) \triangleq -\rho \log \sum_{x, \bar{x}} Q(x)Q(\bar{x}) \left(\sum_y W(y|x) \left(\frac{q(\bar{x}, y)}{q(x, y)} \right)^s \right)^{1/\rho} \quad (3.52)$$

denote the objective in (2.123) with a fixed value of s . We define the tiled distribution

$$\tilde{P}_s(y|x, \bar{x}) \triangleq \frac{W(y|x) \left(\frac{q(\bar{x}, y)}{q(x, y)} \right)^s}{\sum_{y'} W(y'|x) \left(\frac{q(\bar{x}, y')}{q(x, y')} \right)^s} \quad (3.53)$$

$$\tilde{P}_s^n(\mathbf{y}|\mathbf{x}, \bar{\mathbf{x}}) \triangleq \prod_{i=1}^n \tilde{P}_s(y_i|x_i, \bar{x}_i), \quad (3.54)$$

and the generalized information density

$$j_s(x, \bar{x}, y) \triangleq \log \frac{\tilde{P}_s(y|x, \bar{x})}{W(y|x)} \quad (3.55)$$

$$j_s^n(\mathbf{x}, \bar{\mathbf{x}}, \mathbf{y}) \triangleq \sum_{i=1}^n j_s(x_i, \bar{x}_i, y_i). \quad (3.56)$$

Furthermore, we define the joint tilted distribution

$$P_{\rho, s}^*(x, \bar{x}) = \frac{Q(x)Q(\bar{x}) \left(\sum_y W(y|x) \left(\frac{q(\bar{x}, y)}{q(x, y)} \right)^s \right)^{1/\rho}}{\sum_{x', \bar{x}'} Q(x')Q(\bar{x}') \left(\sum_{y'} W(y'|x') \left(\frac{q(\bar{x}', y')}{q(x', y')} \right)^s \right)^{1/\rho}}, \quad (3.57)$$

and the conditional variance

$$c_0(Q, \rho, s) \triangleq \mathbb{E} \left[\text{Var} [j_s(X^*, \bar{X}^*, Y^*) | X^*, \bar{X}^*] \right], \quad (3.58)$$

where $(X^*, \bar{X}^*, Y^*) \sim P_{\rho, s}^*(x, \bar{x}) \tilde{P}_s(y|x, \bar{x})$. The arguments to c_0 will be omitted when their values are understood from the context. Writing $Y_s \sim \tilde{P}_s(\cdot|x, \bar{x})$, the following arguments show that the assumptions in (3.50)–(3.51) imply that $c_0 > 0$ whenever $s > 0$:

$$\text{Var}[j_s(x, \bar{x}, Y_s)] = 0 \iff j_s(x, \bar{x}, y) \text{ is independent of } y \text{ wherever } \tilde{P}_s(y|x, \bar{x}) > 0 \quad (3.59)$$

$$\iff \frac{q(\bar{x}, y)}{q(x, y)} \text{ is independent of } y \text{ wherever } W(y|x)q(\bar{x}, y) > 0 \quad (3.60)$$

$$\iff (x, \bar{x}) \notin \mathcal{A}(Q), \quad (3.61)$$

where (3.60) follows from the definitions of j_s and $\tilde{P}_s$, and (3.61) follows from (3.50) and the definition of $\mathcal{A}(Q)$. Using the assumption in (3.51), it follows that $c_0 > 0$.

Finally, we define the set

$$\mathcal{I}_s \triangleq \bigcup_{(x, \bar{x}, y) : W(y|x) > 0, (x, \bar{x}) \in \mathcal{A}(Q)} \{j_s(x, \bar{x}, y)\}, \quad (3.62)$$

and the constant

$$\psi_s \triangleq \begin{cases} 1 & \mathcal{I}_s \text{ does not lie on a lattice} \\ \frac{\bar{h}}{1 - e^{-\bar{h}}} & \mathcal{I}_s \text{ lies on a lattice with span } \bar{h}. \end{cases} \quad (3.63)$$

3.3.2 Main Result

Theorem 3.3. *Fix the input distribution $Q \in \mathcal{P}(\mathcal{X})$, rate $R > 0$, and parameters $\rho \geq 1$ and $s > 0$. Under the assumptions in (3.50)–(3.51), there exists a sequence of codebooks $\mathcal{C}_n$ with $M \geq e^{nR}$ codewords whose maximal error probability satisfies*

$$p_{e,\max}(\mathcal{C}_n) \leq \frac{4^\rho \psi_s}{\sqrt{2\pi n c_0}} e^{-n(E_{\mathbf{x}}^{\text{iid}}(Q, \rho, s) - \rho R)} (1 + o(1)). \quad (3.64)$$

Proof. See Section 3.5.5. □

We observe that at any rate where the expurgated exponent and random-coding exponent coincide (i.e. $\rho = 1$ in both cases), the growth rate of the prefactor in Theorem 3.3 coincides with that of non-expurgated random coding (cf. (3.45)). Of course, Theorem 3.3 is primarily of interest at low rates, where the expurgated exponent exceeds the random-coding exponent.

3.4 Constant-Composition Random Coding

Thus far, we have focused on obtaining refined asymptotics bounds for i.i.d. random coding. There is substantial motivation to obtain similar refinements for the constant-composition ensemble, since it yields improved achievable rates and error exponents when the decoding rule is suboptimal (cf. Chapter 2). Moreover, even in the case of matched decoding, constant-composition codes are of interest for channels with input constraints, as well as multi-user settings [66, 97]. In this section, we study the fixed-error third-order term and the fixed-rate asymptotic prefactor for constant-composition random coding.

3.4.1 Setup and Definitions

Similarly to Sections 3.2.3 and 3.3, we assume throughout this section that

$$q(x, y) > 0 \iff W(y|x) > 0. \quad (3.65)$$

We use constant-composition random coding with an input distribution Q :

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{|T^n(Q_n)|} \mathbb{1}\{\mathbf{x} \in T^n(Q_n)\}, \quad (3.66)$$

where Q_n is a type with the same support as Q such that $\|Q - Q_n\|_\infty \leq \frac{1}{n}$. To simplify the statements of the definitions and results, we assume that $Q(x) > 0$ for all x . This is without

loss of generality for the results we provide, since in the general case the remaining inputs can be removed. We similarly assume that all outputs are reachable.

As in Section 2.5, we define the generalized information density

$$i_{s,a}(x, y) \triangleq \log \frac{q(x, y)^s e^{a(x)}}{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{a(\bar{x})}}, \quad (3.67)$$

and its n -letter extension

$$i_{s,a}^n(\mathbf{x}, \mathbf{y}) \triangleq \sum_{i=1}^n i_{s,a}(x_i, y_i). \quad (3.68)$$

The mean and conditional variance of $i_{s,a}$ are denoted by

$$I_{s,a}(Q) \triangleq \mathbb{E}[i_{s,a}(X, Y)] \quad (3.69)$$

$$V_{s,a}(Q) \triangleq \mathbb{E}[\text{Var}[i_{s,a}(X, Y) | X]], \quad (3.70)$$

where $(X, Y) \sim Q \times W$. The relevant E_0 function is given in (2.40), and can be written as

$$E_0^{\text{cc}}(Q, \rho) = \sup_{s \geq 0, a(\cdot)} - \sum_x Q(x) \log \sum_y W(y|x) e^{-\rho i_{s,a}(x,y)}. \quad (3.71)$$

It will again be useful to define the objective with fixed values in place of the supremum:

$$E_0^{\text{cc}}(Q, \rho, s, a) \triangleq \sum_x Q(x) E_{0,x}^{\text{cc}}(Q, \rho, s, a), \quad (3.72)$$

where

$$E_{0,x}^{\text{cc}}(Q, \rho, s, a) \triangleq - \log \sum_y W(y|x) e^{-\rho i_{s,a}(x,y)}. \quad (3.73)$$

The optimal value of ρ and the critical rate are given by

$$\hat{\rho}(Q, R, s, a) \triangleq \arg \max_{\rho \in [0,1]} E_0^{\text{cc}}(Q, \rho, s, a) - \rho R \quad (3.74)$$

$$R_{s,a}^{\text{cr}}(Q) \triangleq \sup \{R : \hat{\rho}(Q, R, s, a) = 1\}. \quad (3.75)$$

The arguments to $\hat{\rho}$ will be omitted when their values are understood from the context.

For a given channel W and decoding metric q , we define $M_Q^*(n, \epsilon)$ to be the maximum number of messages for a given block length n and target error probability ϵ for constant-composition codebooks with type Q_n , and we define $M^*(n, \epsilon)$ similarly for arbitrary codebooks (not necessarily constant-composition). Similarly, we let $p_{e,Q}^*(n, M)$ and $p_e^*(n, M)$ de-

note the smallest error probability for a given block length and number of messages, where the former is restricted to constant-composition codebooks with type Q_n .

3.4.2 Main Results

Further Definitions

We define the following sets of decoding metrics $q(x, y)$ on $\mathcal{X} \times \mathcal{Y}$:

$$\mathcal{Q}_0^{\text{cc}} \triangleq \left\{ q : \text{There exist functions } g(x), h(y) \text{ s.t. } q(x, y) = g(x)h(y) \text{ wherever } q(x, y) > 0 \right\} \quad (3.76)$$

$$\mathcal{Q}_0 \triangleq \left\{ q : \text{There exists a function } h(y) \text{ s.t. } q(x, y) = h(y) \text{ wherever } q(x, y) > 0 \right\}. \quad (3.77)$$

Moreover, we define the sets of channels $\mathcal{W}_0^{\text{cc}}$ and $\mathcal{W}_0$ in the same way, with the added condition that each element is a conditional distribution $W(y|x)$. We clearly have $\mathcal{Q}_0 \subseteq \mathcal{Q}_0^{\text{cc}}$ and $\mathcal{W}_0 \subseteq \mathcal{W}_0^{\text{cc}}$. Observe that the condition $q \notin \mathcal{Q}_0$ coincides with the condition $\mathcal{Y}_1(Q) \neq \emptyset$ in (3.27) (recall that we have assumed $Q(x) > 0$ for all x).

It was shown by Csiszár and Narayan [8] that, if the codewords have a fixed composition, then any $q \in \mathcal{Q}_0^{\text{cc}}$ is equivalent to the erasures-only metric $q(x, y) = \mathbb{1}\{W(y|x) > 0\}$. Thus, in the case of ML decoding (i.e. $q = W$), the condition $W \notin \mathcal{W}_0^{\text{cc}}$ is a natural generalization of the *feasibility decoding is suboptimal* (FDIS) condition from [96], namely $W \notin \mathcal{W}_0$. Due to the additional structure of constant-composition codes, there is a wider class of metrics that are equivalent to the erasures-only metric, and a wider class of channels for which feasibility decoding is optimal.

An alternative formulation of the condition $q \in \mathcal{Q}_0^{\text{cc}}$ is given in Lemma B.1 in Appendix B. This formulation shows that a sufficient (but not necessary) condition for $q \notin \mathcal{Q}_0^{\text{cc}}$ is that (3.51) holds. Conversely, combining (3.65) with [8, Thm. 1], we have $q \in \mathcal{Q}_0^{\text{cc}}$ whenever the channel graph is acyclic (e.g. the binary erasure channel).

Our results also vary depending on whether W is a sum channel [1] according to the following definition.

Definition 3.1. A discrete memoryless channel W is said to be a sum of N subchannels $W_1, \dots, W_N$ if the alphabets $\mathcal{X}$ and $\mathcal{Y}$ can be partitioned into disjoint non-empty subsets $\mathcal{X}_1, \dots, \mathcal{X}_N$ and $\mathcal{Y}_1, \dots, \mathcal{Y}_N$ such that

$$W(y|x) = \begin{cases} W_i(y|x) & x \in \mathcal{X}_i \text{ and } y \in \mathcal{Y}_i \ (i \in \{1, \dots, N\}) \\ 0 & \text{otherwise,} \end{cases} \quad (3.78)$$

and the same is not true for any larger value of N . We say that W is a sum channel if it is a sum of $N \geq 2$ subchannels.

Fixed-Error Asymptotics

Previous works on the fixed-error asymptotics of constant-composition codes in the matched setting include those of Hayashi [62], Kostina-Verdú [98], and Moulin [87]. Each of these works provides a dispersion term of the form $V = \mathbb{E}[\text{Var}[i(X, Y) | X]]$, where $i(\cdot, \cdot)$ is defined in (1.29). The third-order term is $o(\sqrt{n})$ in [62], $O(\log n)$ in [98], and $\frac{1}{2} \log n$ under various technical conditions in [87].

The following theorem provides a $\frac{1}{2} \log n$ third-order term under various technical conditions using different techniques to those of [86], as well as considering the mismatched setting and providing refined results for the case that the technical conditions fail.

Theorem 3.4. *Fix an input distribution $Q \in \mathcal{P}(\mathcal{X})$ such that $Q(x) > 0$ for all x , and the parameters $s > 0$ and $a(\cdot)$. If (3.65) holds and W is the sum of N subchannels, then we have for any $\epsilon \in (0, 1)$ that*

$$\log M_Q^*(n, \epsilon) \geq nI_{s,a}(Q) - \sqrt{nV_{s,a}(Q)} Q^{-1}(\epsilon) + \frac{1}{2}(\mathbb{1}\{Q \notin \mathcal{Q}_0^{\text{cc}}\} - (N-1)) \log n + O(1). \quad (3.79)$$

Proof. See Section 3.5.6. □

In the special case that W is not a sum channel (i.e. $N = 1$), the coefficient to $\log n$ in (3.79) is $\frac{1}{2}$ when $q \notin \mathcal{Q}_0^{\text{cc}}$, and 0 when $q \in \mathcal{Q}_0^{\text{cc}}$. In the case of ML decoding (i.e. $q = W$), a matching converse for the case $W \notin \mathcal{W}_0^{\text{cc}}$ was given in [61, Thm. 48], and a matching converse for the case $W \in \mathcal{W}_0$ with an optimized input distribution was given in [88, Prop. 2]. However, it is unclear whether our result is tight for ML decoding with $W \in \mathcal{W}_0^{\text{cc}} \setminus \mathcal{W}_0$. It is also unclear whether our result is tight for sum channels, but there are cases where the answer is affirmative. Before stating this formally, we present the following general theorem relating $M_Q^*(n, \epsilon)$ and $M^*(n, \epsilon)$ (cf. Section 3.4.1).

Theorem 3.5. *Fix an input distribution Q such that $Q(x) > 0$ for all x . If (3.65) holds and W is the sum of N subchannels, each of which has an identical transition law, then*

$$\log M_Q^*(n, \epsilon) \leq \log M^*(n, \epsilon) - \frac{1}{2}(N-1) \log n + O(1). \quad (3.80)$$

Proof. See Section 3.5.8 □

In the special case of ML decoding, we can combine Theorem 3.5 with existing asymptotic upper bounds on $M^*(n, \epsilon)$ to obtain more concrete asymptotic expansions of $M_Q^*(n, \epsilon)$. To this end, we let $\mathcal{W}^{\text{symm}}$ be the set of channels that are symmetric in the sense of Gallager, i.e. the output alphabet $\mathcal{Y}$ can be partitioned into subsets such that the resulting submatrices of the transition matrix $\{W(y|x)\}$ contain rows that are permutations of each other, and columns that are permutations of each other. Furthermore, we define $\mathcal{W}_0^{\text{symm}} \triangleq \mathcal{W}^{\text{symm}} \cap \mathcal{W}_0$, and we let C and V denote the channel capacity and dispersion, as given in (1.5) and (1.30).

Corollary 1. *Let $q(x, y) = W(y|x)$, and suppose that W is the sum of N subchannels, each of which has an identical transition law. Fix $\epsilon \in (0, 1)$, and suppose that $V > 0$. For any input distribution Q such that $Q(x) > 0$ for all x , we have*

$$\log M_Q^*(n, \epsilon) \leq nC - \sqrt{nV} Q^{-1}(\epsilon) + \frac{1}{2}(\mathbb{1}\{W \notin \mathcal{W}_0^{\text{symm}}\} - (N - 1)) \log n + O(1). \quad (3.81)$$

Proof. This follows by combining (3.80) with the upper bounds on $M^*(n, \epsilon)$ given in [88] ($W \in \mathcal{W}_0^{\text{symm}}$) and [89] ($W \notin \mathcal{W}_0^{\text{symm}}$). The assumption $V > 0$ suffices to obtain a coefficient to $\log n$ of 0 and $\frac{1}{2}$ in [88] and [89] respectively. $\square$

Consider the expansion in (3.79) with $q = W$, $s = 1$, $a(\cdot) = 0$, and Q chosen to achieve the capacity and dispersion. Provided that $Q(x) > 0$ for all x , we see that the expansion coincides with (3.81) whenever the two indicator functions yield the same value. Examples include W equal to the sum of N identical binary symmetric channels (BSCs), or N identical binary erasure channels (BECs). To our knowledge, these are the first known cases in which i.i.d. random coding provably outperforms constant-composition random coding; the former yields the optimal third-order expansion of $M^*(n, \epsilon)$ without the $\frac{1}{2}(N - 1) \log n$ term [61, 88].

Fixed-Rate Asymptotics

To our knowledge, no previous works have attempted to accurately characterize the fixed-rate prefactor for constant-composition random coding. Here we provide refined bounds that improve significantly on standard prefactors containing polynomials depending on the alphabet sizes (e.g. see [27, Ch. 10]), and that will be seen to be tight in some cases.

Theorem 3.6. *Fix an input distribution Q such that $Q(x) > 0$ for all x , the parameters $s > 0$ and $a(\cdot)$, and the rate $R \in (0, I_{s,a}(Q))$. If (3.65) holds and W is the sum of N subchannels, then the following holds for sufficiently large n under the codeword distribution in (3.66):*

$$p_{e,Q}^*(n, \lfloor e^{nR} \rfloor) \leq \alpha_n(Q, R, s, a) e^{-n(E_0^{\text{cc}}(Q, \hat{p}, s, a) - \hat{p}R)}, \quad (3.82)$$

where

$$\alpha_n(Q, R, s, a) = \begin{cases} \frac{K_0}{n^{\frac{1}{2}(1+\tau\hat{\rho}(Q, R, s, a))}} & R \in (R_{s,a}^{\text{cr}}(Q), I_{s,a}(Q)) \\ \frac{K_0}{n^{\frac{1}{2}\tau}}, & R \in (0, R_{s,a}^{\text{cr}}(Q)] \end{cases} \quad (3.83)$$

for some constant K_0 depending on W , q , Q , s , $a(\cdot)$ and R , and where

$$\tau \triangleq \mathbb{1}\{q \notin \mathcal{Q}_0^{\text{cc}}\} - (N - 1). \quad (3.84)$$

Proof. See Section 3.5.6. □

In the case that W is not a sum channel (i.e. $N = 1$) and $q \notin \mathcal{Q}_0^{\text{cc}}$ (respectively, $q \in \mathcal{Q}_0^{\text{cc}}$), Theorem 3.6 yields a prefactor of $O(\frac{1}{n^{\frac{1}{2}(1+\hat{\rho})}})$ (respectively, $O(\frac{1}{\sqrt{n}})$) for rates above the critical rate, and a prefactor of $O(\frac{1}{\sqrt{n}})$ (respectively, $O(1)$) for rates below the critical rate. This is analogous to the behavior of i.i.d. random coding [81, 91] (see also (3.21)–(3.24) and (3.45)–(3.47)). A matching converse is given in [82] for the case that $q = W$, $W \notin \mathcal{W}_0^{\text{cc}}$, the input distribution Q is optimized, and the rate is above the critical rate.

In order to better understand the case that W is a sum channel, we proceed by presenting an analog of Corollary 1. We focus on symmetric channels, since the behavior of the fixed-rate prefactor is currently less well-understood than that of the third-order term in the fixed-error asymptotics, and proving a result in the same generality as that of Theorem 3.5 appears to be more difficult. We let $\hat{\rho}(R)$, $E_0^{\text{cc}}(\rho)$ and R_{cr} respectively denote the values of $\hat{\rho}(Q, R, s, a)$, $E_0^{\text{cc}}(Q, \rho, s, a)$ and $R_{s,a}^{\text{cr}}(Q)$ with $s = \frac{1}{1+\rho}$, $a(\cdot) = 0$, and Q uniform on $\mathcal{X}$ (i.e. the parameters maximizing the exponent under ML decoding [2, Sec. 5.6]). The channel capacity is denoted by C , and we use the definition of $\mathcal{W}^{\text{symm}}$ following Theorem 3.5.

Theorem 3.7. *Let $q(x, y) = W(y|x)$, and suppose that $W \in \mathcal{W}^{\text{symm}}$, and that W is the sum of N subchannels, each of which has an identical transition law. Let Q be the uniform distribution on $\mathcal{X}$. For any rate $R \in (R_{\text{cr}}, C)$, we have for sufficiently large n that*

$$p_{\text{e},Q}^*(n, \lfloor e^{nR} \rfloor) \geq \frac{K'_0}{n^{\frac{1}{2}(1+\tau'\hat{\rho})}} e^{-n(E_0^{\text{cc}}(\hat{\rho}) - \hat{\rho}R)} \quad (3.85)$$

for some constant $K'_0 > 0$ depending on W and R , where

$$\tau' \triangleq \mathbb{1}\{W \notin \mathcal{W}_0^{\text{symm}}\} - (N - 1). \quad (3.86)$$

Proof. See Section 3.5.8. □

Theorem 3.7 provides a matching converse to Theorem 3.6 (with $q = W$, $s = \frac{1}{1+\rho}$, $a(\cdot) = 0$,

and Q uniform on $\mathcal{X}$) whenever the indicator functions in (3.84) and (3.86) coincide. Analogously to the discussion following Corollary 1, we conclude that i.i.d. random coding can outperform constant-composition random coding when $N > 1$ (see (3.21)–(3.24) and (3.45)–(3.47)); notable examples again include sum BSCs and sum BECs.

3.5 Proofs

3.5.1 Properties of $E_0^{\text{iid}}(Q, \rho, s)$

Here we prove the properties of $E_0^{\text{iid}}(Q, \rho, s)$ given following (3.10), by following the arguments of Gallager [2, pp. 141–143].

The continuity of $\hat{\rho}$ in R follows from the continuity of E_0^{iid} in ρ . A direct differentiation reveals that the first two partial derivatives of E_0^{iid} are also continuous in ρ , and that $\frac{\partial E_0^{\text{iid}}}{\partial \rho} \big|_{\rho=0} = I_s(Q)$ and $-\frac{\partial^2 E_0^{\text{iid}}}{\partial \rho^2} \big|_{\rho=0} = U_s(Q)$.

A function is said to be log-convex if its logarithm is convex. Since the logarithm of $e^{\rho i_s(x,y)}$ is linear in ρ and a sum of log-convex functions is log-convex [38, Sec. 3.5], we conclude from (3.6) that $E_0^{\text{iid}}(Q, \rho, s)$ is concave in ρ . Moreover, the analysis of [38, Sec. 3.5] reveals that the concavity is strict (and hence $c_2 > 0$) unless $i_s(X, Y)$ is deterministic, or equivalently $U_s(Q) = 0$.

The fact that $R_s^{\text{cr}}(Q) = I_s(Q)$ when $U_s(Q) = 0$ follows since in this case we have from (3.6) that $E_0^{\text{iid}} = \rho I_s(Q)$, and hence the maximum of $E_0^{\text{iid}} - \rho R$ over $\rho \in [0, 1]$ is achieved at one of the endpoints. More specifically, we have $\hat{\rho} = 1$ if $R < I_s(Q)$, and $\hat{\rho} = 0$ if $R > I_s(Q)$.

Suppose now that $U_s(Q) > 0$. The *global* minimizer ρ^* of $E_0^{\text{iid}} - \rho R$ is the unique value such that $R = \frac{\partial E_0^{\text{iid}}}{\partial \rho}$. Clearly, if $\rho^* \in [0, 1]$ then $\hat{\rho} = \rho^*$ and $c_1 = 0$. Since $\frac{\partial E_0^{\text{iid}}}{\partial \rho}$ is continuous and strictly decreasing in ρ (the latter of which follows from the strict concavity established above), this occurs if and only if $\frac{\partial E_0^{\text{iid}}}{\partial \rho} \big|_{\rho=1} \leq R \leq \frac{\partial E_0^{\text{iid}}}{\partial \rho} \big|_{\rho=0}$. It is then seen from the definition of $\hat{\rho}$ that $\hat{\rho} = 1$ for rates below the lower limit and $\hat{\rho} = 0$ for rates above the upper limit, and these cases readily yield $c_1 < 0$ and $c_1 > 0$ respectively.

3.5.2 Asymptotic Behavior of the Saddlepoint Approximation

Here we prove the asymptotic relations given in (3.21)–(3.24). We will make frequent use of the properties of $\hat{\rho}$, c_1 and c_2 given in Section 3.2.1.

We first prove (3.21)–(3.22) in the non-lattice case. Suppose that $R < R_s^{\text{cr}}(Q)$, and hence $\hat{\rho} = 1$, $c_1 < 0$ and $c_2 > 0$. Using (3.18) and the identity $Q(z) \leq \frac{1}{2}e^{-\frac{z^2}{2}}$ for $z > 0$, it is easily verified that the first term in (3.14) decays to zero exponentially fast. The second term is given

by $Q(-c_1\sqrt{\frac{n}{c_2}})$, which tends to one since $\lim_{z \rightarrow \infty} Q(z) = 1$. We thus obtain (3.21). When $R = R_s^{\text{cr}}(Q)$, the argument is similar except that $c_1 = 0$, yielding the following: (i) From (3.18), the first term in (3.14) equals $Q(\sqrt{nc_2})e^{nc_2/2} \asymp \frac{1}{\sqrt{2\pi nc_2}}$, rather than decaying exponentially fast; (ii) The second term in (3.14) equals $Q(0) = \frac{1}{2}$, rather than one. For $R > I_s(Q)$ (respectively, $R = I_s(Q)$) the argument is similar with the roles of the two terms in (3.14) reversed, and with $\hat{\rho} = 0$ and $c_1 > 0$ (respectively, $c_1 = 0$).

In the lattice case, the arguments in proving (3.21)–(3.22) are similar to the non-lattice case, so we focus on (3.21) with $R < R_s^{\text{cr}}(Q)$. Similarly to the non-lattice case, it is easily shown that the first summation in (3.16) decays to zero exponentially fast. Since $\hat{\rho} = 1$, the second summation is given by

$$\sum_{i=-\infty}^{-1} \phi_h(\gamma_n + ih; nc_1, nc_2) = (1 + o(1)) \sum_{i=-\infty}^{\infty} \phi_h(\gamma_n + ih; nc_1, nc_2) \quad (3.87)$$

$$= 1 + o(1), \quad (3.88)$$

where (3.87) follows since the added terms from $i = 0$ to ∞ contribute an exponentially small amount to the sum since $c_1 < 0$, and (3.88) is easily understood by interpreting the right-hand side of (3.87) as approximating the integral of a Gaussian density function via discrete sampling. Since the sampling is done using intervals of a fixed size h but the variance nc_2 increases, the approximation improves with n and approaches one.

Finally, we consider the case that $R \in (R_s^{\text{cr}}(Q), I_s(Q))$, and hence $\hat{\rho} \in (0, 1)$, $c_1 = 0$ and $c_2 > 0$. In the non-lattice case, we can substitute $c_1 = 0$ and (3.18) into (3.14) to obtain

$$\alpha_n = e^{\frac{1}{2}nc_2\hat{\rho}^2} Q(\hat{\rho}\sqrt{nc_2}) + e^{\frac{1}{2}nc_2(1-\hat{\rho})^2} Q((1-\hat{\rho})\sqrt{nc_2}). \quad (3.89)$$

Using the fact that $Q(z)e^{z^2/2} \asymp \frac{1}{z\sqrt{2\pi}}$ as $z \rightarrow \infty$, along with the identity $\frac{1}{\rho} + \frac{1}{1-\rho} = \frac{1}{\rho(1-\rho)}$, we obtain (3.23).

We now turn to the lattice case. Setting $c_1 = 0$ in (3.16) yields

$$\alpha_n = \frac{h}{\sqrt{2\pi nc_2}} \left(\sum_{i=0}^{\infty} e^{-\hat{\rho}(\gamma_n + ih) - \frac{(\gamma_n + ih)^2}{2nc_2}} + \sum_{i=-\infty}^{-1} e^{(1-\hat{\rho})(\gamma_n + ih) - \frac{(\gamma_n + ih)^2}{2nc_2}} \right). \quad (3.90)$$

The two summations are handled in a nearly identical fashion, so we focus on the first. Using

the identity $1 - x \leq e^{-x} \leq 1$, we can write

$$\left| \sum_{i=0}^{\infty} e^{-\hat{\rho}(\gamma_n + ih) - \frac{(\gamma_n + ih)^2}{2nc_2}} - \sum_{i=0}^{\infty} e^{-\hat{\rho}(\gamma_n + ih)} \right| \leq \sum_{i=0}^{\infty} e^{-\hat{\rho}(\gamma_n + ih)} \frac{(\gamma_n + ih)^2}{2nc_2} \quad (3.91)$$

$$= O\left(\frac{1}{n}\right), \quad (3.92)$$

where (3.92) follows since the summation $\sum_{i=0}^{\infty} e^{-\zeta i} p(i)$ is convergent for any polynomial $p(i)$ and constant $\zeta > 0$. Furthermore, we have from the geometric series that

$$\sum_{i=0}^{\infty} e^{-\hat{\rho}(\gamma_n + ih)} = e^{-\hat{\rho}\gamma_n} \left(\frac{1}{1 - e^{-\hat{\rho}h}} \right), \quad (3.93)$$

We have thus weakened the first summation in (3.90) to the first term in the sum in (3.24) (up to an $O(\frac{1}{n})$ remainder term). The second term is handled in a similar fashion.

3.5.3 Proof of Theorem 3.1

Since $M_n \rightarrow \infty$ by assumption, we have $\frac{M_n - 1}{M_n} \rightarrow 1$, and we can consider $\text{rcu}_s(n, M + 1)$ in place of $\text{rcu}_s(n, M)$ without affecting the theorem statement. We begin by considering fixed values of n , M , and $R = \frac{1}{n} \log M$.

Alternative Expressions for rcu_s

Let $F(t)$ denote the cumulative distribution function (CDF) of $R - i_s(X, Y)$ with $(X, Y) \sim Q \times W$, and let $Z_1, \dots, Z_n$ be i.i.d. according to the tilted CDF

$$F_Z(z) = e^{E_0^{\text{iid}}(Q, \hat{\rho}, s) - \hat{\rho}R} \int_{-\infty}^z e^{\hat{\rho}t} dF(t). \quad (3.94)$$

It is seen that this is indeed a CDF by writing

$$\int_{-\infty}^{\infty} e^{\hat{\rho}t} dF(t) = \mathbb{E}[e^{\hat{\rho}(R - i_s(X, Y))}] = e^{-(E_0^{\text{iid}}(Q, \hat{\rho}, s) - \hat{\rho}R)}, \quad (3.95)$$

where we have used (3.6). The moment generating function (MGF) of Z is given by

$$M_Z(\tau) \triangleq \mathbb{E}[e^{\tau Z}] \quad (3.96)$$

$$= e^{E_0^{\text{iid}}(Q, \hat{\rho}, s) - \hat{\rho}R} \mathbb{E}[e^{(\hat{\rho} + \tau)(R - i_s(X, Y))}] \quad (3.97)$$

$$= e^{E_0^{\text{iid}}(Q, \hat{\rho}, s)} e^{-(E_0^{\text{iid}}(Q, \hat{\rho} + \tau, s) - \tau R)}, \quad (3.98)$$

where (3.97) follows from (3.94), and (3.98) follows from (3.6). We can now compute the mean and variance of Z in terms of the derivatives of the MGF, namely

$$\mathbb{E}[Z] = \frac{dM_Z}{d\tau} \Big|_{\tau=0} = c_1 \quad (3.99)$$

$$\text{Var}[Z] = \frac{d^2 M_Z}{d\tau^2} \Big|_{\tau=0} - \mathbb{E}[Z]^2 = c_2, \quad (3.100)$$

where c_1 and c_2 are defined in (3.9)–(3.10). Recall that $U_s(Q) > 0$ by assumption, which implies that $c_2 > 0$ (see Section 3.2.1).

In the remainder of the proof, we omit the arguments $(Q, \hat{\rho}, s)$ to E_0^{iid} . Following [95, Lemma 2], we can use (3.94) to write (3.5) as follows:

$$\text{rcu}_s(n, M+1) = \int \dots \int_{\mathbb{R}^n} \min \{1, e^{\sum_{i=1}^n t_i}\} dF(t_1) \dots dF(t_n) \quad (3.101)$$

$$= e^{-n(E_0^{\text{iid}} - \hat{\rho}R)} \int \dots \int_{\mathbb{R}^n} e^{-\hat{\rho} \sum_{i=1}^n z_i} \min \{1, e^{\sum_{i=1}^n z_i}\} dF_Z(z_1) \dots dF_Z(z_n) \quad (3.102)$$

$$\triangleq I_n e^{-n(E_0^{\text{iid}} - \hat{\rho}R)}. \quad (3.103)$$

Defining F_n to be the CDF of $\sum_{i=1}^n Z_i$, the prefactor I_n can be written as

$$I_n = \int_{-\infty}^{\infty} e^{-\hat{\rho}z} \min\{1, e^z\} dF_n(z) \quad (3.104)$$

$$= \int_0^{\infty} e^{-\hat{\rho}z} dF_n(z) + \int_{-\infty}^0 e^{(1-\hat{\rho})z} dF_n(z), \quad (3.105)$$

where (3.105) follows by splitting the integral according to which value achieves the $\min\{\cdot, \cdot\}$ in (3.104). Letting $\hat{F}_n$ denote the CDF of $\frac{\sum_{i=1}^n Z_i - nc_1}{\sqrt{nc_2}}$, we can write (3.105) as

$$I_n = \int_{-\frac{c_1\sqrt{n}}{\sqrt{c_2}}}^{\infty} e^{-\hat{\rho}(z\sqrt{nc_2}+nc_1)} d\hat{F}_n(z) + \int_{-\infty}^{-\frac{c_1\sqrt{n}}{\sqrt{c_2}}} e^{(1-\hat{\rho})(z\sqrt{nc_2}+nc_1)} d\hat{F}_n(z). \quad (3.106)$$

Non-Lattice Case

Let $\Phi(z)$ denote the CDF of a zero-mean unit-variance Gaussian random variable. In the case that $R - i_s(X, Y)$ is non-lattice, the same is true of Z , since the two have the same support. Hence, and using the fact that $\mathbb{E}[Z] = c_1$ and $\text{Var}[Z] = c_2 > 0$ (see (3.99)–(3.100)), we have from the Edgeworth expansion in [76, Sec. XVI.4, Thm. 1] that

$$\hat{F}_n(z) = \Phi(z) + G_n(z) + \tilde{F}_n(z), \quad (3.107)$$

where $\tilde{F}_n(z) = o(n^{-\frac{1}{2}})$ uniformly in z , and

$$G_n(z) \triangleq \frac{K}{\sqrt{n}}(1 - z^2)e^{-\frac{1}{2}z^2} \quad (3.108)$$

for some constant K depending only on the variance and third absolute moment of Z , the latter of which is finite since we are considering finite alphabets. Substituting (3.107) into (3.106), we obtain

$$I_n = I_{1,n} + I_{2,n} + I_{3,n}, \quad (3.109)$$

where the three terms denote the right-hand side of (3.106) with Φ , G_n and $\tilde{F}_n$ respectively in place of $\hat{F}_n$. By reversing the step from (3.105) to (3.106), we see that $I_{1,n}$ is precisely α_n in (3.14). Furthermore, using $\frac{dG_n}{dz} = \frac{K}{\sqrt{n}}(z^3 - 3z)e^{-\frac{1}{2}z^2}$, we obtain

$$\begin{aligned} I_{2,n} = \frac{K}{\sqrt{n}} & \left(\int_{-\frac{c_1\sqrt{n}}{\sqrt{c_2}}}^{\infty} e^{-\hat{\rho}(z\sqrt{nc_2}+nc_1)}(z^3 - 3z)e^{-\frac{1}{2}z^2} dz \right. \\ & \left. + \int_{-\infty}^{-\frac{c_1\sqrt{n}}{\sqrt{c_2}}} e^{(1-\hat{\rho})(z\sqrt{nc_2}+nc_1)}(z^3 - 3z)e^{-\frac{1}{2}z^2} dz \right). \end{aligned} \quad (3.110)$$

In accordance with the theorem statement, we must show that $I_{2,n} = o(\alpha_n)$ and $I_{3,n} = o(\alpha_n)$ even when R and $\hat{\rho}$ vary with n . Let $R_n \triangleq \frac{1}{n} \log M_n$, $\hat{\rho}_n \triangleq \hat{\rho}(Q, R_n, s)$, $c_{1,n} \triangleq c_1(Q, R_n, s)$, and $c_{2,n} \triangleq c_2(Q, R_n, s)$. We assume with no real loss of generality that

$$\lim_{n \rightarrow \infty} R_n = R^* \quad (3.111)$$

for some $R^* \geq 0$ possibly equal to ∞ . Once the theorem is proved for all such R^* , the same will follow for R_n varying with n .

Table 3.1 summarizes the growth rates α_n , $I_{2,n}$ and $I_{3,n}$ for various ranges of R^* , and indicates whether the first or second integral (see (3.106) and (3.110)) dominates the behavior of each. We see that $I_{2,n} = o(\alpha_n)$ and $I_{3,n} = o(\alpha_n)$ for all values of R^* , as desired.

The derivations of the growth rates in Table 3.1 when $R^* \notin \{R_s^{\text{cr}}(Q), I_s(Q)\}$ are done in a similar fashion to Section 3.5.2. To avoid repetition, we provide details only for $R^* = R_s^{\text{cr}}(Q)$; this is a less straightforward case whose analysis differs slightly from the others. From Section 3.2.1, we have $\hat{\rho} \rightarrow 1$ and $c_{1,n} \rightarrow 0$ from below, with $c_{1,n} < 0$ only if $\hat{\rho}_n = 1$.

For any $\hat{\rho} \in [0, 1]$, the terms $e^{-\hat{\rho}(\cdot)}$ and $e^{(1-\hat{\rho})(\cdot)}$ in (3.110) are both upper bounded by one across their respective ranges of integration. Since the moments of a Gaussian random variable are finite, it follows that both integrals are $O(1)$, and thus $I_{2,n} = O(\frac{1}{\sqrt{n}})$. The term $I_{3,n}$ is handled similarly, so it only remains to show that $\alpha_n = \omega(\frac{1}{\sqrt{n}})$. In the case that $\hat{\rho} = 1$, the

Table 3.1 Growth rates of α_n , $I_{2,n}$ and $I_{3,n}$ when the rate converges to R^* .

	$\hat{\rho}$	c_1	Dominant Term(s)	α_n	$I_{2,n}$	$I_{3,n}$
$R^* \in [0, R_s^{\text{cr}}(Q))$	1	< 0	2	$\Theta(1)$	$\Theta\left(\frac{1}{\sqrt{n}}\right)$	$o\left(\frac{1}{\sqrt{n}}\right)$
$R^* = R_s^{\text{cr}}(Q)$	$\rightarrow 1$	$\rightarrow 0$	2	$\omega\left(\frac{1}{\sqrt{n}}\right)$	$O\left(\frac{1}{\sqrt{n}}\right)$	$o\left(\frac{1}{\sqrt{n}}\right)$
$R^* \in (R_s^{\text{cr}}(Q), I_s(Q))$	$\in (0, 1)$	0	1,2	$\Theta\left(\frac{1}{\sqrt{n}}\right)$	$\Theta\left(\frac{1}{n\sqrt{n}}\right)$	$o\left(\frac{1}{\sqrt{n}}\right)$
$R^* = I_s(Q)$	$\rightarrow 0$	$\rightarrow 0$	1	$\omega\left(\frac{1}{\sqrt{n}}\right)$	$O\left(\frac{1}{\sqrt{n}}\right)$	$o\left(\frac{1}{\sqrt{n}}\right)$
$R^* > I_s(Q)$	0	> 0	1	$\Theta(1)$	$\Theta\left(\frac{1}{\sqrt{n}}\right)$	$o\left(\frac{1}{\sqrt{n}}\right)$

second integral in (3.14) is at least $\frac{1}{2}$, since $c_1 \leq 0$. Considering now the case that $c_1 = 0$ and $\hat{\rho}_n \rightarrow 1$ with $\hat{\rho}_n < 1$, we observe that for any $\delta > 0$, we have $\hat{\rho}_n \geq 1 - \delta$ for sufficiently large n . Lower bounding α_n by replacing $1 - \hat{\rho}$ by δ in the second term of (3.14), we have similarly to (3.89) that

$$\alpha_n \geq e^{\frac{1}{2}nc_2\delta^2} Q(\delta\sqrt{nc_2}) \asymp \frac{1}{\sqrt{2\pi nc_2}\delta}. \quad (3.112)$$

Since δ can be arbitrarily small, we obtain $\alpha_n = \omega\left(\frac{1}{\sqrt{n}}\right)$, as desired.

Lattice case

The arguments following (3.109) are similar in the lattice case, so we focus our attention on obtaining the analogous expression to (3.109). Letting $P_n(z)$ denote the probability mass function (PMF) of $\sum_{i=1}^n Z_i$, we can write (3.105) as

$$I_n = \sum_{z \geq 0} P_n(z) e^{-\hat{\rho}z} + \sum_{z < 0} P_n(z) e^{(1-\hat{\rho})z}. \quad (3.113)$$

If $R - i_s(X, Y)$ has span h , then the same is true of Z . Hence, and using the fact that $\mathbb{E}[Z] = c_1$ and $\text{Var}[Z] = c_2 > 0$, we have from the local limit theorem in [2, Eq. (5A.12)] that

$$P_n(z) = \phi_h(z; nc_1, nc_2) + \tilde{P}_n(z), \quad (3.114)$$

where ϕ_h is defined in (3.12), and $\tilde{P}_n(z) = o\left(\frac{1}{\sqrt{n}}\right)$ uniformly in z . Thus, analogously to (3.109), we can write

$$I_n = I_{1,n} + I_{2,n}, \quad (3.115)$$

where the two terms denote the right-hand side of (3.113) with ϕ_h and $\tilde{P}_n$ respectively in place of P_n . Using the definition of γ_n in (3.15) and the fact that $\sum_i Z_i$ has the same support as $nR - i_s^n(\mathbf{X}, \mathbf{Y})$ (cf. (3.94)), we see that the first summation in (3.113) is over the set $\{\gamma_n + ih :$

$i \in \mathbb{Z}, i \geq 0\}$, and the second summation is over the set $\{\gamma_n + ih : i \in \mathbb{Z}, i < 0\}$. It follows that $I_{1,n} = \alpha_n$, and similar arguments to the non-lattice case show that $I_{2,n} = o(\alpha_n)$.

3.5.4 Proof of Theorem 3.2

Throughout this subsection, we make use of the same notation as Section 3.5.3. We first discuss the proof of (3.44). Using the definition of rcu_s^* , we can follow identical arguments to those following (3.101) to conclude that

$$\text{rcu}_s^*(n, M) = I_n e^{-n(E_0^{\text{iid}}(Q, \hat{\rho}, s) - \hat{\rho}R)}, \quad (3.116)$$

where analogously to (3.105), we have

$$I_n = \int_{\log \frac{\sqrt{2\pi n c_3}}{\psi_s}}^{\infty} e^{-\hat{\rho}z} dF_n(z) + \frac{\psi_s}{\sqrt{2\pi n c_3}} \int_{-\infty}^{\log \frac{\sqrt{2\pi n c_3}}{\psi_s}} e^{(1-\hat{\rho})z} dF_n(z). \quad (3.117)$$

The remaining arguments in proving (3.44) follow those in Section 3.5.3, and are omitted.

To prove (3.42), we make use of two technical lemmas, whose proofs are postponed until the end of the subsection. The following lemma is a refinement of [61, Lemma 47].

Lemma 3.1. *Fix an integer K , and for each n , let $(n_1, \dots, n_K)$ be integers such that $\sum_j n_j = n$. Fix the PMFs $Q_1, \dots, Q_K$ on a finite subset of $\mathbb{R}$, and let $\sigma_1^2, \dots, \sigma_K^2$ be the corresponding variances. Let $Z_1, \dots, Z_n$ be independent random variables, n_j of which are distributed according to Q_j for each j . Suppose that $\max_j \sigma_j > 0$ and $\min_j n_j = \Theta(n)$. Defining*

$$\mathcal{I}_0 \triangleq \bigcup_{j: \sigma_j > 0} \{z : Q_j(z) > 0\} \quad (3.118)$$

$$\psi_0 \triangleq \begin{cases} 1 & \mathcal{I}_0 \text{ does not lie on a lattice} \\ \frac{h_0}{1-e^{-h_0}} & \mathcal{I}_0 \text{ lies on a lattice with span } h_0, \end{cases} \quad (3.119)$$

the summation $S_n \triangleq \sum_i Z_i$ satisfies the following uniformly in t :

$$\mathbb{E}\left[e^{-S_n} \mathbf{1}\{S_n \geq t\}\right] \leq e^{-t} \left(\frac{\psi_0}{\sqrt{2\pi V_n}} + o\left(\frac{1}{\sqrt{n}}\right) \right), \quad (3.120)$$

where $V_n \triangleq \text{Var}[S_n]$.

The following lemma ensures the existence of a sufficiently high probability set in which Lemma 3.1 can be applied to the inner probability in (2.7). We make use of the definitions in

(3.28)–(3.30), and we define the random variables

$$(\mathbf{X}, \mathbf{Y}, \bar{\mathbf{X}}, \tilde{\mathbf{X}}) \sim Q^n(\mathbf{x})W^n(\mathbf{y}|\mathbf{x})Q^n(\bar{\mathbf{x}})\tilde{P}_s^n(\tilde{\mathbf{x}}|\mathbf{y}), \quad (3.121)$$

where $\tilde{P}_s^n(\mathbf{x}|\mathbf{y}) \triangleq \prod_{i=1}^n \tilde{P}_s(x_i|y_i)$. Furthermore, we write the empirical distribution of $\mathbf{y}$ as $\hat{P}_{\mathbf{y}}$, and we let $P_{\mathbf{Y}}$ denote the PMF of $\mathbf{Y}$.

Lemma 3.2. *Fix the parameter $s > 0$ and the sequence $\{R_n\}$, and let $\hat{\rho} = \hat{\rho}(Q, R_n, s)$ be defined according to (3.7). Under the assumptions in (3.26)–(3.27), the set*

$$\mathcal{F}_{\hat{\rho},s}^n(\delta) \triangleq \left\{ \mathbf{y} : P_{\mathbf{Y}}(\mathbf{y}) > 0, \max_{\mathbf{y}} |\hat{P}_{\mathbf{y}}(\mathbf{y}) - P_{\hat{\rho},s}^*(\mathbf{y})| \leq \delta \right\} \quad (3.122)$$

satisfies the following properties:

1. For any $\mathbf{y} \in \mathcal{F}_{\hat{\rho},s}^n(\delta)$, we have

$$\text{Var}[i_s^n(\tilde{\mathbf{X}}, \mathbf{Y}) | \mathbf{Y} = \mathbf{y}] \geq n(c_3 - r(\delta)), \quad (3.123)$$

where $r(\delta) \rightarrow 0$ as $\delta \rightarrow 0$.

2. For any $\delta > 0$, we have

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log \frac{\sum_{\mathbf{x}, \mathbf{y} \notin \mathcal{F}_{\hat{\rho},s}^n(\delta)} Q^n(\mathbf{x})W^n(\mathbf{y}|\mathbf{x})e^{-\hat{\rho}i_s^n(\mathbf{x}, \mathbf{y})}}{\sum_{\mathbf{x}, \mathbf{y}} Q^n(\mathbf{x})W^n(\mathbf{y}|\mathbf{x})e^{-\hat{\rho}i_s^n(\mathbf{x}, \mathbf{y})}} > 0. \quad (3.124)$$

Before proving the lemmas, we show how they are used to obtain the desired result.

Proof of (3.42) based on Lemmas 3.1–3.2

By upper bounding $M - 1$ by M and splitting rcu (see (2.7)) according to whether or not $\mathbf{y} \in \mathcal{F}_{\hat{\rho},s}^n(\delta)$, we obtain

$$\begin{aligned} \text{rcu}(n, M) &\leq \sum_{\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\hat{\rho},s}^n(\delta)} Q^n(\mathbf{x})W^n(\mathbf{y}|\mathbf{x}) \min \left\{ 1, M\mathbb{P}[i_s^n(\bar{\mathbf{X}}, \mathbf{y}) \geq i_s^n(\mathbf{x}, \mathbf{y})] \right\} \\ &\quad + M^{\hat{\rho}} \sum_{\mathbf{x}, \mathbf{y} \notin \mathcal{F}_{\hat{\rho},s}^n(\delta)} Q^n(\mathbf{x})W^n(\mathbf{y}|\mathbf{x})e^{-\hat{\rho}i_s^n(\mathbf{x}, \mathbf{y})}, \end{aligned} \quad (3.125)$$

where we have replaced q^n by i_s^n since each is a monotonically increasing function of the other, and in the summation over $\mathbf{y} \notin \mathcal{F}_{\hat{\rho},s}^n(\delta)$ we further weakened the bound using Markov's

inequality and $\min\{1, \zeta\} \leq \zeta^{\hat{\rho}}$. In order to make the inner probability in the first term of (3.125) more amenable to an application of Lemma 3.1, we follow [85, Sec. 3.4.5] and note that the following holds whenever $\tilde{P}_s^n(\bar{\mathbf{x}}|\mathbf{y}) > 0$:

$$Q^n(\bar{\mathbf{x}}) = Q^n(\bar{\mathbf{x}}) \frac{\tilde{P}_s^n(\bar{\mathbf{x}}|\mathbf{y})}{\tilde{P}_s^n(\bar{\mathbf{x}}|\mathbf{y})} = \tilde{P}_s^n(\bar{\mathbf{x}}|\mathbf{y}) e^{-i_s^n(\bar{\mathbf{x}}, \mathbf{y})}. \quad (3.126)$$

For fixed $\mathbf{y}$ and t , summing (3.126) over all $\bar{\mathbf{x}}$ such that $i_s^n(\bar{\mathbf{x}}, \mathbf{y}) \geq t$ yields

$$\mathbb{P}[i_s^n(\bar{\mathbf{X}}, \mathbf{y}) \geq t] = \mathbb{E}\left[e^{-i_s^n(\tilde{\mathbf{X}}, \mathbf{Y})} \mathbb{1}\{i_s^n(\tilde{\mathbf{X}}, \mathbf{Y}) \geq t\} \mid \mathbf{Y} = \mathbf{y}\right] \quad (3.127)$$

under the joint distribution in (3.121).

We now observe that (3.127) is of the same form as the left-hand side of (3.120). We apply Lemma 3.1 with Q_j given by the PMF of $i_s(\tilde{X}, y_j)$ under $\tilde{X} \sim \tilde{P}_s(\cdot | y_j)$, where y_j is the j -th output symbol for which $\sum_x Q(x)W(y|x) > 0$. The conditions of the lemma are seen to be satisfied for sufficiently small δ due to the definition of $\mathcal{F}_{\hat{\rho}, s}^n(\delta)$ in (3.122), the assumption in (3.27), and (3.35)–(3.36). We have from (3.120), (3.123) and (3.127) that

$$\mathbb{P}[i_s^n(\bar{\mathbf{X}}, \mathbf{y}) \geq t] \leq \frac{\psi_s}{\sqrt{2\pi n(c_3 - r(\delta))}} e^{-t}(1 + o(1)) \quad (3.128)$$

for all $\mathbf{y} \in \mathcal{F}_{\hat{\rho}, s}^n(\delta)$, provided that δ is sufficiently small so that $r(\delta) < c_3$. Here we have used the fact that ψ_0 in (3.119) coincides with ψ_s in (3.32), which follows from (3.35) and the fact that $\tilde{P}_s(x|y) > 0$ if and only if $Q(x)W(y|x) > 0$ (see (3.26) and (3.28)).

Using the uniformity of the $o(1)$ term in t in (3.128) (see Lemma 3.1), taking $\delta \rightarrow 0$ (and hence $r(\delta) \rightarrow 0$), and using the identity

$$\min\{1, \zeta_1(1 + \zeta_2)\} \leq (1 + |\zeta_2|) \min\{1, \zeta_1\}, \quad (3.129)$$

we see that the first term in (3.125) is upper bounded by $\text{rcu}_s^*(n, M)(1 + o(1))$. To complete the proof of (3.42), we must show that the second term in (3.125) can be incorporated into the multiplicative $1 + o(1)$ term. To see this, we note from (3.37) and (3.44) that the exponent of rcu_s^* is given by $E_0^{\text{iid}}(Q, \hat{\rho}, s) - \hat{\rho}R$. From (3.6), the denominator in the logarithm in (3.124) equals $e^{-nE_0^{\text{iid}}(Q, \hat{\rho}, s)}$. Combining these observations, the second part of Lemma 3.2 shows that the second term in (3.125) decays at a faster exponential rate than rcu_s^* , thus yielding the desired result.

Proof of Lemma 3.1

The proof makes use of the local limit theorems given in [99, Thm. 1] and [100, Sec. VII.1, Thm. 2] for the non-lattice and lattice cases respectively. We first consider the summation $S'_n \triangleq \sum_{i=1}^{n'} Z_i$, where we assume without loss of generality that the first $n' = \Theta(n)$ indices correspond to positive variances, and the remaining $n - n'$ correspond to zero variances. We similarly assume that $\sigma_j > 0$ for $j = 1, \dots, K'$, and $\sigma_j = 0$ for $j = K' + 1, \dots, K$, where $K' \geq 1$. We clearly have $\text{Var}[S'_n] = \text{Var}[S_n] = V_n$.

We first consider the non-lattice case. We claim that the conditions of the lemma imply the following local limit theorem given in [99, Thm. 1]:

$$\mathbb{P}[S'_n \in [z, z + \eta)] = \frac{\eta}{\sqrt{2\pi V_n}} e^{-\frac{(z - \mu'_n)^2}{2V_n}} + o\left(\frac{1}{\sqrt{n}}\right) \quad (3.130)$$

uniformly in z , where $\mu'_n \triangleq \mathbb{E}[S'_n]$, and $\eta > 0$ is arbitrary. To show this, we must verify the technical assumptions given in [99, p. 593]. First, [99, Cond. (α)] states that there exists $Z_{\max} < \infty$ and $c > 0$ such that

$$\frac{1}{\text{Var}[Z]} \mathbb{E}[(Z - \mathbb{E}[Z])^2 \mathbf{1}\{|Z - \mathbb{E}[Z]| \leq Z_{\max}\}] > c \quad (3.131)$$

under $Z \sim Q_j$ and each $j = 1, \dots, K'$. This is trivially satisfied since we are considering finite alphabets, which implies that the support of each Q_j is bounded. The Lindeberg condition is stated in [99, Cond. (γ)], and is trivially satisfied since we are considering the case of finitely many distributions $\{Q_j\}$. The only non-trivial condition is [99, Cond. (β)], which can be written as follows in the finite alphabet case: For any given lattice, there exists $\delta > 0$ such that

$$\frac{1}{\log V_n} \sum_{i=1}^{n'} \mathbb{P}[Z_i \text{ is not } \delta\text{-close to a lattice point}] \rightarrow \infty. \quad (3.132)$$

Since we are considering the case that $\mathcal{I}_0$ does not lie on a lattice, we have for sufficiently small δ that the summation grows linearly in n , whereas $\log V_n$ only grows as $\log n$. We have thus shown that each of the technical conditions in [99] is satisfied, and hence (3.130) holds.

Upper bounding the exponential term in (3.130) by one, and noting that $S_n - S'_n$ has variance zero, we obtain

$$\mathbb{P}[S_n \in [z, z + \eta)] \leq \frac{\eta}{\sqrt{2\pi V_n}} + o\left(\frac{1}{\sqrt{n}}\right) \quad (3.133)$$

uniformly in z . We can now prove the lemma similarly to [61, Lemma 47] by writing

$$\mathbb{E}\left[e^{-S_n} \mathbb{1}\{S_n \geq t\}\right] \leq \sum_{l=0}^{\infty} e^{-t-l\eta} \mathbb{P}\left[t + l\eta \leq S_n \leq t + (l+1)\eta\right] \quad (3.134)$$

$$\leq \sum_{l=0}^{\infty} e^{-t-l\eta} \left(\frac{\eta}{\sqrt{2\pi V_n}} + o\left(\frac{1}{\sqrt{n}}\right) \right) \quad (3.135)$$

$$= e^{-t} \left(\frac{\eta}{(1 - e^{-\eta})\sqrt{2\pi V_n}} + o\left(\frac{1}{\sqrt{n}}\right) \right), \quad (3.136)$$

where (3.136) follows by evaluating the summation using the geometric series. The proof is concluded by taking $\eta \rightarrow 0$ and using the identity $\lim_{\eta \rightarrow 0} \frac{\eta}{1 - e^{-\eta}} = 1$. The uniformity of (3.120) in t follows from the uniformity of (3.133) in z .

In the lattice case, the argument is essentially unchanged, but we instead use the local limit theorem given in [100, Sec. VII.1, Thm. 2], which yields

$$\mathbb{P}[S'_n = z] = \frac{h_0}{\sqrt{2\pi V_n}} e^{-\frac{(z - \mu'_n)^2}{2V_n}} + o\left(\frac{1}{\sqrt{n}}\right) \quad (3.137)$$

uniformly in z on the lattice corresponding to S'_n (with span h_0). The remaining arguments are identical to the non-lattice case, with $\eta = h_0$ instead of $\eta \rightarrow 0$.

Proof of Lemma 3.2

We present the proof for the case that R_n , and hence $\hat{\rho}(Q, R_n, s)$, does not vary with n . Once this case is established, the general case with $\hat{\rho}$ varying in the range $[0, 1]$ readily follows.

We obtain (3.123) by using the definitions of c_3 and $\mathcal{F}_{\hat{\rho},s}^n$ (see (3.30) and (3.122)) to write

$$\text{Var}[i_s^n(\tilde{X}, Y) | Y = y] = \sum_y n \hat{P}_y(y) \text{Var}[i_s(\tilde{X}, Y) | Y = y] \quad (3.138)$$

$$\geq \sum_y n (P_{\hat{\rho},s}^*(y) - \delta) \text{Var}[i_s(\tilde{X}, Y) | Y = y] \quad (3.139)$$

$$= n(c_3 - o(\delta)), \quad (3.140)$$

where $(\tilde{X} | Y = y) \sim \tilde{P}_s(\cdot | y)$. To prove the second property, we perform an expansion in terms of types in the same way as Section 2.10.1 to conclude that the exponent of the denominator in the logarithm in (3.124) is given by

$$\min_{P_{XY}} \sum_{x,y} P_{XY}(x, y) \log \left(\frac{P_{XY}(x, y)}{Q(x)W(y|x)} e^{\hat{\rho} i_s(x, y)} \right). \quad (3.141)$$

Similarly, and from the definition of $\mathcal{F}_{\hat{\rho},s}^n(\delta)$ in (3.122), the exponent of the numerator in the logarithm in (3.124) is given by

$$\min_{P_{XY} : \max_y |P_Y(y) - P_{\hat{\rho},s}^*(y)| > \delta} \sum_{x,y} P_{XY}(x,y) \log \left(\frac{P_{XY}(x,y)}{Q(x)W(y|x)} e^{\hat{\rho} i_s(x,y)} \right). \quad (3.142)$$

A straightforward evaluation of the Karush-Kuhn-Tucker (KKT) conditions [38, Sec. 5.5.3] yields that the minimum (3.141) is uniquely achieved by $P_{\hat{\rho},s}^*$, defined in (3.29). On the other hand, $P_{\hat{\rho},s}^*$ does not satisfy the constraint in (3.142), and thus (3.142) is strictly greater than (3.141). This concludes the proof of (3.124).

3.5.5 Proof of Theorem 3.3

The proof is similar to that of Theorem 3.2. We make use of Lemma 3.1 in Section 3.5.4, as well as the following lemma, which is analogous to Lemma 3.2. We postpone the proof of the lemma until the end of the subsection.

Lemma 3.3. *Fix the parameters $\rho \geq 1$ and $s > 0$. Under the assumptions in (3.50)–(3.51), the sequence of sets*

$$\mathcal{F}_{\rho,s}^n(\delta) \triangleq \left\{ (x, \bar{x}) : Q^n(x)Q^n(\bar{x}) > 0, \max_{x,\bar{x}} \left| \hat{P}_{x\bar{x}}(x, \bar{x}) - P_{\rho,s}^*(x, \bar{x}) \right| \leq \delta \right\} \quad (3.143)$$

satisfies the following properties:

1. *For any $\delta > 0$ and $(x, \bar{x}) \in \mathcal{F}_{\rho,s}^n(\delta)$, the random variable $\tilde{\mathbf{Y}} \sim \tilde{P}_s^n(\cdot | x, \bar{x})$ satisfies*

$$\text{Var}[j_s^n(x, \bar{x}, \tilde{\mathbf{Y}})] \geq n(c_0 - r(\delta)), \quad (3.144)$$

where $r(\delta) \rightarrow 0$ as $\delta \rightarrow 0$.

2. *For any $\delta > 0$, we have*

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \frac{\sum_{(x,\bar{x}) \notin \mathcal{F}_{\rho,s}^n(\delta)} Q^n(x)Q^n(\bar{x}) \left(\sum_{\mathbf{y}} W^n(\mathbf{y}|x) \left(\frac{q(\bar{x}, \mathbf{y})}{q(x, \mathbf{y})} \right)^s \right)^{1/\rho}}{\sum_{x,\bar{x}} Q^n(x)Q^n(\bar{x}) \left(\sum_{\mathbf{y}} W^n(\mathbf{y}|x) \left(\frac{q(\bar{x}, \mathbf{y})}{q(x, \mathbf{y})} \right)^s \right)^{1/\rho}} > 0. \quad (3.145)$$

Proof of Theorem 3.3 Based on Lemma 3.3

Using the bound rcux_ρ in Theorem 2.2 with the i.i.d. codeword distribution $P_X = Q^n$, we have for any $\delta > 0$ that

$$\frac{1}{4(M-1)} \text{rcux}_\rho(n, M)^{1/\rho} = \sum_{\mathbf{x}, \bar{\mathbf{x}}} Q^n(\mathbf{x}) Q^n(\bar{\mathbf{x}}) \mathbb{P} \left[q^n(\bar{\mathbf{x}}, \mathbf{Y}) \geq q^n(\mathbf{x}, \mathbf{Y}) \right]^{1/\rho} \quad (3.146)$$

$$= \sum_{(\mathbf{x}, \bar{\mathbf{x}}) \in \mathcal{F}_{\rho, s}^n(\delta)} Q^n(\mathbf{x}) Q^n(\bar{\mathbf{x}}) \mathbb{P} \left[q^n(\bar{\mathbf{x}}, \mathbf{Y}) \geq q^n(\mathbf{x}, \mathbf{Y}) \right]^{1/\rho} \\ + \sum_{(\mathbf{x}, \bar{\mathbf{x}}) \notin \mathcal{F}_{\rho, s}^n(\delta)} Q^n(\mathbf{x}) Q^n(\bar{\mathbf{x}}) \mathbb{P} \left[q^n(\bar{\mathbf{x}}, \mathbf{Y}) \geq q^n(\mathbf{x}, \mathbf{Y}) \right]^{1/\rho}, \quad (3.147)$$

where each probability is implicitly conditioned on $\mathbf{X} = \mathbf{x}$.

We first analyze the summation over $\mathcal{F}_{\rho, s}^n(\delta)$ in (3.147). In order to make the inner probability more amenable to an application of Lemma 3.1, we write it as

$$\mathbb{P} \left[q^n(\bar{\mathbf{x}}, \mathbf{Y}) \geq q^n(\mathbf{x}, \mathbf{Y}) \right] = \mathbb{P} \left[\left(\frac{q^n(\bar{\mathbf{x}}, \mathbf{Y})}{q^n(\mathbf{x}, \mathbf{Y})} \right)^s \geq 1 \right] \quad (3.148)$$

$$= \mathbb{P} \left[\frac{\left(\frac{q^n(\bar{\mathbf{x}}, \mathbf{Y})}{q^n(\mathbf{x}, \mathbf{Y})} \right)^s}{\sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \left(\frac{q^n(\bar{\mathbf{x}}, \mathbf{y})}{q^n(\mathbf{x}, \mathbf{y})} \right)^s} \geq \frac{1}{\sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \left(\frac{q^n(\bar{\mathbf{x}}, \mathbf{y})}{q^n(\mathbf{x}, \mathbf{y})} \right)^s} \right] \quad (3.149)$$

$$= \mathbb{P} \left[j_s^n(\mathbf{x}, \bar{\mathbf{x}}, \mathbf{Y}) \geq -\log \sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \left(\frac{q(\bar{\mathbf{x}}, \mathbf{y})}{q(\mathbf{x}, \mathbf{y})} \right)^s \right], \quad (3.150)$$

where j_s^n is defined in (3.56). Next, following [85, Sec. 3.4.5], we note that the following holds when $\tilde{P}_s^n(\mathbf{y}|\mathbf{x}, \bar{\mathbf{x}}) \neq 0$:

$$W^n(\mathbf{y}|\mathbf{x}) = W^n(\mathbf{y}|\mathbf{x}) \frac{\tilde{P}_s^n(\mathbf{y}|\mathbf{x}, \bar{\mathbf{x}})}{\tilde{P}_s^n(\mathbf{y}|\mathbf{x}, \bar{\mathbf{x}})} \quad (3.151)$$

$$= \tilde{P}_s^n(\mathbf{y}|\mathbf{x}, \bar{\mathbf{x}}) e^{-nj_s(\mathbf{x}, \bar{\mathbf{x}}, \mathbf{y})}. \quad (3.152)$$

Summing (3.152) over all $\mathbf{y}$ such that $j_s(\mathbf{x}, \bar{\mathbf{x}}, \mathbf{y}) \geq t$, we obtain

$$\mathbb{P} \left[j_s^n(\mathbf{x}, \bar{\mathbf{x}}, \mathbf{Y}) \geq t \right] = \mathbb{E} \left[e^{-nj_s^n(\mathbf{x}, \bar{\mathbf{x}}, \tilde{\mathbf{Y}})} \mathbf{1} \{ j_s^n(\mathbf{x}, \bar{\mathbf{x}}, \tilde{\mathbf{Y}}) \geq t \} \right], \quad (3.153)$$

where $\tilde{\mathbf{Y}} \sim \tilde{P}_s^n(\cdot|\mathbf{x}, \bar{\mathbf{x}})$. For any $(\mathbf{x}, \bar{\mathbf{x}}) \in \mathcal{F}_{\rho, s}^n(\delta)$, we obtain the following using Lemma 3.1 in Section 3.5.4, the first part of Lemma 3.3, and the fact that $c_0 > 0$ (see the arguments following

(3.59)):

$$\mathbb{P}[j_s^n(\mathbf{x}, \bar{\mathbf{x}}, \mathbf{Y}) \geq t] \leq \frac{\psi_s(1 + o(1))}{\sqrt{2\pi n(c_0 - r(\delta))}} e^{-t} \quad (3.154)$$

uniformly in t , provided that δ is sufficiently small so that $r(\delta) < c_0$. Substituting (3.154) into (3.150), we obtain

$$\mathbb{P}\left[q^n(\bar{\mathbf{x}}, \mathbf{Y}) \geq q^n(\mathbf{x}, \mathbf{Y})\right] \leq \frac{\psi_s(1 + o(1))}{\sqrt{2\pi n(c_0 - r(\delta))}} \sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \left(\frac{q(\bar{\mathbf{x}}, \mathbf{y})}{q(\mathbf{x}, \mathbf{y})}\right)^s, \quad (3.155)$$

and hence

$$\begin{aligned} & \sum_{(\mathbf{x}, \bar{\mathbf{x}}) \in \mathcal{F}_{\rho, s}^n(\delta)} Q^n(\mathbf{x}) Q^n(\bar{\mathbf{x}}) \mathbb{P}\left[q^n(\bar{\mathbf{x}}, \mathbf{Y}) \geq q^n(\mathbf{x}, \mathbf{Y})\right]^{1/\rho} \\ & \leq \sum_{\mathbf{x}, \bar{\mathbf{x}}} Q^n(\mathbf{x}) Q^n(\bar{\mathbf{x}}) \left(\frac{\psi_s(1 + o(1))}{\sqrt{2\pi n(c_0 - r(\delta))}} \sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \left(\frac{q(\bar{\mathbf{x}}, \mathbf{y})}{q(\mathbf{x}, \mathbf{y})}\right)^s \right)^{1/\rho}. \end{aligned} \quad (3.156)$$

We observe that the right-hand side of (3.156) has the same exponent as the denominator of (3.145). Using Markov's inequality, the summation over $(\mathbf{x}, \bar{\mathbf{x}}) \notin \mathcal{F}_{\rho, s}^n(\delta)$ in (3.147) can be upper bounded by the numerator of (3.145), and thus the second part of Lemma 3.3 implies

$$\begin{aligned} & \frac{1}{4(M-1)} \text{rcux}_{\rho, s}(n, M)^{1/\rho} \\ & \leq (1 + o(1)) \sum_{\mathbf{x}, \bar{\mathbf{x}}} Q^n(\mathbf{x}) Q^n(\bar{\mathbf{x}}) \left(\frac{\psi_s(1 + o(1))}{\sqrt{2\pi n(c_0 - r(\delta))}} \sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \left(\frac{q(\bar{\mathbf{x}}, \mathbf{y})}{q(\mathbf{x}, \mathbf{y})}\right)^s \right)^{1/\rho}, \end{aligned} \quad (3.157)$$

and hence

$$\text{rcux}_{\rho, s}(n, M) \leq \frac{4^p \psi_s(1 + o(1))}{\sqrt{2\pi n(c_0 - r(\delta))}} M^\rho \left(\sum_{\mathbf{x}, \bar{\mathbf{x}}} Q^n(\mathbf{x}) Q^n(\bar{\mathbf{x}}) \left(\sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \left(\frac{q(\bar{\mathbf{x}}, \mathbf{y})}{q(\mathbf{x}, \mathbf{y})}\right)^s \right)^{1/\rho} \right)^\rho \quad (3.158)$$

$$= \frac{4^p \psi_s(1 + o(1))}{\sqrt{2\pi n(c_0 - r(\delta))}} e^{-n(E_{\mathbf{x}}^{\text{iid}}(Q, \rho, s) - \rho R)}, \quad (3.159)$$

where (3.159) follows by expanding each term as a product from 1 to n and using the definition of $E_{\mathbf{x}}^{\text{iid}}$. The proof is concluded by taking $\delta \rightarrow 0$ (and hence $r(\delta) \rightarrow 0$).

Proof of Lemma 3.3

We obtain (3.144) by expanding the variance as

$$\text{Var}[j_s^n(\mathbf{x}, \bar{\mathbf{x}}, \tilde{\mathbf{Y}})] = \sum_{\mathbf{x}, \bar{\mathbf{x}}} n \hat{P}_{\mathbf{x}\bar{\mathbf{x}}}(x, \bar{x}) \text{Var}_{\tilde{P}_s(\cdot|\mathbf{x}, \bar{\mathbf{x}})}[j_s(x, \bar{x}, \tilde{\mathbf{Y}})] \quad (3.160)$$

and substituting the bound in the definition of $\mathcal{F}_{\rho,s}^n(\delta)$ in (3.143). To prove the second property, we note that a nearly identical argument to Section 2.10.4 (see (2.167)–(2.169)) yields that the exponent of the denominator of (3.145) is equal to

$$\min_{P_{X\bar{X}}} D(P_{X\bar{X}} \| Q \times Q) + \frac{1}{\rho} \mathbb{E}_P[d_s(X, \bar{X})], \quad (3.161)$$

where d_s is defined in (2.117). By a similar argument, the exponent of the numerator of (3.145) is given by

$$\min_{P_{X\bar{X}} : \max_{\mathbf{x}, \bar{\mathbf{x}}} |P_{X\bar{X}}(x, \bar{x}) - P_{\rho,s}^*(x, \bar{x})| > \delta} D(P_{X\bar{X}} \| Q \times Q) + \frac{1}{\rho} \mathbb{E}_P[d_s(X, \bar{X})]. \quad (3.162)$$

By a straightforward analysis of the Karush-Kuhn-Tucker (KKT) conditions [38, Sec. 5.5.3], (3.161) is uniquely minimized by $P_{\rho,s}^*$, defined in (3.57). On the other hand, $P_{\rho,s}^*$ does not satisfy the constraint in (3.162), and thus (3.162) is strictly higher than (3.161).

3.5.6 Proofs of Theorems 3.4 and 3.6

We present the proof under the assumption that Q is itself a type, and thus $Q_n = Q$. The results for the general case follow using the fact that $\|Q - Q_n\| = O(\frac{1}{n})$; this contributes only to the $O(1)$ term to (3.79), and affects only the constant K_0 in (3.83). We similarly assume that e^{nR} is an integer in Theorem 3.6, and hence $M \triangleq \lfloor e^{nR} \rfloor = e^{nR}$.

As usual, we define the joint distribution

$$(\mathbf{X}, \mathbf{Y}, \bar{\mathbf{X}}) \sim P_{\mathbf{X}}(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) P_{\mathbf{X}}(\bar{\mathbf{x}}). \quad (3.163)$$

The main difference in the proofs of the theorems compared to the i.i.d. case (cf. [91], [85, Sec. 3.4.5]) is in the first step, the result of which is given in the following lemma. We recall the non-asymptotic bound $\text{rcu}(n, M)$ given in Theorem 2.1.

Lemma 3.4. *Fix the parameters (s, a) and the sequence $\{M_n\}$. Under the setup of Theorems 3.4 and 3.6, there exist constants K_1 and $\gamma > 0$ such that*

$$\text{rcu}(n, M_n) \leq \text{rcu}_{s,a}^*(n, M_n, K_1)(1 + e^{-\gamma n}) \quad (3.164)$$

for sufficiently large n , where

$$\text{rcu}_{s,a}^*(n, M, K_1) \triangleq \mathbb{E} \left[\min \left\{ 1, MK_1 n^{-\frac{\tau}{2}} e^{-i_{s,a}^n(\mathbf{X}, \mathbf{Y})} \right\} \right] \quad (3.165)$$

with τ defined in (3.84).

Proof. See Section 3.5.7. □

Proof of Theorem 3.4 Based on Lemma 3.4

Using the identity in (2.77), we can write $\text{rcu}_{s,a}^*$ in (3.165) as

$$\text{rcu}_{s,a}^*(n, M, K_1) = \mathbb{P} \left[i_{s,a}^n(\mathbf{X}, \mathbf{Y}) + \log U \leq \log(MK_1 n^{-\frac{\tau}{2}}) \right]. \quad (3.166)$$

Moreover, we have for all $\mathbf{x} \in T^n(Q)$ that

$$\mathbb{E}[i_{s,a}^n(\mathbf{x}, \mathbf{Y}) \mid \mathbf{X} = \mathbf{x}] = nI_{s,a}(Q) \quad (3.167)$$

$$\text{Var}[i_{s,a}^n(\mathbf{x}, \mathbf{Y}) \mid \mathbf{X} = \mathbf{x}] = nV_{s,a}(Q). \quad (3.168)$$

Combining these observations, we can prove Theorem 3.4 in an identical fashion to Section 2.5 using the Berry-Esseen theorem (respectively, Chebyshev's inequality) when $V(Q) > 0$ (respectively, $V(Q) = 0$). The remaining details are omitted to avoid repetition.

Proof of Theorem 3.6 for $R \leq R_{s,a}^{\text{cr}}(Q)$ Based on Lemma 3.4

Using the identity $\min\{1, \zeta\} \leq \zeta$, we obtain from (3.165) that

$$\text{rcu}_{s,a}^*(n, M, K_1) \leq MK_1 n^{-\frac{\tau}{2}} \mathbb{E} \left[e^{-i_{s,a}^n(\mathbf{X}, \mathbf{Y})} \right] \quad (3.169)$$

$$= MK_1 n^{-\frac{\tau}{2}} e^{-n(E_0^{\text{cc}}(Q, 1, s, a) - R)}, \quad (3.170)$$

where (3.170) follows from (3.72)–(3.73) and the fact that $\mathbf{X}$ has composition Q . Since $\hat{\rho} = 1$ for rates below $R_{s,a}^{\text{cr}}(Q)$, (3.170) shows that $\text{rcu}_{s,a}^*$ has the desired error exponent and prefactor (cf. (3.82)–(3.83)). The proof is concluded using (3.164).

Proof of Theorem 3.6 for $R > R_{s,a}^{\text{cr}}(Q)$ Based on Lemma 3.4

Using (2.77) and the identity $\min\{1, K\zeta\} \leq K \min\{1, \zeta\}$, we obtain from (3.165) that

$$\text{rcu}_{s,a}^*(n, e^{nR}) \leq K_1 \mathbb{P} \left[\sum_{i=1}^n (R - i_{s,a}(X_i, Y_i)) \geq \log(U n^{\frac{\tau}{2}}) \right] \quad (3.171)$$

$$= K_1 \sum_{\mathbf{x}} P_{\mathbf{X}}(\mathbf{x}) \int_0^1 \sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \mathbb{1} \left\{ \sum_{i=1}^n (R - i_{s,a}(x_i, y_i)) \geq \log(un^{\frac{\tau}{2}}) \right\} du \quad (3.172)$$

$$= K_1 \int_0^1 \sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \mathbb{1} \left\{ \sum_{i=1}^n (R - i_{s,a}(x_i, y_i)) \geq \log(un^{\frac{\tau}{2}}) \right\} du, \quad (3.173)$$

where (3.173) holds for an arbitrary sequence $\mathbf{x} \in T^n(Q)$ by symmetry. For $i = 1, \dots, n$, we define the CDF $F_i(t) \triangleq \mathbb{P}[R - i_{s,a}(x_i, Y_i) \leq t]$ (with $Y_i \sim W(\cdot|x_i)$), and we let $\{Z_i\}_{i=1}^n$ be independent random variables whose CDFs are given by

$$F_{Z_i}(z) = e^{E_{0,x_i}^{\text{cc}}(Q, \hat{\rho}, s, a) - \hat{\rho}R} \int_{-\infty}^z e^{\hat{\rho}t} dF_i(t). \quad (3.174)$$

These are seen to be valid CDFs since

$$\int_{-\infty}^{\infty} e^{\hat{\rho}t} dF_i(t) = \mathbb{E}[e^{\hat{\rho}(R - i_{s,a}(x_i, Y_i))}] = e^{\hat{\rho}R - E_{0,x_i}^{\text{cc}}(Q, \hat{\rho}, s, a)}, \quad (3.175)$$

where we have used (3.73). We apply a change of measure to the summation over $\mathbf{y}$ in (3.173) in the same way as (3.101)–(3.103):

$$\sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \mathbb{1} \left\{ \sum_{i=1}^n (R - i_{s,a}(x_i, y_i)) \geq t \right\} \quad (3.176)$$

$$= \int \dots \int_{\sum_{i=1}^n t_i \geq t} dF_1(t_1) \dots dF_n(t_n) \quad (3.177)$$

$$= e^{-\sum_{i=1}^n (E_{0,x_i}^{\text{cc}}(Q, \hat{\rho}, s, a) - \hat{\rho}R)} \int \dots \int_{\sum_{i=1}^n z_i \geq t} e^{-\hat{\rho} \sum_{i=1}^n z_i} dF_{Z_1}(z_1) \dots dF_{Z_n}(z_n) \quad (3.178)$$

$$= e^{-n(E_0^{\text{cc}}(Q, \hat{\rho}, s, a) - \hat{\rho}R)} \mathbb{E} \left[e^{-\hat{\rho} \sum_{i=1}^n Z_i} \mathbb{1} \left\{ \hat{\rho} \sum_{i=1}^n Z_i \geq \hat{\rho}t \right\} \right], \quad (3.179)$$

where in (3.179) we have used (3.72) and the fact that $\mathbf{x}$ has type Q .

Before proceeding, we state the following lemma by Polyanskiy *et al.* [61].

Lemma 3.5. [61, Lemma 47] *Let $Z_1, \dots, Z_n$ be independent random variables with $\sigma^2 = \sum_{i=1}^n \text{Var}[Z_i] > 0$ and $T = \sum_{i=1}^n \mathbb{E}[|Z_i - \mathbb{E}[Z_i]|^3] < \infty$. Then for any t ,*

$$\mathbb{E}\left[e^{-\sum_i Z_i} \mathbb{1}\left\{\sum_i Z_i \geq t\right\}\right] \leq 2\left(\frac{\log 2}{\sqrt{2\pi}} + \frac{12T}{\sigma^2}\right) \frac{1}{\sigma} e^{-t}. \quad (3.180)$$

Since F_{Z_i} is a tilted version of F_i , the variance of Z_i is positive if and only if the same is true of $R - i_{s,a}(x_i, Y_i)$. The latter is guaranteed provided that x_i coincides a value of x satisfying $\text{Var}[i_{s,a}(x, Y)] > 0$, and such a symbol occurring $\Theta(n)$ times in x is guaranteed to exist provided that $V_{s,a}(Q) > 0$. By a similar argument to Gallager [2, pp. 142-143] (see also Section 3.5.1) $V_{s,a}(Q)$ equaling zero would imply $R_{s,a}^{\text{cr}}(Q) = I_{s,a}(Q)$, so the fact that we are operating in the range $R \in (R_{s,a}^{\text{cr}}(Q), I_{s,a}(Q))$ means that $V_{s,a}(Q) > 0$. We conclude that the variance of $\sum_{i=1}^n Z_i$ grows linearly in n , thus yielding the following upon applying Lemma 3.5 to (3.179).

$$\begin{aligned} & \int_0^1 \sum_{\mathbf{y}} W^n(\mathbf{y}|\mathbf{x}) \mathbb{1}\left\{\sum_{i=1}^n (R - i_{s,a}(x_i, y_i)) \geq \log(un^{\frac{\tau}{2}})\right\} du \\ & \leq \frac{K_2}{\sqrt{n}} e^{-n(E_0^{\text{cc}}(Q, \hat{\rho}, s, a) - \rho R)} \mathbb{E}\left[e^{-\hat{\rho} \log(Un^{\frac{\tau}{2}})}\right] \end{aligned} \quad (3.181)$$

$$= \frac{K_3}{n^{\frac{1}{2}(1+\tau\hat{\rho})}} e^{-n(E_0^{\text{cc}}(Q, \hat{\rho}, s, a) - \rho R)}, \quad (3.182)$$

where K_2 and $K_3 \triangleq K_2 \mathbb{E}[U^{-\hat{\rho}}]$ are constants. Note that we have used the fact that T in Lemma 3.5 grows linearly in n since we are considering the case of finite alphabets.

The proof is concluded by substituting (3.182) into (3.173) and using (3.164).

3.5.7 Proof of Lemma 3.4

Throughout this subsection, we assume without loss of generality that $\mathcal{X} = \{1, \dots, |\mathcal{X}|\}$ and $\mathcal{Y} = \{1, \dots, |\mathcal{Y}|\}$. Recall that we are considering the case that $Q(x) > 0$ for all x , all outputs are reachable, and Q is a type (i.e. $Q = Q_n$).

Alternative Forms of the Codeword Distribution

We begin by writing the codeword distribution in (3.66) in various alternative forms. A straightforward symmetry argument yields

$$P_{\mathbf{X}}(\mathbf{x}) = \frac{1}{\mu_n} \prod_{i=1}^n Q(x_i) \mathbb{1}\{\mathbf{x} \in T^n(Q)\}, \quad (3.183)$$

where μ_n is a normalizing constant. Using the refined bounds on the size of $T^n(Q)$ in [27, Ex 2.2], it is readily verified that

$$\mu_n = \Theta\left(n^{-\frac{|\mathcal{X}|-1}{2}}\right). \quad (3.184)$$

As noted in Section 2.3, $T^n(Q)$ can be written as

$$T^n(Q) \triangleq \left\{ \mathbf{x} : |\hat{P}_{\mathbf{x}}(x) - Q(x)| \leq \frac{\delta}{n}, x = 1, \dots, |\mathcal{X}| \right\}, \quad (3.185)$$

where $\delta \in [0, 1)$. We will use this formulation with δ chosen to be strictly positive.

It will prove useful to give another equivalent definition in terms of the product alphabet $\mathcal{X}^k$, where k is a fixed integer. We divide the sequence $\mathbf{x}$ into n/k blocks of length k :

$$\begin{aligned} x_1^{(k)} &\triangleq (x_1, \dots, x_k) \\ x_2^{(k)} &\triangleq (x_{k+1}, \dots, x_{2k}) \\ &\vdots \\ x_{n/k}^{(k)} &\triangleq (x_{n-k+1}, \dots, x_n). \end{aligned} \quad (3.186)$$

For clarity of exposition, it is assumed here that n/k is an integer; the extension to the general case is straightforward. For a given product symbol $x_j^{(k)}$, we define the following function that counts the number of occurrences of a given symbol x :

$$\xi_x^k(x_j^{(k)}) \triangleq \sum_{i=(j-1)k+1}^{jk} \mathbb{1}\{x_i = x\}. \quad (3.187)$$

We can now write (3.185) as

$$T^n(Q) = \left\{ \mathbf{x} : \left| \sum_{i=1}^{n/k} \xi_x^k(x_i^{(k)}) - nQ(x) \right| \leq \delta, x = 1, \dots, |\mathcal{X}| \right\}. \quad (3.188)$$

We provide some reasoning behind this block decomposition in the proof of Lemma 3.7 below.

We will also make use of a vector $\boldsymbol{\xi}^k(x^{(k)})$ containing $|\mathcal{X}| - N$ of the functions $\{\xi_x^{(k)}\}_{x \in \mathcal{X}}$. Specifically, we write

$$\boldsymbol{\xi}^k(x^{(k)}) \triangleq \begin{bmatrix} \xi_{\tilde{x}_1}^k(x^{(k)}) \\ \vdots \\ \xi_{\tilde{x}_{|\mathcal{X}|-N}}^k(x^{(k)}) \end{bmatrix}, \quad (3.189)$$

where $\tilde{x}_1, \dots, \tilde{x}_{|\mathcal{X}|-N}$ are chosen to include all of the symbols except one (e.g. the one with the highest index) from each of the N subchannels.

Further Auxiliary Lemmas

Along with the definitions in Section 3.4.1, we define the tilted conditional distribution

$$\tilde{P}_{s,a}(x|y) \triangleq \frac{Q(x)q(x,y)^s e^{a(x)}}{\sum_{\bar{x}} Q(\bar{x})q(\bar{x},y)^s e^{a(\bar{x})}}, \quad (3.190)$$

the tilted joint distribution

$$P_{\hat{\rho},s,a}^*(x,y) \triangleq Q(x) \frac{W(y|x) e^{-\hat{\rho} i_{s,a}(x,y)}}{\sum_{y'} W(y'|x) e^{-\hat{\rho} i_{s,a}(x,y')}} \quad (3.191)$$

and its Y -marginal $P_{\hat{\rho},s,a}^*(y)$, and the random variables

$$(\mathbf{X}, \mathbf{Y}, \bar{\mathbf{X}}, \tilde{\mathbf{X}}) \sim P_{\mathbf{X}}(x) W^n(\mathbf{y}|x) P_{\mathbf{X}}(\bar{x}) \tilde{P}_{s,a}^n(\tilde{\mathbf{x}}|\mathbf{y}), \quad (3.192)$$

where $\tilde{P}_{s,a}^n$ is the n -fold product of $\tilde{P}_{s,a}$. The following two lemmas play the role of Lemma 3.2 from the corresponding analysis of i.i.d. random coding in Section 3.5.4.

Lemma 3.6. *Fix the parameters (s, a) and the sequence $\{R_n\}$, and let $\hat{\rho} = \hat{\rho}(Q, R_n, s, a)$ be defined according to (3.74). Under the setup of Theorems 3.4 and 3.6, the set²*

$$\mathcal{F}_{\hat{\rho},s,a}^n(\delta') \triangleq \left\{ \mathbf{y} : \max_y |\hat{P}_{\mathbf{y}}(y) - P_{\hat{\rho},s,a}^*(y)| \leq \delta' \right\} \quad (3.193)$$

satisfies the following properties:

1. *For sufficiently small δ' and any sequence of vectors $\mathbf{y} \in \mathcal{F}_{\hat{\rho},s,a}^n(\delta')$ (indexed by n), we have $\min_y \hat{P}_{\mathbf{y}}(y) = \Theta(n)$.*

²In the absence of the assumptions that all inputs are used and all outputs are reachable, the constraint $P_{\mathbf{Y}}(\mathbf{y}) > 0$ should be added to this set, as was done in Lemma 3.2.

2. For any $\delta' > 0$ and $K_1 > 0$, we have

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log \frac{e^{n\hat{\rho}R_n} \sum_{\mathbf{x}, \mathbf{y} \notin \mathcal{F}_{\hat{\rho}, s, a}^n(\delta')} P_{\mathbf{X}}(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) e^{-\hat{\rho} i_{s,a}^n(\mathbf{x}, \mathbf{y})}}{\text{rcu}_{s,a}^*(n, e^{nR_n}, K_1)} > 0. \quad (3.194)$$

Lemma 3.7. Let $k = |\mathcal{Y}|$, and fix the product symbol $y^{(k)} = (1, \dots, |\mathcal{Y}|)$. The covariance matrix of $\xi^k(\tilde{X}^{(k)})$ in (3.189) has full rank under $\tilde{X}^{(k)} \sim \tilde{P}_{s,a}^k(\cdot|y^{(k)})$ (i.e. the k -fold product of $\tilde{P}_{s,a}$). Moreover, if $q \notin \mathcal{Q}_0^{\text{cc}}$, then the covariance matrix of

$$\begin{bmatrix} i_{s,a}^k(\tilde{X}^{(k)}, y^{(k)}) \\ \xi^k(\tilde{X}^{(k)}) \end{bmatrix} \quad (3.195)$$

has full rank, where $i_{s,a}^k$ is defined analogously to (3.68).

Next, we provide a strong large deviations result that can be thought of as generalizing Lemma 3.1 to the multivariate setting.

Lemma 3.8. Fix the integers $K \geq 1$ and $d \geq 2$, and for each n , let $(n_1, \dots, n_K)$ be integers such that $\sum_j n_j = n$ and $\min_j n_j = \Theta(n)$. Fix the probability mass functions (PMFs) $Q_1, \dots, Q_K$ on a finite subset of $\mathbb{R}^d$, let $\{\Sigma_j\}_{j=1}^K$ be the corresponding $d \times d$ covariance matrices, and let $\{\mathbf{Z}_i\}_{i=1}^n$ be independent d -dimensional random vectors, n_j of which are distributed according to Q_j for each j . Furthermore, let $\{\Sigma'_j\}_{j=1}^K$ be the submatrices of $\{\Sigma_j\}_{j=1}^K$ obtained by removing the first row and column of each, and let $\{\mathbf{Z}'_i\}_{i=1}^n$ be obtained from $\{\mathbf{Z}_i\}_{i=1}^n$ by removing the first entry of each. If $\det(\Sigma'_j) > 0$ for some j , then for any constants t and $\delta > 0$, and any sequence of vectors $\phi_n \in \mathbb{R}^{d-1}$, we have

$$\mathbb{E} \left[e^{-\sum_{i=1}^n \mathbf{Z}_{1,i}} \mathbb{1} \left\{ \sum_{i=1}^n \mathbf{Z}_{1,i} \geq t \cap \left\| \sum_{i=1}^n \mathbf{Z}'_i - \phi_n \right\|_{\infty} \leq \delta \right\} \right] \leq \beta_n e^{-t}, \quad (3.196)$$

where

$$\beta_n = O\left(n^{-\frac{d-1}{2}}\right) \quad (3.197)$$

uniformly in t and ϕ_n . Furthermore, if $\det(\Sigma_j) > 0$ for some j , this can be strengthened to

$$\beta_n = O\left(n^{-\frac{d}{2}}\right) \quad (3.198)$$

uniformly in t and ϕ_n .

Proof of Lemma 3.4 Based on Lemmas 3.6–3.8

We write the RCU bound in (2.7) as

$$\text{rcu}(n, M) = \mathbb{E} \left[\min \left\{ 1, \mathbb{P} \left[i_{s,a}^n(\bar{\mathbf{X}}, \mathbf{Y}) \geq i_{s,a}^n(\mathbf{X}, \mathbf{Y}) \mid \mathbf{X}, \mathbf{Y} \right] \right\} \right]. \quad (3.199)$$

This is equivalent to (2.7) due to the fact that $a^n(\mathbf{X}) = a^n(\bar{\mathbf{X}})$ almost surely, which follows since $\mathbf{X}$ and $\bar{\mathbf{X}}$ have the same composition.

We now analyze the inner probability in (3.199). By a standard change of measure from constant-composition to i.i.d. [62] (see also (2.71)–(2.73)), we have for any sequence $\mathbf{y}$ that

$$\mathbb{P} \left[i_{s,a}^n(\bar{\mathbf{X}}, \mathbf{y}) \geq t \right] \leq p_0(n) e^{-t}, \quad (3.200)$$

where $p_0(n)$ grows polynomially in n (cf. (3.184)). The main step in the proof is to show the following refined bound for some constants δ' and K_1 , all $\mathbf{y} \in \mathcal{F}_{\hat{\rho},s,a}^n(\delta')$ (see (3.193)), and sufficiently large n :

$$\mathbb{P} \left[i_{s,a}^n(\bar{\mathbf{X}}, \mathbf{y}) \geq t \right] \leq K_1 n^{-\frac{\tau}{2}} e^{-t}. \quad (3.201)$$

To show this, we again follow [61, Sec. 3.4.5] and note that the following holds whenever $\tilde{P}_{s,a}^n(\bar{\mathbf{x}}|\mathbf{y}) \neq 0$:

$$P_{\mathbf{X}}(\bar{\mathbf{x}}) = \frac{1}{\mu_n} Q^n(\bar{\mathbf{x}}) \frac{\tilde{P}_{s,a}^n(\bar{\mathbf{x}}|\mathbf{y})}{\tilde{P}_{s,a}^n(\bar{\mathbf{x}}|\mathbf{y})} \mathbb{1}_{\{\bar{\mathbf{x}} \in T^n(Q)\}} \quad (3.202)$$

$$= \frac{1}{\mu_n} \tilde{P}_{s,a}^n(\bar{\mathbf{x}}|\mathbf{y}) e^{-i_{s,a}^n(\bar{\mathbf{x}}, \mathbf{y})} \mathbb{1}_{\{\bar{\mathbf{x}} \in T^n(Q)\}}. \quad (3.203)$$

Summing both sides over all $\bar{\mathbf{x}}$ such that $i_{s,a}^n(\bar{\mathbf{x}}, \mathbf{y}) \geq t$ yields

$$\mathbb{P} \left[i_{s,a}^n(\bar{\mathbf{X}}, \mathbf{y}) \geq t \right] = \frac{1}{\mu_n} \mathbb{E} \left[e^{-i_{s,a}^n(\tilde{\mathbf{X}}, \mathbf{y})} \mathbb{1}_{\{i_{s,a}^n(\tilde{\mathbf{X}}, \mathbf{y}) \geq t \cap \tilde{\mathbf{X}} \in T^n(Q)\}} \mid \mathbf{Y} = \mathbf{y} \right] \quad (3.204)$$

under the joint distribution in (3.192).

We set $k = |\mathcal{Y}|$ and let $y_1^{(k)}, \dots, y_{n/k}^{(k)}$ denote the decomposition of $\mathbf{y}$ into blocks in the same way as (3.186), and similarly for $\tilde{X}_1^{(k)}, \dots, \tilde{X}_{n/k}^{(k)}$. From the first part of Lemma 3.6, any sequence of vectors $\mathbf{y} \in \mathcal{F}_{\hat{\rho},s,a}^n(\delta')$ (indexed by n) must contain $\Theta(n)$ appearances of each output symbol. Moreover, by the symmetry of $P_{\mathbf{X}}$, the probability on the left-hand side of (3.201) is invariant under permutations of $\mathbf{y}$, and we may therefore assume that $\mathbf{y}$ starts with $\Theta(n)$ repetitions of $(1, \dots, |\mathcal{Y}|)$. Thus, the symbol $y^{(k)} = (1, \dots, |\mathcal{Y}|)$ occurs $\Theta(n)$ times in the block decomposition.

Using the alternative form of $T^n(Q)$ in (3.188), we can write (3.204) as

$$\begin{aligned} \mathbb{P}\left[i_{s,a}^n(\bar{\mathbf{X}}, \mathbf{y}) \geq t\right] &= \frac{1}{\mu_n} \mathbb{E} \left[e^{-\sum_{i=1}^{n/k} i_{s,a}^k(\tilde{X}_i^{(k)}, y^{(k)})} \mathbb{1} \left\{ \sum_{i=1}^{n/k} i_{s,a}^k(\tilde{X}_i^{(k)}, y^{(k)}) \geq t \right. \right. \\ &\quad \left. \left. \cap \left| \sum_{i=1}^{n/k} \xi_x^k(x_i^{(k)}) - nQ(x) \right| \leq \delta, x = 1, \dots, |\mathcal{X}| \right\} \middle| \mathbf{Y} = \mathbf{y} \right]. \end{aligned} \quad (3.205)$$

We are now in a position to prove (3.201) using Lemma 3.8. We set $d = |\mathcal{X}| - N + 1$, let each Q_j therein correspond to the vector in (3.195) for some $y^{(k)}$, and replace n in the lemma statement by n/k .

In the case that $q \notin \mathcal{Q}_0^{\text{cc}}$, we upper bound (3.205) by requiring the second event in the indicator function to hold only for $x \in \{\tilde{x}_1, \dots, \tilde{x}_{|\mathcal{X}|-N}\}$ (see (3.189)), rather than $x \in \mathcal{X}$. Recalling the above observation that the symbol $y^{(k)} = (1, \dots, |\mathcal{Y}|)$ occurs $\Theta(n)$ times, the second part of Lemma 3.7 reveals that Σ_j (defined in Lemma 3.8) has full rank for the corresponding value of j . Combining (3.196), (3.198) (with $d = |\mathcal{X}| - N + 1$) and (3.184), we can upper bound (3.205) by $O(n^{\frac{1}{2}(N-2)})e^{-t}$. This coincides with (3.201), since the condition $q \notin \mathcal{Q}_0^{\text{cc}}$ implies that the indicator function in (3.84) equals one.

In the remaining case, i.e. $q \in \mathcal{Q}_0^{\text{cc}}$, we follow similar steps, except that we also remove the first constraint (containing $i_{s,a}^k$) from the indicator function in (3.205), and we use the first part of Lemma 3.7 instead of the second part, leading to a bound of the form $O(n^{\frac{1}{2}(N-1)})e^{-t}$. This completes the proof of (3.201).

Combining (3.200) and (3.201) with (3.199), we obtain

$$\begin{aligned} \text{rcu}(n, M) &\leq \sum_{\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\hat{\rho}, s, a}^n(\delta')} P_{\mathbf{X}}(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) \min \left\{ 1, MK_1 n^{-\frac{\tau}{2}} e^{-i_{s,a}^n(\mathbf{x}, \mathbf{y})} \right\} \\ &\quad + \sum_{\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\hat{\rho}, s, a}^n(\delta')} P_{\mathbf{X}}(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) \min \left\{ 1, Mp_0(n) e^{-i_{s,a}^n(\mathbf{x}, \mathbf{y})} \right\} \end{aligned} \quad (3.206)$$

$$\leq \text{rcu}_{s,a}^*(n, M, K_1) + (Mp_0(n))^{\hat{\rho}} \sum_{\mathbf{x}, \mathbf{y} \in \mathcal{F}_{\hat{\rho}, s, a}^n(\delta')} P_{\mathbf{X}}(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x}) e^{-\hat{\rho} i_{s,a}^n(\mathbf{x}, \mathbf{y})}, \quad (3.207)$$

where (3.207) follows since $\min\{1, \zeta\} \leq \zeta^{\hat{\rho}}$. The second part of Lemma 3.6 implies that the second term in (3.207) has a strictly faster exponential rate of decay than the first, thus completing the proof of Lemma 3.4.

Proof of Lemma 3.6

Similarly to the proof of Lemma 3.2, it suffices to prove the case that R_n (and hence $\hat{\rho}(Q, R_n, s, a)$) is fixed, and we thus omit the dependence on n .

Observe from (3.191) and the definition of $i_{s,a}$ in (3.67) that $P_{\hat{\rho},s,a}^*(x, y) > 0$ if and only if $Q(x)W(y|x)q(x, y) > 0$. Using the assumption in (3.65) and the fact that we are considering the case that $Q(x) > 0$ for all x , it follows that $P_{\hat{\rho},s,a}^*(x, y) > 0$ if and only if $W(y|x) > 0$ (or equivalently $q(x, y) > 0$). Since all outputs are reachable by assumption, we obtain $P_{\hat{\rho},s,a}^*(y) > 0$ for all y . The first part of the lemma follows by choosing δ' in (3.193) to be any positive value smaller than $\min_y P_{\hat{\rho},s,a}^*(y)$.

The proof of (3.194) is similar to that of the corresponding statement in Lemma 3.2. Using expansions in terms of types and applying standard properties [27, Ch. 2] (e.g. see Section 2.10.1), we have $\text{rcu}_{s,a}^*(n, e^{nR}, K_1) \doteq e^{-n\hat{E}_r(Q, R, s, a)}$, where

$$\hat{E}_r(Q, R, s, a) \triangleq \min_{P_{XY}: P_X=Q} D(P_{XY} \| Q \times W) + [\mathbb{E}_P[i_{s,a}(X, Y)] - R]^+. \quad (3.208)$$

Writing $[\zeta]^+ = \max_{\rho \in [0,1]} \rho \zeta$ and using Fan's minimax theorem [78], we obtain

$$\hat{E}_r(Q, R, s, a) = \max_{\rho \in [0,1]} \min_{P_{XY}: P_X=Q} D(P_{XY} \| Q \times W) + \rho \mathbb{E}_P[i_{s,a}(X, Y)] - \rho R \quad (3.209)$$

$$= \max_{\rho \in [0,1]} \min_{P_{XY}: P_X=Q} \sum_{x,y} P_{XY}(x, y) \log \left(\frac{P_{XY}(x, y)}{Q(x)W(y|x)} e^{\rho i_{s,a}(x,y)} \right) - \rho R. \quad (3.210)$$

An evaluation of the KKT conditions reveals that the minimum is uniquely achieved by $P_{\hat{\rho},s,a}^*$ (see (3.191)), and yields an objective of $E_0(Q, \rho, s, a) - \rho R$. Using the definition of $\hat{\rho}$ in (3.74), we conclude that $\hat{E}_r(Q, R, s, a) = E_0(Q, \hat{\rho}, s, a) - \hat{\rho}R$.

Using similar steps to those used to obtain (3.210), along with the definition of $\mathcal{F}_{\hat{\rho},s,a}^n(\delta')$ in (3.193), we find that the numerator in (3.194) has an exponent given by

$$\hat{E}'_r(Q, R, \hat{\rho}, s, a) \triangleq \min_{\substack{P_{XY}: P_X=Q \\ \max_y |P_Y(y) - P_{\hat{\rho},s,a}^*(y)| > \delta}} \sum_{x,y} P_{XY}(x, y) \log \left(\frac{P_{XY}(x, y)}{Q(x)W(y|x)} e^{\hat{\rho} i_{s,a}(x,y)} \right) - \hat{\rho}R. \quad (3.211)$$

As we established following (3.210), in the absence of the constraint $\max_y |P_Y(y) - P_{\hat{\rho},s,a}^*(y)| > \delta$, the unique minimizer is given by $P_{\hat{\rho},s,a}^*$. However, $P_{\hat{\rho},s,a}^*$ is clearly infeasible in (3.211), and we therefore conclude that $\hat{E}'_r > \hat{E}_r$, thus concluding the proof of (3.194).

Proof of Lemma 3.7

From (3.189), the covariance matrix of $\xi^k(\tilde{X}^{(k)})$ has determinant zero if and only if

$$\text{Var}\left[\gamma_{\tilde{x}_1}\xi_{\tilde{x}_1}^k(\tilde{X}^{(k)}) + \cdots + \gamma_{\tilde{x}_{|\mathcal{X}|-N}}\xi_{\tilde{x}_{|\mathcal{X}|-N}}^k(\tilde{X}^{(k)})\right] = 0 \quad (3.212)$$

for some $(\gamma_{\tilde{x}_1}, \dots, \gamma_{\tilde{x}_{|\mathcal{X}|-N}}) \neq (0, \dots, 0)$. Since $\tilde{X}^{(k)} \sim \tilde{P}_{s,a}^k(\cdot|y^{(k)})$ is a product distribution and $y^{(k)} = (1, \dots, |\mathcal{Y}|)$ (see the lemma statement), we can write (3.212) as

$$\sum_y \text{Var}_{\tilde{P}_{s,a}(\cdot|y)}\left[\gamma_{\tilde{x}_1}\xi_{\tilde{x}_1}(\tilde{X}) + \cdots + \gamma_{\tilde{x}_{|\mathcal{X}|-N}}\xi_{\tilde{x}_{|\mathcal{X}|-N}}(\tilde{X})\right] = 0, \quad (3.213)$$

where $\xi_x(\tilde{x}) \triangleq \mathbb{1}\{\tilde{x} = x\}$ by (3.187). For notational convenience, we rewrite (3.213) as

$$\sum_y \text{Var}_{\tilde{P}_{s,a}(\cdot|y)}\left[\gamma_1\xi_1(\tilde{X}) + \cdots + \gamma_{|\mathcal{X}|}\xi_{|\mathcal{X}|}(\tilde{X})\right] = 0, \quad (3.214)$$

where $\gamma_x = 0$ for all $x \notin \{\tilde{x}_1, \dots, \tilde{x}_{|\mathcal{X}|-N}\}$. We proceed by assuming that (3.214) is true, and then arriving at a contradiction. By an identical argument to that given at the start of the proof of Lemma 3.6 above, we have $\tilde{P}_{s,a}(x|y) > 0$ if and only if $W(y|x) > 0$ (or equivalently $q(x, y) > 0$). It follows from (3.214) that $\gamma_x = \gamma_{\bar{x}}$ for any pair $(x, \bar{x})$ sharing a common output (i.e. $W(y|x)W(y|\bar{x}) > 0$ for some y), since otherwise the argument to $\text{Var}_{\tilde{P}_{s,a}(\cdot|y)}$ would differ depending on whether $\tilde{X} = x$ or $\tilde{X} = \bar{x}$, thus yielding a positive variance.

Since $\gamma_x = \gamma_{\bar{x}}$ for all $(x, \bar{x})$ sharing a common output, it also holds that $\gamma_x = \gamma_{\bar{x}}$ for all $\bar{x}$ that can be reached from x by following a path in the channel graph (i.e. by moving between inputs and outputs such that $W(\cdot|\cdot) \neq 0$). Since $\gamma_x = 0$ for at least one symbol of each of the N subchannels (see (3.189)), it follows that $\gamma_x = 0$ for all x , thus contradicting the assumption that $(\gamma_1, \dots, \gamma_{|\mathcal{X}|-1}) \neq (0, \dots, 0)$.

The claim regarding the covariance matrix of (3.195) follows similarly using a proof by contradiction. By the preceding argument, the submatrix of the covariance matrix obtained by removing the first row and column has full rank. The only remaining possibility for the matrix (3.195) to have determinant zero is that there exists a function $g(x)$ (equal to weighted linear combinations of $\{\xi_x\}_{x \in \mathcal{X}}$) such that $\text{Var}_{\tilde{P}_{s,a}(\cdot|y)}[i_{s,a}(\tilde{X}, y) + g(\tilde{X})] = 0$ for all y . Recalling that $\tilde{P}_{s,a}(x|y) > 0$ if and only if $q(x, y) > 0$, this condition is equivalent to $i_{s,a}(x, y) + g(x) + h(y)$ taking a constant value wherever $q(x, y) > 0$, for some function $h(y)$. We conclude from the definition of $i_{s,a}$ in (3.67) that

$$q(x, y) = g'(x)h'(y)\mathbb{1}\{q(x, y) > 0\} \quad (3.215)$$

for some functions $g'(x)$ and $h'(y)$. This contradicts the assumption $q \notin \mathcal{Q}_0^{\text{cc}}$ (see (3.76)), and we conclude that the covariance matrix of (3.195) has full rank.

We can now provide some intuition as to why it is beneficial to work with blocks of size $k = |\mathcal{Y}|$ (cf. (3.186)). It is this decomposition that allowed us to write (3.213) with a summation over y , but with values of $\{\gamma_x\}$ not depending on y . In contrast, a similar argument considering one symbol at a time would lead to an analogous expression to (3.213) with γ_x replaced by $\gamma_{x,y}$, which is not useful.

Proof of Lemma 3.8

We begin by proving the second part of the lemma. We write

$$\begin{aligned} & \mathbb{E} \left[e^{-\sum_{i=1}^n Z_{1,i}} \mathbb{1} \left\{ \sum_{i=1}^n Z_{1,i} \geq t \cap \left\| \sum_{i=1}^n \mathbf{Z}'_i - \phi_n \right\|_{\infty} \leq \delta \right\} \right] \\ & \leq \sum_{l=0}^{\infty} e^{-t-2l\delta} \mathbb{P} \left[t + 2l\delta \leq \sum_{i=1}^n Z_{1,i} \leq t + 2(l+1)\delta \cap \left\| \sum_{i=1}^n \mathbf{Z}'_i - \phi_n \right\|_{\infty} \leq \delta \right]. \end{aligned} \quad (3.216)$$

The probability on the right-hand side of (3.219) is that of a sum of independent vectors falling within a d -dimensional cube of side length 2δ , and can therefore be upper bounded by the probability of the same sum falling within a sphere of radius $\delta\sqrt{d}$. We claim that the concentration function result of [101, Thm. 6.2] implies that this probability behaves as $O(n^{-\frac{d}{2}})$ uniformly in the location of the sphere (i.e. the vector ϕ_n in (3.219)), provided that $\det(\Sigma_j) > 0$ for some $j \in \{1, \dots, K\}$ corresponding to $\Theta(n)$ of the terms in the sum. Since the latter statement is true by assumption and $\sum_{l=1}^{\infty} e^{-2l\delta} = \frac{1}{1-e^{-2\delta}}$ is a constant, this will complete the proof of the second part of the lemma.

The required $O(n^{-\frac{d}{2}})$ behavior for the probability of a falling within a sphere is proved in a nearly identical fashion to [96, Appendix F], so we only explain the differences. According to the concentration function result of Esseen [101, Thm. 6.2] (and choosing $\rho = \tau$ according to the notation therein), the probability under consideration is upper bounded by

$$\inf_{u \geq \delta\sqrt{d}} b \left(\frac{1}{u^2} \sum_{j=1}^K n_j \chi_j(u) \right)^{-\frac{d}{2}} \quad (3.217)$$

for some constant b , where

$$\chi_j(u) \triangleq \inf_{\|v\|_2=1} \sum_{\|z\|_2 < u} \tilde{Q}_j(z) (v^T z)^2, \quad (3.218)$$

and where $\tilde{Q}_j$ is the symmetrized version of Q_j (i.e. if A_1 and A_2 are independent random variables with distribution Q_j , then $A_1 - A_2$ has PMF $\tilde{Q}_j$). Since we are considering the case of finite alphabets, each $\tilde{Q}_j$ has bounded support, and thus there exists a finite constant u^* such that, for all $u > u^*$, the summation in (3.218) is unchanged when it is taken to be over all z sequences. For any such u , we have $\chi_j(u) = 2\lambda_{\min}(\Sigma_j)$, where $\lambda_{\min}$ denotes the smallest eigenvalue of a square matrix [96]. The desired result now follows in the same way as [96] by upper bounding u^2 by $\max\{(u^*)^2, \delta^2 d\}$ in (3.217), upper bounding the infimum over u by the specific choice $u = u^*$, and noting that the summation grows linearly in n due to the fact that $\lambda_{\min}(\Sigma_j) > 0$ for at least one value of j (this follows directly from the corresponding assumption $\det(\Sigma_j) > 0$).

The first part of the lemma is proved similarly by writing

$$\mathbb{E} \left[e^{-\sum_{i=1}^n Z_{1,i}} \mathbb{1} \left\{ \sum_{i=1}^n Z_{1,i} \geq t \cap \left\| \sum_{i=1}^n \mathbf{Z}'_i - \phi_n \right\|_{\infty} \leq \delta \right\} \right] \leq e^{-t} \mathbb{P} \left[\left\| \sum_{i=1}^n \mathbf{Z}'_i - \phi_n \right\|_{\infty} \leq \delta \right], \quad (3.219)$$

which follows immediately from the constraint $\sum_{i=1}^n Z_{1,i} \geq t$. The probability on the right-hand side of (3.219) is that of a sum of independent vectors falling within a $(d-1)$ -dimensional cube of side length 2δ . Using the same argument as the first part of the lemma and the assumption that $\det(\Sigma'_j) > 0$ for some j , the probability behaves as $\Theta(n^{-\frac{d-1}{2}})$ uniformly in t and ϕ_n .

3.5.8 Proofs of Theorems 3.5 and 3.7

Let $\mathcal{U} = \{1, \dots, N\}$ be the alphabet indexing the N subchannels (cf. Definition 3.1), and let W_1 be one such subchannel. Recall that the subchannels are identical by assumption. Let $\mathcal{C} = \{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(M)}\}$ be an arbitrary codebook (not necessarily constant-composition), and for a given codeword $\mathbf{x}$, let $\mathbf{u}(\mathbf{x})$ be a sequence of symbols in $\mathcal{U}$ indicating which subchannel is used for each channel use. By (3.65), the decoding metric $q^n(\mathbf{x}, \mathbf{y})$ is zero unless the values $\mathbf{y}$ are consistent with those of $\mathbf{u}$ (i.e. $y_i \in \mathcal{Y}_{u_i}$ for $i = 1, \dots, n$), and hence the maximum-metric decoding rule can be interpreted as follows: Given the output sequence $\mathbf{y}$, deterministically construct the corresponding auxiliary sequence $\mathbf{u}$, and then choose

$$\hat{\mathbf{m}} = \arg \max_{j: \mathbf{u}(\mathbf{x}^{(j)}) = \mathbf{u}} q^n(\mathbf{x}^{(j)}, \mathbf{y}). \quad (3.220)$$

We claim the following: If the total fraction of $\mathbf{u}$ sequences used (i.e. such that $\mathbf{u}(\mathbf{x}^{(j)}) = \mathbf{u}$ for some j) is less than $\frac{1}{\zeta}$ for some integer $\zeta \geq 2$, then the number of messages can be increased by a factor of ζ without changing the error probability. To see this, first note that since the

$\mathbf{u}$ sequence is known with certainty at the decoder, the determination of the message among $\{j : \mathbf{u}(\mathbf{x}^{(j)}) = \mathbf{u}\}$ simply corresponds to coding over W_1 ; the precise values in $\mathbf{u}$ play no further role. Thus, since less than a proportion $\frac{1}{\zeta}$ of the $\mathbf{u}$ sequences are utilized, we may take the original codebook structure and repeat it ζ times across the remaining $\mathbf{u}$ sequences, without affecting the error probability.

We now specialize these observations to constant-composition codes. If all the codewords $\{\mathbf{x}^{(j)}\}$ have the same composition, the same is true for the auxiliary sequences $\{\mathbf{u}(\mathbf{x}^{(j)})\}$. Thus, the total number of $\mathbf{u}$ sequences utilized is not N^n , but instead the size of some type class $T^n(Q_U)$ on $\mathcal{U}^n$. From [27, Ex 2.2], the largest type class (at least asymptotically) corresponds to a uniform distribution Q_U , and yields $|T^n(Q_U)| = \Theta(N^n n^{-\frac{N-1}{2}})$. Thus, at most a fraction $\Theta(n^{-\frac{N-1}{2}})$ of the $\mathbf{u}$ sequences are utilized, and we may increase the number of messages by a factor of $\Theta(n^{\frac{N-1}{2}})$ without affecting the error probability. It follows that for any sequence of constant-composition codes, there exists a sequence of general codes with an addition of $\frac{1}{2}(N-1)\log n + O(1)$ to the third-order term, as stated in (3.80).

We now turn to the proof of Theorem 3.7. Recall that we are considering symmetric channels and rates above the critical rate. Using Theorem 2.5 and the fact that the i.i.d. exponent with a uniform input distribution equals the reliability function $E(R)$ (cf. Section 1.2.1) above the critical rate [2, Ch. 5], we have $E(R) = E_0^{\text{cc}}(\hat{\rho}) - \hat{\rho}R$. Assuming for notational convenience that e^{nR} is an integer, the preceding arguments yield

$$p_{e,Q}^*(n, e^{nR}) \geq p_e^*(n, g_n e^{nR}), \quad (3.221)$$

where $g_n = \Theta(n^{-\frac{N-1}{2}})$. Moreover, the results given in [84, Ch. 5, Thms. 7-8] state that

$$p_e^*(n, g_n e^{nR}) \geq \frac{K'_1}{n^{\frac{1}{2}(1 + \mathbb{1}\{W \notin \mathcal{W}_0^{\text{symm}}\})\hat{\rho}(R - \Delta_n)}} e^{-nE(R - \Delta_n)}, \quad (3.222)$$

where $\Delta_n \triangleq \frac{1}{n} \log g_n$, and K'_1 is a constant depending on W and R . The proof is concluded by Taylor expanding $E(R)$ and $\hat{\rho}(R)$ in the same way as [84, Lem. 27]. In particular, letting $E'(\cdot)$ denote the derivative of $E(\cdot)$, we have

$$e^{-nE(R - \Delta_n)} = \Theta\left(n^{\frac{N-1}{2}|E'(R)|}\right) e^{-nE(R)}, \quad (3.223)$$

thus yielding the term $(N-1)$ in (3.84) upon identifying $|E'(R)| = \hat{\rho}(R)$ [84, Lem. 22].

Chapter 4

Multiple-Access Channels

4.1 Introduction

Network information theory problems with mismatched decoding are not only of independent interest, but can also provide valuable insight into the single-user mismatched decoding problem. In particular, it was shown by Lapidoth [35] that rates exceeding the LM rate (cf. Section 1.1.4) can be achieved by treating the single-user channel as a multiple-access channel (MAC) and generating multiple codebooks in parallel.

In this chapter, we revisit the mismatched MAC and build on the work of [35]. We derive ensemble-tight random-coding error exponents that recover the achievable rate region of [35], and we provide equivalent dual expressions and extensions to infinite and continuous alphabets. Furthermore, we study a mismatched multi-letter successive decoding rule, as well as second-order coding rates (cf. Section 1.2.2) for maximum-likelihood decoding.

4.1.1 System Setup

We consider a two-user memoryless MAC $W(y|x_1, x_2)$ with input alphabets $\mathcal{X}_1$ and $\mathcal{X}_2$ and output alphabet $\mathcal{Y}$. In the case that each alphabet is finite, the MAC is referred to as a discrete memoryless MAC (DM-MAC). The decoding metric is denoted by $q(x_1, x_2, y)$, and we write $W^n(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \triangleq \prod_{i=1}^n W(y_i|x_{1,i}, x_{2,i})$ and $q^n(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \triangleq \prod_{i=1}^n q(x_{1,i}, x_{2,i}, y_i)$.

Encoder $\nu = 1, 2$ takes as input a message m_ν uniformly distributed on the set $\{1, \dots, M_\nu\}$, and transmits the corresponding codeword $\mathbf{x}_\nu^{(m_\nu)}$ from the codebook $\mathcal{C}_\nu = \{\mathbf{x}_\nu^{(1)}, \dots, \mathbf{x}_\nu^{(M_\nu)}\}$. Given the output sequence $\mathbf{y}$, the decoder forms an estimate $(\hat{m}_1, \hat{m}_2)$ of the message pair, given by

$$(\hat{m}_1, \hat{m}_2) = \arg \max_{(i,j) : i \in \{1, \dots, M_1\}, j \in \{1, \dots, M_2\}} q^n(\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(j)}, \mathbf{y}). \quad (4.1)$$

We assume that ties are broken uniformly at random. The optimal maximum-likelihood (ML) decoding rule is recovered from (4.1) by setting $q(x_1, x_2, y) = W(y|x_1, x_2)$.

An error is said to have occurred if the estimate $(\hat{m}_1, \hat{m}_2)$ differs from (m_1, m_2) . The error probability for a given pair of codebooks (C_1, C_2) is denoted by $p_e(C_1, C_2)$. Given two codeword distributions P_{X_1} and P_{X_2} , we define the random-coding error probability

$$\bar{p}_e(n, M_1, M_2) \triangleq \mathbb{E}[p_e(C_1, C_2)], \quad (4.2)$$

where C_1 and C_2 contain independent codewords drawn from P_{X_1} and P_{X_2} respectively.

A rate pair $(\tilde{R}_1, \tilde{R}_2)$ is said to be achievable if, for all $\delta > 0$, there exist sequences of codebooks $C_{1,n}$ and $C_{2,n}$ of sizes at least $e^{n(\tilde{R}_1 - \delta)}$ and $e^{n(\tilde{R}_2 - \delta)}$ respectively such that $p_e(C_{1,n}, C_{2,n}) \rightarrow 0$. The capacity region $\mathcal{R}^*$ is defined to be the closure of the set of achievable rate pairs. An error exponent $\tilde{E}(R_1, R_2)$ is said to be achievable if there exist sequences of codebooks $C_{1,n}$ and $C_{2,n}$ of sizes at least e^{nR_1} and e^{nR_2} such that

$$\liminf_{n \rightarrow \infty} -\frac{1}{n} \log p_e(C_{1,n}, C_{2,n}) \geq \tilde{E}(R_1, R_2). \quad (4.3)$$

For a given pair of random-coding distributions (P_{X_1}, P_{X_2}) , we say that $\tilde{E}(R_1, R_2)$ exhibits ensemble tightness if

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \bar{p}_e(n, \lfloor e^{nR_1} \rfloor, \lfloor e^{nR_2} \rfloor) = \tilde{E}(R_1, R_2). \quad (4.4)$$

4.1.2 Previous Work

The capacity region of the matched MAC (corresponding to $q(x_1, x_2, y) = W(y|x_1, x_2)$) is given by [102, 103]

$$\mathcal{R}^* \triangleq \text{cl} \left\{ \bigcup_{P_{X_1}, P_{X_2}} \left\{ (R_1, R_2) : R_1 \leq I(X_1; Y|X_2), \right. \right. \\ \left. \left. R_2 \leq I(X_2; Y|X_1), R_1 + R_2 \leq I(X_1, X_2; Y) \right\} \right\}, \quad (4.5)$$

where $\text{cl}\{\cdot\}$ denotes the convex closure, and the mutual informations are taken with respect to $P_{X_1} \times P_{X_2} \times W$.

Error exponents for the matched MAC were presented in [66, 104–107]. In particular, Liu and Hughes [66] derived an achievable exponent using constant-composition random coding, and showed that in contrast to the single-user setting, the exponent can improve on that of

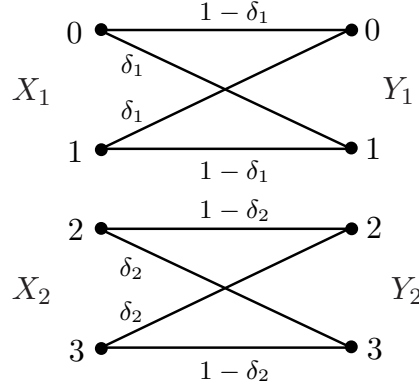


Figure 4.1 Parallel BSC example [35].

i.i.d. random coding [105] even after the optimization of the input distributions. An ensemble tightness result was proved for constant-composition random coding in [107] in the case that both rates are sufficiently small; to our knowledge, the general case has not been handled previously.

Lapidoth [35] obtained an achievable rate region for the mismatched MAC, and showed the surprising fact that the single-user LM rate can be improved by treating the single-user channel as a MAC. Moreover, it was shown that the random-coding error probability tends to one for rate pairs outside the region, thus establishing ensemble tightness. As an example, Lapidoth considered the channel in Figure 4.1 consisting of two parallel binary symmetric channels (BSCs) with crossover probabilities $\delta_1 \in (0, 0.5)$ and $\delta_2 \in (0, 0.5)$. The mismatched decoder assumes that both crossover probabilities are equal to $\delta \in (0, 0.5)$ (all such choices of δ are equivalent). By treating the channel as a mismatched single-user channel from (X_1, X_2) to (Y_1, Y_2) and using random coding with a uniform distribution on the quaternary input alphabet, one can only achieve rates R satisfying

$$R \leq 2 \left(1 - H_2 \left(\frac{\delta_1 + \delta_2}{2} \right) \right) \quad \text{bits/use}, \quad (4.6)$$

where $H_2(\cdot)$ is the binary entropy function in bits. On the other hand, by treating the channel as a mismatched MAC from X_1 and X_2 to (Y_1, Y_2) and using random coding with equiprobable input distributions on each binary input alphabet, one can achieve any rate R satisfying

$$R \leq (1 - H_2(\delta_1)) + (1 - H_2(\delta_2)) \quad \text{bits/use}. \quad (4.7)$$

This is the best rate possible even under maximum-likelihood (ML) decoding.

Error exponents for the mismatched MAC were presented in [107], but they are not tight enough to prove the achievability of Lapidath's rate region. For example, for the parallel BSC example in Figure 4.1 with uniform input distributions, the exponents of [107] are only positive when the sum-rate R satisfies (4.6).

4.1.3 Contributions

The main contributions of this chapter are as follows:

- In Section 4.2, we present our main results for the mismatched MAC. Considering constant-composition random coding, and making use of various bounds on the probability of a multiply-indexed union, we obtain an ensemble-tight error exponent for the DM-MAC which is positive in the interior of Lapidath's achievable rate region [35]. By specializing our results to the case of ML decoding, we prove the ensemble tightness of the exponent given in [66] for constant-composition random coding.
- Using the method of Lagrange duality [38], we provide alternative expressions for the achievable rate region and error exponent. We show that the dual expressions can be obtained directly using cost-constrained random coding, thus proving their validity in the case of infinite and continuous alphabets. Furthermore, we study the number of costs needed for the cost-constrained ensemble to match the performance of constant-composition codes for the DM-MAC.
- In Section 4.3, we consider a multi-letter successive decoding rule which differs from (4.1), but which still depends on the codewords only through the metrics $q^n(x_1^{(i)}, x_2^{(j)}, \mathbf{y})$. We derive an achievable rate region and error exponent, and show that the rate region can yield a strictly higher sum rate than the maximum-metric rule for a given pair of input distributions.
- In Section 4.4, we consider second-order coding rates (cf. Section 1.2.2) for maximum-likelihood decoding. Using constant-composition random coding, coded time sharing, and a variant of Hoeffding's combinatorial central limit theorem [108], we provide an inner bound on the set of locally achievable second-order coding rates for any given point on the boundary of the capacity region. Using numerical examples, we demonstrate the improved performance compared to i.i.d. random coding. Furthermore, we provide an extension of our main result to the Gaussian MAC via an increasingly fine quantization of the inputs.

The required results on multiply-indexed unions and Lagrange duality are presented in Appendices D and E respectively.

Parts of this chapter were presented in [109–113].

4.2 Random-Coding Bounds, Exponents and Rates

We consider the random coding setup described in Section 4.1.1. We let $\mathbf{X}_\nu^{(i)}$ be the random variable corresponding to the i -th codeword of user ν , yielding

$$\left(\{\mathbf{X}_1^{(i)}\}_{i=1}^{M_1}, \{\mathbf{X}_2^{(j)}\}_{j=1}^{M_2}\right) \sim \prod_{i=1}^{M_1} P_{\mathbf{X}_1}(\mathbf{x}_1^{(i)}) \prod_{j=1}^{M_2} P_{\mathbf{X}_2}(\mathbf{x}_2^{(j)}). \quad (4.8)$$

We assume without loss of generality that $m_1 = m_2 = 1$, and we write $\mathbf{X}_1$ and $\mathbf{X}_2$ in place of $\mathbf{X}_1^{(1)}$ and $\mathbf{X}_2^{(1)}$. We write $\bar{\mathbf{X}}_1$ and $\bar{\mathbf{X}}_2$ to denote arbitrary codewords that are generated independently of $\mathbf{X}_1$ and $\mathbf{X}_2$, and the random sequence at the output of the channel is denoted by $\mathbf{Y}$. It follows that

$$(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}, \bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2) \sim P_{\mathbf{X}_1}(\mathbf{x}_1) P_{\mathbf{X}_2}(\mathbf{x}_2) W^n(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) P_{\mathbf{X}_1}(\bar{\mathbf{x}}_1) P_{\mathbf{X}_2}(\bar{\mathbf{x}}_2). \quad (4.9)$$

For clarity of exposition, we focus primarily on the case that there is no time sharing (e.g. see [66]). In Section 4.2.5, we state the corresponding results with time sharing, and discuss the differences between the matched and mismatched settings.

We study the random-coding error probability by considering the following events:

- (Type 1) $q^n(\mathbf{X}_1^{(i)}, \mathbf{X}_2, \mathbf{Y}) \geq q^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$ for some $i \neq 1$;
- (Type 2) $q^n(\mathbf{X}_1, \mathbf{X}_2^{(j)}, \mathbf{Y}) \geq q^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$ for some $j \neq 1$;
- (Type 12) $q^n(\mathbf{X}_1^{(i)}, \mathbf{X}_2^{(j)}, \mathbf{Y}) \geq q^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$ for some $i \neq 1, j \neq 1$.

We refer to these as error events, but it should be noted that they do not necessarily correspond to decoder errors when the inequalities hold with equality, since we have assumed that the decoder breaks ties randomly.

The probabilities of the above error events are denoted by $\bar{p}_{e,1}(n, M_1)$, $\bar{p}_{e,2}(n, M_2)$ and $\bar{p}_{e,12}(n, M_1, M_2)$, and the overall random-coding error probability is denoted by $\bar{p}_e(n, M_1, M_2)$. Since breaking ties as errors at most doubles the error probability [12], we have

$$\frac{1}{2} \max\{\bar{p}_{e,1}, \bar{p}_{e,2}, \bar{p}_{e,12}\} \leq \bar{p}_e \leq \bar{p}_{e,1} + \bar{p}_{e,2} + \bar{p}_{e,12}. \quad (4.10)$$

4.2.1 Non-Asymptotic Bounds

The following theorem extends the random-coding union (RCU) bound for mismatched decoders (see Theorem 2.1) to the MAC.

Theorem 4.1. *For any mismatched MAC (possibly having infinite or continuous alphabets), we have under any random-coding distributions $P_{\mathbf{X}_1}$ and $P_{\mathbf{X}_2}$ that*

$$\bar{p}_{e,1}(n, M_1) \leq \text{rcu}_1(n, M_1) \quad (4.11)$$

$$\bar{p}_{e,2}(n, M_2) \leq \text{rcu}_2(n, M_2) \quad (4.12)$$

$$\bar{p}_{e,12}(n, M_1, M_2) \leq \min \{ \text{rcu}_{12,1}(n, M_1, M_2), \text{rcu}_{12,2}(n, M_1, M_2) \}, \quad (4.13)$$

where

$$\text{rcu}_1(n, M_1) \triangleq \mathbb{E} \left[\min \left\{ 1, (M_1 - 1) \mathbb{P} \left[\frac{q^n(\bar{\mathbf{X}}_1, \mathbf{X}_2, \mathbf{Y})}{q^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})} \geq 1 \mid \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y} \right] \right\} \right] \quad (4.14)$$

$$\text{rcu}_2(n, M_2) \triangleq \mathbb{E} \left[\min \left\{ 1, (M_2 - 1) \mathbb{P} \left[\frac{q^n(\mathbf{X}_1, \bar{\mathbf{X}}_2, \mathbf{Y})}{q^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})} \geq 1 \mid \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y} \right] \right\} \right] \quad (4.15)$$

$$\begin{aligned} \text{rcu}_{12,1}(n, M_1, M_2) &\triangleq \\ &\mathbb{E} \left[\min \left\{ 1, (M_1 - 1) \mathbb{E} \left[\min \left\{ 1, (M_2 - 1) \mathbb{P} \left[\frac{q^n(\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \mathbf{Y})}{q^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})} \geq 1 \mid \bar{\mathbf{X}}_1 \right] \right\} \mid \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y} \right] \right\} \right] \end{aligned} \quad (4.16)$$

$$\begin{aligned} \text{rcu}_{12,2}(n, M_1, M_2) &\triangleq \\ &\mathbb{E} \left[\min \left\{ 1, (M_2 - 1) \mathbb{E} \left[\min \left\{ 1, (M_1 - 1) \mathbb{P} \left[\frac{q^n(\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \mathbf{Y})}{q^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})} \geq 1 \mid \bar{\mathbf{X}}_2 \right] \right\} \mid \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y} \right] \right\} \right] \end{aligned} \quad (4.17)$$

under the joint distribution in (4.9).

Proof. For the type-1 and type-2 error events, the bounds are obtained using the same steps as Theorem 2.1. For the type-12 event, we write

$$\bar{p}_{e,12} = \mathbb{P} \left[\bigcup_{i \neq 1, j \neq 1} \left\{ \frac{q^n(\mathbf{X}_1^{(i)}, \mathbf{X}_2^{(j)}, \mathbf{Y})}{q^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})} \geq 1 \right\} \right] \quad (4.18)$$

$$= \mathbb{E} \left[\mathbb{P} \left[\bigcup_{i \neq 1, j \neq 1} \left\{ \frac{q^n(\mathbf{X}_1^{(i)}, \mathbf{X}_2^{(j)}, \mathbf{Y})}{q^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})} \geq 1 \right\} \middle| \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y} \right] \right]. \quad (4.19)$$

We obtain (4.16)–(4.17) by applying Lemma D.1 to the union in (4.19) (with $Z_1(i) = \mathbf{X}_1^{(i)}$ and $Z_2(j) = \mathbf{X}_2^{(j)}$), and upper bounding the minimum of three values in (D.4) in two different ways, each yielding a minimum of two values. $\square$

The type-1 and type-2 RCU bounds in (4.14)–(4.15) have appeared previously [114], along with the type-12 RCU bound

$$\text{rcu}'_{12}(n, M_1, M_2) \triangleq \mathbb{E} \left[\min \left\{ 1, (M_1 - 1)(M_2 - 1) \mathbb{P} \left[\frac{q^n(\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \mathbf{Y})}{q^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})} \geq 1 \middle| \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y} \right] \right\} \right]. \quad (4.20)$$

Both $\text{rcu}_{12,1}$ and $\text{rcu}_{12,2}$ are always less than or equal to rcu'_{12} , and in general they give better bounds on the error probability. We will see that in the mismatched setting, rcu'_{12} yields not only a worse non-asymptotic bound, but also a worse achievable rate region and error exponent. While no such losses in the rates and exponents are observed under ML decoding (see Section 4.2.4), the refined RCU bounds $\text{rcu}_{12,1}$ and $\text{rcu}_{12,2}$ are still of independent interest for characterizing the non-asymptotic performance.

4.2.2 Exponents and Rates for the DM-MAC

In this subsection, we study the DM-MAC using the constant-composition ensemble, which was studied by Liu and Hughes [66], among others. For $\nu = 1, 2$, we fix $Q_\nu \in \mathcal{P}(\mathcal{X}_\nu)$ and let $P_{\mathbf{X}_\nu}$ be the uniform distribution on the type class $T^n(Q_{\nu,n})$, where $Q_{\nu,n} \in \mathcal{P}_n(\mathcal{X}_\nu)$ is a type with the same support as Q_ν such that $\max_{x_\nu} |Q_{\nu,n}(x_\nu) - Q_\nu(x_\nu)| \leq \frac{1}{n}$. Thus,

$$P_{\mathbf{X}_\nu}(x_\nu) = \frac{1}{|T^n(Q_{\nu,n})|} \mathbb{1}\{x_\nu \in T^n(Q_{\nu,n})\}. \quad (4.21)$$

Our analysis is based on the method of types; the properties used are given in Appendix A. Throughout the section, we write $f(\mathbf{Q})$ to denote a quantity f that depends on Q_1 and Q_2 . Similarly, we write $f(\mathbf{Q}_n)$ to denote a quantity that depends on $Q_{1,n}$ and $Q_{2,n}$.

Error Exponents

The error exponents and achievable rates are expressed in terms of the sets

$$\mathcal{S}(\mathbf{Q}) \triangleq \left\{ P_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : P_{X_1} = Q_1, P_{X_2} = Q_2 \right\} \quad (4.22)$$

$$\begin{aligned} \mathcal{T}_1(P_{X_1 X_2 Y}) &\triangleq \left\{ \tilde{P}_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \right. \\ &\quad \left. \tilde{P}_{X_1} = P_{X_1}, \tilde{P}_{X_2 Y} = P_{X_2 Y}, \mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)] \right\} \end{aligned} \quad (4.23)$$

$$\begin{aligned} \mathcal{T}_2(P_{X_1 X_2 Y}) &\triangleq \left\{ \tilde{P}_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \right. \\ &\quad \left. \tilde{P}_{X_2} = P_{X_2}, \tilde{P}_{X_1 Y} = P_{X_1 Y}, \mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)] \right\} \end{aligned} \quad (4.24)$$

$$\begin{aligned} \mathcal{T}_{12}(P_{X_1 X_2 Y}) &\triangleq \left\{ \tilde{P}_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \right. \\ &\quad \left. \tilde{P}_{X_1} = P_{X_1}, \tilde{P}_{X_2} = P_{X_2}, \tilde{P}_Y = P_Y, \mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)] \right\}. \end{aligned} \quad (4.25)$$

Theorem 4.2. *For any input distributions Q_1 and Q_2 , and rates $R_1 > 0$ and $R_2 > 0$, we have the following under the codeword distributions described by (4.21):*

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \bar{p}_{e,1}(n, \lfloor e^{nR_1} \rfloor) = E_{r,1}^{\text{cc}}(\mathbf{Q}, R_1) \quad (4.26)$$

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \bar{p}_{e,2}(n, \lfloor e^{nR_2} \rfloor) = E_{r,2}^{\text{cc}}(\mathbf{Q}, R_2) \quad (4.27)$$

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \bar{p}_{e,12}(n, \lfloor e^{nR_1} \rfloor, \lfloor e^{nR_2} \rfloor) = E_{r,12}^{\text{cc}}(\mathbf{Q}, R_1, R_2), \quad (4.28)$$

where

$$\begin{aligned} E_{r,1}^{\text{cc}}(\mathbf{Q}, R_1) &\triangleq \min_{P_{X_1 X_2 Y} \in \mathcal{S}(\mathbf{Q})} \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_1(P_{X_1 X_2 Y})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) \\ &\quad + [I_{\tilde{P}}(X_1; X_2, Y) - R_1]^+ \end{aligned} \quad (4.29)$$

$$\begin{aligned} E_{r,2}^{\text{cc}}(\mathbf{Q}, R_2) &\triangleq \min_{P_{X_1 X_2 Y} \in \mathcal{S}(\mathbf{Q})} \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_2(P_{X_1 X_2 Y})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) \\ &\quad + [I_{\tilde{P}}(X_2; X_1, Y) - R_2]^+ \end{aligned} \quad (4.30)$$

$$\begin{aligned} E_{r,12}^{\text{cc}}(\mathbf{Q}, R_1, R_2) &\triangleq \min_{P_{X_1 X_2 Y} \in \mathcal{S}(\mathbf{Q})} \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}(P_{X_1 X_2 Y})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) \\ &\quad + \left[\max \left\{ I_{\tilde{P}}(X_1; Y) - R_1, I_{\tilde{P}}(X_2; Y) - R_2, D(\tilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \tilde{P}_Y) - R_1 - R_2 \right\} \right]^+. \end{aligned} \quad (4.31)$$

Proof. The random-coding error probabilities $\bar{p}_{e,1}$ and $\bar{p}_{e,2}$ can be handled similarly to the single-user setting (cf. Theorem 2.3). Furthermore, equivalent error exponents to (4.29)–(4.30)

were given in [107]. We therefore focus on $\bar{p}_{e,12}$, which requires a more careful analysis. We first rewrite

$$\bar{p}_{e,12} = \mathbb{E} \left[\mathbb{P} \left[\bigcup_{i \neq 1, j \neq 1} \left\{ \frac{q^n(\mathbf{X}_1^{(i)}, \mathbf{X}_2^{(j)}, \mathbf{Y})}{q^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})} \geq 1 \right\} \middle| \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y} \right] \right]. \quad (4.32)$$

in terms of the possible joint types of $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$ and $(\mathbf{X}_1^{(i)}, \mathbf{X}_2^{(j)}, \mathbf{Y})$. To this end, we define

$$\mathcal{S}_n(\mathbf{Q}_n) \triangleq \left\{ P_{X_1 X_2 Y} \in \mathcal{P}_n(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : P_{X_1} = Q_{1,n}, P_{X_2} = Q_{2,n} \right\} \quad (4.33)$$

$$\mathcal{T}_{12,n}(P_{X_1 X_2 Y}) \triangleq \mathcal{T}_{12}(P_{X_1 X_2 Y}) \cap \mathcal{P}_n(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}). \quad (4.34)$$

Roughly speaking, $\mathcal{S}_n$ is the set of possible joint types of $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$, and $\mathcal{T}_{12,n}(P_{X_1 X_2 Y})$ is the set of types of $(\mathbf{X}_1^{(i)}, \mathbf{X}_2^{(j)}, \mathbf{Y})$ that lead to decoding errors when $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \in T^n(P_{X_1 X_2 Y})$. The constraints on P_{X_ν} and $\tilde{P}_{X_\nu}$ arise from the fact that we are using constant-composition random coding, and the constraint $\mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)]$ holds if and only if $q^n(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{y}) \geq q^n(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$ for $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \in T^n(P_{X_1 X_2 Y})$ and $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{y}) \in T^n(\tilde{P}_{X_1 X_2 Y})$. Fixing $P_{X_1 X_2 Y} \in \mathcal{S}_n(\mathbf{Q}_n)$ and letting $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$ be an arbitrary triplet of sequences such that $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \in T^n(P_{X_1 X_2 Y})$, it follows that the event

$$\bigcup_{i \neq 1, j \neq 1} \left\{ \frac{q^n(\mathbf{X}_1^{(i)}, \mathbf{X}_2^{(j)}, \mathbf{y})}{q^n(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})} \geq 1 \right\} \quad (4.35)$$

can be written as

$$\bigcup_{i \neq 1, j \neq 1} \bigcup_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12,n}(P_{X_1 X_2 Y})} \left\{ (\mathbf{X}_1^{(i)}, \mathbf{X}_2^{(j)}, \mathbf{y}) \in T^n(\tilde{P}_{X_1 X_2 Y}) \right\}. \quad (4.36)$$

Expanding the probability and expectation in (4.32) in terms of types, substituting (4.36), and interchanging the order of the unions, we obtain

$$\begin{aligned} \bar{p}_{e,12} = & \sum_{P_{X_1 X_2 Y} \in \mathcal{S}_n(\mathbf{Q}_n)} \mathbb{P}[(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \in T^n(P_{X_1 X_2 Y})] \\ & \times \mathbb{P} \left[\bigcup_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12,n}(P_{X_1 X_2 Y})} \bigcup_{i \neq 1, j \neq 1} \left\{ (\mathbf{X}_1^{(i)}, \mathbf{X}_2^{(j)}, \mathbf{y}) \in T^n(\tilde{P}_{X_1 X_2 Y}) \right\} \right] \end{aligned} \quad (4.37)$$

$$\begin{aligned} &\stackrel{\text{def}}{=} \max_{P_{X_1 X_2 Y} \in \mathcal{S}_n(\mathbf{Q}_n)} \mathbb{P}[(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \in T^n(P_{X_1 X_2 Y})] \\ &\quad \times \max_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12,n}(P_{X_1 X_2 Y})} \mathbb{P}\left[\bigcup_{i \neq 1, j \neq 1} \left\{(\mathbf{X}_1^{(i)}, \mathbf{X}_2^{(j)}, \mathbf{y}) \in T^n(\tilde{P}_{X_1 X_2 Y})\right\}\right], \end{aligned} \quad (4.38)$$

where $\mathbf{y}$ is an arbitrary element of $T^n(P_Y)$, and (4.38) follows from the union bound and since the number of joint types grows polynomially in n .

The exponential behavior of the first probability in (4.38) is given by the property of types in (A.16), so it only remains to determine the exponential behavior of the second probability. To this end, we make use of Lemma D.3 with $Z_1(i) = \mathbf{X}_1^{(i)}$, $Z_2(j) = \mathbf{X}_2^{(j)}$, $\mathcal{A} = T_{\mathbf{y}}^n(\tilde{P}_{X_1 X_2 Y})$, $\mathcal{A}_1 = T_{\mathbf{y}}^n(\tilde{P}_{X_1 Y})$ and $\mathcal{A}_2 = T_{\mathbf{y}}^n(\tilde{P}_{X_2 Y})$. Using (D.12)–(D.13) and the properties of types in (A.17) and (A.18), it follows that the second probability in (4.38) has an exponent of

$$\left[\max \left\{ I_{\tilde{P}}(X_1; Y) - R_1, I_{\tilde{P}}(X_2; Y) - R_2, D(\tilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \tilde{P}_Y) - R_1 - R_2 \right\} \right]^+. \quad (4.39)$$

Substituting (4.39) and the property of types in (A.16) into (4.38), and using a standard continuity argument (e.g. see [58, Eq. (30)]) to replace the sets $\mathcal{S}_n$ and $\mathcal{T}_{12,n}$ by $\mathcal{S}$ and $\mathcal{T}_{12}$ respectively, we recover the exponent in (4.31). $\square$

From Theorem 4.2 and (4.10), the overall error exponent is given by

$$E_r^{\text{cc}}(\mathbf{Q}, R_1, R_2) \triangleq \min \left\{ E_{r,1}^{\text{cc}}(\mathbf{Q}, R_1), E_{r,2}^{\text{cc}}(\mathbf{Q}, R_2), E_{r,12}^{\text{cc}}(\mathbf{Q}, R_1, R_2) \right\}, \quad (4.40)$$

and is tight with respect to the ensemble average. The preceding proof made use of the refined union bound given in Lemma D.3; if we had instead used the standard truncated union bound in (D.2), we would have obtained the type-12 exponent

$$\begin{aligned} E_{r,12}^{\text{cc}'}(\mathbf{Q}, R_1, R_2) &\triangleq \min_{P_{X_1 X_2 Y} \in \mathcal{S}(\mathbf{Q})} \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}(P_{X_1 X_2 Y})} \\ &\quad D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + \left[D(\tilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \tilde{P}_Y) - (R_1 + R_2) \right]^+, \end{aligned} \quad (4.41)$$

which coincides with an achievable exponent given in [107]. Following the proof of Theorem 2.3, one can easily show that this is the exponent of the quantity rcu'_{12} defined in (4.20). In Section 4.2.6, we will see that (4.31) provides no improvement over (4.41) for maximum-likelihood decoding, even though it can provide an improvement in the mismatched case.

In the case that there are $K > 2$ users, one can use Lemma D.4 in place of Lemma D.3. For example, for the joint error event between all K users, the minimum of four terms in (4.31) is replaced by a minimum of 2^K terms, corresponding to the subsets of $\{1, \dots, K\}$

Achievable Rate Region

The following achievable rate region follows from Theorem 4.2 in a straightforward fashion, and coincides with the ensemble-tight achievable rate region of Lapidoth [35].

Theorem 4.3. *For any pair (Q_1, Q_2) , the overall error exponent $E_r^{\text{cc}}(\mathbf{Q}, R_1, R_2)$ in (4.40) is positive for all rate pairs (R_1, R_2) in the interior of $\mathcal{R}_{\text{LM}}(\mathbf{Q})$, defined to be the set of all rate pairs (R_1, R_2) satisfying*

$$R_1 \leq \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_1(Q_1 \times Q_2 \times W)} I_{\tilde{P}}(X_1; X_2, Y) \quad (4.42)$$

$$R_2 \leq \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_2(Q_1 \times Q_2 \times W)} I_{\tilde{P}}(X_2; X_1, Y) \quad (4.43)$$

$$R_1 + R_2 \leq \min_{\substack{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}(Q_1 \times Q_2 \times W) \\ I_{\tilde{P}}(X_1; Y) \leq R_1, I_{\tilde{P}}(X_2; Y) \leq R_2}} D(\tilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \tilde{P}_Y). \quad (4.44)$$

Proof. The conditions in (4.42)–(4.44) are obtained from (4.29)–(4.31) respectively. Focusing on (4.44), we see that the objective in (4.31) is always positive when $D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) > 0$, $I_{\tilde{P}}(X_1; Y) > R_1$ or $I_{\tilde{P}}(X_2; Y) > R_2$. Thus, by using the continuity argument of [35, p. 1444, Step 3], taking $D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) \rightarrow 0$, and introducing the constraints $I_{\tilde{P}}(X_1; Y) \leq R_1$ and $I_{\tilde{P}}(X_2; Y) \leq R_2$, we obtain the condition in (4.44). $\square$

Using the usual time sharing argument [35, 115] (see also Section 4.2.5), it follows from Theorem 4.3 that we can achieve any rate pair in the convex hull of $\bigcup_Q \mathcal{R}_{\text{LM}}(\mathbf{Q})$, where the union is over all distributions Q_1 and Q_2 on $\mathcal{X}_1$ and $\mathcal{X}_2$ respectively.

By a similar argument to the proof of Theorem 4.3, (4.41) yields the rate condition

$$R_1 + R_2 \leq \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}(Q_1 \times Q_2 \times W)} D(\tilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \tilde{P}_Y). \quad (4.45)$$

We proceed by numerically comparing the ensemble-tight expressions in (4.31) and (4.44) with the weaker expressions in (4.41) and (4.45).

Numerical Examples

Here we return to the parallel BSC example given in Figure 4.1, setting $Y = (Y_1, Y_2)$. As mentioned in Section 4.1.2, the decoder assumes that both crossover probabilities are equal. It is straightforward to show that the corresponding decoding rule is equivalent to minimizing sum of t_1 and t_2 , where t_ν is the number of bit flips from the input sequence x_ν to the output sequence y_ν . As noted in [35], this decision rule is in fact equivalent to ML. This channel

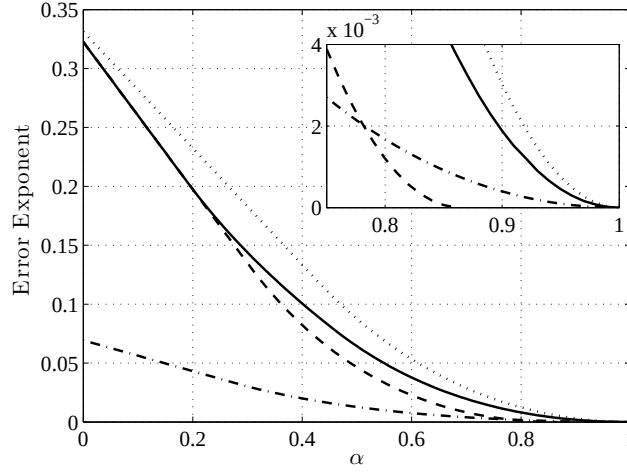


Figure 4.2 Error exponents $E_{r,1}^{cc}$ (dotted), $E_{r,2}^{cc}$ (dash-dot), $E_{r,12}^{cc}$ (solid) and $E_{r,12}^{cc'}$ (dashed) for the parallel channel shown in Figure 4.1 using $\delta_1 = 0.05$, $\delta_2 = 0.25$ and equiprobable input distributions on $\{0, 1\}$. The rate pair is given by $(R_1, R_2) = (\alpha C_1, \alpha C_2)$.

could easily be analyzed by treating the two subchannels separately, but we treat it as a MAC because it serves as a good example for comparing the ensemble-tight achievability results with (4.41) and (4.45).

We let both Q_1 and Q_2 be equiprobable on $\{0, 1\}$. With this choice, it was shown in [35] that the right-hand side of (4.45) is no greater than

$$2 \left(1 - H_2 \left(\frac{\delta_1 + \delta_2}{2} \right) \right) \text{ bits/use.} \quad (4.46)$$

On the other hand, the refined condition in (4.44) can be used to prove the achievability of *any* (R_1, R_2) within the rectangle with corners $(0, 0)$ and (C_1, C_2) , where $C_\nu \triangleq 1 - H_2(\delta_\nu)$ [35]. This observation is analogous to the comparison between (4.6) and (4.7) in the introduction to the chapter. The main difference is that the weakness in (4.6) is in the random-coding ensemble itself, whereas the weakness in (4.46) is in the bounding techniques used in the analysis.

We evaluate the error exponents using the optimization software YALMIP [77]. Figure 4.2 plots each of the exponents as a function of α , where the rate pair is $(R_1, R_2) = (\alpha C_1, \alpha C_2)$. While the overall error exponent $E_r^{cc}(\mathbf{Q}, R_1, R_2)$ in (4.40) is unchanged at low to moderate values of α when $E_{r,12}^{cc'}$ is used in place of $E_{r,12}^{cc}$, this is not true for high values of α . Furthermore, consistent with the preceding discussion, $E_{r,12}^{cc'}$ is non-zero only for $\alpha < 0.865$, whereas $E_{r,12}^{cc}$ is positive for all $\alpha < 1$. The fact that the curves $E_{r,12}^{cc}$ and $E_{r,12}^{cc'}$ coincide at low values of α is consistent with [107, Cor. 5], which states that $E_{r,12}^{cc'}$ is ensemble-tight at low rates.

Next, we show that in contrast to the matched case, the achievable rate region $\mathcal{R}_{LM}(\mathbf{Q})$ can

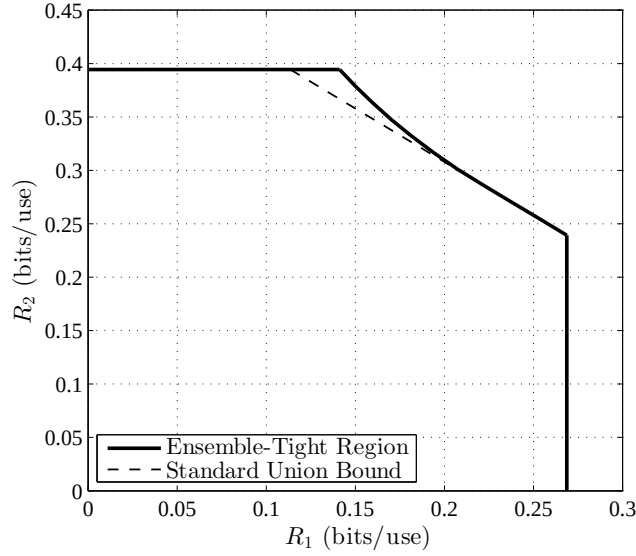


Figure 4.3 Achievable rate region for the MAC associated with (4.47) using the corresponding decoding metric in which all values of δ are equal. The crossover probabilities are $\delta_0 = 0.02$, $\delta_1 = 0.2$, $\delta_2 = 0.025$ and $\delta_3 = 0.25$, and both input distributions are equiprobable on $\{0, 1\}$. The ensemble-tight region is given in Theorem 4.3, whereas the naive union bound replaces (4.44) by (4.45).

be non-convex even for a given pair (Q_1, Q_2) . Let $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$ and $\mathcal{Y} = \{0, 1, 2, 3\}$. The channel is described by $W(y|x_1, x_2) = W'(y|2x_1 + x_2)$, where $W'(y|x)$ is defined by the entries of the matrix

$$W' = \begin{bmatrix} 1 - 3\delta_0 & \delta_0 & \delta_0 & \delta_0 \\ \delta_1 & 1 - 3\delta_1 & \delta_1 & \delta_1 \\ \delta_2 & \delta_2 & 1 - 3\delta_2 & \delta_2 \\ \delta_3 & \delta_3 & \delta_3 & 1 - 3\delta_3 \end{bmatrix}. \quad (4.47)$$

The decoding metric is defined similarly, but with each δ_i replaced by a single value $\delta \in (0, 0.25)$ (all such choices of δ are equivalent).

Figure 4.3 plots the achievable rate region $\mathcal{R}_{\text{LM}}(\mathbf{Q})$ with $\delta_0 = 0.02$, $\delta_1 = 0.2$, $\delta_2 = 0.025$, $\delta_3 = 0.25$, and $Q_1 = Q_2 = (0.5, 0.5)$. The boundary between the horizontal and vertical limits is not a straight line, but instead a convex curve. To gain some insight as to how this behavior arises, we discuss the condition (4.44). In this example, the constraint $I_{\tilde{P}}(X_2; Y) \leq R_2$ is inactive, whereas the constraint $I_{\tilde{P}}(X_1; Y) \leq R_1$ is sometimes active. As R_1 increases, the constraint is satisfied by a wider range of distributions $\tilde{P}_{X_1 X_2 Y}$, and the sum rate $R_1 + R_2$ decreases. The corner point $(0.2687, 0.2392)$ is the same whether the ensemble-tight bound or the weakened bound is used, indicating that the constraint $I_{\tilde{P}}(X_1; Y) \leq R_1$ is also inactive at this point. However, at lower values of R_1 , the constraint is active, and the ensemble-tight bound

is better. In particular, the upper corner point $(0.1412, 0.3944)$ improves on $(0.1136, 0.3944)$, which corresponds to the weaker rate condition in (4.45).

Alternative Expressions for the Exponents and Rates

In this subsection, we derive equivalent dual expressions for the exponents and rates given in Theorems 4.2 and 4.3. The proofs are a non-trivial extension of the single-user setting [9], and make use of Lemmas E.1 and E.3 in Appendix E. In Section 4.2.3, we will show that these dual expressions generalize to infinite and continuous alphabets. As discussed in Section 1.1.4, another advantage is that achievable performance bounds can be obtained without performing the full optimizations.

While dual expressions for (4.29)–(4.30) can easily be obtained similarly to the single-user setting (cf. Theorem 2.4), it is more difficult to handle (4.31) directly, due to the additional terms containing $I_{\tilde{P}}(X_1; Y)$ and $I_{\tilde{P}}(X_2; Y)$. The following lemma provides an alternative expression for (4.31) which turns out to resolve this issue.

Lemma 4.1. *The exponent $E_{r,12}^{\text{cc}}$ in (4.31) can be expressed as*

$$E_{r,12}^{\text{cc}}(\mathbf{Q}, R_1, R_2) = \max \{ E_{r,12,1}^{\text{cc}}(\mathbf{Q}, R_1, R_2), E_{r,12,2}^{\text{cc}}(\mathbf{Q}, R_1, R_2) \}, \quad (4.48)$$

where

$$\begin{aligned} E_{r,12,1}^{\text{cc}}(\mathbf{Q}, R_1, R_2) \triangleq & \min_{P_{X_1 X_2 Y} \in \mathcal{S}(\mathbf{Q})} \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}(P_{X_1 X_2 Y})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) \\ & + \left[\max \left\{ I_{\tilde{P}}(X_1; Y) - R_1, D(\tilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \tilde{P}_Y) - R_1 - R_2 \right\} \right]^+ \end{aligned} \quad (4.49)$$

$$\begin{aligned} E_{r,12,2}^{\text{cc}}(\mathbf{Q}, R_1, R_2) \triangleq & \min_{P_{X_1 X_2 Y} \in \mathcal{S}(\mathbf{Q})} \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}(P_{X_1 X_2 Y})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) \\ & + \left[\max \left\{ I_{\tilde{P}}(X_2; Y) - R_2, D(\tilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \tilde{P}_Y) - R_1 - R_2 \right\} \right]^+. \end{aligned} \quad (4.50)$$

Proof. This is a special case of Lemma E.3 with the following identifications:

$$f(\mathbf{z}) = D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) \quad (4.51)$$

$$g(\mathbf{z}) = D(\tilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \tilde{P}_Y) - R_1 - R_2 \quad (4.52)$$

$$g_1(\mathbf{z}) = I_{\tilde{P}}(X_1; Y) - R_1 \quad (4.53)$$

$$g_2(\mathbf{z}) = I_{\tilde{P}}(X_2; Y) - R_2. \quad (4.54)$$

From the latter three choices and the identity

$$D(\tilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \tilde{P}_Y) = I_{\tilde{P}}(X_1; Y) + I_{\tilde{P}}(X_2; Y) + I_{\tilde{P}}(X_1; X_2 | Y), \quad (4.55)$$

we see that the condition in (E.12) is satisfied. $\square$

The exponents in (4.49)–(4.50) are precisely those that are obtained by applying properties of types directly to the RCU bounds in (4.16)–(4.17), rather than directly analyzing the random-coding error probability as in the proof of Theorem 4.2. Thus, Lemma 4.1 shows that the RCU bounds yield the true error exponent for the constant-composition ensemble.

Lemma 4.2. *The condition in (4.44) holds if and only if at least one of the following hold:*

$$R_1 + R_2 \leq \min_{\substack{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}(Q_1 \times Q_2 \times W) \\ I_{\tilde{P}}(X_1; Y) \leq R_1}} D(\tilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \tilde{P}_Y) \quad (4.56)$$

$$R_1 + R_2 \leq \min_{\substack{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}(Q_1 \times Q_2 \times W) \\ I_{\tilde{P}}(X_2; Y) \leq R_2}} D(\tilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times \tilde{P}_Y). \quad (4.57)$$

Proof. These conditions can be derived from (4.49)–(4.50) in an identical fashion to (4.44). Specifically, $E_{r,12}^{\text{cc}}$ is positive whenever one of (4.49) and (4.50) hold, and zero otherwise. A similar statement holds for (4.44), and thus the two formulations must be equivalent. $\square$

The conditions in (4.56)–(4.57) yield a notable advantage over the equivalent condition in (4.44) in terms of computation. Based on (4.44), one may expect that the achievable rate region must be generated by evaluating the minimization over a grid of (R_1, R_2) values. However, from (4.56)–(4.57), we can avoid such a computationally complex search by instead evaluating the minimization in (4.56) over a range of R_1 values, and the minimization in (4.57) over a range of R_2 values. The number of minimizations performed is therefore reduced from $O(N_1 N_2)$ to $O(N_1 + N_2)$, where for $\nu = 1, 2$, N_ν is the number of points used in the quantization of the R_ν axis.

The dual expressions for the error exponents in (4.29)–(4.30) and (4.49)–(4.50) are given in the following theorem.

Theorem 4.4. *The exponents $E_{r,1}^{\text{cc}}$, $E_{r,2}^{\text{cc}}$, $E_{r,12,1}^{\text{cc}}$ and $E_{r,12,2}^{\text{cc}}$ can be expressed as*

$$E_{r,1}^{\text{cc}}(\mathbf{Q}, R_1) = \max_{\rho \in [0,1]} E_{0,1}^{\text{cc}}(\mathbf{Q}, \rho) - \rho R_1 \quad (4.58)$$

$$E_{r,2}^{\text{cc}}(\mathbf{Q}, R_2) = \max_{\rho \in [0,1]} E_{0,2}^{\text{cc}}(\mathbf{Q}, \rho) - \rho R_2 \quad (4.59)$$

$$E_{r,12,1}^{\text{cc}}(\mathbf{Q}, R_1, R_2) = \max_{\rho_1 \in [0,1], \rho_2 \in [0,1]} E_{0,12,1}^{\text{cc}}(\mathbf{Q}, \rho_1, \rho_2) - \rho_1(R_1 + \rho_2 R_2) \quad (4.60)$$

$$E_{r,12,2}^{\text{cc}}(\mathbf{Q}, R_1, R_2) = \max_{\rho_1 \in [0,1], \rho_2 \in [0,1]} E_{0,12,2}^{\text{cc}}(\mathbf{Q}, \rho_1, \rho_2) - \rho_2(R_2 + \rho_1 R_1), \quad (4.61)$$

where

$$E_{0,1}^{\text{cc}}(\mathbf{Q}, \rho) \triangleq \sup_{s \geq 0, a_1(\cdot), a'_1(\cdot), a'_2(\cdot)} -\log \mathbb{E} \left[\left(\frac{\mathbb{E}[q(\bar{X}_1, X_2, Y)^s e^{a_1(\bar{X}_1)} | X_2, Y]}{q(X_1, X_2, Y)^s e^{a_1(X_1)}} \right)^\rho e^{\sum_{\nu=1}^2 a'_\nu(X_\nu) - \phi'_\nu} \right] \quad (4.62)$$

$$E_{0,2}^{\text{cc}}(\mathbf{Q}, \rho) \triangleq \sup_{s \geq 0, a_2(\cdot), a'_1(\cdot), a'_2(\cdot)} -\log \mathbb{E} \left[\left(\frac{\mathbb{E}[q(X_1, \bar{X}_2, Y)^s e^{a_2(\bar{X}_2)} | X_1, Y]}{q(X_1, X_2, Y)^s e^{a_2(X_2)}} \right)^\rho e^{\sum_{\nu=1}^2 a'_\nu(X_\nu) - \phi'_\nu} \right] \quad (4.63)$$

$$E_{0,12,1}^{\text{cc}}(\mathbf{Q}, \rho_1, \rho_2) \triangleq \sup_{s \geq 0, a_1(\cdot), a_2(\cdot), a'_1(\cdot), a'_2(\cdot)} -\log \mathbb{E} \left[\left(\mathbb{E} \left[\left(\frac{\mathbb{E}[q(\bar{X}_1, \bar{X}_2, Y)^s e^{a_2(\bar{X}_2)} | \bar{X}_1]}{q(X_1, X_2, Y)^s e^{a_2(X_2)}} \right)^{\rho_2} \frac{e^{a_1(\bar{X}_1)}}{e^{a_1(X_1)}} \middle| X_1, X_2, Y \right] \right)^{\rho_1} e^{\sum_{\nu=1}^2 a'_\nu(X_\nu) - \phi'_\nu} \right] \quad (4.64)$$

$$E_{0,12,2}^{\text{cc}}(\mathbf{Q}, \rho_1, \rho_2) \triangleq \sup_{s \geq 0, a_1(\cdot), a_2(\cdot), a'_1(\cdot), a'_2(\cdot)} -\log \mathbb{E} \left[\left(\mathbb{E} \left[\left(\frac{\mathbb{E}[q(\bar{X}_1, \bar{X}_2, Y)^s e^{a_1(\bar{X}_1)} | \bar{X}_2]}{q(X_1, X_2, Y)^s e^{a_1(X_1)}} \right)^{\rho_1} \frac{e^{a_2(\bar{X}_2)}}{e^{a_2(X_2)}} \middle| X_1, X_2, Y \right] \right)^{\rho_2} e^{\sum_{\nu=1}^2 a'_\nu(X_\nu) - \phi'_\nu} \right], \quad (4.65)$$

and where $(X_1, X_2, Y, \bar{X}_1, \bar{X}_2) \sim Q_1(x_1)Q_2(x_2)W(y|x_1, x_2)Q_1(\bar{x}_1)Q_2(\bar{x}_2)$ and $\phi'_\nu \triangleq \mathbb{E}_{Q_\nu}[a'_\nu(X_\nu)]$ ($\nu = 1, 2$).

Proof. Each expression follows from its counterpart in (4.29)–(4.30) and (4.49)–(4.50) using Lemma E.1 and Fan's minimax theorem [78]; details are given in Section 4.5.1. $\square$

The expressions in (4.60)–(4.61) contain two parameters ρ_1 and ρ_2 , thus differing from the usual error exponents containing a single value of ρ (see Section 2.4). To our knowledge, the first exponents of this form were by Poltyrev [116] in the setting of broadcast channels with degraded message sets.

The following dual expression for the achievable rate region follows from Theorem 4.4.

Theorem 4.5. *The region $\mathcal{R}_{\text{LM}}(\mathbf{Q})$ described by (4.42)–(4.44) can be expressed as the set of rate pairs (R_1, R_2) satisfying*

$$R_1 \leq \sup_{s \geq 0, a_1(\cdot)} \mathbb{E} \left[\log \frac{q(X_1, X_2, Y)^s e^{a_1(X_1)}}{\mathbb{E}[q(\bar{X}_1, X_2, Y)^s e^{a_1(\bar{X}_1)} | X_2, Y]} \right] \quad (4.66)$$

$$R_2 \leq \sup_{s \geq 0, a_2(\cdot)} \mathbb{E} \left[\log \frac{q(X_1, X_2, Y)^s e^{a_2(X_2)}}{\mathbb{E}[q(X_1, \bar{X}_2, Y)^s e^{a_2(\bar{X}_2)} | X_1, Y]} \right], \quad (4.67)$$

and at least one of

$$R_1 \leq \sup_{\rho_2 \in [0,1], s \geq 0, a_1(\cdot), a_2(\cdot)} \mathbb{E} \left[\log \frac{(q(X_1, X_2, Y)^s e^{a_2(X_2)})^{\rho_2} e^{a_1(X_1)}}{\mathbb{E} \left[\left(\mathbb{E}[q(\bar{X}_1, \bar{X}_2, Y)^s e^{a_2(\bar{X}_2)} | \bar{X}_1] \right)^{\rho_2} e^{a_1(\bar{X}_1)} | Y \right]} \right] - \rho_2 R_2 \quad (4.68)$$

$$R_2 \leq \sup_{\rho_1 \in [0,1], s \geq 0, a_1(\cdot), a_2(\cdot)} \mathbb{E} \left[\log \frac{(q(X_1, X_2, Y)^s e^{a_1(X_1)})^{\rho_1} e^{a_2(X_2)}}{\mathbb{E} \left[\left(\mathbb{E}[q(\bar{X}_1, \bar{X}_2, Y)^s e^{a_1(\bar{X}_1)} | \bar{X}_2] \right)^{\rho_1} e^{a_2(\bar{X}_2)} | Y \right]} \right] - \rho_1 R_1. \quad (4.69)$$

where $(X_1, X_2, Y, \bar{X}_1, \bar{X}_2) \sim Q_1(x_1)Q_2(x_2)W(y|x_1, x_2)Q_1(\bar{x}_1)Q_2(\bar{x}_2)$.

Proof. The conditions in (4.66)–(4.69) follow from (4.58)–(4.61) respectively using a similar argument to Gallager’s single-user analysis [52]. For example, (4.68) follows from (4.60) by setting $a'_1(\cdot) = 0$ and $a'_2(\cdot) = 0$, fixing $s, a_1(\cdot), a_2(\cdot)$ and ρ_2 , and evaluating the partial derivative of the objective at $\rho_1 = 0$. The equivalence to (4.42)–(4.44) follows by using Lemma 4.2 and applying Lagrange duality techniques analogously to Theorem 4.4. For example, the equivalence of (4.56) and (4.68) follows from Steps 1–2 of the proof of Theorem 4.4 in Section 4.5.1, using $Q_1 \times Q_2 \times W$ in place of $P_{X_1 X_2 Y}$. $\square$

4.2.3 Exponents and Rates for More General Alphabets

In this subsection, we show that the expressions given in Theorems 4.4 and 4.5 remain valid for the memoryless MAC with more general alphabets. To this end, we use cost-constrained random coding, and we avoid the use of types. Our analysis holds for input distributions (Q_1, Q_2) and channels W representing discrete transition laws, density functions, or a mixture of the two.

We make use of the cost-constrained ensemble defined in the same way as the single-user

setting (see Section 2.3): We fix $Q_1 \in \mathcal{P}(\mathcal{X}_1)$ and $Q_2 \in \mathcal{P}(\mathcal{X}_1)$, and choose

$$P_{\mathbf{X}_\nu}(\mathbf{x}_\nu) = \frac{1}{\mu_{\nu,n}} \prod_{i=1}^n Q_\nu(x_{\nu,i}) \mathbb{1}\{\mathbf{x}_\nu \in \mathcal{D}_{\nu,n}\}, \quad (4.70)$$

where for $\nu = 1, 2$, $\mu_{\nu,n}$ is a normalizing constant, and

$$\mathcal{D}_{\nu,n} \triangleq \left\{ \mathbf{x}_\nu : \left| \frac{1}{n} \sum_{i=1}^n a_{\nu,l}(x_{\nu,i}) - \phi_{\nu,l} \right| \leq \frac{\delta}{n}, l = 1, \dots, L_\nu \right\}, \quad (4.71)$$

where $\{a_{\nu,l}\}_{l=1}^{L_\nu}$ are auxiliary cost functions, δ is a positive constant, and $\phi_{\nu,l} \triangleq \mathbb{E}_{Q_\nu}[a_{\nu,l}(X_\nu)]$. The case $L_\nu = 0$ should be understood as corresponding to the case that $\mathcal{D}_{\nu,n}$ contains all $\mathbf{x}_\nu$ sequences, thus recovering the i.i.d. distribution studied in [105].

While our focus is on unconstrained channels, a similar discussion to that of Section 2.9 applies here, and our results readily extend to multiple-access channels with input constraints. We recall from Proposition 2.1 that $\mu_{\nu,n} = \Omega(n^{-L_\nu/2})$ provided that $\mathbb{E}_{Q_\nu}[a_{\nu,l}(X_\nu)^2] < \infty$ for $l = 1, \dots, L_\nu$.

Theorem 4.6. *For any mismatched memoryless MAC (possibly having infinite or continuous alphabets) and pair (Q_1, Q_2) , the error exponent*

$$E_r^{\text{cc}}(\mathbf{Q}, R_1, R_2) = \min \left\{ E_{r,1}^{\text{cc}}(\mathbf{Q}, R_1), E_{r,2}^{\text{cc}}(\mathbf{Q}, R_2), \max \left\{ E_{r,12,1}^{\text{cc}}(\mathbf{Q}, R_1, R_2), E_{r,12,2}^{\text{cc}}(\mathbf{Q}, R_1, R_2) \right\} \right\} \quad (4.72)$$

is achievable for $E_{r,1}^{\text{cc}}$, $E_{r,2}^{\text{cc}}$, $E_{r,12,1}^{\text{cc}}$ and $E_{r,12,2}^{\text{cc}}$ defined in (4.58)–(4.65), provided that (i) each maximum over $\rho \in [0, 1]$ is replaced by a supremum, and (ii) each supremum over $a_\nu(\cdot)$ and $a'_\nu(\cdot)$ is subject to $\mathbb{E}_{Q_\nu}[a_\nu(X_\nu)^2] < \infty$ and $\mathbb{E}_{Q_\nu}[a'_\nu(X_\nu)^2] < \infty$ ($\nu = 1, 2$). Furthermore, this exponent can be achieved using the cost-constrained ensemble with $L_1 = L_2 = 5$.

Proof. We first consider $E_{r,12,1}^{\text{cc}}$. We define $Q_\nu^n(\mathbf{x}_\nu) \triangleq \prod_{i=1}^n Q_\nu(x_{\nu,i})$ for $\nu = 1, 2$. Expanding (4.16) and applying Markov's inequality and $\min\{1, \alpha\} \leq \alpha^\rho$ ($\rho \in [0, 1]$), we obtain¹

$$\begin{aligned} \text{rcu}_{12,1}(n, M_1) &\leq \sum_{\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}} P_{\mathbf{X}_1}(\mathbf{x}_1) P_{\mathbf{X}_2}(\mathbf{x}_2) W^n(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \\ &\quad \times \left(M_1 \sum_{\bar{\mathbf{x}}_1} P_{\mathbf{X}_1}(\bar{\mathbf{x}}_1) \left(M_2 \frac{\sum_{\bar{\mathbf{x}}_2} P_{\mathbf{X}_2}(\bar{\mathbf{x}}_2) q^n(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{y})^s}{q^n(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})^s} \right)^{\rho_2} \right)^{\rho_1} \end{aligned} \quad (4.73)$$

¹In the case of continuous alphabets, the summations should be replaced by integrals.

for any $\rho_1 \in [0, 1]$, $\rho_2 \in [0, 1]$ and $s \geq 0$. For $\nu = 1, 2$, we let $a_\nu(x)$ and $a'_\nu(x)$ be two of the five cost functions in the ensemble, and we define $a_\nu^n(\mathbf{x}_\nu) \triangleq \sum_{i=1}^n a_\nu(x_{\nu,i})$, $a'_\nu^n(\mathbf{x}_\nu) \triangleq \sum_{i=1}^n a'_\nu(x_{\nu,i})$, $\phi_\nu \triangleq \mathbb{E}_{Q_\nu}[a_\nu(X_\nu)]$, and $\phi'_\nu \triangleq \mathbb{E}_{Q_\nu}[a'_\nu(X_\nu)]$. In accordance with the theorem statement, we assume that $\mathbb{E}_{Q_\nu}[a_\nu(X_\nu)^2] < \infty$ and $\mathbb{E}_{Q_\nu}[a'_\nu(X_\nu)^2] < \infty$. Using the bounds on the cost functions in (4.71), we can weaken (4.73) to

$$\begin{aligned} \text{rcu}_{12,1}(n, M_1) &\leq e^{2\delta(\rho_1 + \rho_1\rho_2 + 1)} \sum_{\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}} P_{\mathbf{X}_1}(\mathbf{x}_1) P_{\mathbf{X}_2}(\mathbf{x}_2) W^n(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) \left(M_1 \sum_{\bar{\mathbf{x}}_1} P_{\mathbf{X}_1}(\bar{\mathbf{x}}_1) \right. \\ &\quad \times \left(M_2 \frac{\sum_{\bar{\mathbf{x}}_2} P_{\mathbf{X}_2}(\bar{\mathbf{x}}_2) q^n(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{y})^s e^{a_2^n(\bar{\mathbf{x}}_2)}}{q^n(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})^s e^{a_2^n(\mathbf{x}_2)}} \right)^{\rho_2} \frac{e^{a_1^n(\bar{\mathbf{x}}_1)}}{e^{a_1^n(\mathbf{x}_1)}} \Big)^{\rho_1} e^{\sum_{\nu=1}^2 a'_\nu^n(\mathbf{x}_\nu) - n\phi_\nu}. \end{aligned} \quad (4.74)$$

We further upper bound (4.74) by substituting (4.70) and replacing the summations over $\mathcal{D}_{\nu,n}$ by summations over all sequences on $\mathcal{X}_\nu^n$. Writing the resulting terms (e.g. $W^n(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)$) as a product from 1 to n and taking the supremum over (s, ρ_1, ρ_2) and the cost functions, we obtain a bound whose exponent is given by (4.60), with a prefactor of $\frac{e^{2\delta(\rho_1 + \rho_1\rho_2 + 1)}}{\mu_{1,n}^{1+\rho_1} \mu_{2,n}^{1+\rho_1\rho_2}}$. From Proposition 2.1 and the assumptions on the second moments of the cost functions, this prefactor is subexponential in n .

Dealing with the other three exponents in (4.58)–(4.65) similarly, we conclude that it suffices to choose $L_1 = L_2 = 5$. To see this, one can simply count the number of occurrences of each $a_\nu(\cdot)$ and $a'_\nu(\cdot)$ in (4.62)–(4.65). It suffices to let the cost functions for $E_{0,12,1}^{\text{cc}}$ and $E_{0,12,2}^{\text{cc}}$ coincide, since the error exponent for the type-12 event is the maximum of the two. $\square$

From Theorem 4.6, we immediately obtain the following generalization of Theorem 4.5.

Theorem 4.7. *For any mismatched MAC (possibly having infinite or continuous alphabets) and pair (Q_1, Q_2) , the rate region $\mathcal{R}_{\text{LM}}(\mathbf{Q})$ described by (4.66)–(4.69) is achievable, where each supremum is subject to $\mathbb{E}_{Q_\nu}[a_\nu(X_\nu)^2] < \infty$ ($\nu = 1, 2$). Furthermore, any point in the region can be achieved using cost-constrained random coding with $L_1 = L_2 = 3$.*

Proof. The first part is proved similarly to Theorem 4.5, and the sufficiency of $L_1 = L_2 = 3$ follows since $a'_1(\cdot)$ and $a'_2(\cdot)$ are set to zero therein. $\square$

4.2.4 Application to the Matched MAC

In this subsection, we apply our results to the setting of ML decoding, where $q(x_1, x_2, y) = W(y|x_1, x_2)$. As mentioned in Section 4.1.2, an exponent for the constant-composition ensemble was derived by Liu and Hughes [66], and was shown to yield a strict improvement

over Gallager's exponent for the i.i.d. ensemble [105] even after the optimization of the input distributions.

Simplified Primal Expressions for the Exponents

We will show that, in contrast to the mismatched setting, the overall error exponent E_r^{cc} given in (4.40) is unchanged when $E_{r,12}^{\text{cc}'}$ in (4.41) is used in place of $E_{r,12}^{\text{cc}}$. We begin with the following proposition, which shows that the exponents $(E_{r,1}^{\text{cc}}, E_{r,2}^{\text{cc}}, E_{r,12}^{\text{cc}'})$ (see (4.29)–(4.30) and (4.41)) under ML decoding coincide with those of Liu and Hughes in the absence of time sharing [66].

Proposition 4.1. *Under ML decoding (i.e. $q = W$), $E_{r,1}^{\text{cc}}$, $E_{r,2}^{\text{cc}}$ and $E_{r,12}^{\text{cc}'}$ can be expressed as*

$$E_{r,1}^{\text{cc}}(\mathbf{Q}, R_1) = \min_{P_{X_1 X_2 Y} \in \mathcal{S}(\mathbf{Q})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + [I_P(X_1; X_2, Y) - R_1]^+ \quad (4.75)$$

$$E_{r,2}^{\text{cc}}(\mathbf{Q}, R_2) = \min_{P_{X_1 X_2 Y} \in \mathcal{S}(\mathbf{Q})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + [I_P(X_2; X_1, Y) - R_2]^+ \quad (4.76)$$

$$E_{r,12}^{\text{cc}'}(\mathbf{Q}, R_1, R_2) = \min_{P_{X_1 X_2 Y} \in \mathcal{S}(\mathbf{Q})} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + [D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times P_Y) - (R_1 + R_2)]^+. \quad (4.77)$$

Proof. The proof is nearly identical to that of [56, Lemma 9], so we provide only an outline, and we focus on the type-12 exponent. Consider any pair $(P_{X_1 X_2 Y}, \tilde{P}_{X_1 X_2 Y})$ satisfying the constraints of (4.41). If $D(\tilde{P}_{X_1 X_2 Y} \| Q_1 \times Q_2 \times P_Y) \geq D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times P_Y)$, we can lower bound the objective of (4.41) by that of (4.77). In the remaining case, we may use the constraint $\mathbb{E}_{\tilde{P}}[\log W] \geq \mathbb{E}_P[\log W]$ to lower bound the objective in (4.41) by that of (4.77) with $\tilde{P}_{X_1 X_2 Y}$ in place of $P_{X_1 X_2 Y}$. This proves that (4.77) lower bounds (4.41), and the matching upper bound follows immediately from the fact that $\tilde{P}_{X_1 X_2 Y} = P_{X_1 X_2 Y}$ satisfies the constraints of the minimization in (4.41). $\square$

The statement that $E_{r,12}^{\text{cc}}$ can be replaced by $E_{r,12}^{\text{cc}'}$ is formally given as follows.

Theorem 4.8. *Under ML decoding (i.e. $q = W$), we have for any input distributions (Q_1, Q_2) and rates (R_1, R_2) that*

$$\begin{aligned} & \min \{ E_{r,1}^{\text{cc}}(\mathbf{Q}, R_1), E_{r,2}^{\text{cc}}(\mathbf{Q}, R_2), E_{r,12}^{\text{cc}}(\mathbf{Q}, R_1, R_2) \} \\ &= \min \{ E_{r,1}^{\text{cc}}(\mathbf{Q}, R_1), E_{r,2}^{\text{cc}}(\mathbf{Q}, R_2), E_{r,12}^{\text{cc}'}(\mathbf{Q}, R_1, R_2) \}. \end{aligned} \quad (4.78)$$

Proof. We know that $E_{r,12}^{\text{cc}} \geq E_{r,12}^{\text{cc}'}$, and hence the left-hand side of (4.78) is greater than or equal to the right-hand side. It remains to prove the converse. From the definition of $\mathcal{T}_{12}(P_{X_1X_2Y})$, $\tilde{P}_{X_1X_2Y} = P_{X_1X_2Y}$ always satisfies the constraints of (4.31), and hence

$$E_{r,12}^{\text{cc}}(\mathbf{Q}, R_1, R_2) \leq F_{12}(\mathbf{Q}, R_1, R_2), \quad (4.79)$$

where

$$\begin{aligned} F_{12}(\mathbf{Q}, R_1, R_2) &\triangleq \min_{P_{X_1X_2Y} \in \mathcal{S}(\mathbf{Q})} D(P_{X_1X_2Y} \| Q_1 \times Q_2 \times W) \\ &+ \left[\max \left\{ I_P(X_1; Y) - R_1, I_P(X_2; Y) - R_2, D(P_{X_1X_2Y} \| Q_1 \times Q_2 \times P_Y) - R_1 - R_2 \right\} \right]^+. \end{aligned} \quad (4.80)$$

We will prove (4.78) by showing that

$$\begin{aligned} &\min \{ E_{r,1}^{\text{cc}}(\mathbf{Q}, R_1), E_{r,2}^{\text{cc}}(\mathbf{Q}, R_2), F_{12}(\mathbf{Q}, R_1, R_2) \} \\ &\leq \min \{ E_{r,1}^{\text{cc}}(\mathbf{Q}, R_1), E_{r,2}^{\text{cc}}(\mathbf{Q}, R_2), E_{r,12}^{\text{cc}'}(\mathbf{Q}, R_1, R_2) \}. \end{aligned} \quad (4.81)$$

It suffices to show that whenever F_{12} exceeds $E_{r,12}^{\text{cc}'}$, F_{12} also greater than or equal to either $E_{r,1}^{\text{cc}}$ or $E_{r,2}^{\text{cc}}$. Comparing (4.77) and (4.80), the objective in (4.80) only exceeds that of (4.77) when the maximum in (4.80) is achieved by $I_P(X_1; Y) - R_1$ or $I_P(X_2; Y) - R_2$. We show that the former implies $F_{12} \geq E_{r,1}^{\text{cc}}$; it can similarly be shown that the latter implies $F_{12} \geq E_{r,2}^{\text{cc}}$. If $I_P(X_1; Y) - R_1$ achieves the maximum, we have

$$I_P(X_1; Y) - R_1 \geq D(P_{X_1X_2Y} \| Q_1 \times Q_2 \times P_Y) - R_1 - R_2. \quad (4.82)$$

Using the identity

$$D(P_{X_1X_2Y} \| Q_1 \times Q_2 \times P_Y) = I_P(X_1; Y) + I_P(X_2; X_1, Y), \quad (4.83)$$

we can equivalently write (4.82) as $I_P(X_2; X_1, Y) \leq R_2$. For any $P_{X_1X_2Y}$ satisfying this property, the objective in (4.76) equals $D(P_{X_1X_2Y} \| Q_1 \times Q_2 \times W)$, and thus cannot exceed the objective in (4.80). It follows that $F_{12} \geq E_{r,1}^{\text{cc}}$. $\square$

While it is possible that $E_{r,12}^{\text{cc}} > E_{r,12}^{\text{cc}'}$ under ML decoding, Theorem 4.8 shows that this never occurs in the region where $E_{r,12}^{\text{cc}}$ achieves the minimum in (4.40). Thus, combining Theorem 4.8 and Theorem 4.2, we conclude that the exponent given in [66] is the true exponent

for the constant-composition ensemble under ML decoding. While we have only proved this in the absence of time sharing, the same holds true when time sharing is used; see Section 4.2.5 for further discussion.

Dual Expressions for the Exponents

The following theorem gives dual expressions for the exponents of Liu and Hughes [66] in a form closely resembling Gallager's exponents for the i.i.d. ensemble [105].

Theorem 4.9. *Under ML decoding (i.e. $q = W$), the error exponents in (4.75)–(4.77) can be expressed as*

$$E_{r,1}^{\text{cc}}(\mathbf{Q}, R_1) = \max_{\rho \in [0,1]} E_{0,1}^{\text{cc}}(\mathbf{Q}, \rho) - \rho R_1 \quad (4.84)$$

$$E_{r,2}^{\text{cc}}(\mathbf{Q}, R_2) = \max_{\rho \in [0,1]} E_{0,2}^{\text{cc}}(\mathbf{Q}, \rho) - \rho R_2 \quad (4.85)$$

$$E_{r,12}^{\text{cc}'}(\mathbf{Q}, R_1, R_2) = \max_{\rho \in [0,1]} E_{0,12}^{\text{cc}'}(\mathbf{Q}, \rho) - \rho(R_1 + R_2), \quad (4.86)$$

where

$$E_{0,1}^{\text{cc}}(\mathbf{Q}, \rho) = \sup_{a_1(\cdot), a_2(\cdot)} -\log \sum_{x_2, y} Q_2(x_2) \left(\sum_{x_1} Q_1(x_1) W(y|x_1, x_2)^{\frac{1}{1+\rho}} e^{\sum_{\nu=1}^2 a_\nu(x_\nu) - \phi_\nu} \right)^{1+\rho} \quad (4.87)$$

$$E_{0,2}^{\text{cc}}(\mathbf{Q}, \rho) = \sup_{a_1(\cdot), a_2(\cdot)} -\log \sum_{x_1, y} Q_1(x_1) \left(\sum_{x_2} Q_2(x_2) W(y|x_1, x_2)^{\frac{1}{1+\rho}} e^{\sum_{\nu=1}^2 a_\nu(x_\nu) - \phi_\nu} \right)^{1+\rho} \quad (4.88)$$

$$E_{0,12}^{\text{cc}'}(\mathbf{Q}, \rho) = \sup_{a_1(\cdot), a_2(\cdot)} -\log \sum_y \left(\sum_{x_1, x_2} Q_1(x_1) Q_2(x_2) W(y|x_1, x_2)^{\frac{1}{1+\rho}} e^{\sum_{\nu=1}^2 a_\nu(x_\nu) - \phi_\nu} \right)^{1+\rho}, \quad (4.89)$$

and $\phi_\nu \triangleq \mathbb{E}_{Q_\nu}[a_\nu(X_\nu)]$. Furthermore, these exponents are achievable in the case of infinite and continuous alphabets (with the summations replaced by integrals where necessary) using the cost-constrained ensemble with $L_1 = L_2 = 3$, provided that (i) each maximum over $\rho \in [0, 1]$ is replaced by a supremum, and (ii) each supremum over $a_\nu(\cdot)$ ($\nu = 1, 2$) is subject to $\mathbb{E}_{Q_\nu}[a_\nu(X_\nu)^2] < \infty$.

Proof. By definition, $E_{r,12}^{\text{cc}'}$ is obtained from either of (4.49)–(4.50) by removing the first term in the $\max\{\cdot, \cdot\}$. From the proof of Theorem 4.4, we conclude that the analogous statement in the dual expressions (4.58)–(4.65) is that the supremum over ρ_2 in (4.64) (respectively, over

ρ_1 in (4.65)) is replaced by $\rho_2 = 1$ (respectively, $\rho_1 = 1$). With these substitutions, (4.64) and (4.65) become equal. Thus, it suffices to show that (4.62)–(4.64) (with $\rho_2 = 1$ in (4.64)) coincide with (4.87)–(4.89) under ML decoding. We focus on (4.64) and (4.89), since the others are handled similarly.

Setting $\rho_2 = 1$ in (4.64) and renaming ρ_1 as ρ , we obtain the E_0 function

$$\begin{aligned}
 E_{0,12}^{\text{cc}''}(\mathbf{Q}, \rho) &\triangleq -\log \sum_{x_1, x_2, y} Q_1(x_1) Q_2(x_2) W(y|x_1, x_2) \\
 &\quad \times \left(\frac{\sum_{\bar{x}_1, \bar{x}_2} Q_1(\bar{x}_1) Q_2(\bar{x}_2) W(y|\bar{x}_1, \bar{x}_2)^s e^{\sum_{\nu=1}^2 a_\nu(\bar{x}_\nu) - \phi_\nu}}{W(y|x_1, x_2)^s e^{\sum_{\nu=1}^2 a_\nu(x_\nu) - \phi_\nu}} \right)^\rho e^{\sum_{\nu=1}^2 a'_\nu(x_\nu) - \phi'_\nu} \quad (4.90) \\
 &= -\log \sum_y \left(\sum_{x_1, x_2} Q_1(x_1) Q_2(x_2) W(y|x_1, x_2)^{1-\rho s} e^{\sum_{\nu=1}^2 a'_\nu(x_\nu) - \phi'_\nu - \rho(a_\nu(x_\nu) - \phi_\nu)} \right) \\
 &\quad \times \left(\sum_{\bar{x}_1, \bar{x}_2} Q_1(\bar{x}_1) Q_2(\bar{x}_2) W(y|\bar{x}_1, \bar{x}_2)^s e^{\sum_{\nu=1}^2 a_\nu(\bar{x}_\nu) - \phi_\nu} \right)^\rho, \quad (4.91)
 \end{aligned}$$

where in (4.90) we multiplied the numerator and denominator of the fraction by $e^{-(\phi_1 + \phi_2)}$. Using Hölder's inequality in an identical fashion to (2.58)–(2.60) (see also [2, Ex. 5.6]), the summation over y in (4.91) is lower bounded by

$$\sum_y \left(\sum_{x_1, x_2} Q_1(x_1) Q_2(x_2) W(y|x_1, x_2)^{\frac{1}{1+\rho}} e^{\frac{1}{1+\rho} \sum_{\nu=1}^2 a'_\nu(x_\nu) - \phi'_\nu} \right)^{1+\rho} \quad (4.92)$$

and equality holds when $s = \frac{1}{1+\rho}$ and $a'_\nu(x_\nu) = (1 + \rho)a_\nu(x_\nu)$. By handling the other E_0 functions similarly, we obtain the first part of the theorem. For the second part, it remains to show that the choice $L_1 = L_2 = 3$ suffices. This is seen by noting that whenever two cost functions are scalar multiples of each other (in this case, $a_\nu(\cdot)$ and $a'_\nu(\cdot)$ for $\nu = 1, 2$) one of them can be removed without affecting the exponent obtained. $\square$

It is possible to derive Theorem 4.9 more directly, without using the mismatched decoding results of the previous sections; see [71, Sec. V] for details. Gallager's exponents [105] are recovered by setting $a_1(\cdot) = a_2(\cdot) = 0$ in (4.87)–(4.89).

4.2.5 Time-Sharing

In this subsection, we discuss time sharing for the MAC. We begin with a brief overview of time sharing for the matched case, which is included in order to elucidate the subtle issues that arise in the mismatched setting.

Matched Setting

For any pair (Q_1, Q_2) , the achievable rate region for ML decoding is pentagonal (or possibly rectangular) and therefore convex. However, the region obtained by taking the union over all (Q_1, Q_2) is non-convex in general [117]. To prove that the rate pairs in the convex hull of this union are also achievable, one can use the standard time sharing argument in which, for some $\lambda \in (0, 1)$, one pair of codebooks is used for λn symbols of the block, and another pair of codebooks is used for the remaining $(1 - \lambda)n$ symbols. We refer to this scheme as explicit time sharing, or simply time sharing.

A drawback of explicit time sharing is that the corresponding error exponents can be low. If $E_r^{(1)}$ is the exponent for some pair $(R_1^{(1)}, R_2^{(1)})$, and $E_r^{(2)}$ is the exponent for another pair $(R_1^{(2)}, R_2^{(2)})$, then the resulting exponent for the weighted rate pair

$$R_\nu = \lambda R_\nu^{(1)} + (1 - \lambda) R_\nu^{(2)}, \quad \nu = 1, 2 \quad (4.93)$$

is given by

$$E_r(R_1, R_2) = \min \left\{ \lambda E_r^{(1)}(R_1^{(1)}, R_2^{(1)}), (1 - \lambda) E_r^{(2)}(R_1^{(2)}, R_2^{(2)}) \right\}. \quad (4.94)$$

This may be particularly low if λ is close to zero or one. Nevertheless, we have $E_r(R_1, R_2) > 0$ provided that $E_r^{(1)} > 0$ and $E_r^{(2)} > 0$.

Higher error exponents can be obtained by considering the coded time sharing ensemble [115, Sec. 4.5.3], which directly achieves any point within the capacity region. Each codeword is generated conditionally on a (possibly random) time sharing sequence U in $\mathcal{U}^n$, where $\mathcal{U}$ is a finite time sharing alphabet. For example, one may consider conditional i.i.d. random coding:

$$P_{\mathbf{X}_\nu | U}(\mathbf{x}_\nu | \mathbf{u}) = \prod_{i=1}^n Q_\nu(x_{\nu,i} | u_i), \quad (4.95)$$

where Q_1 and Q_2 are (conditional) input distributions. Let $E_0^{(u)}$ denote the E_0 function for some error type under i.i.d. random coding on $Q_1(\cdot | u)$. Using coded time sharing with $\mathcal{U} = \{1, 2\}$, $Q_U = (\lambda, 1 - \lambda)$, and U equal to a deterministic sequence $\mathbf{u}$ whose empirical distribution is that closest to Q_U , we obtain the E_0 function [105]

$$E_0 = \lambda E_0^{(1)} + (1 - \lambda) E_0^{(2)}. \quad (4.96)$$

Thus, in contrast to explicit time sharing, the exponent does not tend to zero as λ approaches zero or one.

Similar improvements apply for the constant-composition and cost-constrained ensembles. For the former, we write

$$P_{\mathbf{X}_\nu|U}(\mathbf{x}_\nu|\mathbf{u}) = \frac{1}{|T_{\mathbf{u}}(Q_{\nu,n})|} \mathbb{1}\{\mathbf{x}_\nu \in T_{\mathbf{u}}(Q_{\nu,n})\}, \quad (4.97)$$

where $Q_{\nu,n}$ is the closest conditional type to Q_ν for $\nu = 1, 2$, and $\mathbf{u}$ is the same as in the i.i.d. ensemble. The cost-constrained ensemble is described by

$$P_{\mathbf{X}_\nu|U}(\mathbf{x}_\nu|\mathbf{u}) = \frac{1}{\mu_{\nu,n}} \prod_{i=1}^n Q_\nu(x_{\nu,i}|u_i) \mathbb{1}\{\mathbf{x}_\nu \in \mathcal{D}_{\nu,n}\}, \quad (4.98)$$

where

$$\mathcal{D}_{\nu,n} \triangleq \left\{ \mathbf{x}_\nu : \left| \sum_{i=1}^n a_{\nu,l}(u_i, x_{\nu,i}) - \phi_{\nu,l} \right| \leq \frac{\delta}{n}, l = 1, \dots, L_\nu \right\}, \quad (4.99)$$

and where each quantity is defined similarly to (4.71) (e.g. $\phi_{\nu,l} \triangleq \mathbb{E}_{Q_U \times Q_\nu}[a_{\nu,l}(U, X_\nu)]$).

Mismatched Setting

Under mismatched decoding, explicit time sharing behaves similarly to the matched setting. For time sharing between two achievable pairs $(R_1^{(1)}, R_2^{(1)})$ and $(R_1^{(2)}, R_2^{(2)})$ with exponents $E_r^{(1)}$ and $E_r^{(2)}$, one can achieve the rate pair of (4.93), with the exponent given by (4.94).

It may seem reasonable to believe that one can use coded time sharing with $|\mathcal{U}| = 2$ to prove the achievability of any point on the line segment between two achievable points $(R_1^{(1)}, R_2^{(1)})$ and $(R_1^{(2)}, R_2^{(2)})$. However, this is not true in general in the mismatched setting. To see this, we outline some of the results for coded time sharing. Since the changes compared to the case $\mathcal{U} = \emptyset$ are similar for each error type, we focus on the type-12 event.

For general conditional codeword distributions $P_{\mathbf{X}_\nu|U}$ with a (possibly random) time sharing sequence $U \sim P_U$, the RCU bounds in (4.14)–(4.17) should be modified so that the outer expectations are over $(U, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$. Following the analysis of Section 4.2.2, the analogous exponent to $E_{r,12}^{\text{cc}}$ in (4.31) is given by

$$\begin{aligned} E_{r,12}^{\text{cc-ts}}(\mathbf{Q}, R_1, R_2) &\triangleq \min_{P_{UX_1X_2Y} \in \mathcal{S}^{\text{ts}}(\mathbf{Q})} \min_{\tilde{P}_{X_1X_2Y} \in \mathcal{T}_{12}^{\text{ts}}(P_{UX_1X_2Y})} \\ &\quad D(P_{UX_1X_2Y} \| Q_U \times Q_1 \times Q_2 \times W) + \left[\max \left\{ I_{\tilde{P}}(X_1; Y|U) - R_1, \right. \right. \\ &\quad \left. \left. I_{\tilde{P}}(X_2; Y|U) - R_2, D(\tilde{P}_{UX_1X_2Y} \| Q_U \times Q_1 \times Q_2 \times \tilde{P}_Y) - R_1 - R_2 \right\} \right]^+, \end{aligned} \quad (4.100)$$

where

$$\mathcal{S}^{\text{ts}}(\mathbf{Q}) \triangleq \left\{ P_{UX_1X_2Y} \in \mathcal{P}(\mathcal{U} \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : P_U = Q_U, P_{X_1|U} = Q_1, P_{X_2|U} = Q_2 \right\} \quad (4.101)$$

$$\begin{aligned} \mathcal{T}_{12}^{\text{ts}}(P_{UX_1X_2Y}) \triangleq & \left\{ \tilde{P}_{UX_1X_2Y} \in \mathcal{P}(\mathcal{U} \times \mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \tilde{P}_{X_1|U} = P_{X_1|U}, \tilde{P}_{X_2|U} = P_{X_2|U}, \right. \\ & \left. \tilde{P}_{UY} = P_{UY}, \mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)] \right\}. \end{aligned} \quad (4.102)$$

In the dual domain, the analogous quantity to $E_{0,12,1}^{\text{cc}}$ in (4.64) is given by

$$\begin{aligned} E_{0,12,1}^{\text{cc-ts}}(\mathbf{Q}, \rho_1, \rho_2) \triangleq & \sup_{s \geq 0, a_1(\cdot, \cdot), a_2(\cdot, \cdot), a'_1(\cdot, \cdot), a'_2(\cdot, \cdot)} - \sum_u Q_U(u) \log \mathbb{E} \left[\right. \\ & \left. \left(\mathbb{E} \left[\left(\frac{\mathbb{E}[q(\bar{X}_1, \bar{X}_2, Y)^s e^{a_2(u, \bar{X}_2)} | \bar{X}_1]}{q(X_1, X_2, Y)^s e^{a_2(u, X_2)}} \right)^{\rho_2} \frac{e^{a_1(u, \bar{X}_1)}}{e^{a_1(u, X_1)}} \middle| X_1, X_2, Y \right] \right)^{\rho_1} e^{\sum_{\nu=1}^2 a'_\nu(u, X_\nu) - \phi'_\nu} \right], \end{aligned} \quad (4.103)$$

where for each u , the expectations are taken with respect to the conditional distribution $(X_1, X_2, Y, \bar{X}_1, \bar{X}_2 | U = u) \sim Q_1(x_1|u)Q_2(x_2|u)W(y|x_1, x_2)Q_1(\bar{x}_1|u)Q_2(\bar{x}_2|u)$. Similarly, the rate condition corresponding to (4.44) is

$$R_1 + R_2 \leq \min_{\substack{\tilde{P}_{UX_1X_2Y} \in \mathcal{T}_{12}^{\text{ts}}(Q_U \times Q_1 \times Q_2 \times W) \\ I_{\tilde{P}}(X_1; Y|U) \leq R_1, I_{\tilde{P}}(X_2; Y|U) \leq R_2}} D(\tilde{P}_{UX_1X_2Y} \| Q_U \times Q_1 \times Q_2 \times \tilde{P}_Y), \quad (4.104)$$

and the rate condition corresponding to (4.68) is

$$\begin{aligned} R_1 \leq & \sup_{\rho_2 \in [0,1], s \geq 0, a_1(\cdot, \cdot), a_2(\cdot, \cdot)} \\ & \mathbb{E} \left[\log \frac{(q(X_1, X_2, Y)^s e^{a_2(U, X_2)})^{\rho_2} e^{a_1(U, X_1)}}{\mathbb{E} \left[\left(\mathbb{E}[q(\bar{X}_1, \bar{X}_2, Y)^s e^{a_2(U, \bar{X}_2)} | \bar{X}_1] \right)^{\rho_2} e^{a_1(U, \bar{X}_1)} \middle| U, Y \right]} \right] - \rho_2 R_2, \end{aligned} \quad (4.105)$$

where $U \sim Q_U$.

We now compare the above rate conditions with those obtained by explicit time sharing, focusing on the case that $\mathcal{U} = \{1, 2\}$. We again write $Q_U(1) = \lambda$ and $Q_U(2) = 1 - \lambda$, and we assume that the time sharing is done using the input distribution pairs $\mathbf{Q}^{(1)}$ and $\mathbf{Q}^{(2)}$, and the rate pairs $(R_1^{(1)}, R_2^{(1)})$ and $(R_1^{(2)}, R_2^{(2)})$. We make use of R_1 and R_2 defined in (4.93).

We first consider the primal domain expressions. We have from (4.94) that under explicit

time sharing, the condition corresponding to (4.44) is given by

$$R_1 \leq \lambda \min_{\tilde{P}_{X_1 X_2 Y}} D(\tilde{P}_{X_1 X_2 Y} \| Q_1^{(1)} \times Q_2^{(1)} \times \tilde{P}_Y) + (1 - \lambda) \min_{\tilde{P}_{X_1 X_2 Y}} D(\tilde{P}_{X_1 X_2 Y} \| Q_1^{(2)} \times Q_2^{(2)} \times \tilde{P}_Y), \quad (4.106)$$

where the minimizations are subject to constraints of the form in (4.44), the first using $Q^{(1)}$ and the second using $Q^{(2)}$. By writing the first minimization variable in (4.106) as $\tilde{P}_{X_1 X_2 Y|U}(\cdot, \cdot, \cdot | 1)$ and the second as $\tilde{P}_{X_1 X_2 Y|U}(\cdot, \cdot, \cdot | 2)$, we can combine these into a single minimization over $\tilde{P}_{X_1 X_2 Y}$, where $\tilde{P}_U = Q_U$. The resulting condition has the same form as (4.104), but with the constraint $\mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)]$ replaced by

$$\mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y) | U = u] \geq \mathbb{E}_P[\log q(X_1, X_2, Y) | U = u], \quad u = 1, 2, \quad (4.107)$$

and each constraint $I_{\tilde{P}}(X_\nu; Y|U) \leq R_\nu$ ($\nu = 1, 2$) replaced by

$$I_{\tilde{P}}(X_\nu; Y|U = u) \leq R_\nu^{(u)}, \quad u = 1, 2. \quad (4.108)$$

It follows that the constraints on $\tilde{P}_{X_1 X_2 Y}$ are stricter for explicit time sharing than for coded time sharing; the former must hold for each value of u , while the latter only need to hold for the average over u with respect to Q_U . Since the minimization for the coded time sharing ensemble is over a larger set, its rate cannot be higher, and can in fact be strictly lower.

A similar observation applies in the dual domain. Let us write $I_{12,1}(\mathbf{Q}, s, \rho_2)$ to denote (4.68) with fixed values of s and ρ_2 in place of the suprema. Using explicit time sharing, the condition corresponding to (4.68) is given by

$$R_1 \leq \lambda \sup_{s \geq 0, \rho_2 \in [0,1]} I_{12,1}(\mathbf{Q}^{(1)}, s, \rho_2) + (1 - \lambda) \sup_{s \geq 0, \rho_2 \in [0,1]} I_{12,1}(\mathbf{Q}^{(2)}, s, \rho_2), \quad (4.109)$$

whereas from (4.105), coded time sharing only permits

$$R_1 \leq \sup_{s \geq 0, \rho_2 \in [0,1]} \left(\lambda I_{12,1}(\mathbf{Q}^{(1)}, s, \rho_2) + (1 - \lambda) I_{12,1}(\mathbf{Q}^{(2)}, s, \rho_2) \right). \quad (4.110)$$

In other words, the coded time sharing ensemble only allows a single choice of (s, ρ_2) , whereas explicit time sharing allows (s, ρ_2) to vary with u . An analogy to the discussion regarding the primal expressions can be seen by interpreting (s, ρ_2) as Lagrange multipliers.

In summary, the achievable rate regions of the coded time sharing ensembles can be strictly smaller than those of explicit time sharing in the mismatched setting, in contrast to the matched setting. However, from the discussion preceding (4.94), we expect coded time sharing to yield

higher error exponents for many channels, metrics, and rate pairs. It remains an open question as to whether the rate region of explicit time sharing can be achieved using a different approach that yields better exponents.

4.2.6 Application to Single-User Mismatched Channels

The parallel BSC example [35] given in Section 4.1.2 shows that the achievable rate region for the MAC can be used to improve on the LM rate in the single-user setting. In [35, Thm. 4], Lapidoth showed that we can in fact do even better by expurgating all codeword pairs except for those having a fixed joint type approximately equal to $Q_1 \times Q_2$. Such expurgation is not possible for the MAC, since it requires cooperation between the users.

We state the achievable rate region of [35, Thm. 4] in the following theorem, making use of the sets

$$\mathcal{T}_\nu^{\text{ex}}(Q_1 \times Q_2 \times W) \triangleq \mathcal{T}_\nu(Q_1 \times Q_2 \times W) \cap \left\{ \tilde{P}_{X_1 X_2 Y} : \tilde{P}_{X_1 X_2} = Q_1 \times Q_2 \right\} \quad (4.111)$$

for $\nu = 1, 2, 12$, where the sets $\mathcal{T}_\nu$ are defined in (4.23)–(4.25).

Theorem 4.10. [35, Thm. 4] *Let the DMC $W'(y|x)$ and metric $q'(x, y)$ be given, and let $W(y|x_1, x_2) = W'(y|\psi(x_1, x_2))$ and $q(x_1, x_2, y) = q'(\psi(x_1, x_2), y)$ for some finite alphabets $\mathcal{X}_1, \mathcal{X}_2$ and function $\psi : \mathcal{X}_1 \times \mathcal{X}_2 \rightarrow \mathcal{X}$. For any pair (Q_1, Q_2) , the rate $R = R_1 + R_2$ is achievable for W' with the metric q' provided that*

$$R_1 \leq \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_1^{\text{ex}}(Q_1 \times Q_2 \times W)} I_{\tilde{P}}(X_1; Y|X_2) \quad (4.112)$$

$$R_2 \leq \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_2^{\text{ex}}(Q_1 \times Q_2 \times W)} I_{\tilde{P}}(X_2; Y|X_1) \quad (4.113)$$

$$R_1 + R_2 \leq \min_{\substack{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\text{ex}}(Q_1 \times Q_2 \times W) \\ I_{\tilde{P}}(X_1; Y) \leq R_1, I_{\tilde{P}}(X_2; Y) \leq R_2}} I_{\tilde{P}}(X_1, X_2; Y). \quad (4.114)$$

Analogously to Lemma 4.2, we have the following.

Lemma 4.3. *The condition in (4.114) holds if and only if at least one of the following hold:*

$$R_1 + R_2 \leq \min_{\substack{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\text{ex}}(Q_1 \times Q_2 \times W) \\ I_{\tilde{P}}(X_1; Y) \leq R_1}} I_{\tilde{P}}(X_1, X_2; Y) \quad (4.115)$$

$$R_1 + R_2 \leq \min_{\substack{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\text{ex}}(Q_1 \times Q_2 \times W) \\ I_{\tilde{P}}(X_2; Y) \leq R_2}} I_{\tilde{P}}(X_1, X_2; Y). \quad (4.116)$$

Proof. The condition in (4.114) can equivalently be written as

$$0 \leq \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}^{\text{ex}}(Q_1 \times Q_2 \times W)} \max \left\{ I_{\tilde{P}}(X_1; Y) - R_1, I_{\tilde{P}}(X_2; Y) - R_2, I_{\tilde{P}}(X_1, X_2; Y) - R_1 - R_2 \right\}. \quad (4.117)$$

The lemma follows by writing (4.115)–(4.116) in a similar form and applying Lemma E.3, analogously to the proof of Lemma 4.1. $\square$

The following theorem is analogous to Theorem 4.4, and can be proved similarly using Lagrange duality techniques. The only significant differences in the derivation are the additional constraint $\tilde{P}_{X_1 X_2} = Q_1 \times Q_2$ in each primal expression, and the presence of joint functions $a(x_1, x_2)$ in each dual expression rather than separate functions $a_1(x_1)$ and $a_2(x_2)$. The proof is omitted to avoid repetition.

Theorem 4.11. *Under the setup of Theorem 4.10, the pair (R_1, R_2) satisfies (4.112)–(4.114) if and only if it satisfies*

$$R_1 \leq \sup_{s \geq 0, a(\cdot, \cdot)} \mathbb{E} \left[\log \frac{q(X_1, X_2, Y)^s e^{a(X_1, X_2)}}{\mathbb{E}[q(\bar{X}_1, X_2, Y)^s e^{a(\bar{X}_1, X_2)} | X_2, Y]} \right] \quad (4.118)$$

$$R_2 \leq \sup_{s \geq 0, a(\cdot, \cdot)} \mathbb{E} \left[\log \frac{q(X_1, X_2, Y)^s e^{a(X_1, X_2)}}{\mathbb{E}[q(X_1, \bar{X}_2, Y)^s e^{a(X_1, \bar{X}_2)} | X_1, Y]} \right], \quad (4.119)$$

and at least one of

$$R_1 \leq \sup_{\rho_2 \in [0, 1], s \geq 0, a(\cdot, \cdot)} \mathbb{E} \left[\log \frac{(q(X_1, X_2, Y)^s e^{a(X_1, X_2)})^{\rho_2}}{\mathbb{E} \left[\left(\mathbb{E}[q(\bar{X}_1, \bar{X}_2, Y)^s e^{a(\bar{X}_1, \bar{X}_2)} | \bar{X}_1] \right)^{\rho_2} | Y \right]} \right] - \rho_2 R_2 \quad (4.120)$$

$$R_2 \leq \sup_{\rho_1 \in [0, 1], s \geq 0, a(\cdot, \cdot)} \mathbb{E} \left[\log \frac{(q(X_1, X_2, Y)^s e^{a(X_1, X_2)})^{\rho_1}}{\mathbb{E} \left[\left(\mathbb{E}[q(\bar{X}_1, \bar{X}_2, Y)^s e^{a(\bar{X}_1, \bar{X}_2)} | \bar{X}_2] \right)^{\rho_1} | Y \right]} \right] - \rho_1 R_1, \quad (4.121)$$

where $(X_1, X_2, Y, \bar{X}_1, \bar{X}_2) \sim Q_1(x_1)Q_2(x_2)W(y|x_1, x_2)Q_1(\bar{x}_1)Q_2(\bar{x}_2)$.

A possible approach to generalizing (4.118)–(4.121) to more general alphabets is to combine the expurgation arguments of [35, Thm. 4] with the cost-constrained coding ideas of Section 4.2.3, using joint costs $a(x_1, x_2)$ in place of the separate costs $a_1(x_1)$ and $a_2(x_2)$. We instead take a different approach: We show in Chapter 5 that superposition coding yields an achievable rate at least as high as that of Theorem 4.10 for any DMC, and we provide a straightforward extension to more general alphabets without the need for expurgation.

4.3 Multi-Letter Successive Decoding

Given that the decoder only knows the metric $q^n(\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(j)}, \mathbf{y})$ corresponding to each code-word pair, one may question whether there exists a decoding rule with better performance than the maximum-metric rule in (4.1). In general, this question is only interesting if “reasonable” decoding rules are considered. For example, if the values $\{\log q(x_1, x_2, y)\}$ are rationally independent (i.e. no value can be written as linear combinations of the others with rational coefficients), then there is a one-to-one correspondence between the joint empirical distribution of (x_1, x_2, y) and the possible values of $q^n(x_1, x_2, y)$, and hence the decoder can implement the maximum-likelihood (ML) rule.

In this section, we consider successive decoding of the form

$$\hat{m}_1 = \arg \max_i \sum_j q^n(\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(j)}, \mathbf{y}) \quad (4.122)$$

$$\hat{m}_2 = \arg \max_j q^n(\mathbf{x}_1^{(\hat{m}_1)}, \mathbf{x}_2^{(j)}, \mathbf{y}). \quad (4.123)$$

We define the error probability, achievable rate pairs, and achievable error exponents in the same way as maximum-metric decoding.

In the case that $q(x_1, x_2, y) = W(y|x_1, x_2)$, then (4.122) is the decision rule that minimizes $\mathbb{P}[\hat{m}_1 \neq m_1]$. We can therefore view (4.122) as being a mismatched version of the optimal decoding rule for (one user of) the interference channel. Thus, as well as giving an achievable rate region for the MAC with mismatched successive decoding, our results will quantify the loss due to mismatch for the interference channel, at least in the setting of random coding with independent codewords.

The following proposition shows that successive decoding and simultaneous decoding behave similarly when $q = W$. In contrast, we will see that when $q \neq W$, the two can behave significantly differently.

Proposition 4.2. *For any MAC, a rate pair or error exponent is achievable under maximum-likelihood (ML) decoding if and only if it is achievable under the successive rule in (4.122)–(4.123) with $q(x_1, x_2, y) = W(y|x_1, x_2)$.*

Proof. See Section 4.5.2. □

4.3.1 Achievable Rate Region

Our achievable rate region is written in terms of the functions

$$\begin{aligned} \bar{F}(\tilde{P}_{X_1X_2Y}, \tilde{P}'_{X_1X_2Y}, R_2) \triangleq \max \Big\{ & \mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)], \\ & \mathbb{E}_{\tilde{P}'}[\log q(X_1, X_2, Y)] + [R_2 - I_{\tilde{P}'}(X_2; X_1, Y)]^+ \Big\} \end{aligned} \quad (4.124)$$

$$\begin{aligned} \underline{F}(P_{X_1X_2Y}, R_2) \triangleq \max \Big\{ & \mathbb{E}_P[\log q(X_1, X_2, Y)], \\ & \max_{P'_{X_1X_2Y} \in \mathcal{T}'_1(P_{X_1X_2Y}, R_2)} \mathbb{E}_{P'}[\log q(X_1, X_2, Y)] + R_2 - I_{P'}(X_2; X_1, Y) \Big\}, \end{aligned} \quad (4.125)$$

and the sets

$$\begin{aligned} \mathcal{T}_1(P_{X_1X_2Y}, R_2) \triangleq \Big\{ & (\tilde{P}_{X_1X_2Y}, \tilde{P}'_{X_1X_2Y}) : \tilde{P}_{X_2Y} = P_{X_2Y}, \tilde{P}_{X_1} = P_{X_1}, \\ & \tilde{P}'_{X_1Y} = \tilde{P}_{X_1Y}, P'_{X_2} = P_{X_2}, \bar{F}(\tilde{P}_{X_1X_2Y}, \tilde{P}'_{X_1X_2Y}, R_2) \geq \underline{F}(P_{X_1X_2Y}, R_2) \Big\} \end{aligned} \quad (4.126)$$

$$\mathcal{T}'_1(P_{X_1X_2Y}, R_2) \triangleq \Big\{ P'_{X_1X_2Y} : P'_{X_1Y} = P_{X_1Y}, P'_{X_2} = P_{X_2}, I_{P'}(X_2; X_1, Y) \leq R_2 \Big\} \quad (4.127)$$

$$\begin{aligned} \mathcal{T}_2(P_{X_1X_2Y}) \triangleq \Big\{ & \tilde{P}_{X_1X_2Y} : \tilde{P}_{X_2} = P_{X_2}, \tilde{P}_{X_1Y} = P_{X_1Y}, \\ & \mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)] \Big\}. \end{aligned} \quad (4.128)$$

Here we have reused the symbol $\mathcal{T}_1$ (cf. (4.23)), whereas the definition of $\mathcal{T}_2$ in (4.128) coincides with (4.24).

Theorem 4.12. *For any input distributions Q_1 and Q_2 , the pair (R_1, R_2) is achievable for the successive decoder in (4.122)–(4.123) provided that*

$$R_1 \leq \min_{(\tilde{P}_{X_1X_2Y}, \tilde{P}'_{X_1X_2Y}) \in \mathcal{T}_1(Q_1 \times Q_2 \times W, R_2)} I_{\tilde{P}}(X_1; X_2, Y) + [I_{\tilde{P}'}(X_2; X_1, Y) - R_2]^+ \quad (4.129)$$

$$R_2 \leq \min_{\tilde{P}_{X_1X_2Y} \in \mathcal{T}_2(Q_1 \times Q_2 \times W)} I_{\tilde{P}}(X_2; X_1, Y). \quad (4.130)$$

Proof. See Section 4.5.3. □

The minimization in (4.129) is a non-convex optimization problem, but it can be cast in terms of convex optimization problems, thus facilitating its computation. Details are provided in Section 4.5.4.

While our focus is on achievable rates, the proof of Theorem 4.12 also provides error exponents. The exponent corresponding to (4.130) is precisely $E_{r,2}^{\text{cc}}$ in (4.30), and the exponent

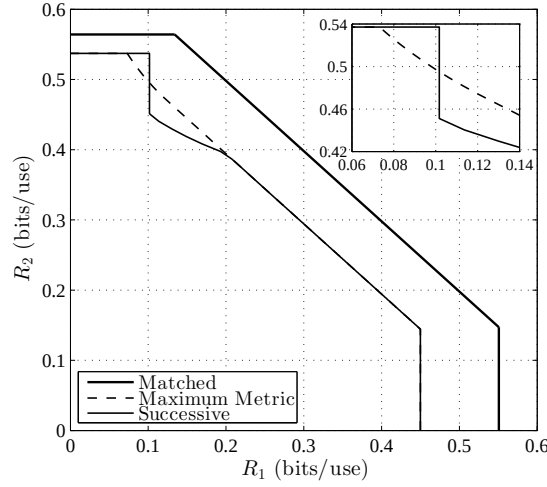


Figure 4.4 Achievable rate regions for the channel given in (4.132).

corresponding to (4.129) is given by

$$\min_{P_{X_1 X_2 Y} : P_{X_1} = Q_1, P_{X_2} = Q_2} D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + [I_0(P_{X_1 X_2 Y}, R_2) - R_1]^+, \quad (4.131)$$

where $I_0(P_{X_1 X_2 Y}, R_2)$ denotes the right-hand side of (4.129) with an arbitrary distribution $P_{X_1 X_2 Y}$ in place of $Q_1 \times Q_2 \times W$. Unlike the rate conditions, we have thus far been unable to express (4.131) in terms of convex optimization problems, thus making a numerical comparison with existing exponents (e.g. [118]) difficult.

4.3.2 Numerical Example

We consider the MAC with $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1\}$, $\mathcal{Y} = \{0, 1, 2\}$, and

$$W(y|x_1, x_2) = \begin{cases} 1 - 2\delta_{x_1 x_2} & y = x_1 + x_2 \\ \delta_{x_1 x_2} & \text{otherwise,} \end{cases} \quad (4.132)$$

where $\{\delta_{x_1 x_2}\}$ are constants. The mismatched decoder uses $q(x_1, x_2, y)$ of a similar form, but with a fixed value $\delta \in (0, \frac{1}{3})$ in place of $\{\delta_{x_1 x_2}\}$. While all such choices of δ are equivalent for maximum-metric decoding, this is not true for successive decoding.

We set $\delta_{00} = 0.01$, $\delta_{01} = 0.1$, $\delta_{10} = 0.01$, $\delta_{11} = 0.3$, $\delta = 0.15$, and $Q_1 = Q_2 = (0.5, 0.5)$. Figure 4.4 plots the achievable rates regions of successive decoding (Theorem 4.12), maximum-metric decoding (Theorem 4.3), and matched decoding (yielding the same region whether successive or maximum-metric).

Interestingly, neither of the mismatched rate regions dominates the other, thus suggesting that the two decoding rules are fundamentally different. For the given input distribution, the sum rate for successive decoding exceeds that of maximum-metric decoding. Furthermore, upon taking the convex hull (which is justified by a time sharing argument, cf. Section 4.2.5), the region for successive decoding is strictly larger. While we observed similar behaviors for other choices of Q_1 and Q_2 , it remains unclear as to whether this is always the case. Furthermore, while the rate region for maximum-metric decoding is tight with respect to the ensemble average, it is unclear whether the same is true of the region given in Theorem 4.12.

The vertical line at $R_1 \approx 0.1$ is analogous to the interference channel, where for R_1 below a certain threshold, R_2 can be arbitrarily large while still ensuring that user 1's message is estimated correctly [118]. Due to the mismatch, this induces a non-pentagonal shape in the present example.

4.4 Second-Order Asymptotics for ML Decoding

In this section, we study second-order coding rates (cf. Section 1.2.2) for the MAC. We limit our attention to maximum-likelihood decoding (i.e. (4.1) with $q = W$), as it provides the most insight into the techniques and ideas presented.

4.4.1 Preliminaries

Notation

Along with the notations given in Section 1.4, we make use of the following. The vectors (or matrices) of all zeros and all ones are denoted by $\mathbf{0}$ and $\mathbf{1}$ respectively, and the identity matrix is denoted by $\mathbb{I}$; the sizes will be clear from the context. The symbols $\prec, \preceq$, etc. denote element-wise inequalities for vectors, and inequalities on the positive semidefinite cone for matrices (e.g. $V \succ \mathbf{0}$ means V is positive definite). We denote the transpose of a vector or matrix by $(\cdot)^T$, the inverse of a matrix by $(\cdot)^{-1}$, the positive definite matrix square root by $(\cdot)^{\frac{1}{2}}$, and its inverse by $(\cdot)^{-\frac{1}{2}}$. The multivariate Gaussian distribution with mean μ and covariance matrix Σ is denoted by $N(\mu, \Sigma)$.

We denote the cross-covariance matrix of two random vectors by $\text{Cov}[\mathbf{Z}_1, \mathbf{Z}_2] = \mathbb{E}[(\mathbf{Z}_1 - \mathbb{E}[\mathbf{Z}_1])(\mathbf{Z}_2 - \mathbb{E}[\mathbf{Z}_2])^T]$, and we write $\text{Cov}[\mathbf{Z}]$ in place of $\text{Cov}[\mathbf{Z}, \mathbf{Z}]$. For a set $\mathcal{S}$ of vectors and a vector c , we write $\mathcal{S} + c$, or equivalently $c + \mathcal{S}$, to denote the set $\{s + c : s \in \mathcal{S}\}$. We similarly write $c\mathcal{S} = \{cs : s \in \mathcal{S}\}$ for a given constant c . A vector or matrix is said to be $O(f_n)$ if all of its entries are $O(f_n)$ in the scalar sense, and similarly for $o(\cdot)$.

Definitions

A rate pair (R_1, R_2) is said to be (n, ϵ) -achievable if there exist codebooks with $M_1 \geq e^{nR_1}$ and $M_2 \geq e^{nR_2}$ codewords of length n for users 1 and 2 respectively such that the average error probability does not exceed ϵ .

We consider constant-composition random coding with coded time sharing, similarly to Section 4.2.5. The precise description of the ensemble is postponed until Section 4.5.7; here we simply provide the definitions required to state the results. We fix a finite time sharing alphabet $\mathcal{U}$, as well as the input distributions $Q_U(u)$, $Q_1(x_1|u)$ and $Q_2(x_2|u)$. We define the joint distribution

$$P_{UX_1X_2Y}(u, x_1, x_2, y) \triangleq Q_U(u)Q_1(x_1|u)Q_2(x_2|u)W(y|x_1, x_2) \quad (4.133)$$

and denote the induced marginal distributions by $P_{Y|X_1U}$, $P_{Y|U}$, etc. Defining the rate vector

$$\mathbf{R} \triangleq \begin{bmatrix} R_1 \\ R_2 \\ R_1 + R_2 \end{bmatrix} \quad (4.134)$$

and the mutual information vector

$$\mathbf{I} \triangleq \begin{bmatrix} I(X_1; Y|X_2, U) \\ I(X_2; Y|X_1, U) \\ I(X_1, X_2; Y|U) \end{bmatrix}, \quad (4.135)$$

we can write the capacity region (see (4.5)) as follows [115, Thm. 4.3]:

$$\mathcal{R}^* = \bigcup_{\mathcal{U}} \bigcup_{Q_U, Q_1, Q_2} \left\{ (R_1, R_2) : \mathbf{R} \preceq \mathbf{I} \right\}. \quad (4.136)$$

Moreover, the union over $\mathcal{U}$ may be restricted to satisfy $|\mathcal{U}| \leq 2$. The three conditions in the element-wise inequality $\mathbf{R} \preceq \mathbf{I}$ correspond to the type-1, type-2 and type-12 conditions respectively (see Section 4.2).

A key quantity in our analysis is the information density vector [97, 119]

$$\mathbf{i}(u, x_1, x_2, y) \triangleq \begin{bmatrix} i_1(u, x_1, x_2, y) \\ i_2(u, x_1, x_2, y) \\ i_{12}(u, x_1, x_2, y) \end{bmatrix}, \quad (4.137)$$

where

$$i_1(u, x_1, x_2, y) \triangleq \log \frac{W(y|x_1, x_2)}{P_{Y|X_2U}(y|x_2, u)} \quad (4.138)$$

$$i_2(u, x_1, x_2, y) \triangleq \log \frac{W(y|x_1, x_2)}{P_{Y|X_1U}(y|x_1, u)} \quad (4.139)$$

$$i_{12}(u, x_1, x_2, y) \triangleq \log \frac{W(y|x_1, x_2)}{P_{Y|U}(y|u)}. \quad (4.140)$$

Averaging i with respect to (4.133) yields the mutual information vector in (4.135).

We consider the local notion of second-order asymptotics introduced by Nomura-Han for the Slepian-Wolf problem [67]; see also Hayashi [62] for the analogous definitions in the single-user setting. We proceed by presenting similar definitions for the present setting, albeit in a slightly different form. A pair (ℓ_1, ℓ_2) is said to be $(n, \epsilon, R_1^*, R_2^*)$ -achievable if there exist codebooks with $M_\nu \geq e^{nR_\nu^* + \sqrt{n}\ell_\nu}$ codewords of length n for $\nu = 1, 2$ such that the average error probability p_e does not exceed ϵ . The *second-order rate region* $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ is defined as the closure of the set of pairs (ℓ_1, ℓ_2) that are $(n, \epsilon, R_1^*, R_2^*)$ -achievable for sufficiently large n . In other words, $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ is the set of all (ℓ_1, ℓ_2) pairs for which there exists an ϵ -reliable code with $\log M_\nu = nR_\nu^* + \sqrt{n}\ell_\nu + o(\sqrt{n})$ for $\nu = 1, 2$. While this definition is valid for any pair (R_1^*, R_2^*) , our focus will be on pairs on the boundary of the capacity region $\mathcal{R}^*$; in all other cases we trivially have either $\mathcal{L} = \emptyset$ or $\mathcal{L} = \mathbb{R}^2$.

Finally, we define the set

$$\mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon) \triangleq \left\{ \mathbf{z} \in \mathbb{R}^d : \mathbb{P}[\mathbf{Z} \preceq \mathbf{z}] \geq 1 - \epsilon \right\}, \quad (4.141)$$

where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{V})$, and $\mathbf{V}$ is a $d \times d$ dispersion matrix. This definition applies for an arbitrary dimension d , which is dictated by the first argument.

Previous Work

One approach to characterizing the second-order asymptotics is to seek global asymptotic expansions, stated as follows: For any triplet (Q_U, Q_1, Q_2) , rate vectors $\mathbf{R}$ satisfying

$$n\mathbf{R} \in n\mathbf{I} - \sqrt{n}\mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon) + g(n)\mathbf{1}, \quad (4.142)$$

are (n, ϵ) -achievable for some *dispersion matrix* $\mathbf{V}$ and function $g(n) = o(\sqrt{n})$, where $\mathbf{I}$ is the 3×1 mutual information vector appearing in (4.136).

The first global result for the MAC was given by Tan and Kosut [97], who used i.i.d. ran-

dom coding to obtain (4.142) with $\mathbf{I} = \mathbb{E}[\mathbf{i}(U, X_1, X_2, Y)]$ and $\mathbf{V} = \text{Cov}[\mathbf{i}(U, X_1, X_2, Y)]$ (see also [120]). Huang and Moulin [119] showed that the improved dispersion matrix

$$\mathbf{V}^{\text{iid}} \triangleq \mathbb{E}[\text{Cov}[\mathbf{i}(U, X_1, X_2, Y) | U]] \quad (4.143)$$

can be obtained by using a constant-composition time sharing sequence.

As discussed by Haim *et al.* [114], expansions of the form (4.142) are more difficult to interpret than the scalar counterpart in (1.28). While the scalar dispersion V in (1.28) corresponds to a concrete operational definition [61], it appears difficult to directly give any such meaning to the matrix $\mathbf{V}$. In fact, non-global asymptotic studies in [97] indicate that, in most cases of interest, entries (1, 2) and (2, 1) of the matrix do not play a fundamental role in characterizing the performance. Furthermore, it may be difficult to compare two dispersion matrices, since the partial positive definite ordering does not guarantee that at least one of $\mathbf{V}_1 \preceq \mathbf{V}_2$ or $\mathbf{V}_2 \preceq \mathbf{V}_1$ hold. These issues are even more troublesome when one considers taking the union over all input distributions, particularly since standard proofs often yield a non-uniform remainder term $g(n)$ in (4.142). These limitations motivate the study of local asymptotics, such as $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ defined above.

The second-order rate region $\mathcal{L}$ has been characterized for very few multi-user problems [67, 121, 122]. The one most relevant to this section is the Gaussian MAC with degraded message sets [121], which has the notable feature of having a curved capacity region, giving rise to a non-standard derivative term in the expression for $\mathcal{L}$. Our analysis will yield a similar term using different techniques.

Other local notions of second-order achievability were considered in [97] and [123]. In particular, both of these works considered the problem of finding the backoff from the rates when a point (R_1^*, R_2^*) is approached from a given angle. As demonstrated in [121], this problem can be solved numerically in a straightforward fashion once $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ is characterized.

4.4.2 Main Results

Further Definitions

Our main result is written in terms of a dispersion matrix of the form

$$\mathbf{V} \triangleq \mathbb{E}[\text{Cov}[\mathbf{i}(U, X_1, X_2, Y) | U] - \text{Cov}[\mathbf{i}^{(1)}(U, X_1) | U] - \text{Cov}[\mathbf{i}^{(2)}(U, X_2) | U]], \quad (4.144)$$

where

$$\mathbf{i}^{(1)}(u, x_1) \triangleq \mathbb{E}[\mathbf{i}(U, X_1, X_2, Y) \mid (U, X_1) = (u, x_1)] \quad (4.145)$$

$$\mathbf{i}^{(2)}(u, x_2) \triangleq \mathbb{E}[\mathbf{i}(U, X_1, X_2, Y) \mid (U, X_2) = (u, x_2)]. \quad (4.146)$$

We can interpret (4.144) as follows: The term $\text{Cov}[\mathbf{i}]$ represents the variations in (X_1, X_2, Y) in the i.i.d. case (see (4.143)), and the terms $\text{Cov}[\mathbf{i}^{(1)}]$ and $\text{Cov}[\mathbf{i}^{(2)}]$ represent the reduced variations in X_1 and X_2 , resulting from the codewords having a fixed composition. We henceforth write the entries of $\mathbf{I}$ (see (4.135)) and $\mathbf{V}$ using subscripts:

$$\mathbf{I} = \begin{bmatrix} I_1 \\ I_2 \\ I_{12} \end{bmatrix}, \quad \mathbf{V} = \begin{bmatrix} V_1 & V_{1,2} & V_{1,12} \\ V_{1,2} & V_2 & V_{2,12} \\ V_{1,12} & V_{2,12} & V_{12} \end{bmatrix}. \quad (4.147)$$

For a fixed point (R_1^*, R_2^*) on the boundary of $\mathcal{R}^*$, we let $\hat{\mathbf{T}}_- \triangleq \hat{\mathbf{T}}_-(R_1^*, R_2^*)$ and $\hat{\mathbf{T}}_+ \triangleq \hat{\mathbf{T}}_+(R_1^*, R_2^*)$ respectively denote the left and right unit tangent vectors along the boundary of $\mathcal{R}^*$ in (R_1, R_2) -space; see Figure 4.5. It is understood that $\hat{\mathbf{T}}_-$ (respectively, $\hat{\mathbf{T}}_+$) is undefined when $R_1^* = 0$ (respectively, $R_2^* = 0$). In all other cases, the vectors are well-defined due to the fact that the capacity region is convex. The case $\hat{\mathbf{T}}_- = -\hat{\mathbf{T}}_+$ corresponds to a smooth or curved or straight-line part of the boundary, whereas $\hat{\mathbf{T}}_- \neq -\hat{\mathbf{T}}_+$ corresponds to a sudden change in slope (e.g. at a corner point).

We construct the following vectors in the same way as (4.134):

$$\mathbf{T}_- \triangleq \begin{bmatrix} \hat{T}_{-,1} \\ \hat{T}_{-,2} \\ \hat{T}_{-,1} + \hat{T}_{-,2} \end{bmatrix}, \quad \mathbf{T}_+ \triangleq \begin{bmatrix} \hat{T}_{+,1} \\ \hat{T}_{+,2} \\ \hat{T}_{+,1} + \hat{T}_{+,2} \end{bmatrix}, \quad (4.148)$$

where $\hat{T}_{(\cdot),i}$ denotes the i -th entry of the corresponding unit tangent vector. To ease some of the subsequent discussions, we define the following scalars that correspond to $\hat{\mathbf{T}}_-$ and $\hat{\mathbf{T}}_+$ in a one-to-one fashion:

$$D_- \triangleq \frac{\hat{T}_{-,2}}{\hat{T}_{-,1}}, \quad D_+ \triangleq \frac{\hat{T}_{+,2}}{\hat{T}_{+,1}} \quad (4.149)$$

These are the left and right derivatives of R_2^* as a function of R_1^* ; they are non-positive, and are understood to equal $-\infty$ when $\hat{T}_{(\cdot),1} = 0$. Observe that $\hat{\mathbf{T}}_-$ is obtained by normalizing $[-1 \ -D_-]$ and $\hat{\mathbf{T}}_+$ is obtained by normalizing $[1 \ D_+]$, and hence $\hat{\mathbf{T}}_- = -\hat{\mathbf{T}}_+$ if and only if $D_- = D_+$.

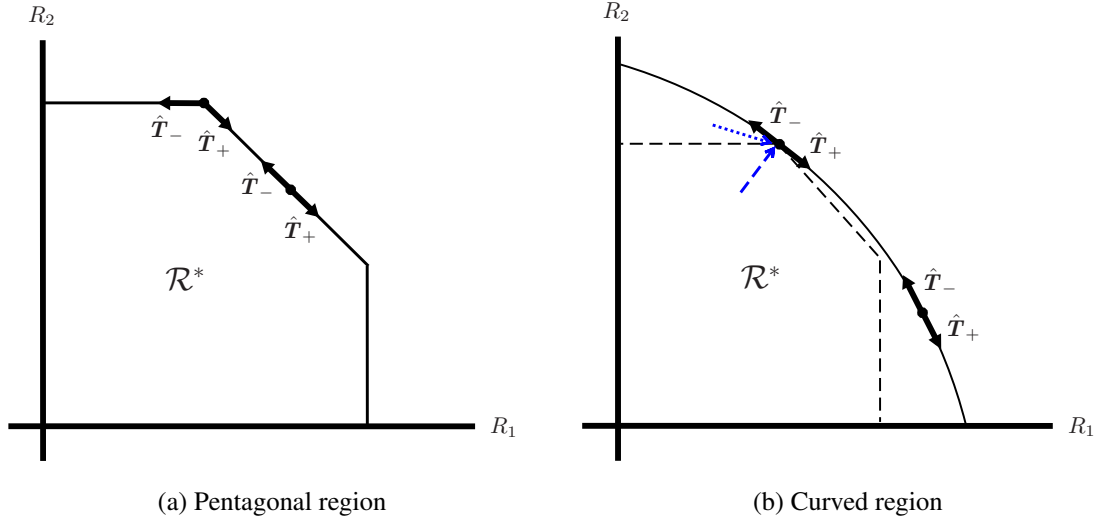


Figure 4.5 Illustration of the vectors $\hat{T}_-$ and $\hat{T}_+$ for various boundary points of two hypothetical capacity regions (solid lines). The dashed pentagon in (b) corresponds to a fixed input distribution achieving one of the boundary points.

Given the pairs (R_1^*, R_2^*) and (ℓ_1, ℓ_2) , we define

$$\mathbf{R}^* \triangleq \begin{bmatrix} R_1^* \\ R_2^* \\ R_1^* + R_2^* \end{bmatrix}, \quad \boldsymbol{\ell} \triangleq \begin{bmatrix} \ell_1 \\ \ell_2 \\ \ell_1 + \ell_2 \end{bmatrix}. \quad (4.150)$$

For a non-empty set $\mathcal{K} \subseteq \{1, 2, 12\}$, we let $\boldsymbol{\ell}^{(\mathcal{K})}$ denote the subvector of $\boldsymbol{\ell}$ where only the indices corresponding to $\mathcal{K}$ are kept; similarly for $\mathbf{R}^{(\mathcal{K})}$, $\mathbf{R}^{*(\mathcal{K})}$, $\mathbf{T}_-^{(\mathcal{K})}$, $\mathbf{T}_+^{(\mathcal{K})}$ and $\mathbf{I}^{(\mathcal{K})}$. Similarly, $\mathbf{V}^{(\mathcal{K})}$ denotes the $|\mathcal{K}| \times |\mathcal{K}|$ submatrix of $\mathbf{V}$ where only the rows and columns indexed by $\mathcal{K}$ are kept.

Statement of Main Result

We use the usual terminology that the triplet (Q_U, Q_1, Q_2) achieves the rate pair (R_1, R_2) if $\mathbf{R} \preceq \mathbf{I}$. In the following theorem, $\mathcal{K}$ can be thought of as the set of error types that are active for a given input distribution and boundary point (e.g. if the boundary point is achieved by the corner point of the pentagonal region corresponding to the type-2 and type-12 conditions, then $\mathcal{K} = \{2, 12\}$).

Theorem 4.13. Fix $\epsilon \in (0, 1)$, let (R_1^*, R_2^*) be a point on the boundary of the capacity region $\mathcal{R}^*$ in (4.136), let (Q_U, Q_1, Q_2) be an arbitrary triplet achieving that point, and consider $\mathbf{I}$ and $\mathbf{V}$ in (4.135) and (4.144) respectively. Letting $\mathcal{K} \subseteq \{1, 2, 12\}$ be the set of indices of the largest

cardinality such that $\mathbf{R}^{*(\mathcal{K})} = \mathbf{I}^{(\mathcal{K})}$, we have

$$\begin{aligned} \mathcal{L}(\epsilon, R_1^*, R_2^*) \supseteq & \left\{ (\ell_1, \ell_2) : \ell^{(\mathcal{K})} \in \bigcup_{\beta \geq 0} \left\{ \beta \mathbf{T}_-^{(\mathcal{K})} - \mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon) \right\} \right\} \\ & \cup \left\{ (\ell_1, \ell_2) : \ell^{(\mathcal{K})} \in \bigcup_{\beta \geq 0} \left\{ \beta \mathbf{T}_+^{(\mathcal{K})} - \mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon) \right\} \right\}, \end{aligned} \quad (4.151)$$

where the first (respectively, second) set is understood to be empty when $R_1^* = 0$ (respectively, $R_2^* = 0$).

Proof. See Section 4.5.5. □

We make the following remarks on Theorem 4.13:

1. To provide some intuition as to why the terms $\beta \mathbf{T}_-^{(\mathcal{K})}$ and $\beta \mathbf{T}_+^{(\mathcal{K})}$ in (4.151) are useful, we discuss the hypothetical example in Figure 4.5b. If $\epsilon < 0.5$ and we consider random coding with the fixed input distribution whose first-order region is the dashed pentagon, then we can only approach the boundary point from directions within that pentagon (e.g. along the dashed arrow). The terms involving the tangent vectors arise from arguments based on coded time sharing, and allow the boundary point to be approached from *all* directions within $\mathcal{R}^*$ (e.g. along the dotted arrow).
2. In the case that $\hat{\mathbf{T}}_- = -\hat{\mathbf{T}}_+$, or equivalently $D_- = D_+$ (i.e. a curved or straight-line part of the boundary), the two sets in (4.151) can be combined into a single set containing a coefficient $\beta \in \mathbb{R}$ (with negative values allowed) and the vector $\mathbf{T}_+^{(\mathcal{K})}$. In this case, the inner bound on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ is a half-space. We will see in Section 4.4.3 that this does not always occur, and combinations other than $(D_-, D_+) = (0, -1)$ and $(D_-, D_+) = (-1, -\infty)$ are possible (these are the combinations that are observed for standard pentagonal regions).
3. In the case that $\mathcal{K}$ contains only a single entry $\nu \in \{1, 2, 12\}$, the unions over β can be replaced by $\beta = 0$, yielding a simpler inner bound given by

$$\mathcal{L}(\epsilon, R_1^*, R_2^*) \supseteq \left\{ (\ell_1, \ell_2) : \ell_\nu \leq -\sqrt{V_\nu} \mathbf{Q}^{-1}(\epsilon) \right\}, \quad (4.152)$$

where $\ell_{12} \triangleq \ell_1 + \ell_2$. The fact that $\beta = 0$ suffices is shown in the same way for each ν , so we consider the case $\nu = 12$. Since (R_1^*, R_2^*) lies on the diagonal part of the pentagonal region corresponding to (Q_U, Q_1, Q_2) (and away from the corners), both $(R_1^* - \delta, R_2^* + \delta)$ and $(R_1^* + \delta, R_2^* - \delta)$ are achievable for sufficiently small δ , and hence $D_- \in [-1, -\infty]$

and $D_+ \in [0, -1]$. The convexity of the capacity region implies that $D_+ \leq D_-$, and it follows that $D_- = D_+ = -1$. From (4.148), we see that $D_- = D_+ = -1$ implies $\mathbf{T}_-^{(\mathcal{K})} = \mathbf{T}_+^{(\mathcal{K})} = 0$, and hence each coefficient β in (4.151) is simply multiplying zero.

4. More generally, if the input distribution achieving (R_1^*, R_2^*) also achieves each boundary point in a neighborhood of (R_1^*, R_2^*) , then the unions over β can be replaced by $\beta = 0$. In particular, this is true when the entire capacity region is achieved by a single input distribution. This will be observed for the Gaussian MAC in Section 4.4.2.
5. All non-empty subsets $\mathcal{K}$ of $\{1, 2, 12\}$ can occur with the exception of $\{1, 2\}$. Focusing on the case that time sharing is absent, the case $\mathcal{K} = \{1, 2\}$ is impossible since

$$I(X_1, X_2; Y) = I(X_1; Y) + I(X_2; Y|X_1) \quad (4.153)$$

$$\leq I(X_1; X_2, Y) + I(X_2; Y|X_1) \quad (4.154)$$

$$= I(X_1; Y|X_2) + I(X_2; Y|X_1), \quad (4.155)$$

where (4.155) follows since X_1 and X_2 are independent. Whenever $\mathcal{K}$ includes $\{1, 2\}$, we have $R_1^* = I(X_1; Y|X_2)$ and $R_2^* = I(X_2; Y|X_1)$, and it follows from (4.155) that $R_1^* + R_2^* = I(X_1, X_2; Y)$. Therefore, we have $\mathcal{K} = \{1, 2, 12\}$, corresponding to a rectangular achievable rate region.

6. The inner bound in (4.151) is of a similar form to the set appearing in [121, Thm. 3] for the Gaussian MAC with degraded message sets. The main differences are (i) The left and right tangent vectors are treated separately here, since unlike in [121], the two do not have the same slope in general (e.g. see Figure 4.5 and the example in Section 4.4.3); (ii) There are six cases here corresponding to the different subsets $\mathcal{K}$ of $\{1, 2, 12\}$ (see the previous item), whereas in [121] there are only two possibilities for the set of active rate conditions.
7. The proof of Theorem 4.13 can be followed using i.i.d. codeword distributions, yielding an analogous result with $\mathbf{V}^{\text{iid}}$ (see (4.143)) in place of $\mathbf{V}$. Using the fact that $\mathbf{V} \preceq \mathbf{V}^{\text{iid}}$, it is not difficult to show that the inner bounds on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ obtained using $\mathbf{V}$ include those obtained using $\mathbf{V}^{\text{iid}}$ whenever $\epsilon < \frac{1}{2}$. In Section 4.4.3, we will see that the inclusion may be strict.
8. It is also of interest to compare $\mathbf{V}$ to a hypothetical dispersion matrix of the form

$$\mathbf{V}^{\text{joint}} \triangleq \mathbb{E} \left[\text{Cov}[\mathbf{i}(U, X_1, X_2, Y) \mid U, X_1, X_2] \right]. \quad (4.156)$$

This is the matrix that would be obtained if the *joint* composition of $(U, \mathbf{X}_1, \mathbf{X}_2)$ were fixed, which is impossible in the absence of cooperation. It should therefore be unsurprising that $\mathbf{V}^{\text{joint}} \preceq \mathbf{V}$; as we show in [124], this can be proved using the matrix version of the law of total variance, along with the identity $\mathbb{E}[\mathbf{Z}\mathbf{Z}^T] \succeq \mathbb{E}[\mathbf{Z}]\mathbb{E}[\mathbf{Z}^T]$.

9. We make no attempt to present analytical expressions for $\hat{T}_-$ and $\hat{T}_+$, but two numerical approaches to their computation are presented in Section 4.4.3. In general, if the capacity region is characterized numerically, then one can easily obtain numerical bounds or approximations for these tangent vectors.

Application to the Gaussian Multiple-Access Channel

The preceding results are concerned with the DM-MAC, thus permitting proofs based on combinatorial arguments. Here we show that Theorem 4.13 can be extended to the Gaussian MAC via an increasingly fine quantization of the input, similarly to Hayashi [62, Thm. 3] and Tan [125].² The channel is described by

$$Y = \sqrt{P_1}X_1 + \sqrt{P_2}X_2 + Z, \quad (4.157)$$

where $Z \sim \mathcal{N}(0, 1)$, and where the codewords for user $\nu = 1, 2$ are constrained to satisfy $\frac{1}{n}\|\mathbf{x}_\nu\|^2 \leq 1$. The quantities P_1 and P_2 represent the signal-to-noise ratios for users 1 and 2 respectively; for convenience, we have factored them into the channel itself. We use the definition of $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ in Section 4.4.1 with the additional condition that the codewords of each user must satisfy the respective power constraints.

The capacity region is pentagonal [3, Sec. 15.1], and is achieved using Gaussian input distributions, namely $Q_1, Q_2 \sim \mathcal{N}(0, 1)$. The quantities $\mathbf{I}$ and $\mathbf{V}$ in (4.135) and (4.144) can be written explicitly as follows:

$$\mathbf{I} = \begin{bmatrix} \frac{1}{2} \log(1 + P_1) \\ \frac{1}{2} \log(1 + P_2) \\ \frac{1}{2} \log(1 + P_1 + P_2) \end{bmatrix} \quad (4.158)$$

$$\mathbf{V} = \begin{bmatrix} \frac{P_1(2+P_1)}{2(1+P_1)^2} & \frac{P_1P_2}{2(1+P_1)(1+P_2)} & \frac{P_1(2+P_1+P_2)}{2(1+P_1)(1+P_1+P_2)} \\ \frac{P_1P_2}{2(1+P_1)(1+P_2)} & \frac{P_2(2+P_2)}{2(1+P_2)^2} & \frac{P_2(2+P_1+P_2)}{2(1+P_2)(1+P_1+P_2)} \\ \frac{P_1(2+P_1+P_2)}{2(1+P_1)(1+P_1+P_2)} & \frac{P_2(2+P_1+P_2)}{2(1+P_2)(1+P_1+P_2)} & \frac{(P_1+P_2)(2+P_1+P_2)+2P_1P_2}{2(1+P_1+P_2)^2} \end{bmatrix}. \quad (4.159)$$

²We thank Vincent Tan for recommending the use of this technique.

We use a quantization argument followed by constant-composition random coding to provide an alternative proof of the following result, which was recently proved by MolavianJazi and Laneman using random coding with a uniform distribution over the surface of a sphere [126]. We use the same notation as that used in Theorem 4.13.

Theorem 4.14. *Fix the parameters $P_1, P_2 > 0$ and target error probability $\epsilon \in (0, 1)$, and let $\mathcal{K} \subset \{1, 2, 12\}$ be the set of indices of the largest cardinality such that $\mathbf{R}^{*(\mathcal{K})} = \mathbf{I}^{(\mathcal{K})}$. Then we have*

$$\mathcal{L}(\epsilon, R_1^*, R_2^*) \supseteq \left\{ (\ell_1, \ell_2) : \mathbf{L}^{(\mathcal{K})} \in -\mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon) \right\}. \quad (4.160)$$

Proof. See Section 4.5.8. □

Unlike Theorem 4.13, no tangent vectors appear in (4.160), and the proof does not use any form of time sharing. This is due to the fact that every point on the boundary of the capacity region is simultaneously achieved by Gaussian inputs.

4.4.3 Examples

The Collision Channel

Here we consider the channel with alphabets $\mathcal{X}_1 = \mathcal{X}_2 = \{0, 1, 2\}$ and $\mathcal{Y} = \{(0, 0), (0, 1), (0, 2), (1, 0), (2, 0), \mathbf{c}\}$, and

$$W(y|x_1, x_2) = \begin{cases} 1 & y = (x_1, x_2) \text{ and } \min\{x_1, x_2\} = 0 \\ 1 & y = \mathbf{c} \text{ and } \min\{x_1, x_2\} \neq 0 \\ 0 & \text{otherwise.} \end{cases} \quad (4.161)$$

In words, if either user transmits the zero symbol then the output is (x_1, x_2) , whereas if both users transmit a non-zero symbol then the output is $\mathbf{c}$, meaning “collision”.

We recall the following observations by Gallager [105]: (i) The capacity region can be obtained without time sharing;³ (ii) By symmetry, the points on the boundary of the capacity region are achieved by input distributions of the form $Q_1 = (1 - 2p_1, p_1, p_1)$ and $Q_2 = (1 - 2p_2, p_2, p_2)$, where $\mathcal{U} = \emptyset$; (iii) The achievable rate region corresponding to any such (Q_1, Q_2) pair is rectangular. The capacity region is shown in Figure 4.6. The left and right tangent vectors coincide (i.e. $\hat{\mathbf{T}}_- = -\hat{\mathbf{T}}_+$ and hence $D_- = D_+$) at all boundary points (R_1^*, R_2^*) with $R_1^* > 0$ and $R_2^* > 0$, and the case of interest in Theorem 4.13 is $\mathcal{K} = \{1, 2, 12\}$.

³In the more general case of K non-zero symbols, time sharing is required for $K \geq 8$ [105].

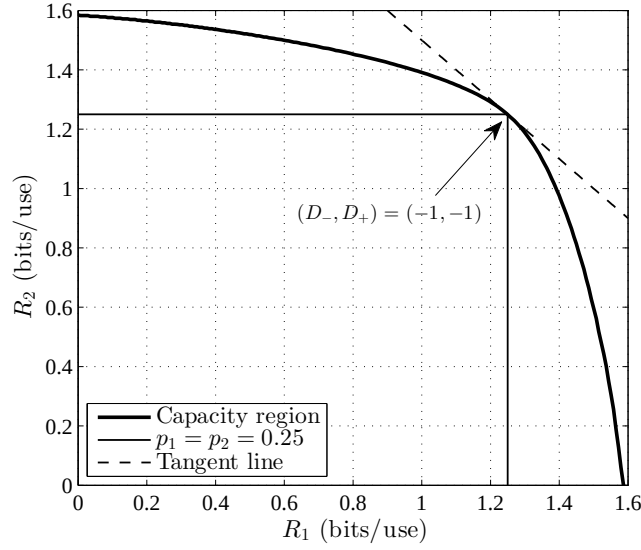


Figure 4.6 Capacity region of the collision channel. The tangent line corresponds to the boundary point $(R_1^*, R_2^*) = (1.25, 1.25)$, and we have $D_- = D_+ = -1$.

One approach to characterizing the inner bound in (4.151) for a given boundary point (R_1^*, R_2^*) is to first find the pair (p_1, p_2) achieving that point, and then calculate $\hat{T}_-$ and $\hat{T}_+$ (e.g. see the example in Section 4.4.3). In this example, the reverse approach turns out to be more convenient: We start with a given derivative $D < 0$, and perform an optimization over (p_1, p_2) to find the corresponding (unique) boundary point (R_1^*, R_2^*) .

As stated above, the achievable rate region for a given pair (p_1, p_2) is a rectangle with a corner point given by (I_1, I_2) . The straight line of slope D passing through this point is given by $R_2 = D(R_1 - I_1) + I_2$. Thus, finding the desired boundary point (R_1^*, R_2^*) simply amounts to maximizing $I_2 - DI_1$ over (p_1, p_2) , which is a straightforward optimization problem.

For concreteness, we provide a numerical example for the case that $D = -1$, which corresponds to $\hat{T}_- = -\hat{T}_+ = \left[\frac{-1}{\sqrt{2}} \quad \frac{1}{\sqrt{2}}\right]^T$. Using a brute force search to three decimal places, we found the optimal parameters to be $p_1 = p_2 = 0.25$, yielding $R_1^* = R_2^* = 1.25$ bits/use. The inner bound on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ from Theorem 4.13, and its counterpart for i.i.d. random coding (cf. (4.143)), are shown in Figure 4.7, where we set $\epsilon = 0.01$. For comparison, we also plot the weaker inner bounds in which β is set to zero and only $-Q_{\text{inv}}$ remains (cf. (4.151)). The boundaries of these regions are shown, and the regions lie to the bottom-left of these boundaries.

We see that the region resulting from constant-composition random coding is strictly larger than that resulting from i.i.d. random coding. This is primarily due to the fact that the diagonal entries of V are strictly smaller than the respective entries of V^{iid} , though the off-diagonal

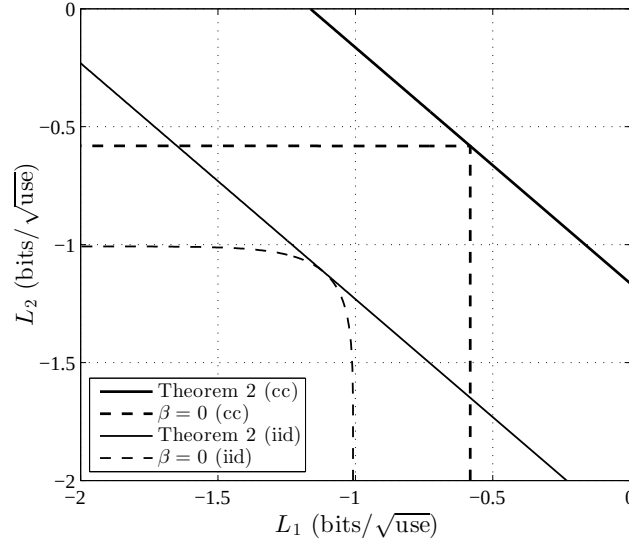


Figure 4.7 Inner bounds on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ for the collision channel, with $R_1^* = R_2^* = 1.25$ and $\epsilon = 0.01$. The regions are to the bottom-left of the boundaries shown.

terms also play a role (see the shapes of the regions corresponding to $\beta = 0$ in Figure 4.7). In this example, the strict inclusion holds for all points on the boundary of the capacity region other than the endpoints corresponding to $R_1^* = 0$ or $R_2^* = 0$.

The improved performance of constant-composition random coding over i.i.d. coding is analogous to a similar gain in the error exponent [66]. These gains are in contrast to the (matched) single-user setting, where the two ensembles yield the same second-order term and error exponent after the optimization of the input distribution [56, 61].

A Non-Deterministic Example

Here we provide an example showing that the two unions in (4.151) cannot, in general, be combined into one. In other words, it is necessary to consider the left and right tangent vectors separately. We set $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{Y} = \{0, 1\}$, and

$$W(y|x_1, x_2) = \begin{cases} 1 & x_1 = x_2 = y \\ 0.5 & x_1 \neq x_2 \\ 0 & \text{otherwise.} \end{cases} \quad (4.162)$$

Thus, the channel is noiseless if $x_1 = x_2$, and completely noisy if $x_1 \neq x_2$. We write the input distributions as $Q_1 = (1 - p_1, p_1)$ and $Q_2 = (1 - p_2, p_2)$.

The capacity region is shown in Figure 4.8. Observe that there are two “corner points”,

but unlike those of standard pentagonal regions, neither of them corresponds to a change in angle of 45 degrees. More precisely, the middle segment shown in the plot has slope -1 , but the other two segments are neither vertical nor horizontal (in fact, they are not even straight line segments, even though they may appear to be). Both corner points are achieved by $p_1 = p_2 = 0.5$.

Here we focus on characterizing the set $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ for the upper corner point $(R_1^*, R_2^*) = (0.272, 0.449)$; identical arguments apply for the lower corner point. The case of interest in Theorem 4.13 is $\mathcal{K} = \{2, 12\}$. Since the middle segment in Figure 4.8 has slope -1 , we have $\hat{T}_+ = [\frac{1}{\sqrt{2}} \frac{-1}{\sqrt{2}}]$. The idea used to compute $\hat{T}_-$ is to shift the point $(p_1, p_2) = (0.5, 0.5)$ by a small amount $(\Delta \cos \theta, \Delta \sin \theta)$, and observe the behavior of the corner point $(I_{12} - I_2, I_2)$ for $\theta \in [0, 2\pi)$, where for each θ we are interested in the limiting behavior as $\Delta \rightarrow 0$. Making the dependence of I_2 and I_{12} on (p_1, p_2) explicit, a second-order Taylor expansion yields

$$I_\nu(p_1 + \Delta \cos \theta, p_2 + \Delta \sin \theta) = I_\nu(p_1, p_2) + f_\nu(\theta) \Delta^2 + O(\Delta^3) \quad (4.163)$$

for $\nu = 2, 12$, where

$$f_\nu(\theta) \triangleq \begin{bmatrix} \cos \theta & \sin \theta \end{bmatrix} \begin{bmatrix} \frac{\partial^2 I_\nu}{\partial p_1^2} & \frac{\partial^2 I_\nu}{\partial p_1 \partial p_2} \\ \frac{\partial^2 I_\nu}{\partial p_1 \partial p_2} & \frac{\partial^2 I_\nu}{\partial p_2^2} \end{bmatrix} \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}. \quad (4.164)$$

Note that the first-order term in (4.163) is absent, since the derivatives $\frac{\partial I_\nu}{\partial p_1}$ and $\frac{\partial I_\nu}{\partial p_2}$ are zero at $(p_1, p_2) = (0.5, 0.5)$ for $\nu = 2, 12$. We conclude from (4.163) that for a fixed choice of θ , $(I_{12} - I_2, I_2)$ moves in the direction $(f_{12}(\theta) - f_2(\theta), f_2(\theta))$ in the limit as $\Delta \rightarrow 0$. Evaluating the direction for 10000 equally spaced angles over the range $\theta \in [0, 2\pi]$, we obtained $\hat{T}_- = [-0.99783 \ 0.06586]$, corresponding to $\theta = 0.221$ radians and $D_- = -0.0660$.

Figure 4.9 shows the resulting inner bound on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ given in Theorem 4.13, with $\epsilon = 0.01$. In this example, the region is identical for i.i.d. random coding and constant-composition random coding. It is interesting to observe the different shape of the region compared to the previous example, resulting from the differing left and right tangent vectors. It is only the former that plays a role in enlarging the achievable region, since the set $-\mathbf{Q}_{\text{inv}}(\mathbf{V}^{(2,12)}, \epsilon)$ already satisfies the property that if a given point is in the set, so are all points on the right-hand side of the line of slope $D_+ = -1$ passing through that point.

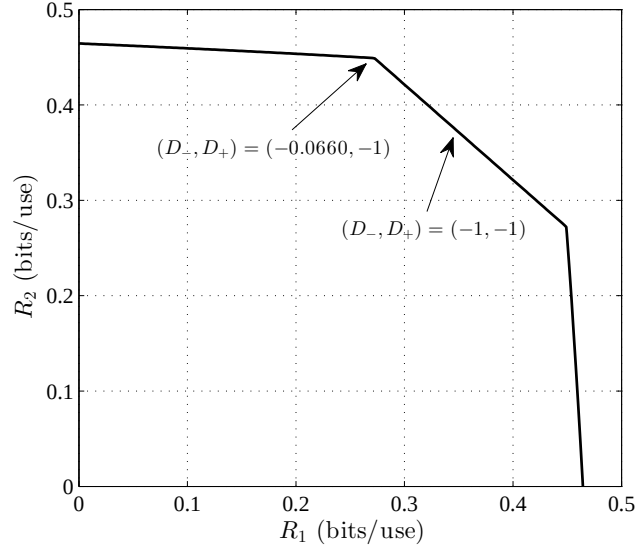
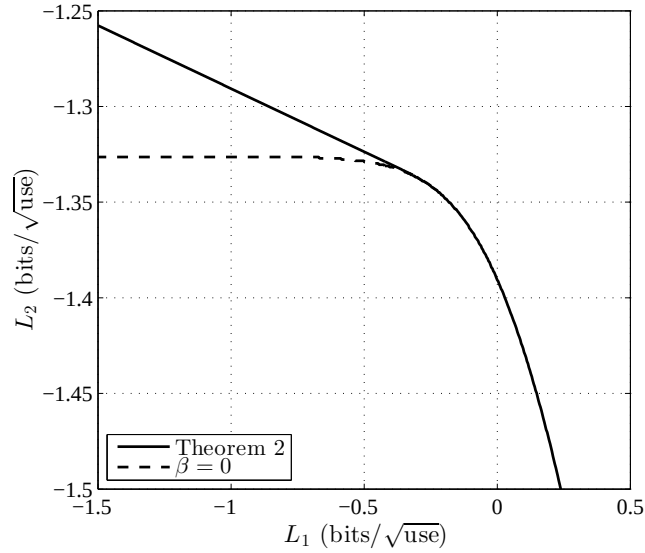


Figure 4.8 Capacity region of the channel described by (4.162).

Figure 4.9 Inner bounds on $\mathcal{L}(\epsilon, R_1^*, R_2^*)$ for the channel described by (4.162), with $(R_1^*, R_2^*) = (0.272, 0.449)$ and $\epsilon = 0.01$. The regions are to the bottom-left of the boundaries shown.

4.5 Proofs

4.5.1 Proof of Theorem 4.4

Each expression in the theorem statement is derived similarly, so we focus on $E_{r,12,1}^{\text{cc}}$. We claim that (4.49) is equivalent to

$$E_{r,12,1}^{\text{cc}}(\mathbf{Q}, R_1, R_2) = \max_{\rho_1 \in [0,1], \rho_2 \in [0,1]} \hat{E}_{0,12,1}^{\text{cc}}(\mathbf{Q}, \rho_1, \rho_2) - \rho_1(R_1 + \rho_2 R_2), \quad (4.165)$$

where

$$\begin{aligned} \hat{E}_{0,12,1}^{\text{cc}}(\mathbf{Q}, \rho_1, \rho_2) \triangleq & \min_{P_{X_1 X_2 Y} \in \mathcal{S}(\mathbf{Q})} \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}(P_{X_1 X_2 Y})} \\ & D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + \rho_1 (I_{\tilde{P}}(X_1; Y) + \rho_2 I_{\tilde{P}}(X_2; X_1, Y)). \end{aligned} \quad (4.166)$$

To see this, we follow the steps of Section 2.10.2. Using (4.83), we can write

$$\begin{aligned} E_{r,12,1}^{\text{cc}}(\mathbf{Q}, R_1, R_2) = & \min_{P_{X_1 X_2 Y} \in \mathcal{S}(\mathbf{Q})} \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}(P_{X_1 X_2 Y})} \\ & D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + [I_{\tilde{P}}(X_1; Y) + [I_{\tilde{P}}(X_2; X_1, Y) - R_2]^+ - R_1]^+ \end{aligned} \quad (4.167)$$

Next, we apply the identity $[\alpha]^+ = \max_{0 \leq \rho_1 \leq 1} \rho_1 \alpha$ to the outer $[\cdot]^+$ function in (4.167). The resulting objective is linear in ρ_1 and jointly convex in $(P_{X_1 X_2 Y}, \tilde{P}_{X_1 X_2 Y})$, so we can apply Fan's minimax theorem [78] to interchange the maximization and minimizations. Applying the same argument to the remaining $[\cdot]^+$ term yields (4.165)–(4.166).

We define the sets

$$\begin{aligned} \mathcal{T}'_{12}(P_{X_1 X_2 Y}, \hat{P}_{X_1 Y}) \triangleq & \left\{ \tilde{P}_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \tilde{P}_{X_2} = P_{X_2}, \right. \\ & \left. \tilde{P}_{X_1 Y} = \hat{P}_{X_1 Y}, \mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)] \geq \mathbb{E}_P[\log q(X_1, X_2, Y)] \right\} \end{aligned} \quad (4.168)$$

$$\mathcal{T}''_{12}(P_{X_1 X_2 Y}) \triangleq \left\{ \hat{P}_{X_1 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{Y}) : \hat{P}_{X_1} = P_{X_1}, \hat{P}_Y = P_Y \right\}. \quad (4.169)$$

It follows that $\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}(P_{X_1 X_2 Y})$ (see (4.25)) if and only if $\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}'_{12}(P_{X_1 X_2 Y}, \hat{P}_{X_1 Y})$ for some $\hat{P}_{X_1 Y} \in \mathcal{T}''_{12}(P_{X_1 X_2 Y})$. We can therefore replace the minimization over $\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}(P_{X_1 X_2 Y})$ in (4.166) with minimizations over $\hat{P}_{X_1 Y} \in \mathcal{T}''_{12}(P_{X_1 X_2 Y})$ and $\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}'_{12}(P_{X_1 X_2 Y}, \hat{P}_{X_1 Y})$.

We prove the theorem by performing the minimization in several steps, and performing multiple applications of Lemma E.1. Each application will yield an overall optimization of

the form $\sup \min \sup \{\cdot\}$, and we will implicitly use Fan's minimax theorem [78] to obtain an equivalent expression of the form $\sup \sup \min \{\cdot\}$. Thus, we will leave the optimization of the dual variables (i.e. the suprema) until the final step.

Step 1

We first consider the minimization of the term $I_{\tilde{P}}(X_1; X_2, Y)$ over $\tilde{P}_{X_1 X_2 Y}$ when $P_{X_1 X_2 Y} \in \mathcal{S}(\mathcal{Q})$ and $\hat{P}_{X_1 Y} \in \mathcal{T}_{12}''(P_{X_1 X_2 Y})$ are fixed, and thus all of the terms in the objective in (4.166) other than $I_{\tilde{P}}(X_1; X_2, Y)$ are fixed. The minimization is given by

$$\min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_{12}'(P_{X_1 X_2 Y}, \hat{P}_{X_1 Y})} I_{\tilde{P}}(X_1; X_2, Y). \quad (4.170)$$

Applying Lemma E.1 with $P_{Z_1} = P_{X_2}$, $P_{Z_2} = \hat{P}_{X_1 Y}$ and $\mu_1(\cdot) = a_2(\cdot)$, we obtain the dual expression

$$\begin{aligned} F_1 \triangleq & - \sum_{x_1, y} \hat{P}_{X_1 Y}(x_1, y) \log \sum_{\bar{x}_2} P_{X_2}(\bar{x}_2) q(x_1, \bar{x}_2, y)^s e^{a_2(\bar{x}_2)} \\ & + s \sum_{x_1, x_2, y} P_{X_1 X_2 Y}(x_1, x_2, y) \log q(x_1, x_2, y) + \sum_{x_2} P_{X_2}(x_2) a_2(x_2). \end{aligned} \quad (4.171)$$

Step 2

After Step 1, the overall objective (see (4.166)) is given by

$$D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + \rho_1(I_{\tilde{P}}(X_1; Y) + \rho_2 F_1), \quad (4.172)$$

where we have replaced $I_{\tilde{P}}(X_1; Y)$ by $I_{\hat{P}}(X_1; Y)$ due to the constraint $\tilde{P}_{X_1 Y} = \hat{P}_{X_1 Y}$ in (4.168). Since the only terms involving $\hat{P}_{X_1 Y}$ are $I_{\hat{P}}(X_1; Y)$ and the first term in (4.171), we consider the minimization

$$\min_{\hat{P}_{X_1 Y} \in \mathcal{T}_{12}''(P_{X_1 X_2 Y})} I_{\hat{P}}(X_1; Y) - \rho_2 \sum_{x_1, y} \hat{P}_{X_1 Y}(x_1, y) \log \sum_{\bar{x}_2} P_{X_2}(\bar{x}_2) q(x_1, \bar{x}_2, y)^s e^{a_2(\bar{x}_2)}. \quad (4.173)$$

Applying Lemma E.1 with $P_{Z_1} = P_{X_1}$, $P_{Z_2} = P_Y$ and $\mu_1(\cdot) = a_1(\cdot)$, we obtain

$$\begin{aligned} F_2 \triangleq & - \sum_y P_Y(y) \log \sum_{\bar{x}_1} P_{X_1}(\bar{x}_1) \left(\sum_{\bar{x}_2} P_{X_2}(\bar{x}_2) q(\bar{x}_1, \bar{x}_2, y)^s e^{a_2(\bar{x}_2)} \right)^{\rho_2} e^{a_1(\bar{x}_1)} \\ & + \sum_{x_1} P_{X_1}(x_1) a_1(x_1). \end{aligned} \quad (4.174)$$

Step 3

From (4.171), (4.172) and (4.174), the overall objective is now given by

$$F_3 \triangleq D(P_{X_1 X_2 Y} \| Q_1 \times Q_2 \times W) + \rho_1 F_2 + \rho_1 \rho_2 \sum_{x_1, x_2, y} P_{X_1 X_2 Y}(x_1, x_2, y) \log q(x_1, x_2, y)^s e^{a_2(x_2)}. \quad (4.175)$$

By expanding the divergence and writing $a_\nu(x_\nu) = \log e^{a_\nu(x_\nu)}$ ($\nu = 1, 2$), (4.175) can equivalently be expressed as

$$F_3 = - \sum_{x_1, x_2, y} P_{X_1 X_2 Y}(x_1, x_2, y) \log \frac{g_3(x_1, x_2, y)}{P_{X_1 X_2 Y}(x_1, x_2, y)}, \quad (4.176)$$

where

$$g_3(x_1, x_2, y) \triangleq Q_1(x_1)Q_2(x_2)W(y|x_1, x_2) \times \left(\sum_{\bar{x}_1} P_{X_1}(\bar{x}_1) \left(\frac{\sum_{\bar{x}_2} P_{X_2}(\bar{x}_2) q(\bar{x}_1, \bar{x}_2, y)^s e^{a_2(\bar{x}_2)}}{q(x_1, x_2, y)^s e^{a_2(x_2)}} \right)^{\rho_2} \frac{e^{a_1(\bar{x}_1)}}{e^{a_1(x_1)}} \right)^{\rho_1}. \quad (4.177)$$

We are interested in minimizing F_3 over $P_{X_1 X_2 Y} \in \mathcal{S}(\mathcal{Q})$ (see (4.166)). We first minimize over $P_{Y|X_1 X_2}$ for a fixed $P_{X_1 X_2}$. Rather than using Lagrange duality, we can simply write $P_{X_1 X_2 Y} = P_{X_1 X_2} \times P_{Y|X_1 X_2}$, and apply Jensen's inequality to (4.176) to obtain

$$F_3 \geq - \sum_{x_1, x_2, y} P_{X_1 X_2}(x_1, x_2) \log \sum_y \frac{g_3(x_1, x_2, y)}{P_{X_1 X_2}(x_1, x_2)}. \quad (4.178)$$

It follows that the minimizing $P_{Y|X_1 X_2}$ is the one that makes $\frac{g_3(x_1, x_2, y)}{P_{X_1 X_2}(x_1, x_2)}$ independent of y for any given x_1 and x_2 , so that equality holds in (4.178).

Step 4

Combining (4.178) and the definition of g_3 , we obtain the optimization

$$\min_{P_{X_1 X_2} : P_{X_1} = Q_1, P_{X_2} = Q_2} I_P(X_1; X_2) - \mathbb{E}_P[\log g_4(X_1, X_2)], \quad (4.179)$$

where

$$g_4(x_1, x_2) \triangleq \sum_y W(y|x_1, x_2) \left(\sum_{\bar{x}_1} P_{X_1}(\bar{x}_1) \left(\frac{\sum_{\bar{x}_2} P_{X_2}(\bar{x}_2) q(\bar{x}_1, \bar{x}_2, y)^s e^{a_2(\bar{x}_2)}}{q(x_1, x_2, y)^s e^{a_2(x_2)}} \right)^{\rho_2} \frac{e^{a_1(\bar{x}_1)}}{e^{a_1(x_1)}} \right)^{\rho_1}. \quad (4.180)$$

Applying Lemma E.1 with $P_{Z_1} = Q_2$, $P_{Z_2} = Q_1$ and $\mu_1(\cdot) = a'_2(\cdot)$, we obtain the dual expression

$$F_4 \triangleq - \sum_{x_1} Q_1(x_1) \log \sum_{x_2} Q_2(x_2) g_4(x_1, x_2) e^{a'_2(x_2)} + \sum_{x_2} Q_2(x_2) a'_2(x_2). \quad (4.181)$$

Step 5

From (4.181), we have

$$F_4 = - \sum_{x_1} Q_1(x_1) \log \left(e^{a'_1(x_1) - \phi'_1} \sum_{x_2} Q_2(x_2) g_4(x_1, x_2) e^{a'_2(x_2) - \phi'_2} \right) \quad (4.182)$$

$$\geq - \log \sum_{x_1, x_2} Q_1(x_1) Q_2(x_2) g_4(x_1, x_2) e^{a'_2(x_2) - \phi'_2} e^{a'_1(x_1) - \phi'_1}, \quad (4.183)$$

where (4.182) follows for any $a'_1(\cdot)$ with $\phi'_\nu \triangleq \mathbb{E}_{Q_\nu}[a'_\nu(X_\nu)]$ ($\nu = 1, 2$) by expanding the logarithm, and (4.183) follows from Jensen's inequality. We observe that equality holds in (4.182) when $a'_1(\cdot)$ is chosen to make the argument to the logarithm in independent of x_1 . Thus, we obtain the desired result by substituting (4.180) into (4.183), and taking the supremum over $s \geq 0$, $a_1(\cdot)$, $a_2(\cdot)$, $a'_1(\cdot)$ and $a'_2(\cdot)$.

4.5.2 Proof of Proposition 4.2

It suffices to show that, for any fixed codebooks $\mathcal{C}_1 = \{\mathbf{x}_1^{(i)}\}_{i=1}^{M_1}$ and $\mathcal{C}_2 = \{\mathbf{x}_2^{(j)}\}_{j=1}^{M_2}$, the error probability under ML decoding is lower bounded by a constant times the error probability under successive decoding. It also suffices to consider the variations of these decoders where ties are broken as errors, since doing so reduces the error probability by at most a factor of two [12]. Formally, we consider the following:

1. The ML decoder maximizing $W^n(\mathbf{y} | \mathbf{x}_1^{(i)}, \mathbf{x}_2^{(j)})$;
2. The successive decoder in (4.122)–(4.123);
3. The genie-aided successive decoder using the true value of m_1 on the second step rather than $\hat{m}_1$ [127]:

$$\hat{m}_1 = \arg \max_i \sum_j W^n(\mathbf{x}_1^{(i)}, \mathbf{x}_2^{(j)}, \mathbf{y}) \quad (4.184)$$

$$\hat{m}_2 = \arg \max_j W^n(\mathbf{x}_1^{(m_1)}, \mathbf{x}_2^{(j)}, \mathbf{y}). \quad (4.185)$$

We denote the probabilities under these decoders by $\mathbb{P}^{(\text{ML})}[\cdot]$, $\mathbb{P}^{(\text{S})}[\cdot]$ and $\mathbb{P}^{(\text{Genie})}[\cdot]$ respectively. Denoting the random message pair by (m_1, m_2) , the resulting estimate by $(\hat{m}_1, \hat{m}_2)$, and the output sequence by $\mathbf{Y}$, we have

$$\begin{aligned} & \mathbb{P}^{(\text{ML})}[(\hat{m}_1, \hat{m}_2) \neq (m_1, m_2)] \\ & \geq \max \left\{ \mathbb{P}^{(\text{ML})}[\hat{m}_1 \neq m_1], \mathbb{P}^{(\text{ML})} \left[\bigcup_{j \neq m_2} \left\{ W^n(\mathbf{x}_1^{(m_1)}, \mathbf{x}_2^{(j)}, \mathbf{Y}) \geq W^n(\mathbf{x}_1^{(m_1)}, \mathbf{x}_2^{(m_2)}, \mathbf{Y}) \right\} \right] \right\} \end{aligned} \quad (4.186)$$

$$\geq \max \left\{ \mathbb{P}^{(\text{Genie})}[\hat{m}_1 \neq m_1], \mathbb{P}^{(\text{Genie})} \left[\bigcup_{j \neq m_2} \left\{ W^n(\mathbf{x}_1^{(m_1)}, \mathbf{x}_2^{(j)}, \mathbf{Y}) \geq W^n(\mathbf{x}_1^{(m_1)}, \mathbf{x}_2^{(m_2)}, \mathbf{Y}) \right\} \right] \right\} \quad (4.187)$$

$$\geq \frac{1}{2} \mathbb{P}^{(\text{Genie})}[(\hat{m}_1, \hat{m}_2) \neq (m_1, m_2)] \quad (4.188)$$

$$= \frac{1}{2} \mathbb{P}^{(\text{S})}[(\hat{m}_1, \hat{m}_2) \neq (m_1, m_2)], \quad (4.189)$$

where (4.187) follows since the two steps of the genie-aided decoder minimize the two terms in the $\max\{\cdot, \cdot\}$, (4.188) follows by writing $\max\{\mathbb{P}[A], \mathbb{P}[B]\} \geq \frac{1}{2}(\mathbb{P}[A] + \mathbb{P}[B]) \geq \frac{1}{2}\mathbb{P}[A \cup B]$, and (4.189) follows since the overall error probability is unchanged by the genie [127].

4.5.3 Proof of Theorem 4.12

Here we present the proof of Theorem 4.12. Our analysis is based on the method of type class enumeration (e.g. see [118, 128, 129]), and is perhaps most similar to that of Somekh-Baruch and Merhav [129]. We use the same random-coding ensemble as that of Section 4.2.2, which we define again here for completeness. For $\nu = 1, 2$ we let $Q_{\nu,n}$ be a type with the same support as Q_ν such that $\max_{x_\nu} |Q_{\nu,n}(x_\nu) - Q_\nu(x_\nu)| \leq \frac{1}{n}$, and we define the constant-composition codeword distribution

$$P_{\mathbf{X}_\nu}(\mathbf{x}_\nu) = \frac{1}{|T^n(Q_{\nu,n})|} \mathbf{1}\{\mathbf{x}_\nu \in T^n(Q_{\nu,n})\}. \quad (4.190)$$

The (independent) random codewords are denoted by $\{\mathbf{X}_\nu^{(i)}\}_{i=1}^{M_\nu}$. We assume without loss of generality that $m_1 = m_2 = 1$, and we write $\mathbf{X}_\nu = \mathbf{X}_\nu^{(1)}$ and let $\bar{\mathbf{X}}_\nu$ denote an arbitrary codeword $\mathbf{X}_\nu^{(j)}$ with $j \neq 1$. Thus,

$$(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}, \bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2) \sim P_{\mathbf{X}_1}(\mathbf{x}_1) P_{\mathbf{X}_2}(\mathbf{x}_2) W^n(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2) P_{\mathbf{X}_1}(\bar{\mathbf{x}}_1) P_{\mathbf{X}_2}(\bar{\mathbf{x}}_2). \quad (4.191)$$

We define the following error events:

- (Type 1) $\sum_j q^n(\mathbf{X}_1^{(i)}, \mathbf{X}_2^{(j)}, \mathbf{Y}) \geq \sum_j q^n(\mathbf{X}_1, \mathbf{X}_2^{(j)}, \mathbf{Y})$ for some $i \neq 1$;
 (Type 2) $q^n(\mathbf{X}_1, \mathbf{X}_2^{(j)}, \mathbf{Y}) \geq q^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$ for some $j \neq 1$.

Denoting the probabilities of these events by $\bar{p}_{e,1}$ and $\bar{p}_{e,2}$ respectively, it follows that the overall random-coding error probability $\bar{p}_e$ is upper bounded by $\bar{p}_{e,1} + \bar{p}_{e,2}$.

The analysis of the type-2 error event is precisely that of Section 4.2, yielding the rate condition in (4.130). We thus focus on the type-1 event. We let $\bar{p}_{e,1}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$ denote the probability of the type-1 event conditioned on $(\mathbf{X}_1^{(1)}, \mathbf{X}_2^{(1)}, \mathbf{Y}) = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$, and we denote the joint type of $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$ by $P_{X_1 X_2 Y}$. We write the objective function in (4.122) as

$$\Xi_{\mathbf{x}_2 \mathbf{y}}(\bar{\mathbf{x}}_1) \triangleq q^n(\bar{\mathbf{x}}_1, \mathbf{x}_2, \mathbf{y}) + \sum_{j \neq 1} q^n(\bar{\mathbf{x}}_1, \mathbf{X}_2^{(j)}, \mathbf{y}). \quad (4.192)$$

This quantity is random due to the randomness of $\{\mathbf{X}_2^{(j)}\}$. The starting point of our analysis is the union bound:

$$\bar{p}_{e,1}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \leq (M_1 - 1) \mathbb{P}[\Xi_{\mathbf{x}_2 \mathbf{y}}(\bar{\mathbf{X}}_1) \geq \Xi_{\mathbf{x}_2 \mathbf{y}}(\mathbf{x}_1)]. \quad (4.193)$$

The difficulty in analyzing (4.193) is that for two different codewords $\mathbf{x}_1$ and $\bar{\mathbf{x}}_1$, $\Xi_{\mathbf{x}_2 \mathbf{y}}(\mathbf{x}_1)$ and $\Xi_{\mathbf{x}_2 \mathbf{y}}(\bar{\mathbf{x}}_1)$ are not independent, and their joint statistics are complicated. We will circumvent this issue by conditioning on high probability events under which these random quantities can be bounded by deterministic values.

From (4.192), we have

$$\Xi_{\mathbf{x}_2 \mathbf{y}}(\bar{\mathbf{x}}_1) = q^n(\tilde{P}_{X_1 X_2 Y}) + \sum_{\tilde{P}'_{X_1 X_2 Y}} N_{\bar{\mathbf{x}}_1 \mathbf{y}}(\tilde{P}'_{X_1 X_2 Y}) q^n(\tilde{P}'_{X_1 X_2 Y}), \quad (4.194)$$

where $N_{\bar{\mathbf{x}}_1 \mathbf{y}}(\tilde{P}'_{X_1 X_2 Y})$ is the random number of $\mathbf{X}_2^{(j)}$ ($j \neq 1$) such that $(\bar{\mathbf{x}}_1, \mathbf{X}_2^{(j)}, \mathbf{y}) \in T^n(\tilde{P}'_{X_1 X_2 Y})$, $\tilde{P}_{X_1 X_2 Y}$ is the joint type of $(\bar{\mathbf{x}}_1, \mathbf{x}_2, \mathbf{y})$, and we write $q^n(\tilde{P}'_{X_1 X_2 Y}) \triangleq q^n(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{y})$ for an arbitrary triplet $(\bar{\mathbf{x}}_1, \bar{\mathbf{x}}_2, \mathbf{y}) \in T^n(\tilde{P}'_{X_1 X_2 Y})$. Since the codewords are generated independently, $N_{\bar{\mathbf{x}}_1 \mathbf{y}}(\tilde{P}'_{X_1 X_2 Y})$ is binomially distributed with $M_2 - 1$ trials and success probability $\mathbb{P}[(\bar{\mathbf{x}}_1, \bar{\mathbf{X}}_2, \mathbf{y}) \in T^n(\tilde{P}'_{X_1 X_2 Y})]$. By construction, we have $N_{\bar{\mathbf{x}}_1 \mathbf{y}}(\tilde{P}'_{X_1 X_2 Y}) = 0$ unless $\tilde{P}'_{X_1 X_2 Y} \in \mathcal{S}'_{1,n}(Q_{2,n}, \tilde{P}_{X_1 Y})$, where

$$\mathcal{S}'_{1,n}(Q_{2,n}, \tilde{P}_{X_1 Y}) \triangleq \left\{ \tilde{P}'_{X_1 X_2 Y} \in \mathcal{P}_n(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \tilde{P}'_{X_1 Y} = \tilde{P}_{X_1 Y}, \tilde{P}'_{X_2} = Q_{2,n} \right\}. \quad (4.195)$$

The following lemma characterizes the behavior of $N_{\bar{\mathbf{x}}_1 \mathbf{y}}(\tilde{P}'_{X_1 X_2 Y})$ for various choices of R_2

and $\tilde{P}'_{X_1X_2Y}$. The proof can be found in [118, 129], and is based on the fact that

$$\mathbb{P}[(\bar{\mathbf{x}}_1, \bar{\mathbf{X}}_2, \mathbf{y}) \in T^n(\tilde{P}'_{X_1X_2Y})] \doteq e^{-nI_{\tilde{P}'}(X_2; X_1, Y)}, \quad (4.196)$$

which follows from the property of types in (A.17).

Lemma 4.4. [118, 129] *Fix the pair $(\bar{\mathbf{x}}_1, \mathbf{y}) \in T^n(\tilde{P}_{X_1Y})$, a constant $\delta > 0$, and a type $\tilde{P}'_{X_1X_2Y} \in \mathcal{S}'_{1,n}(Q_{2,n}, \tilde{P}_{X_1Y})$.*

(i) *If $R_2 \geq I_{\tilde{P}'}(X_2; X_1, Y) + \delta$, then*

$$M_2 e^{-n(I_{\tilde{P}'}(X_2; X_1, Y) + \delta)} \leq N_{\bar{\mathbf{x}}_1 \mathbf{y}}(\tilde{P}'_{X_1X_2Y}) \leq M_2 e^{-n(I_{\tilde{P}'}(X_2; X_1, Y) - \delta)} \quad (4.197)$$

with probability approaching one super-exponentially fast.

(ii) *If $R_2 < I_{\tilde{P}'}(X_2; X_1, Y) + \delta$, then*

$$N_{\bar{\mathbf{x}}_1 \mathbf{y}}(\tilde{P}'_{X_1X_2Y}) \leq e^{-n2\delta} \quad (4.198)$$

with probability approaching one super-exponentially fast.

Roughly speaking, Lemma 4.4 states that if $R_2 > I_{\tilde{P}'}(X_2; X_1, Y)$ then the type enumerator is highly concentrated about its mean, whereas if $R_2 < I_{\tilde{P}'}(X_2; X_1, Y)$ then the type enumerator takes a subexponential value (possibly zero) with overwhelming probability.

Given a joint type $\tilde{P}_{X_1Y}$, let $\mathcal{A}_\delta(\tilde{P}_{X_1Y})$ denote the union of the high-probability events in Lemma 4.4 taken over all $\tilde{P}'_{X_1X_2Y} \in \mathcal{S}'_{1,n}(Q_2, \tilde{P}_{X_1Y})$. By the union bound, $\mathbb{P}[\mathcal{A}_\delta(\tilde{P}_{X_1Y})] \rightarrow 1$ super-exponentially fast, and hence we can safely condition any event on $\mathcal{A}_\delta(\tilde{P}_{X_1Y})$ without changing the exponential behavior of the corresponding probability. This can be seen by writing the following for any event $\mathcal{E}$:

$$\mathbb{P}[\mathcal{E}] = \mathbb{P}[\mathcal{E} \cap \mathcal{A}] + \mathbb{P}[\mathcal{E} \cap \mathcal{A}^c] \quad (4.199)$$

$$\leq \mathbb{P}[\mathcal{E} \mid \mathcal{A}] + \mathbb{P}[\mathcal{A}^c] \quad (4.200)$$

$$\mathbb{P}[\mathcal{E}] \geq \mathbb{P}[\mathcal{E} \cap \mathcal{A}] \quad (4.201)$$

$$= (1 - \mathbb{P}[\mathcal{A}^c])\mathbb{P}[\mathcal{E} \mid \mathcal{A}] \quad (4.202)$$

$$\geq \mathbb{P}[\mathcal{E} \mid \mathcal{A}] - \mathbb{P}[\mathcal{A}^c]. \quad (4.203)$$

Conditioned on $\mathcal{A}_\delta(P_{X_1Y})$, we have the following:

$$\Xi_{\mathbf{x}_2\mathbf{y}}(\mathbf{x}_1) = q^n(P_{X_1X_2Y}) + \sum_{P'_{X_1X_2Y}} N_{\mathbf{x}_1\mathbf{y}}(P'_{X_1X_2Y}) q^n(P'_{X_1X_2Y}) \quad (4.204)$$

$$\geq q^n(P_{X_1X_2Y}) + \max_{\substack{P'_{X_1X_2Y} \in \mathcal{S}'_{1,n}(Q_{2,n}, P_{X_1Y}) \\ R_2 \geq I_{P'}(X_2; X_1, Y) + \delta}} N_{\mathbf{x}_1\mathbf{y}}(P'_{X_1X_2Y}) q^n(P'_{X_1X_2Y}) \quad (4.205)$$

$$\geq q^n(P_{X_1X_2Y}) + \max_{\substack{P'_{X_1X_2Y} \in \mathcal{S}'_{1,n}(Q_{2,n}, P_{X_1Y}) \\ R_2 \geq I_{P'}(X_2; X_1, Y) + \delta}} M_2 e^{-n(I_{P'}(X_2; X_1, Y) + \delta)} q^n(P'_{X_1X_2Y}) \quad (4.206)$$

$$\triangleq \underline{G}_{\delta,n}(P_{X_1X_2Y}), \quad (4.207)$$

where (4.206) follows from (4.197). Unlike $\Xi_{\mathbf{x}_2\mathbf{y}}(\mathbf{x}_1)$, the quantity $\underline{G}_{\delta,n}(P_{X_1X_2Y})$ is deterministic. Substituting (4.207) into (4.193) and using the fact that $\mathbb{P}[\mathcal{A}_\delta(\tilde{P}_{X_1Y})] \rightarrow 1$ super-exponentially fast, we obtain

$$\bar{p}_{e,1}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \leq M_1 \mathbb{P}[\Xi_{\mathbf{x}_2\mathbf{y}}(\bar{\mathbf{X}}_1) \geq \underline{G}_{\delta,n}(P_{X_1X_2Y})]. \quad (4.208)$$

Since the statistics of $\Xi_{\mathbf{x}_2\mathbf{y}}(\bar{\mathbf{x}}_1)$ depend on $\bar{\mathbf{x}}_1$ only through the joint type of $(\bar{\mathbf{x}}_1, \mathbf{x}_2, \mathbf{y})$, we can write (4.208) as follows:

$$\bar{p}_{e,1}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \leq M_1 \sum_{\tilde{P}_{X_1X_2Y}} \mathbb{P}[(\bar{\mathbf{X}}_1, \mathbf{x}_2, \mathbf{y}) \in T^n(\tilde{P}_{X_1X_2Y})] \mathbb{P}[\Xi_{\mathbf{x}_2\mathbf{y}}(\bar{\mathbf{x}}_1) \geq \underline{G}_{\delta,n}(P_{X_1X_2Y})] \quad (4.209)$$

$$\doteq M_1 \max_{\tilde{P}_{X_1X_2Y} \in \mathcal{S}_{1,n}(Q_{1,n}, P_{X_2Y})} e^{-nI_{\tilde{P}}(X_1; X_2, Y)} \mathbb{P}[\Xi_{\mathbf{x}_2\mathbf{y}}(\bar{\mathbf{x}}_1) \geq \underline{G}_{\delta,n}(P_{X_1X_2Y})], \quad (4.210)$$

where $\bar{\mathbf{x}}_1$ denotes an arbitrary sequence such that $(\bar{\mathbf{x}}_1, \mathbf{x}_2, \mathbf{y}) \in T^n(\tilde{P}_{X_1X_2Y})$, and

$$\mathcal{S}_{1,n}(Q_{1,n}, P_{X_2Y}) \triangleq \left\{ \tilde{P}_{X_1X_2Y} \in \mathcal{P}_n(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \tilde{P}_{X_1} = Q_{1,n}, \tilde{P}_{X_2Y} = P_{X_2Y} \right\}. \quad (4.211)$$

In (4.210), we have used an analogous property to (4.196) and the fact that by construction, the joint type of $(\bar{\mathbf{X}}_1, \mathbf{x}_2, \mathbf{y})$ is in $\mathcal{S}_{1,n}(Q_{1,n}, P_{X_2Y})$ with probability one.

Next, we again use Lemma 4.4 in order to replace $\Xi_{\mathbf{x}_2\mathbf{y}}(\bar{\mathbf{x}}_1)$ in (4.210) by a deterministic quantity. We have from (4.194) that

$$\Xi_{\mathbf{x}_2\mathbf{y}}(\bar{\mathbf{x}}_1) \leq q^n(\tilde{P}_{X_1X_2Y}) + p_0(n) \max_{\tilde{P}'_{X_1X_2Y}} N_{\bar{\mathbf{x}}_1\mathbf{y}}(\tilde{P}'_{X_1X_2Y}) q^n(\tilde{P}'_{X_1X_2Y}), \quad (4.212)$$

where $p_0(n)$ is a polynomial corresponding to the total number of joint types. Substituting (4.212) into (4.210), we obtain

$$\begin{aligned} & \bar{p}_{e,1}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \\ & \leq M_1 \max_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{S}_{1,n}(Q_{1,n}, P_{X_2 Y})} \max_{\tilde{P}'_{X_1 X_2 Y} \in \mathcal{S}'_{1,n}(Q_{2,n}, \tilde{P}_{X_1 Y})} e^{-nI_{\tilde{P}}(X_1; X_2, Y)} \mathbb{P}[\mathcal{E}_{P, \tilde{P}}(\tilde{P}'_{X_1 X_2 Y})], \end{aligned} \quad (4.213)$$

where

$$\mathcal{E}_{P, \tilde{P}}(\tilde{P}'_{X_1 X_2 Y}) \triangleq \left\{ q^n(\tilde{P}_{X_1 X_2 Y}) + p_0(n) N_{\bar{\mathbf{x}}_1 \mathbf{y}}(\tilde{P}'_{X_1 X_2 Y}) q^n(\tilde{P}'_{X_1 X_2 Y}) \geq \underline{G}_{\delta, n}(P_{X_1 X_2 Y}) \right\}, \quad (4.214)$$

and we have used the union bound to take the maximum over $\tilde{P}'_{X_1 X_2 Y}$ outside the probability in (4.213). Continuing, we have for any $\tilde{P}_{X_1 X_2 Y} \in \mathcal{S}_{1,n}(Q_{1,n}, P_{X_2 Y})$ that

$$\begin{aligned} \max_{\tilde{P}'_{X_1 X_2 Y} \in \mathcal{S}'_{1,n}(Q_{2,n}, \tilde{P}_{X_1 Y})} \mathbb{P}[\mathcal{E}_{P, \tilde{P}}(\tilde{P}'_{X_1 X_2 Y})] &= \max \left\{ \max_{\substack{\tilde{P}'_{X_1 X_2 Y} \in \mathcal{S}'_{1,n}(Q_{2,n}, \tilde{P}_{X_1 Y}) \\ R_2 \geq I_{\tilde{P}'}(X_2; X_1, Y) + \delta}} \mathbb{P}[\mathcal{E}_{P, \tilde{P}}(\tilde{P}'_{X_1 X_2 Y})], \right. \\ & \quad \left. \max_{\substack{\tilde{P}'_{X_1 X_2 Y} \in \mathcal{S}'_{1,n}(Q_{2,n}, \tilde{P}_{X_1 Y}) \\ R_2 < I_{\tilde{P}'}(X_2; X_1, Y) + \delta}} \mathbb{P}[\mathcal{E}_{P, \tilde{P}}(\tilde{P}'_{X_1 X_2 Y})] \right\}. \end{aligned} \quad (4.215)$$

For the first maximization in (4.215), observe that conditioned on $\mathcal{A}_\delta(\tilde{P}_{X_1 Y})$ (defined following Lemma 4.4), we have for $\tilde{P}'_{X_1 X_2 Y}$ satisfying $R_2 \geq I_{\tilde{P}'}(X_2; X_1, Y) + \delta$ that

$$N_{\bar{\mathbf{x}}_1 \mathbf{y}}(\tilde{P}'_{X_1 X_2 Y}) q^n(\tilde{P}'_{X_1 X_2 Y}) \leq M_2 e^{-n(I_{\tilde{P}'}(X_2; X_1, Y) - \delta)} q^n(\tilde{P}'_{X_1 X_2 Y}), \quad (4.216)$$

where we have used (4.197). Hence, and since $\mathbb{P}[\mathcal{A}_\delta(\tilde{P}_{X_1 Y})] \rightarrow 1$ super-exponentially fast, we have

$$\begin{aligned} & \mathbb{P}[\mathcal{E}_{P, \tilde{P}}(\tilde{P}'_{X_1 X_2 Y})] \\ & \leq \mathbb{1} \left\{ q^n(\tilde{P}_{X_1 X_2 Y}) + M_2 p_0(n) e^{-n(I_{\tilde{P}'}(X_2; X_1, Y) - \delta)} q^n(\tilde{P}'_{X_1 X_2 Y}) \geq \underline{G}_{\delta, n}(P_{X_1 X_2 Y}) \right\}. \end{aligned} \quad (4.217)$$

For the second maximization in (4.215), we define the event $\mathcal{B} \triangleq \{N_{\bar{\mathbf{x}}_1 \mathbf{y}}(\tilde{P}'_{X_1 X_2 Y}) > 0\}$, yielding

$$\mathbb{P}[\mathcal{B}] \leq M_2 e^{-nI_{\tilde{P}'}(X_2; X_1, Y)}, \quad (4.218)$$

which follows from the union bound and (4.196). Whenever $R_2 < I_{\tilde{P}'}(X_2; X_1, Y) + \delta$, we have

$$\begin{aligned} & \mathbb{P}[\mathcal{E}_{P, \tilde{P}}(\tilde{P}'_{X_1 X_2 Y})] \\ & \leq \mathbb{P}[\mathcal{E}_{P, \tilde{P}}(\tilde{P}'_{X_1 X_2 Y}) \mid \mathcal{B}^c] + \mathbb{P}[\mathcal{B}] \mathbb{P}[\mathcal{E}_{P, \tilde{P}}(\tilde{P}'_{X_1 X_2 Y}) \mid \mathcal{B}] \end{aligned} \quad (4.219)$$

$$\leq \mathbb{1}\{q^n(\tilde{P}_{X_1 X_2 Y}) \geq \underline{G}_{\delta, n}(P_{X_1 X_2 Y})\} + M_2 e^{-n I_{\tilde{P}'}(X_2; X_1, Y)} \mathbb{P}[\mathcal{E}_{P, \tilde{P}}(\tilde{P}'_{X_1 X_2 Y}) \mid \mathcal{B}], \quad (4.220)$$

$$\begin{aligned} & \leq \mathbb{1}\{q^n(\tilde{P}_{X_1 X_2 Y}) \geq \underline{G}_{\delta, n}(P_{X_1 X_2 Y})\} \\ & \quad + M_2 e^{-n I_{\tilde{P}'}(X_2; X_1, Y)} \mathbb{1}\{q^n(\tilde{P}_{X_1 X_2 Y}) + p_0(n) e^{-n 2\delta} q^n(\tilde{P}'_{X_1 X_2 Y}) \geq \underline{G}_{\delta, n}(P_{X_1 X_2 Y})\}, \end{aligned} \quad (4.221)$$

where (4.220) follows using (4.218) and (4.214) along with the fact that $\mathcal{B}^c$ implies $N_{\bar{\mathbf{x}}_1 \mathbf{y}}(\tilde{P}'_{X_1 X_2 Y}) = 0$, and (4.221) follows by conditioning on $\mathcal{A}_\delta(\tilde{P}_{X_1 Y})$ and using (4.198).

Observe that $\underline{F}(P_{X_1 X_2 Y}, R_2)$ in (4.125) equals the exponent of $\underline{G}_{\delta, n}$ in (4.207) in the limit as $\delta \rightarrow 0$ and $n \rightarrow \infty$. Similarly, the exponents corresponding to the other quantities appearing in the indicator functions in (4.217) and (4.221) tend toward the following:

$$\begin{aligned} \bar{F}_1(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}, R_2) & \triangleq \max \left\{ \mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)], \right. \\ & \quad \left. \mathbb{E}_{\tilde{P}'}[\log q(X_1, X_2, Y)] + R_2 - I_{\tilde{P}'}(X_2; X_1, Y) \right\} \end{aligned} \quad (4.222)$$

$$\bar{F}_2(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) \triangleq \max \left\{ \mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)], \mathbb{E}_{\tilde{P}'}[\log q(X_1, X_2, Y)] \right\}. \quad (4.223)$$

Combining these expressions with (4.213), (4.215), (4.217) and (4.221), taking $\delta \rightarrow 0$, and using the continuity of the underlying terms in the optimizations, we obtain

$$\begin{aligned} \bar{p}_{e,1}(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) & \leq \max \left\{ \max_{(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) \in \mathcal{T}_1^{(1)}(P_{X_1 X_2 Y}, R_2)} M_1 e^{-n I_{\tilde{P}}(X_1; X_2, Y)}, \right. \\ & \quad \left. \max_{(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) \in \mathcal{T}_1^{(2)}(P_{X_1 X_2 Y}, R_2)} M_1 e^{-n I_{\tilde{P}}(X_1; X_2, Y)} M_2 e^{-n I_{\tilde{P}'}(X_2; X_1, Y)} \right\}, \end{aligned} \quad (4.224)$$

where⁴

$$\begin{aligned} \mathcal{T}_1^{(1)}(P_{X_1 X_2 Y}, R_2) & \triangleq \left\{ (\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) : \tilde{P}_{X_1 X_2 Y} \in \mathcal{S}_1(Q_1, P_{X_2 Y}), \right. \\ & \quad \left. \tilde{P}'_{X_1 X_2 Y} \in \mathcal{S}'_1(Q_2, \tilde{P}_{X_1 Y}), I_{\tilde{P}'}(X_2; X_1, Y) \leq R_2, \bar{F}_1(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}, R_2) \geq \underline{F}(P_{X_1 X_2 Y}, R_2) \right\} \end{aligned} \quad (4.225)$$

⁴Strictly speaking, these sets depend on (Q_1, Q_2) , but this dependence need not be explicit, since we have $P_{X_1} = Q_1$ and $P_{X_2} = Q_2$.

$$\begin{aligned} \mathcal{T}_1^{(2)}(P_{X_1 X_2 Y}, R_2) \triangleq & \left\{ (\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) : \tilde{P}_{X_1 X_2 Y} \in \mathcal{S}_1(Q_1, P_{X_2 Y}), \right. \\ & \left. \tilde{P}'_{X_1 X_2 Y} \in \mathcal{S}'_1(Q_2, \tilde{P}_{X_1 Y}), I_{\tilde{P}'}(X_2; X_1, Y) \geq R_2, \bar{F}_2(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) \geq \underline{F}(P_{X_1 X_2 Y}, R_2) \right\}, \end{aligned} \quad (4.226)$$

and

$$\mathcal{S}_1(Q_1, P_{X_2 Y}) \triangleq \left\{ \tilde{P}_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \tilde{P}_{X_1} = Q_1, \tilde{P}_{X_2 Y} = P_{X_2 Y} \right\} \quad (4.227)$$

$$\mathcal{S}'_1(Q_2, \tilde{P}_{X_1 Y}) \triangleq \left\{ \tilde{P}'_{X_1 X_2 Y} \in \mathcal{P}(\mathcal{X}_1 \times \mathcal{X}_2 \times \mathcal{Y}) : \tilde{P}'_{X_1 Y} = \tilde{P}_{X_1 Y}, \tilde{P}'_{X_2} = Q_2 \right\}. \quad (4.228)$$

The two terms in the maximization in (4.224) respectively correspond to (4.217) and the second term in (4.221). The first term in (4.221) is upper bounded by the right-hand side of (4.217), and thus need not be included in the maximum in (4.224).

The property of types in (A.16) implies that $\mathbb{P}[(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \in T^n(P_{X_1 X_2 Y})]$ decays to zero exponentially fast when $P_{X_1 X_2 Y}$ is bounded away from $Q_1 \times Q_2 \times W$. Therefore, similarly to the proof of Theorem 4.3, we can safely substitute $P_{X_1 X_2 Y} = Q_1 \times Q_2 \times W$ to obtain the following rate conditions for the first decoding step:

$$R_1 \leq \min_{(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) \in \mathcal{T}_1^{(1)}(Q_1 \times Q_2 \times W, R_2)} I_{\tilde{P}}(X_1; X_2, Y) \quad (4.229)$$

$$R_1 + R_2 \leq \min_{(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) \in \mathcal{T}_1^{(2)}(Q_1 \times Q_2 \times W, R_2)} I_{\tilde{P}}(X_1; X_2, Y) + I_{\tilde{P}'}(X_2; X_1, Y). \quad (4.230)$$

Finally, using the definitions of $\bar{F}$, $\mathcal{S}_1$, $\mathcal{S}'_1$, $\mathcal{T}_1^{(1)}$ and $\mathcal{T}_1^{(2)}$ (see (4.124) and (4.225)–(4.228)) to unite (4.229)–(4.230), we obtain (4.129).

4.5.4 Formulation of (4.129) in terms of Convex Optimization Problems

Here we write (4.129) in terms of convex optimization problems, starting with the alternative formulation in (4.229)–(4.230). We first note that (4.230) holds if and only if

$$R_1 \leq \min_{(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) \in \mathcal{T}_1^{(2)}(Q_1 \times Q_2 \times W, R_2)} I_{\tilde{P}}(X_1; X_2, Y) + [I_{\tilde{P}'}(X_2; X_1, Y) - R_2]^+, \quad (4.231)$$

since the argument to the $[\cdot]^+$ is always non-negative due to the constraint $I_{\tilde{P}'}(X_2; X_1, Y) \geq R_2$. Next, we claim that when combining (4.229) and (4.231), the rate region is unchanged if the constraint $I_{\tilde{P}'}(X_2; X_1, Y) \geq R_2$ is omitted from (4.231). This is seen by noting that whenever $I_{\tilde{P}'}(X_2; X_1, Y) < R_2$, the objective in (4.231) coincides with that of (4.229), whereas the latter

has a less restrictive constraint since $\bar{F}_1 > \bar{F}_2$ (see (4.222)–(4.223)).

We now deal with the non-concavity of the functions $\bar{F}_1$ and $\bar{F}_2$ appearing in the sets $\mathcal{T}_1^{(1)}$ and $\mathcal{T}_1^{(2)}$. Using the identity

$$\min_{x \leq \max\{a,b\}} f(x) = \min \left\{ \min_{x \leq a} f(x), \min_{x \leq b} f(x) \right\}, \quad (4.232)$$

we obtain the following rate conditions from (4.229) and (4.231):

$$R_1 \leq \min_{(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) \in \mathcal{T}_1^{(1,1)}(Q_1 \times Q_2 \times W, R_2)} I_{\tilde{P}}(X_1; X_2, Y) \quad (4.233)$$

$$R_1 \leq \min_{(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) \in \mathcal{T}_1^{(1,2)}(Q_1 \times Q_2 \times W, R_2)} I_{\tilde{P}}(X_1; X_2, Y) \quad (4.234)$$

$$R_1 \leq \min_{(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) \in \mathcal{T}_1^{(2,1)}(Q_1 \times Q_2 \times W, R_2)} I_{\tilde{P}}(X_1; X_2, Y) + [I_{\tilde{P}'}(X_2; X_1, Y) - R_2]^+ \quad (4.235)$$

$$R_1 \leq \min_{(\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) \in \mathcal{T}_1^{(2,2)}(Q_1 \times Q_2 \times W, R_2)} I_{\tilde{P}}(X_1; X_2, Y) + [I_{\tilde{P}'}(X_2; X_1, Y) - R_2]^+, \quad (4.236)$$

where

$$\begin{aligned} \mathcal{T}_1^{(1,1)}(P_{X_1 X_2 Y}, R_2) \triangleq & \left\{ (\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) : \tilde{P}_{X_1 X_2 Y} \in \mathcal{S}_1(Q_1, P_{X_2 Y}), \right. \\ & \left. \tilde{P}'_{X_1 X_2 Y} \in \mathcal{S}'_1(Q_2, \tilde{P}_{X_1 Y}), I_{\tilde{P}'}(X_2; X_1, Y) \leq R_2, \mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)] \geq \underline{F}(P_{X_1 X_2 Y}, R_2) \right\} \end{aligned} \quad (4.237)$$

$$\begin{aligned} \mathcal{T}_1^{(1,2)}(P_{X_1 X_2 Y}, R_2) \triangleq & \left\{ (\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) : \tilde{P}_{X_1 X_2 Y} \in \mathcal{S}_1(Q_1, P_{X_2 Y}), \right. \\ & \tilde{P}'_{X_1 X_2 Y} \in \mathcal{S}'_1(Q_2, \tilde{P}_{X_1 Y}), I_{\tilde{P}'}(X_2; X_1, Y) \leq R_2, \\ & \left. \mathbb{E}_{\tilde{P}'}[\log q(X_1, X_2, Y)] + R_2 - I_{\tilde{P}'}(X_2; X_1, Y) \geq \underline{F}(P_{X_1 X_2 Y}, R_2) \right\} \end{aligned} \quad (4.238)$$

$$\begin{aligned} \mathcal{T}_1^{(2,1)}(P_{X_1 X_2 Y}, R_2) \triangleq & \left\{ (\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) : \tilde{P}_{X_1 X_2 Y} \in \mathcal{S}_1(Q_1, P_{X_2 Y}), \right. \\ & \left. \tilde{P}'_{X_1 X_2 Y} \in \mathcal{S}'_1(Q_2, \tilde{P}_{X_1 Y}), \mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)] \geq \underline{F}(P_{X_1 X_2 Y}, R_2) \right\} \end{aligned} \quad (4.239)$$

$$\begin{aligned} \mathcal{T}_1^{(2,2)}(P_{X_1 X_2 Y}, R_2) \triangleq & \left\{ (\tilde{P}_{X_1 X_2 Y}, \tilde{P}'_{X_1 X_2 Y}) : \tilde{P}_{X_1 X_2 Y} \in \mathcal{S}_1(Q_1, P_{X_2 Y}), \right. \\ & \left. \tilde{P}'_{X_1 X_2 Y} \in \mathcal{S}'_1(Q_2, \tilde{P}_{X_1 Y}), \mathbb{E}_{\tilde{P}'}[\log q(X_1, X_2, Y)] \geq \underline{F}(P_{X_1 X_2 Y}, R_2) \right\}. \end{aligned} \quad (4.240)$$

These are obtained from $\mathcal{T}^{(k)}$ ($k = 1, 2$) by keeping only one term in the definition of $\bar{F}_k$ (see (4.222)–(4.223)), and by removing the constraint $I_{\tilde{P}'}(X_2; X_1, Y) \geq R_2$ when $k = 2$ in

accordance with the discussion following (4.231).

The variable $\tilde{P}'_{X_1 X_2 Y}$ can be removed from both (4.233) and (4.235), since in each case the choice $\tilde{P}'_{X_1 X_2 Y}(x_1, x_2, y) = Q_2(x_2) \tilde{P}_{X_1 Y}(x_1, y)$ is feasible and yields $I_{\tilde{P}'}(X_2; X_1, Y) = 0$. It follows that (4.233) and (4.235) yield the same value, and we conclude that (4.129) can equivalently be expressed in terms of three conditions: (4.234), (4.236), and

$$R_1 \leq \min_{\tilde{P}_{X_1 X_2 Y} \in \mathcal{T}_1^{(1,1')}(Q_1 \times Q_2 \times W, R_2)} I_{\tilde{P}}(X_1; X_2, Y), \quad (4.241)$$

where the set

$$\mathcal{T}_1^{(1,1')}(P_{X_1 X_2 Y}, R_2) \triangleq \left\{ \tilde{P}_{X_1 X_2 Y} \in \mathcal{S}_1(Q_1, P_{X_2 Y}) : \mathbb{E}_{\tilde{P}}[\log q(X_1, X_2, Y)] \geq \underline{F}(P_{X_1 X_2 Y}, R_2) \right\} \quad (4.242)$$

is obtained by eliminating $\tilde{P}'_{X_1 X_2 Y}$ from either (4.237) or (4.239). These three conditions are all written as convex optimization problems, as desired.

4.5.5 Proof of Theorem 4.13

While Theorem 4.13 is presented in terms of a single triplet (Q_U, Q_1, Q_2) , the random-coding ensemble used in the proof will, in general, depend on a second triplet (Q'_U, Q'_1, Q'_2) on the same alphabets. We define $P'_{U X_1 X_2 Y}$, $\mathbf{i}'$, $\mathbf{i}'^{(1)}$, $\mathbf{i}'^{(2)}$, $\mathbf{I}'$ and $\mathbf{V}'$ in the same way as Sections 4.4.1 and 4.4.2, with (Q'_U, Q'_1, Q'_2) replacing (Q_U, Q_1, Q_2) . In particular, we have

$$\mathbf{I}' \triangleq \mathbb{E}[\mathbf{i}'(U', X'_1, X'_2, Y')] \quad (4.243)$$

$$\mathbf{V}' \triangleq \mathbb{E}[\text{Cov}[\mathbf{i}'(U', X'_1, X'_2, Y') \mid U'] - \text{Cov}[\mathbf{i}'^{(1)}(U', X'_1) \mid U'] - \text{Cov}[\mathbf{i}'^{(2)}(U', X'_2) \mid U']], \quad (4.244)$$

where $(U', X'_1, X'_2, Y') \sim P'_{U X_1 X_2 Y}$. As an intermediate step towards obtaining our local result, we present the following global result.

Theorem 4.15. *Fix any finite time sharing alphabet $\mathcal{U}$, and the input distributions (Q_U, Q_1, Q_2) and (Q'_U, Q'_1, Q'_2) . For any $\beta \geq 0$ and $\epsilon \in (0, 1)$, there exists a function $g(n) = O(n^{\frac{1}{4}})$ (depending on (Q_U, Q_1, Q_2) , (Q'_U, Q'_1, Q'_2) , β and ϵ) such that all rate pairs (R_1, R_2) satisfying*

$$n\mathbf{R} \in n\mathbf{I} + \sqrt{n}(\beta(\mathbf{I}' - \mathbf{I}) - \mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon)) + g(n)\mathbf{1} \quad (4.245)$$

are (n, ϵ) -achievable.

Proof. For clarity of exposition, we present the proof in two parts. In Section 4.5.6, we handle the case that $\beta = 0$ and $\mathcal{U} = \emptyset$. The changes required to handle the general case are presented in Section 4.5.7. $\square$

In the case that $\beta = 0$, Theorem 4.15 gives the second-order asymptotics for a fixed triplet (Q_U, Q_1, Q_2) . In this case, the proof reveals that the behavior $g(n) = O(n^{\frac{1}{4}})$ can be strengthened to $g(n) = O(\log n)$ if the argument to the expectation in (4.144) has full rank for all u , and $g(n) = O(n^{\frac{1}{6}})$ more generally.

The case $\beta > 0$ is proved by applying an extended version of coded time sharing (cf. Section 4.2.5) in which a fraction $1 - \frac{\beta}{\sqrt{n}}$ of the symbols are generated using (Q_U, Q_1, Q_2) , and the remaining symbols are generated using (Q'_U, Q'_1, Q'_2) .

We proceed by proving Theorem 4.13 using Theorem 4.15. We write $R_{12} \triangleq R_1 + R_2$ and $R_{12}^* \triangleq R_1^* + R_2^*$ (recall that (R_1^*, R_2^*) is the boundary point of interest). We have from the definition of $\mathcal{K}$ that $R_\nu^* < I_\nu$ for all $\nu \in \mathcal{K}^c$, and we can thus weaken (4.245) to

$$n\mathbf{R} \in n\mathbf{R}^* + n\delta\mathbf{1}_{\mathcal{K}^c} + \sqrt{n}\left(\beta(\mathbf{I}' - \mathbf{I}) - \mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon)\right) + O(n^{\frac{1}{4}}), \quad (4.246)$$

where $\delta \triangleq \min_{\nu \in \mathcal{K}^c} I_\nu - R_\nu^* > 0$, and $\mathbf{1}_{\mathcal{K}^c}$ contains ones at the indices corresponding to $\mathcal{K}^c$, and zeros elsewhere.

Let $\mathbf{z}$ be a 3×1 vector, and let $\mathbf{z}^{(\mathcal{K})}$ and $\mathbf{z}^{(\mathcal{K}^c)}$ be the corresponding subvectors indexed by the superscript. From the definition of $\mathbf{Q}_{\text{inv}}$, the set $-\mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon)$ contains the vectors $\mathbf{z}^{(\mathcal{K})}$ such that $\mathbf{z} \in -\mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon)$ in the limit as the entries of $\mathbf{z}^{(\mathcal{K}^c)}$ tend towards $-\infty$ [67]. Moreover, since $n\delta$ grows faster than $O(\sqrt{n})$, the elements of $n\mathbf{R}$ corresponding to $\mathcal{K}^c$ in (4.246) may incur an addition of $\ell\sqrt{n}$ to the first-order term for *any* value of ℓ . Combining these observations with the definition of $\mathcal{L}$, we obtain

$$\mathcal{L}(\epsilon, R_1^*, R_2^*) \supseteq \left\{ (\ell_1, \ell_2) : \ell^{(\mathcal{K})} \in \bigcup_{\beta \geq 0} \left\{ \beta(\mathbf{I}'^{(\mathcal{K})} - \mathbf{I}^{(\mathcal{K})}) - \mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon) \right\} \right\}. \quad (4.247)$$

Suppose for the time being that $R_2^* > 0$, and let (Q'_U, Q'_1, Q'_2) be chosen to achieve a boundary point (R'_1, R'_2) to the right of (R_1^*, R_2^*) (more precisely, one such that $R'_1 \geq R_1^*$ and $R'_2 \leq R_2^*$, with at least one of the inequalities being strict). We define R'_{12} , $\mathbf{R}'$ and $\mathbf{R}'^{(\mathcal{K})}$ in the same way as R_{12} , $\mathbf{R}$ and $\mathbf{R}^{(\mathcal{K})}$. Since (R'_1, R'_2) is achieved by (Q'_U, Q'_1, Q'_2) , we have $\mathbf{R}' \preceq \mathbf{I}'$, and in particular, $\mathbf{R}'^{(\mathcal{K})} \preceq \mathbf{I}'^{(\mathcal{K})}$. Moreover, from the definition of $\mathcal{K}$, we have $\mathbf{I}^{(\mathcal{K})} = \mathbf{R}^{(\mathcal{K})}$.

Combining these, we deduce from (4.247) that

$$\mathcal{L}(\epsilon, R_1^*, R_2^*) \supseteq \left\{ (\ell_1, \ell_2) : \ell^{(\mathcal{K})} \in \bigcup_{\beta \geq 0} \left\{ \beta(\mathbf{R}'^{(\mathcal{K})} - \mathbf{R}^{*(\mathcal{K})}) - \mathbf{Q}_{\text{inv}}(\mathbf{V}^{(\mathcal{K})}, \epsilon) \right\} \right\}. \quad (4.248)$$

By the definition of T_+ (see (4.148)), the direction⁵ of the vector $\mathbf{R}'^{(\mathcal{K})} - \mathbf{R}^{*(\mathcal{K})}$ approaches that of $T_+^{(\mathcal{K})}$ as (R'_1, R'_2) approaches (R_1^*, R_2^*) along the boundary from the right. By taking this limiting choice and using the fact that $\mathcal{L}$ is defined using a closure operation, we obtain the second set in (4.151). When $R_1^* > 0$, the first set is obtained in an identical fashion by letting (Q'_U, Q'_1, Q'_2) achieve a boundary point approaching (R_1^*, R_2^*) from the left.

4.5.6 Proof of Theorem 4.15 ($\beta = 0, \mathcal{U} = \emptyset$)

In this subsection, we consider the case that $\beta = 0$ and $\mathcal{U} = \emptyset$, and we omit the arguments u to the functions defined in Section 4.4 (e.g. $i(u, x_1, x_2, y)$ is replaced by $i(x_1, x_2, y)$).

For $\nu = 1, 2$, we are given the input distribution $Q_\nu \in \mathcal{P}(\mathcal{X}_\nu)$, and we let $Q_{\nu,n} \in \mathcal{P}_n(\mathcal{X}_\nu)$ be a type with the same support as Q_ν such that $\max_{x_\nu} |Q_\nu(x_\nu) - Q_{\nu,n}(x_\nu)| \leq \frac{1}{n}$. We generate the M_ν codewords of user $\nu = 1, 2$ independently according to

$$P_{\mathbf{X}_\nu}(\mathbf{x}_\nu) = \frac{1}{|T^n(Q_\nu)|} \mathbb{1}\{\mathbf{x}_\nu \in T^n(Q_{\nu,n})\}. \quad (4.249)$$

For clarity of exposition, we assume that Q_1 and Q_2 are themselves types, and hence $Q_{1,n} = Q_1$ and $Q_{2,n} = Q_2$; the analysis for the more general case introduces an additive $O(1)$ term that can be incorporated into $g(n)$ in (4.245).

We define the random variables

$$(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}, \bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2) \sim P_{\mathbf{X}_1}(\mathbf{x}_1)P_{\mathbf{X}_2}(\mathbf{x}_2)W^n(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)P_{\mathbf{X}_1}(\bar{\mathbf{x}}_1)P_{\mathbf{X}_2}(\bar{\mathbf{x}}_2). \quad (4.250)$$

Using standard threshold-based arguments (e.g. see [126, Thm. 4]), we have

$$\begin{aligned} \bar{p}_e \leq 1 - \mathbb{P}\left[i^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \succ \gamma\right] &+ M_1 \mathbb{P}\left[i_1^n(\bar{\mathbf{X}}_1, \mathbf{X}_2, \mathbf{Y}) > \gamma_1\right] \\ &+ M_2 \mathbb{P}\left[i_2^n(\mathbf{X}_1, \bar{\mathbf{X}}_2, \mathbf{Y}) > \gamma_2\right] + M_1 M_2 \mathbb{P}\left[i_{12}^n(\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \mathbf{Y}) > \gamma_{12}\right], \end{aligned} \quad (4.251)$$

⁵If $|\mathcal{K}| = 1$, the “direction” should be interpreted as being the sign.

where $\gamma = [\gamma_1 \ \gamma_2 \ \gamma_{12}]^T$ is arbitrary, and

$$\mathbf{i}^n(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \triangleq \sum_{i=1}^n \mathbf{i}(x_{1,i}, x_{2,i}, y_i) \quad (4.252)$$

$$i_\nu^n(\mathbf{x}_1, \mathbf{x}_2, \mathbf{y}) \triangleq \sum_{i=1}^n i_\nu(x_{1,i}, x_{2,i}, y_i). \quad (4.253)$$

By applying a standard change of measure from constant-composition to i.i.d. [27, 62], we can upper bound the second, third and fourth terms of (4.251) by $M_\nu p_0(n) e^{-\gamma_\nu}$ for $\nu = 1, 2, 12$ respectively, where $p_0(n) \triangleq (n+1)^{|\mathcal{X}_1|+|\mathcal{X}_2|-2}$, and $M_{12} \triangleq M_1 M_2$. We thus obtain

$$\bar{p}_e \leq 1 - \mathbb{P}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \succ \gamma] + p_0(n) \sum_{\nu=1,2,12} M_\nu e^{-\gamma_\nu}. \quad (4.254)$$

Using this bound with

$$\gamma_\nu = \log M_\nu + \left(d + \frac{1}{2}\right) \log n, \quad (4.255)$$

where $d \triangleq |\mathcal{X}_1| + |\mathcal{X}_2| - 2$ is the order of the polynomial $p_0(n)$, the desired result in (4.245) (with $\beta = 0$ and $\mathcal{U} = \emptyset$) will follow using identical steps to [97, Thm. 4] once we prove the following:

1. $\mathbb{E}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})] = n\mathbf{I}$ and $\text{Cov}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})] = n\mathbf{V} + O(1)$, where $\mathbf{I}$ and $\mathbf{V}$ are given in (4.135) and (4.144).
2. In the case that $\mathbf{V} \succ \mathbf{0}$, the probability on the right-hand side of (4.254) can be approximated using a multivariate Berry-Esseen theorem with $O(\frac{1}{\sqrt{n}})$ convergence.
3. In the case that $\mathbf{V}$ is singular, the problem can be reduced to a lower dimension using Chebyshev's inequality, and the Berry-Esseen theorem can again be applied.

We formalize and prove these statements in the remainder of this subsection. The remaining details of the proof of Theorem 4.15 are omitted to avoid repetition with [97].

Calculation of Moments

The first moment of $\mathbf{i}^n$ is easily found by writing

$$\mathbb{E}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})] = \sum_{i=1}^n \mathbb{E}[\mathbf{i}(X_{1,i}, X_{2,i}, Y_i)] = n\mathbf{I}, \quad (4.256)$$

where the last equality follows since, by symmetry, $X_{1,i} \sim Q_1$ and $X_{2,i} \sim Q_2$ for all i . To compute the covariance matrix of $\mathbf{i}^n$, we write

$$\text{Cov}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})] = \text{Cov}\left[\sum_{i=1}^n \mathbf{i}(X_{1,i}, X_{2,i}, Y_i)\right] \quad (4.257)$$

$$= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}\left[\mathbf{i}(X_{1,i}, X_{2,i}, Y_i), \mathbf{i}(X_{1,j}, X_{2,j}, Y_j)\right] \quad (4.258)$$

$$= n \text{Cov}\left[\mathbf{i}(X_1, X_2, Y)\right] + (n^2 - n) \text{Cov}\left[\mathbf{i}(X_1, X_2, Y), \mathbf{i}(X_1^\dagger, X_2^\dagger, Y^\dagger)\right], \quad (4.259)$$

where (X_1, X_2, Y) and $(X_1^\dagger, X_2^\dagger, Y^\dagger)$ correspond to two arbitrary different indices in $\{1, \dots, n\}$ (e.g. $(X_1, X_2, Y) = (X_{1,1}, X_{2,1}, Y_1)$ and $(X_1^\dagger, X_2^\dagger, Y^\dagger) = (X_{1,2}, X_{2,2}, Y_2)$). In (4.259), we have used the fact that, by the symmetry of the codebook construction, the n terms in (4.258) with $i = j$ are equal, and similarly for the $n^2 - n$ terms with $i \neq j$.

To compute the cross-covariance matrix in (4.259), we need the joint distribution of (X_1, X_2, Y) and $(X_1^\dagger, X_2^\dagger, Y^\dagger)$. This distribution can be understood by noting that the uniform distribution on $T^n(Q)$ is obtained by randomly permuting the symbols of an arbitrary sequence $\mathbf{x} \in T^n(Q)$. This, in turn, can be interpreted as successively performing uniform sampling from a collection of symbols without replacement (n times in total), where the initial collection contains $nQ(x)$ occurrences of each symbol $x \in \mathcal{X}$. By considering the first two steps of such a procedure, we readily obtain $\mathbb{P}[X_\nu = x_\nu] = Q_\nu(x_\nu)$, and

$$\mathbb{P}[X_\nu^\dagger = x_\nu^\dagger | X_\nu = x_\nu] = \frac{nQ_\nu(x_\nu^\dagger) - \mathbb{1}\{x_\nu = x_\nu^\dagger\}}{n - 1} \quad (4.260)$$

for $\nu = 1, 2$. Letting $P_\nu^\dagger(x_\nu^\dagger | x_\nu)$ denote the right-hand side of (4.260), the cross-covariance matrix in (4.259) is given by

$$\begin{aligned} & \text{Cov}\left[\mathbf{i}(X_1, X_2, Y), \mathbf{i}(X_1^\dagger, X_2^\dagger, Y^\dagger)\right] \\ &= \mathbb{E}\left[\left(\mathbf{i}(X_1, X_2, Y) - \mathbf{I}\right)\left(\mathbf{i}(X_1^\dagger, X_2^\dagger, Y^\dagger) - \mathbf{I}\right)^T\right] \end{aligned} \quad (4.261)$$

$$\begin{aligned} &= \sum_{x_1, x_2, y} Q_1(x_1) Q_2(x_2) W(y | x_1, x_2) \\ &\quad \times \sum_{x_1^\dagger, x_2^\dagger, y^\dagger} P_1^\dagger(x_1^\dagger | x_1) P_2^\dagger(x_2^\dagger | x_2) W(y^\dagger | x_1^\dagger, x_2^\dagger) \left(\mathbf{i}(x_1, x_2, y) - \mathbf{I}\right) \left(\mathbf{i}(x_1^\dagger, x_2^\dagger, y^\dagger) - \mathbf{I}\right)^T \end{aligned} \quad (4.262)$$

$$\triangleq \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{F}_3 + \mathbf{F}_4, \quad (4.263)$$

where the four terms in (4.263) correspond to the four terms in the expansion of $(nQ_1(x_1^\dagger) - \mathbb{1}\{x_1 = x_1^\dagger\})(nQ_2(x_2^\dagger) - \mathbb{1}\{x_2 = x_2^\dagger\})$ resulting from (4.260). These can be written as

$$\mathbf{F}_1 = \frac{n^2}{(n-1)^2} \mathbb{E}[\mathbf{i}(X_1, X_2, Y) - \mathbf{I}] \mathbb{E}[\mathbf{i}(X_1, X_2, Y) - \mathbf{I}]^T \quad (4.264)$$

$$\mathbf{F}_2 = -\frac{n}{(n-1)^2} \mathbb{E}\left[\left(\mathbf{i}(X_1, X_2, Y) - \mathbf{I}\right)\left(\mathbf{i}(\bar{X}_1, X_2, \bar{Y}) - \mathbf{I}\right)^T\right] \quad (4.265)$$

$$\mathbf{F}_3 = -\frac{n}{(n-1)^2} \mathbb{E}\left[\left(\mathbf{i}(X_1, X_2, Y) - \mathbf{I}\right)\left(\mathbf{i}(X_1, \bar{X}_2, \bar{\bar{Y}}) - \mathbf{I}\right)^T\right] \quad (4.266)$$

$$\mathbf{F}_4 = \frac{1}{(n-1)^2} \mathbb{E}\left[\left(\mathbf{i}(X_1, X_2, Y) - \mathbf{I}\right)\left(\mathbf{i}(X_1, X_2, \tilde{Y}) - \mathbf{I}\right)^T\right], \quad (4.267)$$

where

$$(X_1, X_2, Y, \bar{X}_1, \bar{X}_2, \bar{Y}, \bar{\bar{Y}}, \tilde{Y}) \sim Q_1(x_1)Q_2(x_2)W(y|x_1, x_2) \\ \times Q_1(\bar{x}_1)Q_2(\bar{x}_2)W(\bar{y}|\bar{x}_1, x_2)W(\bar{\bar{y}}|x_1, \bar{x}_2)W(\tilde{y}|x_1, x_2). \quad (4.268)$$

Observe that $\mathbf{F}_1$ is the zero matrix, and $\mathbf{F}_4 = O(\frac{1}{n^2})$. Furthermore, recalling the definition and $\mathbf{i}^{(2)}$ (4.146), the expectation in (4.265) can be written as

$$\mathbb{E}\left[\mathbb{E}\left[\left(\mathbf{i}(X_1, X_2, Y) - \mathbf{I}\right)\middle|X_2\right]\mathbb{E}\left[\left(\mathbf{i}(\bar{X}_1, X_2, \bar{Y}) - \mathbf{I}\right)\middle|X_2\right]^T\right] \quad (4.269)$$

$$= \text{Cov}[\mathbf{i}^{(2)}(X_2)]. \quad (4.270)$$

It follows that

$$\mathbf{F}_2 = \frac{-n}{(n-1)^2} \text{Cov}[\mathbf{i}^{(2)}(X_2)], \quad (4.271)$$

and we similarly have

$$\mathbf{F}_3 = \frac{-n}{(n-1)^2} \text{Cov}[\mathbf{i}^{(1)}(X_1)]. \quad (4.272)$$

Using the identity $\frac{n}{(n-1)^2} = \frac{1}{n} + O(\frac{1}{n^2})$ and combining (4.259), (4.263), (4.271) and (4.272), we obtain

$$\text{Cov}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})] = n\mathbf{V} + O(1), \quad (4.273)$$

where $\mathbf{V}$ is defined as in (4.144) with $\mathcal{U} = \emptyset$.

A Combinatorial Berry-Esseen Theorem

The Berry-Esseen theorem required to bound the probability in (4.254) is a special case of a result by Loh [130, Thm. 2] for a problem known as Latin hypercube sampling. This

result builds on a combinatorial central limit theorem due to Hoeffding [108]; for other related works, see [131, 132] and the references therein.

We define the quantities

$$\Sigma_n \triangleq \frac{1}{n} \text{Cov}[\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})], \quad (4.274)$$

$$\widehat{\mathbf{S}}_n \triangleq \Sigma_n^{-\frac{1}{2}} \left(\frac{1}{\sqrt{n}} (\mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) - n\mathbf{I}) \right) \quad (4.275)$$

$$\mathbf{\Lambda}_n(x_1, x_2) \triangleq \Sigma_n^{-\frac{1}{2}} \left(\mathbf{i}(x_1, x_2, Y(x_1, x_2)) - \mathbf{i}^{(1)}(x_1) - \mathbf{i}^{(2)}(x_2) + \mathbf{I} \right) \quad (4.276)$$

$$\xi_n \triangleq \sum_{x_1, x_2} Q_1(x_1) Q_2(x_2) \mathbb{E} \left[\|\mathbf{\Lambda}_n(x_1, x_2)\|^3 \right], \quad (4.277)$$

where $Y(x_1, x_2) \sim W(\cdot | x_1, x_2)$. From (4.273), we have $\Sigma_n = \mathbf{V} + O(\frac{1}{n})$. It follows that whenever $\mathbf{V} \succ \mathbf{0}$, we have for sufficiently large n that $\Sigma_n \succ \mathbf{0}$, and $\Sigma_n^{-\frac{1}{2}}$ is well-defined.

Theorem 4.16. (Corollary of [130, Thm. 2]) *Fix the pair (Q_1, Q_2) , and consider the quantities $(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$, $\mathbf{I}$, $\mathbf{V}$, $\widehat{\mathbf{S}}_n$ and ξ_n respectively defined in (4.250), (4.135), (4.144), (4.275) and (4.277). If $\mathbf{V} \succ \mathbf{0}$, then we have for sufficiently large n that*

$$\left| \mathbb{P}[\widehat{\mathbf{S}}_n \in \mathcal{A}] - \mathbb{P}[\mathbf{Z} \in \mathcal{A}] \right| \leq \frac{1}{\sqrt{n}} \frac{K}{\xi_n} \quad (4.278)$$

for any convex, Borel-measurable set $\mathcal{A} \subseteq \mathbb{R}^3$, where $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, and K is a universal constant.

Proof. We describe how the theorem can be inferred from [130]. Throughout the proof, we write $A_1 \stackrel{d}{=} A_2$ if the random variables A_1 and A_2 have the same distribution.

We set the dimension of the hypercube in [130] to 2, corresponding to the number of users of the MAC. We let $\pi_1(\cdot)$ and $\pi_2(\cdot)$ be independent random permutations of $\{1, \dots, n\}$, uniformly distributed over the $n!$ possible permutations. For $\nu = 1, 2$ and $j_1, j_2 = 1, \dots, n$, we define the random variables $U_\nu(j_1, j_2)$ uniformly distributed on $(0, 1)$ independently of each other and of $\pi_1(\cdot)$ and $\pi_2(\cdot)$, and we set

$$B_\nu(j_1, j_2) \triangleq \frac{j_\nu - U_\nu(j_1, j_2)}{n}, \quad \nu = 1, 2. \quad (4.279)$$

The following sample mean is considered in [130]:

$$\hat{\mu}_n \triangleq \frac{1}{n} \sum_{j=1}^n \mathbf{f} \left(B_1(\pi_1(j), \pi_2(j)), B_2(\pi_1(j), \pi_2(j)) \right), \quad (4.280)$$

where $\mathbf{f}(\cdot)$ is a vector-valued function with two arguments. We wish to choose $\mathbf{f}$ such that, for each (j_1, j_2) , we obtain

$$\mathbf{f}(B_1(j_1, j_2), B_2(j_1, j_2)) \stackrel{d}{=} \mathbf{i}(x_{1,j_1}, x_{2,j_2}, Y(j_1, j_2)), \quad (4.281)$$

where $\mathbf{x}_1 = (x_{1,1}, \dots, x_{1,n})$ and $\mathbf{x}_2 = (x_{2,1}, \dots, x_{2,n})$ are arbitrary sequences of type Q_1 and Q_2 respectively, and $Y(j_1, j_2) \sim W(\cdot | x_{1,j_1}, x_{2,j_2})$ with independence between different (j_1, j_2) pairs. Since drawing a codeword uniformly over a type class is equivalent to randomly permuting any codeword of the given type, such an identification will yield $\hat{\mu}_n$ in (4.280) with the same distribution as $\frac{1}{n} \mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$ in (4.254).

Since $U_\nu(j_1, j_2) \in (0, 1)$ almost surely, any realization b_ν of B_ν uniquely determines both the index j_ν and the variable U_ν in the numerator of (4.279). Thus, overloading the symbol $\mathbf{f}$, we can write $\mathbf{f}(z_1, z_2) = \mathbf{f}(j_1, j_2, u_1, u_2)$ without ambiguity. We choose

$$\mathbf{f}(j_1, j_2, u_1, u_2) = \mathbf{i}(x_{1,j_1}, x_{2,j_2}, F_{Y(j_1, j_2)}^{-1}(u_1 \oplus u_2)), \quad (4.282)$$

where $F_{Y(j_1, j_2)}^{-1}(u) = \inf \{y : F_{Y(j_1, j_2)}(y) \geq u\}$ is the inverse CDF of $Y(j_1, j_2)$, and $\oplus$ denotes real addition modulo one. Clearly $U_1 \oplus U_2$ is uniform on $(0, 1)$, and since $F_X^{-1}(U) \stackrel{d}{=} X$ for any random variable X (discrete or continuous) with CDF F_X , it follows that $F_{Y(j_1, j_2)}^{-1}(U_1 \oplus U_2) \stackrel{d}{=} Y(j_1, j_2)$, as desired. While the choice of the function in (4.282) to recover the desired distribution may not be unique, it yields a simpler analysis when either u_1 or u_2 is fixed. In particular, we have $F_{Y(j_1, j_2)}^{-1}(u_1 \oplus U_2(j_1, j_2)) \stackrel{d}{=} Y(j_1, j_2)$ and $F_{Y(j_1, j_2)}^{-1}(U_1(j_1, j_2) \oplus u_2) \stackrel{d}{=} Y(j_1, j_2)$ for all (u_1, u_2) , thus simplifying the evaluation of several quantities in [130].

Using the choice of $\mathbf{f}$ in (4.282), one can obtain the following equivalences between quantities defined in [130] (left-hand sides) and the present paper (right-hand sides):

$$\hat{\mu} = \frac{1}{n} \mathbf{i}^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \quad (4.283)$$

$$\mu = \mathbf{I} \quad (4.284)$$

$$\int_{[0,1]^2} f_{\text{rem}}(x) f_{\text{rem}}(x)^T dx = \mathbf{V} \quad (4.285)$$

$$\Sigma_{\text{lhs}} = \frac{1}{n} \Sigma_n \quad (4.286)$$

$$\mu(j_1, j_2) = \sum_y W(y | x_{1,j_1}, x_{2,j_2}) \mathbf{i}(x_{1,j_1}, x_{2,j_2}, y) \quad (4.287)$$

$$\mu_{-\nu}(j_\nu) = \mathbf{i}^{(\nu)}(x_{\nu, j_\nu}), \quad \nu = 1, 2 \quad (4.288)$$

$$W = \hat{S}_n \quad (4.289)$$

$$Y(j_1, j_2) = \frac{1}{\sqrt{n}} \Lambda_n(x_{1,j_1}, x_{2,j_2}) \quad (4.290)$$

$$C_{d,p} = K \quad (4.291)$$

$$\beta_3 = \frac{\xi_n}{\sqrt{n}}. \quad (4.292)$$

Using these equivalences, we see that Theorem 4.16 coincides with [130, Thm. 2]. $\square$

In the setting of finite alphabets under consideration, one can show that $\xi_n = \Theta(1)$ using the fact that the relevant third moments are finite (e.g. they can be bounded in terms of the alphabet sizes [97, Appendix D]). Thus, the convergence in (4.278) is $O(\frac{1}{\sqrt{n}})$.

Singular Dispersion Matrices

The dispersion matrix V may not have full rank, in which case Theorem 4.16 does not directly apply. We deal with this case by reducing the problem to a lower dimension, similarly to [97, Sec. VIII-A]. The argument here is more involved, since nV is not necessarily the exact covariance matrix of $i^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$, due to the additional $O(1)$ term in (4.273).

Suppose that V has rank $r < 3$, and consider the matrix Σ_n in (4.274), which equals $V + O(\frac{1}{n})$. Using an eigenvalue decomposition, we obtain

$$\Sigma_n = \Theta_n \Omega_n \Theta_n^T, \quad (4.293)$$

where Θ_n is a unitary matrix, and Ω_n is a diagonal matrix whose first r diagonals are $\Theta(1)$, and whose last $3 - r$ diagonals are $O(\frac{1}{n})$. Noting that Σ_n is the covariance matrix of $\mathbf{A}_n \triangleq \frac{1}{\sqrt{n}}(i^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) - n\mathbf{I})$, we see that Ω_n is the covariance matrix of $\tilde{\mathbf{A}}_n \triangleq \Theta_n^T \mathbf{A}_n$.

Supposing for the time being that $r \geq 1$, we conclude from the structure of Ω_n that

$$\tilde{\mathbf{A}}_n = \begin{bmatrix} \tilde{\mathbf{A}}_n^{(1)} \\ \tilde{\mathbf{A}}_n^{(2)} \end{bmatrix}, \quad (4.294)$$

where $\tilde{\mathbf{A}}_n^{(1)}$ and $\tilde{\mathbf{A}}_n^{(2)}$ have dimension r and $3 - r$ respectively, and the covariance matrix of $\tilde{\mathbf{A}}_n^{(2)}$ has $O(\frac{1}{n})$ entries. Since Θ_n is unitary (i.e. $\Theta_n \Theta_n^T = \mathbb{I}$), we have $\mathbf{A}_n = \Theta_n \tilde{\mathbf{A}}_n$, and hence

$$\mathbf{A}_n = \Theta_n \begin{bmatrix} \tilde{\mathbf{A}}_n^{(1)} \\ \mathbf{0} \end{bmatrix} + \Theta_n \begin{bmatrix} \mathbf{0} \\ \tilde{\mathbf{A}}_n^{(2)} \end{bmatrix} \quad (4.295)$$

$$\triangleq \Theta_n' \tilde{\mathbf{A}}_n^{(1)} + \Delta_n, \quad (4.296)$$

where Θ'_n is obtained from Θ_n by keeping only the first r columns, and Δ_n denotes the second term in (4.295). Since A_n has mean zero by construction, we conclude that the same is true of $\tilde{A}_n$, and hence of Δ_n . Furthermore, since $\text{Cov}[\tilde{A}_n^{(2)}]$ has $O(\frac{1}{n})$ entries, the same is true of $\text{Cov}[\Delta_n]$, and hence Chebyshev's inequality implies for any $\delta_n > 0$ that

$$\mathbb{P}[\|\Delta_n\|_\infty \geq \delta_n] = O\left(\frac{1}{n\delta_n^2}\right). \quad (4.297)$$

We are now in a position to bound the probability appearing in (4.254). The following holds for any $\delta_n > 0$:

$$\mathbb{P}[i^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \succ \gamma] = \mathbb{P}\left[A_n \succ \frac{1}{\sqrt{n}}(\gamma - n\mathbf{I})\right] \quad (4.298)$$

$$= \mathbb{P}\left[\Theta'_n \tilde{A}_n^{(1)} + \Delta_n \succ \frac{1}{\sqrt{n}}(\gamma - n\mathbf{I})\right] \quad (4.299)$$

$$\geq \mathbb{P}\left[\Theta'_n \tilde{A}_n^{(1)} \succ \frac{1}{\sqrt{n}}(\gamma - n\mathbf{I}) + \delta_n \mathbf{1}\right] - \mathbb{P}[\|\Delta_n\|_\infty \geq \delta_n] \quad (4.300)$$

$$= \mathbb{P}\left[\Theta'_n \tilde{A}_n^{(1)} \succ \frac{1}{\sqrt{n}}(\gamma - n\mathbf{I}) + \delta_n \mathbf{1}\right] + O\left(\frac{1}{n\delta_n^2}\right), \quad (4.301)$$

where the final three steps respectively make use of (4.296), [97, Lemma 9], and (4.297). Since the entries of $\tilde{A}_n^{(1)}$ are shifted and weighted sums of the entries of $i^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y})$, and since the corresponding covariance matrix is positive definite by construction, we can analyze (4.301) in the same way as the case $\mathbf{V} \succ \mathbf{0}$ (other than the terms $\delta_n \mathbf{1}$ and $O(\frac{1}{n\delta_n^2})$, which are handled in the following paragraph). The fact that the resulting second-order term can be written as $-\sqrt{n}\mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon)$ follows in the same way as the i.i.d. case [97].

The remainder term in (4.301) contributes an additive $O(\frac{1}{\sqrt{n}\delta_n^2})$ term to the expansion in (4.245), whereas the addition of $\delta_n \mathbf{1}$ in the first probability in (4.301) contributes an additive $O(\delta_n \sqrt{n})$ term to the expansion. The overall contribution $O(\frac{1}{\sqrt{n}\delta_n^2} + \delta_n \sqrt{n})$ is minimized by $\delta_n = \Theta(n^{-\frac{1}{3}})$, yielding $g(n) = O(n^{\frac{1}{6}})$.

The case $r = 0$ (i.e. $\mathbf{V} = \mathbf{0}$) is handled similarly; by following the steps of (4.298)–(4.301), we readily obtain the following analog of (4.301):

$$\mathbb{P}[i^n(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \succ \gamma] \geq \mathbb{1}\left\{\mathbf{0} \succ \frac{1}{\sqrt{n}}(\gamma - n\mathbf{I}) + \delta_n \mathbf{1}\right\} + O\left(\frac{1}{n\delta_n^2}\right) \quad (4.302)$$

Setting $\gamma = n\mathbf{I} - c\sqrt{n}\mathbf{1}$ for some $c > 0$, we see that the indicator function equals one for sufficiently large n . Since c can be arbitrarily small, we obtain (4.245) (with $\beta = 0$ and $\mathcal{U} = \emptyset$) by noting that $\mathbf{Q}_{\text{inv}}(\mathbf{0}, \epsilon)$ equals the set of all vectors with non-negative components.

4.5.7 Proof of Theorem 4.15 (General Case)

Here we provide the changes required in the previous subsection to prove Theorem 4.15 in full generality. We recall the definitions of $P'_{UX_1X_2Y}$, $\mathbf{i}'$, etc. at the start of Section 4.5.5.

Coded Time-Sharing with $\beta = 0$

In the case that $\beta = 0$ but $\mathcal{U} \neq \emptyset$, we modify the constant-composition ensemble (cf. (4.249)) as follows. Given the triplet (Q_U, Q_1, Q_2) , we define $(Q_{U,n}, Q_{1,n}, Q_{2,n})$ to be the corresponding (conditional) types. We fix an arbitrary time sharing sequence $\mathbf{u}$ with type $Q_{U,n}$, and generate the M_ν codewords of user $\nu = 1, 2$ independently according to

$$P_{\mathbf{X}_\nu|U}(\mathbf{x}_\nu|\mathbf{u}) = \frac{1}{|T_{\mathbf{u}}^n(Q_{\nu,n})|} \mathbb{1}\{\mathbf{x}_\nu \in T_{\mathbf{u}}^n(Q_{\nu,n})\}. \quad (4.303)$$

Similarly to (4.250), we define the random variables

$$(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}, \bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2) \sim P_{\mathbf{X}_1|U}(\mathbf{x}_1|\mathbf{u})P_{\mathbf{X}_2|U}(\mathbf{x}_2|\mathbf{u})W^n(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)P_{\mathbf{X}_1|U}(\bar{\mathbf{x}}_1|\mathbf{u})P_{\mathbf{X}_2|U}(\bar{\mathbf{x}}_2|\mathbf{u}). \quad (4.304)$$

The procedure described in Section 4.5.6 for generating a codeword uniformly over the type class is modified as follows. Let $\mathbf{x}$ be an arbitrary element of the conditional type class $T_{\mathbf{u}}(\cdot)$. Instead of randomly permuting the entire sequence $\mathbf{x}$, a random permutation of the subsequence $\mathbf{x}^{(u)}$ corresponding to the indices where $\mathbf{u}$ equals u is applied independently for each value of $u \in \mathcal{U}$. Due to this independence, we can handle the summation

$$\mathbf{i}^n(\mathbf{u}, \mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \triangleq \sum_{i=1}^n \mathbf{i}(u_i, X_{1,i}, X_{2,i}, Y_i) = \sum_u \sum_{i=1}^{nQ_U(u)} \mathbf{i}(u, X_{1,i}^{(u)}, X_{2,i}^{(u)}, Y_i^{(u)}) \quad (4.305)$$

by considering each value of $u \in \mathcal{U}$ separately. For the values of u corresponding to singular dispersion matrices, we can perform a reduction to a lower dimension as discussed following (4.293). From Theorem 4.16, we conclude that each inner summation in (4.305) is asymptotically normal with $O(\frac{1}{\sqrt{nQ_U(u)}}) = O(\frac{1}{\sqrt{n}})$ convergence. It follows that the overall sum is also asymptotically normal with $O(\frac{1}{\sqrt{n}})$ convergence. More concretely, letting $\hat{\mathbf{S}}_{1,n}$ and $\hat{\mathbf{S}}_{2,n}$ be asymptotically normal in the sense of (4.278), and letting $\mathbf{Z}_1$ and $\mathbf{Z}_2$ be the associated

Gaussian random variables, we have

$$\mathbb{P}[\widehat{\mathbf{S}}_{1,n} + \widehat{\mathbf{S}}_{2,n} \in \mathcal{A}] = \mathbb{E}[\mathbb{P}[\widehat{\mathbf{S}}_{1,n} + \widehat{\mathbf{S}}_{2,n} \in \mathcal{A} \mid \widehat{\mathbf{S}}_{2,n}]] \quad (4.306)$$

$$= \mathbb{E}[\mathbb{P}[\mathbf{Z}_1 + \widehat{\mathbf{S}}_{2,n} \in \mathcal{A} \mid \widehat{\mathbf{S}}_{2,n}]] + O\left(\frac{1}{\sqrt{n}}\right) \quad (4.307)$$

$$= \mathbb{P}[\mathbf{Z}_1 + \widehat{\mathbf{S}}_{2,n} \in \mathcal{A}] + O\left(\frac{1}{\sqrt{n}}\right) \quad (4.308)$$

$$= \mathbb{P}[\mathbf{Z}_1 + \mathbf{Z}_2 \in \mathcal{A}] + O\left(\frac{1}{\sqrt{n}}\right), \quad (4.309)$$

where (4.309) follows using similar steps to (4.306)–(4.308). Using this observation and repeating the analysis of Section 4.5.6 and [97], we obtain the more general result of Theorem 4.15 with $\beta = 0$.

Coded Time-Sharing with $\beta > 0$

In the case that $\beta > 0$, we apply a variant of coded time sharing depending on an extended alphabet $\widetilde{\mathcal{U}} \triangleq \mathcal{U} \times \{1, 2\}$ and the triplets (Q_U, Q_1, Q_2) and (Q'_U, Q'_1, Q'_2) . Specifically, we define $(\widetilde{Q}_{U,n}, \widetilde{Q}_{1,n}, \widetilde{Q}_{2,n})$ with $\widetilde{Q}_{U,n} \in \mathcal{P}(\widetilde{\mathcal{U}})$, $\widetilde{Q}_{1,n} \in \mathcal{P}(\mathcal{X}_1 \mid \widetilde{\mathcal{U}})$ and $\widetilde{Q}_{2,n} \in \mathcal{P}(\mathcal{X}_2 \mid \widetilde{\mathcal{U}})$ as follows:

$$\widetilde{Q}_{U,n}(u, 1) = \left(1 - \frac{\beta}{\sqrt{n}}\right) Q_U(u) \quad (4.310)$$

$$\widetilde{Q}_{U,n}(u, 2) = \frac{\beta}{\sqrt{n}} Q'_U(u) \quad (4.311)$$

$$\widetilde{Q}_{\nu,n}(x_\nu \mid u, 1) = Q_\nu(x_\nu \mid u) \quad (4.312)$$

$$\widetilde{Q}_{\nu,n}(x_\nu \mid u, 2) = Q'_\nu(x_\nu \mid u) \quad (4.313)$$

for $\nu = 1, 2$. We consider the constant-composition ensemble in (4.303) with $(\widetilde{Q}_{U,n}, \widetilde{Q}_{1,n}, \widetilde{Q}_{2,n})$ playing the role of (Q_U, Q_1, Q_2) , and we set $\widetilde{P}_{UX_1X_2Y,n} \triangleq \widetilde{Q}_{U,n} \times \widetilde{Q}_{1,n} \times \widetilde{Q}_{2,n} \times W$. We follow the same arguments as the case that $\beta = 0$ and $\mathcal{U} \neq \emptyset$, but with care taken to handle the fact that some of the time sharing values $\widetilde{u} \in \widetilde{\mathcal{U}}$ correspond to subsequences of $\mathbf{u}$ having length $\Theta(\sqrt{n})$, rather than $\Theta(n)$. Since we consider the limit of large n , we may assume that $1 - \frac{\beta}{\sqrt{n}} > 0$.

The role of the information density vector $\mathbf{i}$ is now played by the quantity

$$\tilde{\mathbf{i}}(\tilde{u}, x_1, x_2, y) \triangleq \begin{cases} \mathbf{i}(u, x_1, x_2, y) & \tilde{u} = (u, 1) \\ \mathbf{i}'(u, x_1, x_2, y) & \tilde{u} = (u, 2). \end{cases} \quad (4.314)$$

The corresponding mean vector and dispersion matrix are given by

$$\tilde{\mathbf{I}}_n \triangleq \left(1 - \frac{\beta}{\sqrt{n}}\right) \mathbf{I} + \frac{\beta}{\sqrt{n}} \mathbf{I}' \quad (4.315)$$

$$\tilde{\mathbf{V}}_n \triangleq \left(1 - \frac{\beta}{\sqrt{n}}\right) \mathbf{V} + \frac{\beta}{\sqrt{n}} \mathbf{V}' \quad (4.316)$$

where $\mathbf{I}'$ and $\mathbf{V}'$ are defined in (4.243)–(4.244). Further generalizing the above analysis for $\mathcal{U} \neq \emptyset$ to the case of varying input distributions, we conclude that rate pairs (R_1, R_2) satisfying

$$n\mathbf{R} \in n\tilde{\mathbf{I}}_n - \sqrt{n}\mathbf{Q}_{\text{inv}}(\tilde{\mathbf{V}}_n, \epsilon) + O(n^{\frac{1}{4}}) \quad (4.317)$$

are (n, ϵ) -achievable. Here the remainder term is $O(n^{\frac{1}{4}})$ since, for the subsequences of $\mathbf{u}$ with length $\Theta(\sqrt{n})$, the rate of convergence in the Berry-Esseen theorem is $O(n^{-\frac{1}{4}})$, rather than $O(\frac{1}{\sqrt{n}})$. The proof is concluded by substituting (4.315)–(4.316) into (4.317) and using a Taylor expansion of $\mathbf{Q}_{\text{inv}}$ (in the same way as [97, 121]) to replace $\mathbf{Q}_{\text{inv}}(\tilde{\mathbf{V}}_n, \epsilon)$ by $\mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon)$.

4.5.8 Proof of Theorem 4.14

The key result used in the proof is the following lemma, which states that there exists a sequence of discrete input distributions Q_{m1} and Q_{m2} of cardinality m such that the corresponding vector-matrix pair $(\mathbf{I}_m, \mathbf{V}_m)$ converges to $(\mathbf{I}, \mathbf{V})$ (see (4.158)–(4.159)), with the convergence $\mathbf{I}_m \rightarrow \mathbf{I}$ being exponentially fast in m . This generalizes a result by Wu and Verdú for the scalar case [133], and is proved similarly.

Lemma 4.5. *There exist sequences of discrete input distributions Q_{m1} and Q_{m2} of cardinality m with a corresponding matrix-vector pair $(\mathbf{I}_m, \mathbf{V}_m)$ defined according to (4.135) and (4.144) such that (i) $\|\mathbf{I}_m - \mathbf{I}\|_\infty \leq e^{-\gamma m}$ for some $\gamma > 0$ and sufficiently large m , (ii) $\|\mathbf{V}_m - \mathbf{V}\|_\infty \rightarrow 0$, and (iii) the third absolute moment of each entry of $\mathbf{i}(X_{m1}, X_{m2}, Y)$ under $Q_{m1} \times Q_{m2} \times Y$ is uniformly bounded in m .*

Proof. The proof closely follows that of [133, Thm. 8], so we only explain the differences. We choose Q_{m1} and Q_{m2} according the Gauss quadrature rule Q_g [133, Sec. II], which satisfies the property of having the same moments as those of a standard Gaussian random variable up to order $2m - 1$ [133, Thm. 2]. Since Q_g converges weakly to $N(0, 1)$ [133], we immediately obtain parts (ii) and (iii) of Lemma 4.5, so it remains to prove part (i).

Define $(X_{m1}, X_{m2}, Y_m) \sim Q_{m1} \times Q_{m2} \times W$ and $(X_1, X_2, Y) \sim Q_1 \times Q_2 \times W$, where $Q_1, Q_2 \sim$

$N(0, 1)$. Using the identity [3, Eq. (15.142)]

$$I(X_{m1}; Y_m | X_{m2}) = H(\sqrt{P_1}X_{m1} + Z) - H(Z), \quad (4.318)$$

we see that the convergence of the first entry of I_m to that of I is precisely that studied in [133], and similarly for the second entry. It remains to study the third entry, i.e. to show that $I(X_{m1}, X_{m2}; Y_m) \rightarrow I(X_1, X_2; Y)$ exponentially fast. Analogously to [133, Eq. (5)], we have

$$I(X_1, X_2; Y) - I(X_{m1}, X_{m2}; Y_m) = D(\sqrt{P_1}X_1 + \sqrt{P_2}X_2 + Z \| \sqrt{P_1}X_{m1} + \sqrt{P_2}X_{m2} + Z) \quad (4.319)$$

$$\triangleq D_m. \quad (4.320)$$

Using nearly identical arguments to [133, Sec. V] with an “optimal” output distribution of $N(0, 1 + P_1 + P_2)$, we obtain analogously to [133, Eq. (54)] that

$$D_m \leq \sum_{k \geq 1} \frac{1}{k!} \left(\frac{P_1 + P_2}{1 + P_1 + P_2} \right)^k \left| \mathbb{E} \left[H_k \left(\frac{\sqrt{P_1}X_{m1} + \sqrt{P_2}X_{m2}}{\sqrt{P_1 + P_2}} \right) \right] \right|, \quad (4.321)$$

where H_k is the Hermite polynomial of degree k (see [133, Eq. (15)]). As shown in [133], we obtain the desired exponential convergence rate of the mutual information provided that the expectation appearing in (4.321) is zero for odd values of k , and also for $k \leq 2m - 1$. For odd values of k , we use the same symmetry argument as that of [133]; since the distributions of X_{m1} and X_{m2} are both symmetric, so is that of their weighted sum. To handle the remaining values $k \leq 2m - 1$, we write

$$H_k(a + b) = \sum_{i=0}^k \sum_{j=0}^k c_{ij} a^i b^j \quad (4.322)$$

for some constants c_{ij} , which follows since H_k has degree k . By the independence of X_{m1} and X_{m2} , the expectation $\mathbb{E} \left[H_k \left(\frac{\sqrt{P_1}X_{m1} + \sqrt{P_2}X_{m2}}{\sqrt{P_1 + P_2}} \right) \right]$ depends only on the first k moments of X_{m1} and X_{m2} . Since the i -th moment of $X_{m\nu}$ coincides with the corresponding moment of $X_\nu \sim N(0, P_\nu)$ for $i = 1, \dots, 2m - 1$ [133, Thm. 2], we have for $k \leq 2m - 1$ that

$$\mathbb{E} \left[H_k \left(\frac{\sqrt{P_1}X_{m1} + \sqrt{P_2}X_{m2}}{\sqrt{P_1 + P_2}} \right) \right] = \mathbb{E} \left[H_k \left(\frac{\sqrt{P_1}X_1 + \sqrt{P_2}X_2}{\sqrt{P_1 + P_2}} \right) \right] \quad (4.323)$$

$$= 0, \quad (4.324)$$

where (4.324) follows since for any k , we have $H_k(X) = 0$ under $X \sim N(0, 1)$ [133]. $\square$

We proceed by proving that, analogously to Theorem 4.15, there exists $g(n) = o(\sqrt{n})$ such that all rate pairs (R_1, R_2) satisfying

$$\mathbf{R} \in n\mathbf{I} - \sqrt{n} \mathbf{Q}_{\text{inv}}(\mathbf{V}, \epsilon) + g(n)\mathbf{1} \quad (4.325)$$

are (n, ϵ) -achievable. This is done by following the analysis of Section 4.5.6: We consider random coding with the constant-composition codeword distribution in (4.249), using (Q_{m1}, Q_{m2}) as the input distribution pair. As was done in [62, Thm. 3], [125], we set $m = n^{\frac{1}{4}}$. By part (i) of Lemma 4.5, we have $\|n\mathbf{I}_m - n\mathbf{I}\|_{\infty} \leq ne^{-\gamma m^{\frac{1}{4}}}$, which behaves as $o(\sqrt{n})$. Similarly, parts (ii) and (iii) of Lemma 4.5 show that $\mathbf{V}_m \rightarrow \mathbf{V}$ and the relevant third moments associated with i are bounded. The analysis of Section 4.5.6 reveals the remainder term $g(n)$ in (4.325) depends on the alphabet sizes through $(|\mathcal{X}_1| + |\mathcal{X}_2| + 2) \log n$ (see the choice of d following (4.255)), which is again $o(\sqrt{n})$ due to the fact that $|\mathcal{X}_1| = |\mathcal{X}_2| = n^{\frac{1}{4}}$.

Finally, using (4.325), we obtain (4.160) using identical steps to Section 4.5.6 with $\beta = 0$ (see also [67]).

Chapter 5

Superposition Coding Techniques

5.1 Introduction

In this chapter, we return to the single-user mismatched decoding problem described in Section 1.1.1. Building on the techniques of Section 4.2, we study achievable rates, as well as error exponents and non-asymptotic bounds, for two types of superposition coding (e.g. see [134, 135]). In particular, we introduce a refined version of superposition coding which, to our knowledge, has not appeared previously.

In a work that developed independently of ours, Somekh-Baruch [42] gave error exponents and rate regions for the cognitive MAC (i.e. the MAC where one user knows both messages and the other only knows its own) using two multi-user coding schemes: superposition coding and random binning. When applied to single-user mismatched channels, these yield achievable rates that can improve on those of Lapidot [35] (see Theorem 4.10) when certain auxiliary variables are fixed. There is some overlap in our work and the superposition coding results given in [42], but throughout the chapter we focus primarily on the results for which there is no overlap.

The main contributions of this chapter are as follows:

- In Section 5.2, we present achievable rates and error exponents for the standard version of superposition coding, and show that the rate is at least as that of Lapidot's expurgated MAC ensemble [35, Thm. 4] (see Section 4.2.6) after the optimization of the parameters.
- In Section 5.3, we present a refined version of superposition coding, and show that it yields rates at least as high as the standard version for any input distribution. Using a numerical example, it is shown that the improvement can be strict.

- The rates for each ensemble are expressed in both primal and dual forms, and the latter are shown to remain valid in the case of infinite and continuous alphabets.

Our analysis will make use of the upper and lower bounds on multiply indexed unions in Appendix D, as well as the results on convex optimization problems in Appendix E.

The results of this chapter can be summarized by the following list of random-coding constructions, in decreasing order of rate (see Section 5.3.3 for numerical examples):

1. Refined superposition coding (Theorems 5.6 and 5.8),
2. Standard superposition coding (Theorems 5.3 and 5.5),
3. Expurgated parallel coding (Theorems 4.11 and 4.11),
4. Constant-composition or cost-constrained coding with independent codewords (LM rate, cf. (1.9)),
5. i.i.d. coding with independent codewords (GMI, cf. (1.11)).

Parts of this chapter were presented in [109, 136].

5.2 Standard Superposition Coding

In this section, we study the standard version of superposition coding (SC) [134, 135, 137]. The parameters of the ensemble are an auxiliary alphabet $\mathcal{U}$, an auxiliary codeword distribution P_U , and a conditional codeword distribution $P_{X|U}$. An auxiliary codebook $\{U^{(i)}\}_{i=1}^{M_0}$ with M_0 codewords is generated, with each auxiliary codeword independently distributed according to P_U . For each $i = 1, \dots, M_0$, a codebook $\{X^{(i,j)}\}_{j=1}^{M_1}$ with M_1 codewords is generated, with each codeword conditionally independently distributed according to $P_{X|U}$. The message m at the input to the encoder is indexed as (m_0, m_1) , and for any such pair, the corresponding transmitted codeword is $X^{(m_0, m_1)}$. Thus, the overall number of messages is $M = M_1 M_2$, yielding a rate of $R = R_1 + R_2$. More compactly, we have

$$\left\{ \left(U^{(i)}, \{X^{(i,j)}\}_{j=1}^{M_1} \right) \right\}_{i=1}^{M_0} \sim \prod_{i=1}^{M_0} \left(P_U(u^{(i)}) \prod_{j=1}^{M_1} P_{X|U}(x^{(i,j)} | u^{(i)}) \right). \quad (5.1)$$

We assume without loss of generality that message $(1, 1)$ is transmitted, and we write U and X in place of $U^{(1)}$ and $X^{(1,1)}$ respectively. We write $\tilde{X}$ to denote an arbitrary codeword $X^{(1,j)}$ with $j \neq 1$. Moreover, we let $\bar{U}$ denote an arbitrary auxiliary codeword $U^{(i)}$ with $i \neq 1$, let

$\bar{\mathbf{X}}^{(j)}$ ($j = 1, \dots, M_1$) denote the corresponding codeword $\mathbf{X}^{(i,j)}$, and let $\bar{\mathbf{X}}$ denote $\bar{\mathbf{X}}^{(j)}$ for an arbitrary choice of j . Thus, defining $\mathbf{Y}$ to be the channel output, we have

$$(U, \mathbf{X}, \mathbf{Y}, \tilde{\mathbf{X}}, \bar{U}, \bar{\mathbf{X}}) \sim P_U(u)P_{\mathbf{X}|U}(\mathbf{x}|u)W^n(\mathbf{y}|\mathbf{x})P_{\mathbf{X}|U}(\tilde{\mathbf{x}}|u)P_U(\bar{u})P_{\mathbf{X}|U}(\bar{\mathbf{x}}|\bar{u}). \quad (5.2)$$

The decoder estimates $\hat{m} = (\hat{m}_0, \hat{m}_1)$ according to the decoding rule in (1.2). We define the following error events:

$$\begin{aligned} (\text{Type 0}) \quad & q^n(\mathbf{X}^{(i,j)}, \mathbf{Y}) \geq q^n(\mathbf{X}, \mathbf{Y}) \text{ for some } i \neq 1, j; \\ (\text{Type 1}) \quad & q^n(\mathbf{X}^{(1,j)}, \mathbf{Y}) \geq q^n(\mathbf{X}, \mathbf{Y}) \text{ for some } j \neq 1. \end{aligned}$$

The probabilities of these events are denoted by $\bar{p}_{e,0}(n, M_0, M_1)$ and $\bar{p}_{e,1}(n, M_1)$ respectively. Analogously to (4.10), the overall random-coding error probability $\bar{p}_e(n, M_0, M_1)$ satisfies

$$\frac{1}{2} \max\{\bar{p}_{e,0}, \bar{p}_{e,1}\} \leq \bar{p}_e \leq \bar{p}_{e,0} + \bar{p}_{e,1}. \quad (5.3)$$

5.2.1 Non-Asymptotic Bounds

In Section 4.2, we used Lemma D.3 to obtain ensemble-tight error exponents for the mismatched MAC. Since the proof used the method of types, the corresponding subexponential prefactors differ by a polynomial factor depending on the alphabet sizes. For superposition coding, we can in fact obtain non-asymptotic upper and lower bounds that coincide to within a constant factor.

Theorem 5.1. *For any mismatched memoryless channel (possibly having infinite or continuous alphabets) and codeword distributions P_U and $P_{\mathbf{X}|U}$, we have*

$$\frac{1}{4} \text{rcu}_0(n, M_0, M_1) \leq \bar{p}_{e,0}(n, M_0, M_1) \leq \text{rcu}_0(n, M_0, M_1), \quad (5.4)$$

$$\frac{1}{2} \text{rcu}_1(n, M_1) \leq \bar{p}_{e,1}(n, M_1) \leq \text{rcu}_1(n, M_1), \quad (5.5)$$

where

$$\text{rcu}_0(n, M_0, M_1) \triangleq \mathbb{E} \left[\min \left\{ 1, (M_0 - 1) \mathbb{E} \left[\min \left\{ 1, M_1 \mathbb{P} \left[\frac{q^n(\bar{\mathbf{X}}, \mathbf{Y})}{q^n(\mathbf{X}, \mathbf{Y})} \geq 1 \mid \bar{U} \right] \right\} \mid U, \mathbf{X}, \mathbf{Y} \right] \right\} \right] \quad (5.6)$$

$$\text{rcu}_1(n, M_1) \triangleq \mathbb{E} \left[\min \left\{ 1, (M_1 - 1) \mathbb{P} \left[\frac{q^n(\tilde{\mathbf{X}}, \mathbf{Y})}{q^n(\mathbf{X}, \mathbf{Y})} \geq 1 \mid U, \mathbf{X}, \mathbf{Y} \right] \right\} \right]. \quad (5.7)$$

Proof. Equation (5.5) follows using an identical argument to that of Theorem 2.1, so we focus on (5.4). We have

$$\bar{p}_{e,0} = \mathbb{P} \left[\bigcup_{i \neq 1, j \neq 1} \left\{ \frac{q^n(\mathbf{X}^{(i,j)}, \mathbf{Y})}{q^n(\mathbf{X}, \mathbf{Y})} \geq 1 \right\} \right]. \quad (5.8)$$

Writing the probability as an expectation given $(U, \mathbf{X}, \mathbf{Y})$ and applying the truncated union bound to the union over i , we obtain

$$\bar{p}_{e,0} \leq \mathbb{E} \left[\min \left\{ 1, (M_0 - 1) \mathbb{P} \left[\bigcup_{j \neq 1} \left\{ \frac{q^n(\bar{\mathbf{X}}^{(j)}, \mathbf{Y})}{q^n(\mathbf{X}, \mathbf{Y})} \geq 1 \right\} \mid U, \mathbf{X}, \mathbf{Y} \right] \right\} \right]. \quad (5.9)$$

Applying the same argument to the union over j , we obtain the upper bound in (5.4). The lower bound follows since each truncated union bound was applied to independent events, thus guaranteeing tightness to within a factor of $\frac{1}{2}$ [73, Lemma A.2]. $\square$

5.2.2 Exponents and Rates for DMCs

In this subsection, we assume that $|\mathcal{U}|$, $|\mathcal{X}|$ and $|\mathcal{Y}|$ are finite. We consider the constant-composition superposition coding ensemble, which depends on a joint input distribution $Q_{UX} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$. We let $Q_{UX,n}$ be a type with the same support as Q_{UX} such that $|Q_{UX,n}(u, x) - Q_{UX}(u, x)| \leq \frac{1}{n}$ for all (u, x) , and we denote the marginals by $Q_{U,n}$ and $Q_{X|U,n}$. The codeword distributions are given by

$$P_U(\mathbf{u}) = \frac{1}{|T^n(Q_{U,n})|} \mathbb{1} \left\{ \mathbf{u} \in T^n(Q_{U,n}) \right\} \quad (5.10)$$

$$P_{\mathbf{X}|\mathbf{U}}(\mathbf{x}|\mathbf{u}) = \frac{1}{|T_{\mathbf{u}}^n(Q_{X|U,n})|} \mathbb{1} \left\{ \mathbf{x} \in T_{\mathbf{u}}^n(Q_{X|U,n}) \right\}. \quad (5.11)$$

We state the exponents and rates corresponding to the non-asymptotic bounds in Theorem 5.1 without proof, since the proofs are similar to those in Section 4.2, and since equivalent exponents and rates in a slightly different form were given in [42]. We define the sets

$$\mathcal{S}(Q_{UX}) \triangleq \{ P_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}) : P_{UX} = Q_{UX} \} \quad (5.12)$$

$$\begin{aligned} \mathcal{T}_0(P_{UXY}) \triangleq \{ & \tilde{P}_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}) : \\ & \tilde{P}_{UX} = P_{UX}, \tilde{P}_Y = P_Y, \mathbb{E}_{\tilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)] \} \end{aligned} \quad (5.13)$$

$$\begin{aligned} \mathcal{T}_1(P_{UXY}) \triangleq \{ & \tilde{P}_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}) : \\ & \tilde{P}_{UX} = P_{UX}, \tilde{P}_{UY} = P_{UY}, \mathbb{E}_{\tilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)] \}. \end{aligned} \quad (5.14)$$

Theorem 5.2. *For any input distribution Q_{UX} and pair (R_0, R_1) , we have under the constant-composition superposition coding ensemble in (5.10)–(5.11) that*

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \bar{p}_{e,0}(n, \lfloor e^{nR_0} \rfloor, \lfloor e^{nR_1} \rfloor) = E_{r,0}^{\text{cc}}(Q_{UX}, R_0, R_1) \quad (5.15)$$

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \bar{p}_{e,1}(n, \lfloor e^{nR_1} \rfloor) = E_{r,1}^{\text{cc}}(Q_{UX}, R_1), \quad (5.16)$$

where

$$E_{r,0}^{\text{cc}}(Q_{UX}, R_0, R_1) \triangleq \min_{P_{UXY} \in \mathcal{S}(Q_{UX})} \min_{\tilde{P}_{UXY} \in \mathcal{T}_0(P_{UXY})} D(P_{UXY} \| Q_{UX} \times W) + \left[I_{\tilde{P}}(U; Y) + [I_{\tilde{P}}(X; Y|U) - R_1]^+ - R_0 \right]^+ \quad (5.17)$$

$$E_{r,1}^{\text{cc}}(Q_{UX}, R_1) \triangleq \min_{P_{UXY} \in \mathcal{S}(Q_{UX})} \min_{\tilde{P}_{UXY} \in \mathcal{T}_1(P_{UXY})} D(P_{UXY} \| Q_{UX} \times W) + [I_{\tilde{P}}(X; Y|U) - R_1]^+. \quad (5.18)$$

From (5.3), the overall error exponent is given by

$$E_r^{\text{cc}}(Q_{UX}, R_0, R_1) \triangleq \min \left\{ E_{r,0}^{\text{cc}}(Q_{UX}, R_0, R_1), E_{r,1}^{\text{cc}}(Q_{UX}, R_1) \right\}, \quad (5.19)$$

and is tight with respect to the ensemble average. Following the steps of Section 4.2.4, it can be shown that in the case of ML decoding (i.e. $q(x, y) = W(y|x)$), we have

$$E_r^{\text{cc}}(Q_{UX}, R_0, R_1) = \min \left\{ E_{r,0}^{\text{cc}'}(Q_{UX}, R_0, R_1), E_{r,1}^{\text{cc}}(Q_{UX}, R_1) \right\}, \quad (5.20)$$

where

$$E_{r,0}^{\text{cc}'}(Q_{UX}, R_0, R_1) \triangleq \min_{P_{UXY} \in \mathcal{S}(Q_{UX})} D(P_{UXY} \| Q_{UX} \times W) + [I_P(U, X; Y) - (R_0 + R_1)]^+ \quad (5.21)$$

$$E_{r,1}^{\text{cc}}(Q_{UX}, R_1) = \min_{P_{UXY} \in \mathcal{S}(Q_{UX})} D(P_{UXY} \| Q_{UX} \times W) + [I_P(X; Y|U) - R_1]^+. \quad (5.22)$$

These exponents coincide with those of the primary user of the broadcast channel with degraded message sets given in [137]. Thus, the exponents given in [137] are tight with respect to the ensemble average for ML decoding.

Next, we provide the achievable rates resulting from Theorem 5.2.

Theorem 5.3. *For any input distribution Q_{UX} , the overall error exponent $E_{\text{r}}^{\text{cc}}(Q_{UX}, R_0, R_1)$ is positive for all rate pairs (R_1, R_2) in the interior of the region of pairs (R_0, R_1) satisfying*

$$R_1 \leq \min_{\tilde{P}_{UXY} \in \mathcal{T}_1(Q_{UX} \times W)} I_{\tilde{P}}(X; Y|U) \quad (5.23)$$

$$R_0 + R_1 \leq \min_{\substack{\tilde{P}_{UXY} \in \mathcal{T}_0(Q_{UX} \times W) \\ I_{\tilde{P}}(U; Y) \leq R_0}} I_{\tilde{P}}(U, X; Y). \quad (5.24)$$

It follows from Theorem 5.3 that the rate

$$R = R_0 + R_1 \quad (5.25)$$

is achievable for any (R_0, R_1) satisfying (5.23)–(5.24). The corresponding ensemble tightness result was proved in [42] using the techniques of [35, Thm. 3]. Using (5.21)–(5.22), we recover the rate conditions for the primary user of the broadcast channel with degraded message sets [134] when $q(x, y) = W(y|x)$.

The following theorem gives dual expressions for the rates. In the following subsection, we will use cost-constrained coding to show that these expressions remain valid for more general alphabets. Similar arguments can be applied to the exponents, but we focus on the rates the sake of brevity.

Theorem 5.4. *The conditions in (5.23)–(5.24) can be expressed as*

$$R_1 \leq \sup_{s \geq 0, a(\cdot, \cdot)} \mathbb{E} \left[\log \frac{q(X, Y)^s e^{a(U, X)}}{\mathbb{E}[q(\tilde{X}, Y)^s e^{a(U, \tilde{X})} | U, Y]} \right] \quad (5.26)$$

$$R_0 \leq \sup_{\rho_1 \in [0, 1], s \geq 0, a(\cdot, \cdot)} \mathbb{E} \left[\log \frac{(q(X, Y)^s e^{a(U, X)})^{\rho_1}}{\mathbb{E} \left[\left(\mathbb{E}[q(\bar{X}, Y)^s e^{a(\bar{U}, \bar{X})} | \bar{U}] \right)^{\rho_1} | Y \right]} \right] - \rho_1 R_1, \quad (5.27)$$

where $(U, X, Y, \tilde{X}, \bar{U}, \bar{X}) \sim Q_{UX}(u, x)W(y|x)Q_{X|U}(\tilde{x}|u)Q_{UX}(\bar{u}, \bar{x})$.

Proof. The proof is nearly identical to that of Theorem 4.5, and we thus omit the details. Equation (5.26) follows from (5.23) analogously to the steps in obtaining (4.66)–(4.67) from (4.42)–(4.43), and (5.27) follows from (5.24) analogously to the steps in obtaining (4.68)–(4.69) from (4.56)–(4.57). $\square$

5.2.3 Comparison to Expurgated Parallel Coding

The achievable rate of Theorem 5.3 has a similar form to that of Theorem 4.10. The following proposition states that the former can be weakened to the latter.

Proposition 5.1. *The achievable rate $R^{\text{sc}} \triangleq \max_{R_0, R_1} R_0 + R_1$ described by (5.23)–(5.24) with optimized parameters $\mathcal{U}$ and Q_{UX} is at least as high as the achievable rate $R^{\text{ex}} \triangleq \max_{R_1, R_2} R_1 + R_2$ described by (4.112)–(4.114) with optimized parameters $\mathcal{X}_1, \mathcal{X}_2, Q_1, Q_2$ and $\psi(\cdot, \cdot)$.*

Proof. We make use of the equivalent type-12 conditions in (4.115)–(4.116) for the expurgated parallel coding rate. To avoid ambiguity between overloaded symbols (e.g. R_1), we use a superscript to specify the relevant ensemble for some quantities (e.g. $R_1^{\text{sc}}, R_1^{\text{ex}}$). We will show that by identifying $U^{\text{sc}} = X_2^{\text{ex}}, R_1^{\text{sc}} = R_1^{\text{ex}}$ and $R_0^{\text{sc}} = R_2^{\text{ex}}$, we can weaken (5.23) to (4.112) and (5.24) to (4.116). A similar argument with $U^{\text{sc}} = X_1^{\text{ex}}, R_1^{\text{sc}} = R_2^{\text{ex}}$ and $R_0^{\text{sc}} = R_1^{\text{ex}}$ shows that we can weaken (5.23) to (4.113) and (5.24) to (4.115). Combining these results, it follows that Theorem 5.3 implies the achievability of any $R = R_1 + R_2$ such that (R_1, R_2) satisfies (4.112)–(4.113) and at least one of (4.115)–(4.116), as desired.

The process of weakening (5.23) to (4.112) is similar to that of weakening (5.24) to (4.116), so we focus on the former. Recalling that $U^{\text{sc}} = X_2^{\text{ex}}$, the optimization in (5.23) can be written as

$$\min_{\tilde{P}_{Y|X_2X}} \sum_{x_2, x, y} Q_{X_2X}(x_2, x) \tilde{P}_{Y|X_2X}(y|x_2, x) \log \frac{\tilde{P}_{Y|X_2X}(y|x_2, x)}{\sum_x Q_{X|X_2}(x|x_2) W(y|x)} \quad (5.28)$$

subject to the constraints

$$\sum_x Q_{X_2X}(x_2, x) \tilde{P}_{Y|X_2X}(y|x_2, x) = \sum_x Q_{X_2X}(x_2, x) W(y|x) \quad (5.29)$$

$$\sum_{x_2, x, y} Q_{X_2X}(x_2, x) \tilde{P}_{Y|X_2X}(y|x_2, x) \log q(x, y) \geq \sum_{x, y} Q_X(x) W(y|x) \log q(x, y). \quad (5.30)$$

Since $Q_{X_2X} = Q_{UX}$ can be chosen arbitrarily, we set

$$Q_{X_2X}(x_2, x) = \sum_{x_1} Q_1(x_1) Q_2(x_2) \mathbb{1}\{x = \psi(x_1, x_2)\} \quad (5.31)$$

for some Q_1, Q_2 and $\psi(\cdot, \cdot)$. Under this choice, the objective in (5.28) is given by

$$\sum_{x_1, x_2, y} Q_1(x_1)Q_2(x_2)\tilde{P}_{Y|X_2X}(y|x_2, \psi(x_1, x_2)) \log \frac{\tilde{P}_{Y|X_2X}(y|x_2, \psi(x_1, x_2))}{\sum_{\bar{x}_1} Q_1(\bar{x}_1)W(y|\psi(\bar{x}_1, x_2))}, \quad (5.32)$$

and the constraints in (5.29)–(5.30) are written similarly with $\psi(x_1, x_2)$ playing the role of x . Since the objective and constraints depend on $\tilde{P}_{Y|X_1X}(y|x_1, x)$ only through $\tilde{P}_{Y|X_1X}(y|x_1, \psi(x_1, x_2))$, we can lower bound the minimization by instead taking it over all conditional distributions $\tilde{P}_{Y|X_1X_2}$. Thus, we recover the minimization

$$\min_{\tilde{P}_{Y|X_1X_2}} \sum_{x_1, x_2, y} Q_1(x_1)Q_2(x_2)\tilde{P}_{Y|X_1X_2}(y|x_1, x_2) \log \frac{\tilde{P}_{Y|X_1X_2}(y|x_1, x_2)}{\sum_{\bar{x}_1} Q_1(\bar{x}_1)W(y|\psi(\bar{x}_1, x_2))} \quad (5.33)$$

subject to the constraints

$$\sum_{x_1} Q_1(x_1)Q_2(x_2)\tilde{P}_{Y|X_1X_2}(y|x_1, x_2) = \sum_{x_1} Q_1(x_1)Q_2(x_2)W(y|\psi(x_1, x_2)) \quad (5.34)$$

$$\begin{aligned} \sum_{x_1, x_2, y} Q_1(x_1)Q_2(x_2)\tilde{P}_{Y|X_1X_2}(y|x_1, x_2) \log q(\psi(x_1, x_2), y) \\ \geq \sum_{x_1, x_2, y} Q_1(x_1)Q_2(x_2)W(y|\psi(x_1, x_2)) \log q(\psi(x_1, x_2), y). \end{aligned} \quad (5.35)$$

Rewriting this minimization as a minimization over joint distributions $\tilde{P}_{X_1X_2Y}$ subject to $\tilde{P}_{X_1X_2} = Q_1 \times Q_2$, we recover the right-hand side of (4.112). $\square$

The rate condition in (4.44) may appear to improve on (5.24) via the inclusion of two constraints of the form $I_{\tilde{P}}(X_\nu; Y) \leq R_\nu$ instead of just one, but Lemma 4.3 shows that at most one of these constraints is active for any given rate pair. Conversely, it may appear that the ability to choose Q_{UX} in (5.23)–(5.24) provides more freedom than the ability to choose Q_1 and Q_2 in (4.42)–(4.44), but it is known that in the latter setting, any joint distribution of (X_1, X) (or (X_2, X)) can be induced by a suitable choice of $\psi(\cdot, \cdot)$ [115, p. 626]. This suggests that it may be possible to obtain a matching converse to Proposition 5.1.

Regardless of whether the two rates are equivalent, the SC rate appears to be more amenable to local optimization techniques, since the optimization over $(\mathcal{X}_1, \mathcal{X}_2, Q_1, Q_2, \psi)$ is replaced by a seemingly simpler optimization over $(\mathcal{U}, Q_{UX})$. The optimization of the input distribution is a non-convex problem in general even for the single-user LM rate [11], and thus the ability to perform local optimizations is of significant interest.

5.2.4 Rates for More General Alphabets

The following theorem shows that the expressions in Theorem 5.4 remain valid in the setting of infinite and continuous alphabets.

Theorem 5.5. *For any mismatched channel (possibly having infinite or continuous alphabets), auxiliary alphabet $\mathcal{U}$, and input distribution $Q_{UX} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$, the rate $R = R_0 + R_1$ is achievable for any pair (R_0, R_1) satisfying (5.26)–(5.27), where each supremum over $a(\cdot, \cdot)$ is taken over all functions such that $E_Q[a(U, X)^2] < \infty$.*

To prove Theorem 5.5, we consider superposition coding of the form

$$P_U(\mathbf{u}) = \frac{1}{\mu_{1,n}} \prod_{i=1}^n Q_U(u_i) \mathbb{1}\{\mathbf{u} \in \mathcal{D}_{1,n}\} \quad (5.36)$$

$$P_{\mathbf{X}|\mathbf{U}}(\mathbf{x}|\mathbf{u}) = \frac{1}{\mu_{2,n}(\mathbf{u})} \prod_{i=1}^n Q_{X|U}(x_i|u_i) \mathbb{1}\{\mathbf{x} \in \mathcal{D}_{2,n}(\mathbf{u})\} \quad (5.37)$$

with $\mathcal{D}_{1,n}$ and $\mathcal{D}_{2,n}(\mathbf{u})$ chosen such that (i) $\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_{1,n} = 0$, (ii) $\lim_{n \rightarrow \infty} \frac{1}{n} \log \mu_{2,n}(\mathbf{u}) = 0$ for all $\mathbf{u} \in \mathcal{D}_{1,n}$, and (iii) there exist constants $\{\phi_l\}$ and δ such that for each $l = 1, \dots, L$, we have $|\frac{1}{n} \sum_{i=1}^n a_l(u_i, x_i) - \phi_l| \leq \frac{\delta}{n}$ for all $(\mathbf{u}, \mathbf{x})$ on the support of $P_U \times P_{\mathbf{X}|\mathbf{U}}$. Once this ensemble is established, the proof of Theorem 5.5 will follow using similar steps to Section 4.2.3, starting with the RCU bounds given in Theorem 5.1. These remaining details are omitted to avoid repetition.

We fix a sequence g_n and choose

$$\mathcal{D}_{1,n} \triangleq \left\{ \mathbf{u} : \mathbb{P}[\mathbf{X}'(\mathbf{u}) \in \mathcal{D}_{2,n}(\mathbf{u})] \geq \frac{1}{g_n} \right\} \quad (5.38)$$

$$\mathcal{D}_{2,n}(\mathbf{u}) \triangleq \left\{ \mathbf{x} : \left| \frac{1}{n} \sum_{i=1}^n a_l(u_i, x_i) - \phi_l \right| \leq \frac{\delta}{n}, l = 1, \dots, L \right\}, \quad (5.39)$$

where $\mathbf{X}'(\mathbf{u}) \sim Q_{\mathbf{X}|\mathbf{U}}^n(\mathbf{x}|\mathbf{u})$ and $\phi_l \triangleq \mathbb{E}_Q[a_l(U, X)]$. For any $\mathbf{u} \in \mathcal{D}_{1,n}$, we have from (5.38) that $\mu_{2,n}(\mathbf{u})$ decays subexponentially provided that g_n grows subexponentially. Hence, it only remains to show that there exists a subexponential sequence g_n such that $\mu_{1,n}$ decays subexponentially. Defining

$$(\mathbf{U}', \mathbf{X}') \sim Q_U^n(\mathbf{u}') Q_{\mathbf{X}|\mathbf{U}}^n(\mathbf{x}'|\mathbf{u}'), \quad (5.40)$$

we have from Proposition 2.1 that

$$\mathbb{P}[\mathbf{X}' \in \mathcal{D}_{2,n}(\mathbf{U}')] = \frac{1}{f_n} \quad (5.41)$$

for some f_n growing at most polynomially in n . Since $\mu_{1,n} = \mathbb{P}[U' \in \mathcal{D}_{1,n}]$ by definition, we can expand (5.41) as

$$\frac{1}{f_n} = \mathbb{P}[\mathbf{X}' \in \mathcal{D}_{2,n}(U') | U' \in \mathcal{D}_{1,n}] \mu_{1,n} + \mathbb{P}[\mathbf{X}' \in \mathcal{D}_{2,n}(U') | U' \notin \mathcal{D}_{1,n}] (1 - \mu_{1,n}) \quad (5.42)$$

$$\leq \mu_{1,n} + \frac{1}{g_n}, \quad (5.43)$$

where (5.43) follows by upper bounding the first probability in (5.42) by one, and upper bounding the second probability by $\frac{1}{g_n}$ using (5.38). Choosing $g_n = 2f_n$, we see that g_n grows polynomially fast, and $\mu_{1,n} \geq \frac{1}{2f_n}$ decays polynomially fast, as desired.

Since the conditions (5.26)–(5.27) each contain a supremum over $a(\cdot, \cdot)$, it suffices to choose $L = 2$ in (5.39) to prove Theorem 5.5.

5.3 Refined Superposition Coding

In this section, we introduce a refined version of the superposition coding ensemble that can yield higher achievable rates than the standard version studied in Section 5.2. While our analysis of the error probability will yield non-asymptotic bounds and error exponents as intermediate steps, we focus on the resulting achievable rates for clarity.

The ensemble is defined as follows. We fix a finite alphabet $\mathcal{U}$, an input distribution $Q_U \in \mathcal{P}(\mathcal{U})$, and the integers M_0 and $\{M_{1u}\}_{u \in \mathcal{U}}$. We let $P_U(\mathbf{u})$ be the uniform distribution on the type class $T^n(Q_{U,n})$, where $Q_{U,n} \in \mathcal{P}_n(\mathcal{U})$ approximates Q_U in the same way as Section 5.2.2. As in (5.10), we set

$$P_U(\mathbf{u}) = \frac{1}{|T^n(Q_{U,n})|} \mathbb{1}\{\mathbf{u} \in T^n(Q_{U,n})\} \quad (5.44)$$

and generate the length- n auxiliary codewords $\{U^{(i)}\}_{i=1}^{M_0}$ independently according to P_U . The difference here is that the codewords are not generated conditionally independently given $U^{(i)}$. Instead, we generate a number of partial codewords, and construct the length- n codeword by placing the entries of a partial codeword in the indices where U takes a particular value.

A more precise description is as follows. For each $u \in \mathcal{U}$, we define

$$n_u \triangleq Q_{U,n}(u)n \quad (5.45)$$

and fix a partial codeword distribution $P_{\mathbf{X}_u} \in \mathcal{P}(\mathcal{X}^{n_u})$. For each $i = 1, \dots, M_0$ and $u \in \mathcal{U}$, we generate the length- n_u partial codewords $\{\mathbf{X}_u^{(i,j_u)}\}_{j_u=1}^{M_{1u}}$ independently according to $P_{\mathbf{X}_u}$. For

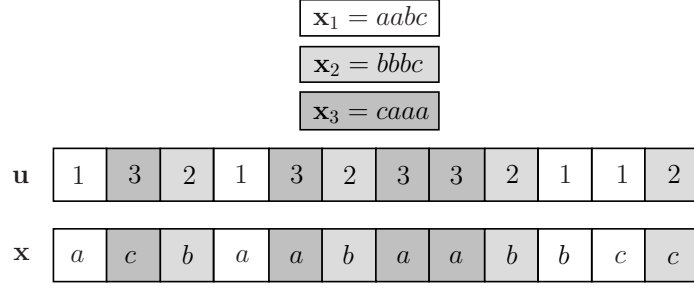


Figure 5.1 The construction of the codeword from the auxiliary sequence u and the partial codewords x_1 , x_2 and x_3 for refined superposition coding. Here we have $n = 12$ and $n_1 = n_2 = n_3 = 4$.

example, when $\mathcal{U} = \{1, 2\}$ we have

$$\begin{aligned}
 & \left\{ \left(U^{(i)}, \{X_1^{(i,j_1)}\}_{j_1=1}^{M_{11}}, \{X_2^{(i,j_2)}\}_{j_2=1}^{M_{12}} \right) \right\}_{i=1}^{M_0} \\
 & \sim \prod_{i=1}^{M_0} \left(P_U(u^{(i)}) \prod_{j_1=1}^{M_{11}} P_{X_1}(x_1^{(i,j_1)}) \prod_{j_2=1}^{M_{12}} P_{X_2}(x_2^{(i,j_2)}) \right). \quad (5.46)
 \end{aligned}$$

The message m at the encoder is indexed as $(m_0, m_{11}, \dots, m_{1|\mathcal{U}|})$. To transmit a given message, we treat $U^{(m_0)}$ as a time sharing sequence; at the indices where $U^{(m_0)}$ equals u , we transmit the symbols of $X_u^{(m_0, m_{1u})}$. The total number of codewords is $M = M_0 \prod_u M_{1u}$. An example of the construction of the codeword x from the auxiliary sequence u and partial codewords x_1 , x_2 and x_3 is shown in Figure 5.1, where we have $\mathcal{U} = \{1, 2, 3\}$ and $\mathcal{X} = \{a, b, c\}$.

While our main result is stated for an arbitrary finite alphabet $\mathcal{U}$, the analysis will be presented for $\mathcal{U} = \{1, 2\}$ for clarity. We proceed by presenting several definitions for this specific choice. We let $\Xi(u, x_1, x_2)$ denote the function for constructing the length- n codeword from the auxiliary sequence and partial codewords, and we write

$$X^{(i,j_1,j_2)} \triangleq \Xi(U^{(i)}, X_1^{(i,j_1)}, X_2^{(i,j_2)}). \quad (5.47)$$

We let $y_u(u)$ denote the subsequence of y corresponding to the indices where u equals u , and similarly for $Y_u(u)$.

We assume without loss of generality that $(m_0, m_1, m_2) = (1, 1, 1)$. We let U , X_1 , X_2 and X be the codewords corresponding to $(1, 1, 1)$, yielding $X = \Xi(U, X_1, X_2)$. We let $\bar{U}$, $\bar{X}_1$ and $\bar{X}_2$ be the codewords corresponding to an arbitrary message with $m_0 \neq 1$. For the index i corresponding to $\bar{U}$, we write $\bar{X}_1^{(j_1)}$, $\bar{X}_2^{(j_2)}$ and $\bar{X}^{(j_1,j_2)}$ in place of $X_1^{(i,j_1)}$, $X_2^{(i,j_2)}$ and $X^{(i,j_1,j_2)}$ respectively. It follows that $\bar{X}^{(j_1,j_2)} = \Xi(\bar{U}, \bar{X}_1^{(j_1)}, \bar{X}_2^{(j_2)})$.

Upon receiving a realization $\mathbf{y}$ of the output sequence $\mathbf{Y}$, the decoder forms the estimate

$$(\hat{m}_0, \hat{m}_1, \hat{m}_2) = \arg \max_{(i, j_1, j_2)} q^n(\mathbf{x}^{(i, j_1, j_2)}, \mathbf{y}) \quad (5.48)$$

$$= \arg \max_{(i, j_1, j_2)} q^{n_1}(\mathbf{x}_1^{(i, j_1)}, \mathbf{y}_1(\mathbf{u}^{(i)})) q^{n_2}(\mathbf{x}_2^{(i, j_2)}, \mathbf{y}_2(\mathbf{u}^{(i)})), \quad (5.49)$$

where the objective in (5.49) follows by separating the indices where $u = 1$ from those where $u = 2$. By writing the objective in this form, it is easily seen that for any given i , the pair (j_1, j_2) with the highest metric is the one for which j_1 maximizes $q^{n_1}(\mathbf{x}_1^{(i, j_1)}, \mathbf{y}_1(\mathbf{u}^{(i)}))$ and j_2 maximizes $q^{n_2}(\mathbf{x}_2^{(i, j_2)}, \mathbf{y}_2(\mathbf{u}^{(i)}))$. We may therefore consider the following error events:

- (Type 0) $q^n(\mathbf{X}^{(i, j_1, j_2)}, \mathbf{Y}) \geq q^n(\mathbf{X}, \mathbf{Y})$ for some $i \neq 1, j_1, j_2$;
- (Type 1) $q^{n_1}(\bar{\mathbf{X}}_1^{(1, j_1)}, \mathbf{Y}_1(\mathbf{U})) \geq q^{n_1}(\mathbf{X}_1, \mathbf{Y}_1(\mathbf{U}))$ for some $j_1 \neq 1$;
- (Type 2) $q^{n_2}(\bar{\mathbf{X}}_2^{(1, j_2)}, \mathbf{Y}_2(\mathbf{U})) \geq q^{n_2}(\mathbf{X}_2, \mathbf{Y}_2(\mathbf{U}))$ for some $j_2 \neq 1$.

The corresponding probabilities are denoted by $\bar{p}_{e,0}(n, M_0, M_{11}, M_{12})$, $\bar{p}_{e,1}(n, M_{11})$ and $\bar{p}_{e,2}(n, M_{12})$ respectively. Analogously to (4.10), the overall random-coding error probability $\bar{p}_e(n, M_0, M_{11}, M_{12})$ satisfies

$$\frac{1}{2} \max\{\bar{p}_{e,0}, \bar{p}_{e,1}, \bar{p}_{e,2}\} \leq \bar{p}_e \leq \bar{p}_{e,0} + \bar{p}_{e,1} + \bar{p}_{e,2}. \quad (5.50)$$

5.3.1 Rates for DMCs

In this subsection, we assume that the channel is a DMC. We fix a joint distribution Q_{UX} , and we use the definitions of $Q_{UX,n}$, $Q_{U,n}$ and $Q_{X|U,n}$ in Section 5.2.2. We let P_{X_u} be the uniform distribution on the type class $T^{n_u}(Q_{X|U,n}(\cdot|u))$, yielding

$$P_{X_u}(\mathbf{x}_u) = \frac{1}{|T^{n_u}(Q_{X|U,n}(\cdot|u))|} \mathbf{1}\{\mathbf{x}_u \in T^{n_u}(Q_{X|U,n}(\cdot|u))\}. \quad (5.51)$$

Combining this with (5.44), a straightforward symmetry argument reveals that each pair $(\mathbf{U}^{(i)}, \mathbf{X}^{(i, j_1, j_2)})$ is uniformly distributed on $T^n(Q_{UX})$, as was the case for standard SC in Section 5.2.2. The difference is in the dependence among the codewords.

The main result of this section is stated in the following theorem, which makes use of the LM rate defined in (1.10) and the set $\mathcal{T}_0$ defined in (5.13):

$$\begin{aligned} \mathcal{T}_0(P_{UXY}) &\triangleq \left\{ \tilde{P}_{UXY} \in \mathcal{P}(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}) : \right. \\ &\quad \left. \tilde{P}_{UX} = P_{UX}, \tilde{P}_Y = P_Y, \mathbb{E}_{\tilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)] \right\}. \end{aligned} \quad (5.52)$$

Theorem 5.6. *For any finite set $\mathcal{U}$ and input distribution $Q_{UX} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$, the rate*

$$R = R_0 + \sum_u Q_U(u) R_{1u} \quad (5.53)$$

is achievable provided that R_0 and $\{R_{1u}\}_{u=1}^{|\mathcal{U}|}$ satisfy

$$R_{1u} \leq I_{\text{LM}}(Q_{X|U}(\cdot|u)), \quad u \in \mathcal{U} \quad (5.54)$$

$$R_0 \leq \min_{\tilde{P}_{UXY} \in \mathcal{T}_0(Q_{UX} \times W)} I_{\tilde{P}}(U; Y) + \left[\max_{\mathcal{K} \subseteq \mathcal{U}, \mathcal{K} \neq \emptyset} \sum_{u \in \mathcal{K}} Q_U(u) \left(I_{\tilde{P}}(X; Y|U = u) - R_{1u} \right) \right]^+. \quad (5.55)$$

Proof. As mentioned above, the proof is presented only for $\mathcal{U} = \{1, 2\}$; the same arguments apply in the general case. Observe that the type-1 error event corresponds to the error event for the standard constant-composition ensemble with rate R_{11} , length $n_1 = nQ_U(1)$, input distribution $Q_{X|U}(\cdot|1)$, and ties broken as errors. A similar statement holds for the type-2 error probability $\bar{p}_{e,2}$, and the analysis for these error events is identical to the derivation of the LM rate (see Section 2.4), yielding (5.54).

The error probability for the type-0 event is given by

$$\bar{p}_{e,0} = \mathbb{P} \left[\bigcup_{i \neq 1} \bigcup_{j_1, j_2} \left\{ \frac{q^n(\mathbf{X}^{(i, j_1, j_2)}, \mathbf{Y})}{q^n(\mathbf{X}, \mathbf{Y})} \geq 1 \right\} \right], \quad (5.56)$$

where $(\mathbf{Y}|\mathbf{X} = \mathbf{x}) \sim W^n(\cdot|\mathbf{x})$. Writing the probability as an expectation given $(U, \mathbf{X}, \mathbf{Y})$ and applying the truncated union bound, we obtain

$$\bar{p}_{e,0} = \zeta_0 \mathbb{E} \left[\min \left\{ 1, (M_0 - 1) \mathbb{E} \left[\mathbb{P} \left[\bigcup_{j_1, j_2} \left\{ \frac{q^n(\bar{\mathbf{X}}^{(j_1, j_2)}, \mathbf{Y})}{q^n(\mathbf{X}, \mathbf{Y})} \geq 1 \right\} \mid \bar{U} \right] \mid U, \mathbf{X}, \mathbf{Y} \right] \right\} \right], \quad (5.57)$$

where $\zeta_0 \in [\frac{1}{2}, 1]$, since for independent events the truncated union bound is tight to within a factor of $\frac{1}{2}$ [73, Lemma A.2]. We have written the probability of the union over j_1 and j_2 as an expectation given $\bar{U}$.

Let the joint types of $(U, \mathbf{X}, \mathbf{Y})$ and $(\bar{U}, \bar{\mathbf{X}}^{(j_1, j_2)}, \mathbf{Y})$ be denoted by P_{UXY} and $\tilde{P}_{UXY}$ respectively. We claim that the condition

$$\frac{q^n(\bar{\mathbf{X}}^{(j_1, j_2)}, \mathbf{Y})}{q^n(\mathbf{X}, \mathbf{Y})} \geq 1 \quad (5.58)$$

can be written as

$$\tilde{P}_{UXY} \in \mathcal{T}_{0,n}(P_{UXY}) \triangleq \mathcal{T}_0(P_{UXY}) \cap \mathcal{P}_n(\mathcal{U} \times \mathcal{X} \times \mathcal{Y}). \quad (5.59)$$

The constraint $\tilde{P}_{UX} = P_{UX}$ follows from the construction of the random coding ensemble, $\tilde{P}_Y = P_Y$ follows since $(U, \mathbf{X}, \mathbf{Y})$ and $(\bar{U}, \bar{\mathbf{X}}^{(j_1, j_2)}, \mathbf{Y})$ share the same $\mathbf{Y}$ sequence, and $\mathbb{E}_{\tilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)]$ coincides with the condition in (5.58). Thus, expanding (5.57) in terms of types, we obtain

$$\begin{aligned} \bar{p}_{e,0} = & \zeta_0 \sum_{P_{UXY}} \mathbb{P}[(U, \mathbf{X}, \mathbf{Y}) \in T^n(P_{UXY})] \min \left\{ 1, (M_0 - 1) \right. \\ & \times \sum_{\tilde{P}_{UXY} \in \mathcal{T}_{0,n}(P_{UXY})} \mathbb{P}[(\bar{U}, \mathbf{y}) \in T^n(\tilde{P}_{UY})] \mathbb{P} \left[\bigcup_{j_1, j_2} \left\{ (\bar{\mathbf{u}}, \bar{\mathbf{X}}^{(j_1, j_2)}, \mathbf{y}) \in T^n(\tilde{P}_{UXY}) \right\} \right] \Big\}, \end{aligned} \quad (5.60)$$

where we write $(\bar{\mathbf{u}}, \mathbf{y})$ to denote an arbitrary pair with $\mathbf{y} \in T^n(P_Y)$ and $(\bar{\mathbf{u}}, \mathbf{y}) \in T^n(\tilde{P}_{UY})$.

Similarly to the discussion following (5.49), we observe that $(\bar{\mathbf{u}}, \bar{\mathbf{X}}^{(j_1, j_2)}, \mathbf{y}) \in T^n(\tilde{P}_{UXY})$ if and only if $(\bar{\mathbf{X}}_u^{(j_u)}, \mathbf{y}_u(\bar{\mathbf{u}})) \in T^{n_u}(\tilde{P}_{XY|U}(\cdot, \cdot|u))$ for $u = 1, 2$. Thus, applying Lemma D.3 with $Z_1(j_1) = \mathbf{X}_1^{(j_1)}$, $Z_2(j_2) = \mathbf{X}_2^{(j_2)}$, $\mathcal{A}_1 = T_{\mathbf{y}_1(\bar{\mathbf{u}})}^{n_1}(\tilde{P}_{XY|U}(\cdot, \cdot|1))$, $\mathcal{A}_2 = T_{\mathbf{y}_2(\bar{\mathbf{u}})}^{n_2}(\tilde{P}_{XY|U}(\cdot, \cdot|2))$, and $\mathcal{A} = \{(\mathbf{X}_1, \mathbf{X}_2) : \mathbf{X}_u \in T_{\mathbf{y}_u(\bar{\mathbf{u}})}^{n_u}(\tilde{P}_{XY|U}(\cdot, \cdot|u)), u = 1, 2\}$, we obtain

$$\begin{aligned} & \mathbb{P} \left[\bigcup_{j_1, j_2} \left\{ (\bar{\mathbf{u}}, \bar{\mathbf{X}}^{(j_1, j_2)}, \mathbf{y}) \in T^n(\tilde{P}_{UXY}) \right\} \right] \\ &= \zeta'_0 \min \left\{ 1, \min_{u=1,2} M_{1u} \mathbb{P}[(\bar{\mathbf{X}}_u, \mathbf{y}_u(\bar{\mathbf{u}})) \in T^{n_u}(\tilde{P}_{XY|U}(\cdot, \cdot|u))] \right\}, \\ & \quad M_{11} M_{12} \mathbb{P} \left[\bigcap_{u=1,2} \left\{ (\bar{\mathbf{X}}_u, \mathbf{y}_u(\bar{\mathbf{u}})) \in T^{n_u}(\tilde{P}_{XY|U}(\cdot, \cdot|u)) \right\} \right], \end{aligned} \quad (5.61)$$

where $\zeta'_0 \in [\frac{1}{4}, 1]$. The four terms in the minimization correspond to the subsets of $\{1, 2\}$.

Fix the rates (R_0, R_{11}, R_{12}) , and let $M_0 = \lfloor e^{nR_0} \rfloor$, and $M_{1u} = \lfloor e^{nR_{1u}} \rfloor$ ($u = 1, 2$). Substituting (5.61) into (5.60), applying the properties of types in (A.16) and (A.17), and using the fact that the number of joint types grows polynomially in n , we obtain

$$\begin{aligned} \lim_{n \rightarrow \infty} -\frac{1}{n} \log \bar{p}_{e,0} = & \min_{P_{UXY} : P_{UX} = Q_{UX}} \min_{\tilde{P}_{UXY} \in \mathcal{T}_0(P_{UXY})} D(P_{UXY} \| Q_{UX} \times W) \\ & + \left[I_{\tilde{P}}(U; Y) + \left[\max_{\mathcal{K} \subseteq \mathcal{U}, \mathcal{K} \neq \emptyset} \sum_{u \in \mathcal{K}} Q_U(u) \left(I_{\tilde{P}}(X; Y|U = u) - R_{1u} \right) \right]^+ - R_0 \right]^+, \end{aligned} \quad (5.62)$$

where we have replaced the minimizations over types by minimizations over all distributions in the same way as the proof of Theorem 4.2. Taking $P_{UXY} \rightarrow Q_{UX} \times W$ and using a similar continuity argument to [35, p. 1444, Step 3], we see that the right-hand side of (5.62) is positive whenever (5.55) holds with strict inequality, thus completing the proof. $\square$

The proof of Theorem 5.6 gives an exponentially tight analysis yielding the exponent in (5.62). However, this does not prove that the resulting rate is ensemble-tight, since a subexponential decay of the error probability to zero is possible in principle. We can prove the ensemble tightness of the rate using (5.60)–(5.61), and dealing with the summation over P_{UXY} more carefully by following [35]. Instead of replacing the first summation in (5.60) by a maximum, we restrict the summation to types that are δ -close to $Q_{UX} \times W$ in terms of the ∞ -norm. The probability of (U, X, Y) having such a joint type approaches one as $n \rightarrow \infty$ for any $\delta > 0$. Since ζ_0 in (5.57) is greater than or equal to $\frac{1}{2}$, the resulting error probability is lower bounded by $\frac{1}{2}(1 + o(1))$ for rates exceeding $R_0^* + \Delta(\delta)$, where R_0^* denotes the right-hand side of (5.55), and $\Delta(\delta) \rightarrow 0$ as $\delta \rightarrow 0$ by a similar continuity argument to [8, Lemma 1]. It follows that the error probability is bounded away from zero whenever (5.55) fails. The analogous claim for (5.54) follows using identical arguments to the single-user setting [9, 16].

5.3.2 Comparison to Standard Superposition Coding

In this subsection, we show that the conditions in (5.54)–(5.55) can be weakened to (5.23)–(5.24) upon identifying

$$R_1 = \sum_u Q_U(u) R_{1u}. \quad (5.63)$$

We show this to hold for *any* auxiliary alphabet and input distribution (in contrast with Proposition 5.1, which states an analogous with optimized parameters).

Proposition 5.2. *For any finite set $\mathcal{U}$ and input distribution $Q_{UX} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$, the rate $\max_{R_0, R_{11}, \dots, R_{1|\mathcal{U}|}} R_0 + \sum_u Q_U(u) R_{1u}$ resulting from Theorem 5.6 is at least as high as the rate $\max_{R_0, R_1} R_0 + R_1$ resulting from Theorem 5.3.*

Proof. We begin by weakening (5.55) to (5.24). We lower bound the right-hand side of (5.55) by replacing the maximum over $\mathcal{K}$ by the particular choice $\mathcal{K} = \mathcal{U}$, yielding

$$R_0 \leq \min_{\tilde{P}_{UXY} \in \mathcal{T}_0(Q_{UX} \times W)} I_{\tilde{P}}(U; Y) + \left[I_{\tilde{P}}(X; Y|U) - R_1 \right]^+, \quad (5.64)$$

where we have used (5.63) and the definition of conditional mutual information. We can weaken (5.64) to (5.24) using the chain rule for mutual information, and noting that (5.64) is

always satisfied when the minimizing $\tilde{P}_{UXY}$ satisfies $I_{\tilde{P}}(U; Y) > R_0$.

Next, we show that highest value of R_1 permitted by the $|\mathcal{U}|$ conditions in (5.54), denoted by R_1^* , is lower bounded by the right-hand side of (5.23). From (5.63) and (5.54), we have

$$R_1^* = \sum_u Q_U(u) I_{\tilde{P}^*}(X; Y | U = u), \quad (5.65)$$

where $\tilde{P}_{XY|U}^*(\cdot, \cdot | u)$ is the distribution that achieves the minimum in (1.10) under $Q_{X|U}(\cdot | u)$. Defining the joint distribution $\tilde{P}_{UXY}^*$ accordingly with $\tilde{P}_U^* = Q_U$, we can write (5.65) as

$$R_1^* = I_{\tilde{P}^*}(X; Y | U). \quad (5.66)$$

Thus, we can lower bound R_1^* by the right-hand side of (5.23) provided that $\tilde{P}_{UXY}^* \in \mathcal{T}_1(Q_{UX} \times W)$. The constraints $\tilde{P}_{UX}^* = Q_{UX}$ and $\tilde{P}_{UY}^* = P_{UY}$ in (5.14) are satisfied since we have chosen $\tilde{P}_U^* = Q_U$, and since the constraints in (1.10) imply $\tilde{P}_{X|U}^*(\cdot | u) = Q_{X|U}(\cdot | u)$ and $\tilde{P}_{Y|U}^*(\cdot | u) = P_{Y|U}(\cdot | u)$ for all $u \in \mathcal{U}$. The constraint $\mathbb{E}_{\tilde{P}^*}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)]$ is satisfied since, from (1.10), we have $\mathbb{E}_{\tilde{P}^*}[\log q(X, Y) | U = u] \geq \mathbb{E}_P[\log q(X, Y) | U = u]$ for all $u \in \mathcal{U}$. $\square$

Intuitively, one can think of the gain of the refined SC ensemble as being due to a stronger dependence among the codewords. For standard SC, the codewords $\{\mathbf{X}^{(i,j)}\}_{j=1}^{M_1}$ are conditionally independent given $U^{(i)}$, whereas for refined SC this is generally not the case. The additional structure leads to further constraints in the minimizations, and further $[\cdot]^+$ functions in the objective functions, both leading to higher overall rates.

It should be noted, however, that the exponents for standard SC may be higher, particularly at low to moderate rates. In particular, we noted in the proof of Theorem 5.6 that the type-1 and type-2 error events are equivalent to a single-user channel, but the corresponding block lengths are only n_1 and n_2 . Thus, if either $Q_U(1)$ or $Q_U(2)$ is close to zero, the corresponding exponent is small. These observations are analogous to the comparison of explicit time sharing and coded time sharing for the MAC, given in Section 4.2.5.

5.3.3 Numerical Examples

In this subsection, we provide examples comparing the two versions of superposition coding and the LM rate. The computations are done using the optimization software YALMIP [77]. We do not explicitly give values for Lapidoth's rate [35] (see Theorem 4.3), since for each example given, we found it to coincide with the SC rate (see Theorem 5.3). This was done by taking the input distribution for SC, and using it to choose the parameters for the MAC using

similar ideas to those given in Section 5.2.3 (e.g. see (5.31)).

Example 1: Sum Channel

Here we provide an example that can be seen as an analog of Lapidot's parallel channel example [35] (see Section 4.1.2).

Given two channels (W_1, W_2) respectively defined on the alphabets $(\mathcal{X}_1, \mathcal{Y}_1)$ and $(\mathcal{X}_2, \mathcal{Y}_2)$, the sum channel is defined to be the channel $W(y|x)$ with $|\mathcal{X}| = |\mathcal{X}_1| + |\mathcal{X}_2|$ and $|\mathcal{Y}| = |\mathcal{Y}_1| + |\mathcal{Y}_2|$ such that one subchannel is used on each transmission [1] (see also Definition 3.1). One can similarly combine two metrics $q_1(x_1, y_1)$ and $q_2(x_2, y_2)$ to form a sum metric $q(x, y)$. Assuming without loss of generality that $\mathcal{X}_1$ and $\mathcal{X}_2$ are disjoint and $\mathcal{Y}_1$ and $\mathcal{Y}_2$ are disjoint, we have

$$q(x, y) = \begin{cases} q_1(x_1, y_1) & x_1 \in \mathcal{X}_1 \text{ and } y_1 \in \mathcal{Y}_1 \\ q_2(x_2, y_2) & x_2 \in \mathcal{X}_2 \text{ and } y_2 \in \mathcal{Y}_2 \\ 0 & \text{otherwise,} \end{cases} \quad (5.67)$$

and similarly for $W(y|x)$. Let $\hat{Q}_1$ and $\hat{Q}_2$ be the distributions that maximize the LM rate in (1.10) on the respective subchannels. We set $\mathcal{U} = \{1, 2\}$, $Q_{X|U}(\cdot|1) = (\hat{Q}_1, \mathbf{0})$ and $Q_{X|U}(\cdot|2) = (\mathbf{0}, \hat{Q}_2)$, where $\mathbf{0}$ denotes the zero vector. We leave Q_U to be specified.

Combining the constraints $\tilde{P}_{UX} = Q_{UX}$ and $\mathbb{E}_{\tilde{P}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)]$ in (5.52), it is straightforward to show that the minimizing $\tilde{P}_{UXY}$ in (5.55) only takes on non-zero values for (u, x, y) such that (i) $u = 1$, $x \in \mathcal{X}_1$ and $y \in \mathcal{Y}_1$, or (ii) $u = 2$, $x \in \mathcal{X}_2$ and $y \in \mathcal{Y}_2$. It follows that U is a deterministic function of Y under the minimizing $\tilde{P}_{UXY}$, and hence

$$I_{\tilde{P}}(U; Y) = H(Q_U) - H_{\tilde{P}}(U|Y) = H(Q_U). \quad (5.68)$$

Therefore, the right-hand side of (5.55) is lower bounded by $H(Q_U)$. Using (5.53), it follows that we can achieve the rate

$$H(Q_U) + Q_U(1)I_1^{\text{LM}}(\hat{Q}_1) + Q_U(2)I_2^{\text{LM}}(\hat{Q}_2), \quad (5.69)$$

$$= \log(e^{I_1^{\text{LM}}(\hat{Q}_1)} + e^{I_2^{\text{LM}}(\hat{Q}_2)}) \quad (5.70)$$

where I_ν^{LM} is the LM rate for subchannel ν , and (5.70) follows by optimizing Q_U in the same way as [1, Sec. 16], yielding $Q_U(1) = \frac{e^{I_1^{\text{LM}}(\hat{Q}_1)}}{e^{I_1^{\text{LM}}(\hat{Q}_1)} + e^{I_2^{\text{LM}}(\hat{Q}_2)}}$. It should be noted that the LM rate with an optimized input distribution can be strictly less than (5.70) even for simple examples (e.g. binary symmetric subchannels).

Example 2: Zero Undetected Error Capacity

As mentioned in Section 1.1.2, the zero-undetected error capacity is a special case of the mismatched capacity, corresponding to the decoding metric $q(x, y) = \mathbb{1}\{W(y|x) > 0\}$ [8]. We consider an example from [138],¹ where $\mathcal{X} = \mathcal{Y} = \{0, 1, 2\}$, and the channel and metric are described by the entries of the matrices

$$\mathbf{W} = \begin{bmatrix} 0.75 & 0.25 & 0 \\ 0 & 0.75 & 0.25 \\ 0.25 & 0 & 0.75 \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{bmatrix}, \quad (5.71)$$

where x indexes the rows and y indexes the columns.

Using an exhaustive search to three decimal places, we found the optimized LM rate to be $R_{\text{LM}}^* = 0.599$ bits/use, using the input distribution $Q = (0.449, 0.551, 0)$.² It was stated in [138] that the rate obtained by considering the second-order product of the channel and metric (see Section 4.1.2) is equal to $R_{\text{LM2}}^* = 0.616$ bits/use. Using local optimization techniques, we verified that this rate is achieved with $Q = (0, 0.250, 0, 0.319, 0, 0, 0, 0.181, 0.250)$, where the order of the inputs is $(0, 0), (0, 1), (0, 2), (1, 0), \dots, (2, 2)$.

The global optimization of (5.23)–(5.24) over $\mathcal{U}$ and Q_{UX} is difficult. Setting $|\mathcal{U}| = 2$ and applying local optimization techniques using a number of starting points, we obtained an achievable rate of $R_{\text{sc}}^* = 0.695$ bits/use, with $Q_U = (0.645, 0.355)$, $Q_{X|U}(\cdot|1) = (0.3, 0.7, 0)$ and $Q_{X|U}(\cdot|2) = (0, 0, 1)$. Thus, SC not only yields an improvement over the single-letter LM rate, but also over the two-letter version. We obtained the same rate and input distribution as that of standard SC when applying local optimization techniques to refined SC.

Example 3

Here we consider the channel and decoding metric described by the entries of the matrices

$$\mathbf{W} = \begin{bmatrix} 0.99 & 0.01 & 0 & 0 \\ 0.01 & 0.99 & 0 & 0 \\ 0.1 & 0.1 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.1 & 0.7 \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} 1 & 0.5 & 0 & 0 \\ 0.5 & 1 & 0 & 0 \\ 0.05 & 0.15 & 1 & 0.05 \\ 0.15 & 0.05 & 0.5 & 1 \end{bmatrix}. \quad (5.72)$$

We have intentionally chosen a highly asymmetric channel and metric, since such examples often yield larger gaps between the various achievable rates. Using an exhaustive search to

¹We thank Christoph Bunte for helpful discussions regarding this example.

²The rate of 0.6128 bits/use reported in [138] appears to be erroneous.

Table 5.1 Achievable rates (bits/use) for the mismatched channel in (5.72).

Input Distribution	Refined SC	Standard SC
$Q_{UX}^{(1)}$	1.313	1.060
$Q_{UX}^{(2)}$	1.236	1.236
$Q_{UX}^{(\text{LM})}$	1.111	1.111

three decimal places, we found the optimized LM rate to be $R_{\text{LM}}^* = 1.111$ bits/use, which is achieved by the input distribution $Q_X^* = (0.403, 0.418, 0, 0.179)$.

Setting $|\mathcal{U}| = 2$ and applying local optimization techniques using a number of starting points, we obtained an achievable rate of $R_{\text{rsc}}^* = 1.313$ bits/use, with $Q_U = (0.698, 0.302)$, $Q_{X|U}(\cdot|1) = (0.5, 0.5, 0, 0)$ and $Q_{X|U}(\cdot|u) = (0, 0, 0.528, 0.472)$. We denote the corresponding input distribution by $Q_{UX}^{(1)}$.

Applying similar techniques to the standard superposition coding rate in (5.23)–(5.24), we obtained an achievable rate of $R_{\text{sc}}^* = 1.236$ bits/use, with $Q_U = (0.830, 0.170)$, $Q_{X|U}(\cdot|1) = (0.435, 0.450, 0.115, 0)$ and $Q_{X|U}(\cdot|2) = (0, 0, 0, 1)$. We denote the corresponding input distribution by $Q_{UX}^{(2)}$.

The achievable rates for this example are summarized in Table 5.1, where $Q_{UX}^{(\text{LM})}$ denotes the distribution in which U is deterministic and the X -marginal maximizes the LM rate. While the achievable rate of Theorem 5.6 coincides with that of Theorem 5.3 under $Q_{UX}^{(2)}$, the former is significantly higher under $Q_{UX}^{(1)}$. Both types of superposition coding yield a strict improvement over the LM rate.

Our parameters may not be globally optimal, and thus we cannot conclude from this example that refined SC yields a strict improvement over standard SC (and hence over Lapidoth's rate [35]) after optimizing $\mathcal{U}$ and Q_{UX} . However, as discussed in Section 5.2.3, improvements for a fixed set of random-coding parameters are still of significant interest due to the fact that global optimizations are prohibitively complex in general.

5.3.4 Dual Expressions and More General Alphabets

In this subsection, we present a dual expression for the rate given in Theorem 5.6 in the case that $|\mathcal{U}| = 2$, as well as extending the result to more general alphabets $\mathcal{X}$ and $\mathcal{Y}$. Generalizations to the case that $|\mathcal{U}| > 2$ are possible, but cumbersome. It should be noted that the random-coding ensemble relies on $|\mathcal{U}|$ being finite.

With $\mathcal{U} = \{1, 2\}$, the condition in (5.55) is given by

$$R_0 \leq \min_{\tilde{P}_{U|XY} \in \mathcal{T}_0(Q_{UX} \times W)} I_{\tilde{P}}(U; Y) + \left[\max \left\{ Q_U(1)(I_{\tilde{P}}(X; Y|U = 1) - R_{11}), \right. \right. \\ \left. \left. Q_U(2)(I_{\tilde{P}}(X; Y|U = 2) - R_{12}), I_{\tilde{P}}(X; Y|U) - R_1 \right\} \right]^+, \quad (5.73)$$

where

$$R_1 \triangleq \sum_u Q_U(u) R_{1u}. \quad (5.74)$$

Similarly to Lemma 4.1 for the MAC, the following lemma gives an alternative expression for (5.73) which is more amenable to Lagrange duality techniques.

Lemma 5.1. *The condition in (5.73) holds if and only if at least one of the following hold:*

$$R_0 \leq \min_{\tilde{P}_{U|XY} \in \mathcal{T}_0(Q_{UX} \times W)} I_{\tilde{P}}(U; Y) \\ + \left[\max \left\{ Q_U(1)(I_{\tilde{P}}(X; Y|U = 1) - R_{11}), I_{\tilde{P}}(X; Y|U) - R_1 \right\} \right]^+ \quad (5.75)$$

$$R_0 \leq \min_{\tilde{P}_{U|XY} \in \mathcal{T}_0(Q_{UX} \times W)} I_{\tilde{P}}(U; Y) \\ + \left[\max \left\{ Q_U(2)(I_{\tilde{P}}(X; Y|U = 2) - R_{12}), I_{\tilde{P}}(X; Y|U) - R_1 \right\} \right]^+. \quad (5.76)$$

Proof. This is a special case of Lemma E.3 with the following identifications:

$$f(z) = I_{\tilde{P}}(U; Y) \quad (5.77)$$

$$g(z) = I_{\tilde{P}}(X; Y|U) - R_1 \quad (5.78)$$

$$g_1(z) = Q_U(1)(I_{\tilde{P}}(X; Y|U = 1) - R_{11}) \quad (5.79)$$

$$g_2(z) = Q_U(2)(I_{\tilde{P}}(X; Y|U = 2) - R_{12}). \quad (5.80)$$

In this case, the condition in (E.12) holds with equality. $\square$

Since the right-hand side of (5.54) is the LM rate, we can make use of the dual expression in (1.9). The following theorem gives the dual expressions for (5.75)–(5.76).

Theorem 5.7. *The conditions in (5.75)–(5.76) can respectively be written as*

$$R_0 \leq \sup_{s \geq 0, \rho_1 \in [0,1], \rho_2 \in [0,1], a(\cdot, \cdot)} \mathbb{E} \left[\log \frac{(q(X, Y)^{s'(u)} e^{a(U, X)})^{\rho'(U)}}{\mathbb{E} \left[\left(\mathbb{E} [q(\bar{X}, Y)^{s'(\bar{U})} e^{a(\bar{U}, \bar{X})} \mid \bar{U}] \right)^{\rho'(\bar{U})} \mid Y \right]} \right] - \sum_{u=1,2} \rho'(u) Q_U(u) R_{1u} \quad (5.81)$$

$$R_0 \leq \sup_{s \geq 0, \rho_1 \in [0,1], \rho_2 \in [0,1], a(\cdot, \cdot)} \mathbb{E} \left[\log \frac{(q(X, Y)^{s''(u)} e^{a(U, X)})^{\rho''(U)}}{\mathbb{E} \left[\left(\mathbb{E} [q(\bar{X}, Y)^{s''(\bar{U})} e^{a(\bar{U}, \bar{X})} \mid \bar{U}] \right)^{\rho''(\bar{U})} \mid Y \right]} \right] - \sum_{u=1,2} \rho''(u) Q_U(u) R_{1u}, \quad (5.82)$$

where

$$\rho'(1) = \rho_1, \rho'(2) = \rho_1 \rho_2, s'(1) = \rho_2 s, s'(2) = s \quad (5.83)$$

$$\rho''(1) = \rho_1 \rho_2, \rho''(2) = \rho_2, s''(1) = s, s''(2) = \rho_1 s \quad (5.84)$$

and $(U, X, Y, \bar{U}, \bar{X}) \sim Q_{UX}(u, x)W(y|x)Q_{UX}(\bar{u}, \bar{x})$.

Proof. See Section 5.5.1. □

The conditions in (5.81)–(5.82) bear a strong resemblance to the standard SC condition in (5.27). In particular, the latter can be recovered by setting $\rho_2 = 1$ in (5.81) or $\rho_1 = 1$ in (5.82), thus providing an alternative proof of Proposition 5.2 for the case that $|\mathcal{U}| = 2$.

In order to extend Theorem 5.7 to more general alphabets, we again use cost-constrained coding. We allow for conditional distributions $Q_{X|U}$ and channels W that are discrete transition laws, conditional density functions, or a mixture of the two. We consider the ensemble given in (5.46), with P_{X_u} given by

$$P_{X_u}(\mathbf{x}_u) = \frac{1}{\mu_{u, n_u}} \prod_{i=1}^{n_u} Q_{X|U}(x_{u,i} | u_i) \mathbb{1}\{\mathbf{x}_u \in \mathcal{D}_{u, n_u}\}, \quad (5.85)$$

where

$$\mathcal{D}_{u, n_u} \triangleq \left\{ \mathbf{x}_u : \left| \frac{1}{n_u} \sum_{i=1}^{n_u} a_{u,l}(x_{u,i}) - \phi_{u,l} \right| \leq \frac{\delta}{n_u}, l = 1, \dots, L_u \right\} \quad (5.86)$$

$$\phi_{u,l} \triangleq \mathbb{E}_Q[a_{u,l}(X) \mid U = u], \quad (5.87)$$

and where μ_{u, n_u} , $\{a_{u,l}\}$ and δ are defined analogously to Section 2.3, and n_u is given in (5.45).

Theorem 5.8. *For any mismatched memoryless channel (possibly having infinite or continuous alphabets) and input distribution $Q_{UX} \in \mathcal{P}(\mathcal{U} \times \mathcal{X})$ with $\mathcal{U} = 1, 2$, the rate $R = R_0 + \sum_{u=1,2} Q_U(u)R_{1u}$ is achievable for any triplet (R_0, R_{11}, R_{12}) satisfying (5.54) (with I_{LM} defined in (1.9)) and at least one of (5.81)–(5.82). The supremum in (1.9) is subject to $\mathbb{E}_Q[a(X)^2] < \infty$, and those in (5.81)–(5.82) are subject to $\mathbb{E}_Q[a(U, X)^2] < \infty$. Furthermore, the rate is achievable using cost-constrained coding in (5.85) with $L_1 = L_2 = 2$.*

Proof. The direct derivation of (5.81)–(5.82) is given in Section 5.5.2. The choice $L_1 = L_2 = 2$ suffices since for $u = 1, 2$, one cost is required for (5.54) and another for (5.81)–(5.82). It suffices to let the cost functions for (5.81)–(5.82) coincide, since the theorem only requires that one of the two hold. $\square$

5.4 Discussion: Primal and Dual Connections

Throughout this chapter and Chapter 4, we considered several variations of constant-composition random coding and cost-constrained random coding. Furthermore, we have given sufficient conditions on the number of costs required for the cost-constrained ensemble to match the performance of the constant-composition ensemble. From Lemma E.1, we see that each dual optimization variable (e.g. $a_1(\cdot)$) corresponds to a primal constraint on a marginal (e.g. $P_{X_1} = Q_1$). Since the dual variables also correspond to cost functions, we can use this observation to relate the number of costs with the number of marginal constraints. For example, consider the exponents for the mismatched MAC (Theorems 4.2 and 4.4) and those for ML decoding (Theorems 4.8 and 4.9). In the primal domain, ML decoding yields two fewer constraints on the X_ν marginals for each $\nu = 1, 2$, due to the absence of the sets $\mathcal{T}_\nu$ for $\nu = 1, 2, 12$ (see (4.23)–(4.25)). Analogously, in the dual domain, the number of cost functions per user is reduced by two.

5.5 Proofs

5.5.1 Proof of Theorem 5.7

We focus on showing the equivalence of (5.75) and (5.81), since identical arguments apply to (5.76) and (5.82). The primal expression is written in terms of a minimization over $\tilde{P}_{UXY}$. It is convenient to split this distribution into three distributions: $\tilde{P}_{UY}$, $\hat{P}_{XY} \triangleq \tilde{P}_{XY|U}(\cdot, \cdot|1)$ and $\hat{\hat{P}}_{XY} \triangleq \tilde{P}_{XY|U}(\cdot, \cdot|2)$. Using an identical argument to the start of Section 4.5.1, we can write

the right-hand side of (5.75) as

$$\sup_{\rho_1 \in [0,1], \rho_2 \in [0,1]} \min_{\tilde{P}_{UY}, \hat{P}_{XY}, \hat{\hat{P}}_{XY}} I_{\tilde{P}}(U; Y) + \rho_1 Q_U(1) I_{\hat{P}}(X; Y) + \rho_1 \rho_2 Q_U(2) I_{\hat{\hat{P}}}(X; Y) - \rho_1 R_{11} - \rho_1 \rho_2 R_{12}. \quad (5.88)$$

Defining $P_{UXY} \triangleq Q_{UX} \times W$, the minimization is subject to the constraints (i) $\tilde{P}_U = Q_U$, (ii) $\hat{P}_X = Q_{X|U}(\cdot|1)$, (iii) $\hat{\hat{P}}_X = Q_{X|U}(\cdot|2)$, (iv) $\tilde{P}_Y = P_Y$, (v) $\hat{P}_Y = \tilde{P}_{Y|U}(\cdot|1)$, (vi) $\hat{\hat{P}}_Y = \tilde{P}_{Y|U}(\cdot|2)$, (vii) $Q_U(1)\mathbb{E}_{\hat{P}}[\log q(X, Y)] + Q_U(2)\mathbb{E}_{\hat{\hat{P}}}[\log q(X, Y)] \geq \mathbb{E}_P[\log q(X, Y)]$.

Similarly to Section 4.5.1, we apply the minimization in several steps, making repeated use of Lemma E.1. We implicitly apply Fan's minimax theorem [78] after each step, so that the supremum over the dual variables can be left until the end. We provide less detail than the amount given in Section 4.5.1, since the general steps are similar.

Step 1

For given joint distributions $\tilde{P}_{UY}$ and $\hat{P}_{XY}$, the minimization $\min_{\hat{\hat{P}}_{XY}} I_{\hat{\hat{P}}}(X; Y)$ subject to the constraints (iii), (vi) and (vii) has a dual expression given by

$$F_1 \triangleq -F_{1,1} + F_{1,2} + F_{1,3} - sQ_U(1)F_{1,4}, \quad (5.89)$$

where

$$F_{1,1} \triangleq \sum_y \tilde{P}_{Y|U}(y|2) \log \sum_{\bar{x}} Q_{X|U}(\bar{x}|2) q(\bar{x}, y)^{sQ_U(2)} e^{a_2(\bar{x})} \quad (5.90)$$

$$F_{1,2} \triangleq \sum_{x_2} Q_{X|U}(x|2) a_2(x) \quad (5.91)$$

$$F_{1,3} \triangleq s \sum_{x,y} P_{XY}(x, y) \log q(x, y) \quad (5.92)$$

$$F_{1,4} \triangleq \sum_{x,y} \hat{P}_{XY}(x, y) \log q(x, y), \quad (5.93)$$

and where $s \geq 0$ and $a_2(\cdot)$ are dual variables.

Step 2

For a given joint distribution $\tilde{P}_{UY}$, the minimization $\min_{\hat{P}_{XY}} I_{\hat{P}}(X; Y) - s\rho_2 Q_U(2)F_{1,4}(\hat{P}_{XY})$ subject to (ii) and (v) has a dual expression given by

$$F_2 \triangleq -F_{2,1} + F_{2,2}, \quad (5.94)$$

where

$$F_{2,1} \triangleq \sum_y \tilde{P}_{Y|U}(y|1) \log \sum_{\bar{x}} Q_{X|U}(\bar{x}|1) q(\bar{x}, y)^{s\rho_2 Q_U(2)} e^{a_1(\bar{x})} \quad (5.95)$$

$$F_{2,2} \triangleq \sum_x Q_{X|U}(x|1) a_1(x), \quad (5.96)$$

and where $a_1(\cdot)$ is a dual variable.

Step 3

Next, we consider the minimization $\min_{\tilde{P}_{UY}} I_{\tilde{P}}(U; Y) - \rho_1 Q_U(1)F_{2,1} - \rho_1 \rho_2 Q_U(2)F_{1,1}$ subject to (i) and (iv). The objective can equivalently be expressed as

$$F_3 \triangleq I_{\tilde{P}}(U; Y) - \sum_u \rho'(u) \sum_y \tilde{P}_{UY}(u, y) \log \sum_{\bar{x}} Q_{X|U}(\bar{x}|u) q(\bar{x}, y)^{s'(u)} e^{a(u, \bar{x})} \quad (5.97)$$

using the definitions in (5.83) along with $a(u, x) \triangleq a_u(x)$. The dual expression is given by

$$F_3 = - \sum_y P_Y(y) \log \sum_{\bar{u}} Q_U(\bar{u}) \left(\sum_{\bar{x}} Q_{X|U}(\bar{x}|\bar{u}) q(\bar{x}, y)^{s'(\bar{u})} e^{a(\bar{u}, \bar{x})} \right)^{\rho'(u)} e^{b(\bar{u})} + \sum_u Q_U(u) b(u), \quad (5.98)$$

where $b(\cdot)$ is a dual variable.

Step 4

The final objective is given by

$$F_4 \triangleq F_3 + \rho_1 Q_U(1)F_{2,1} + \rho_1 \rho_2 Q_U(2)(F_{1,2} + F_{1,3}) - \sum_{u=1,2} \rho'(u) Q_U(u) R_{1u}. \quad (5.99)$$

Applying some algebraic manipulations, we obtain the dual expression

$$\sum_{u,x,y} P_{UY}(u,x,y) \log \frac{\left(q(x,y)^{s'(u)} e^{a(u,x)}\right)^{\rho'(u)} e^{b(u)}}{\sum_{\bar{u}} Q_U(\bar{u}) \left(\sum_{\bar{x}} Q_{X|U}(\bar{x}|\bar{u}) q(\bar{x},y)^{s'(\bar{u})} e^{a(\bar{u},\bar{x})}\right)^{\rho'(\bar{u})} e^{b(\bar{u})}} - \sum_{u=1,2} \rho'(u) Q_U(u) R_{1u}. \quad (5.100)$$

To conclude the proof, we must show that the variable $b(u)$ can be removed from the numerator and denominator in (5.100) without affecting the final dual optimization. For $\rho_1 > 0$ and $\rho_2 > 0$, this follows by simply factoring $b(u)$ into $a(u,x)$. Using the identity $\mathbb{E}[e^{b(U)}] \geq e^{\mathbb{E}[b(U)]}$ (by Jensen's inequality), it is readily verified that the optimal value of the objective is zero when $\rho_1 = 0$, regardless of whether $b(u)$ is present. For the remaining case, namely $\rho_1 > 0$ and $\rho_2 = 0$, the objective depends on $a(u,x)$ only for $u = 1$. Moreover, since (5.100) depends on $b(\cdot)$ only through the difference $b(2) - b(1)$, we may set $b(2) = 0$ without loss of generality. The remaining parameter $b(1)$ can be factored into $a(1,x)$.

5.5.2 Proof of Theorem 5.8

We focus on the derivation of (5.81), since (5.82) can be derived similarly. The ideas used here are similar to those in the proof of Theorem 4.6, but the details are more involved.

Applying Lemma D.1 to (5.57) with $Z_1(i) = \mathbf{X}_1^{(1,i)}$ and $Z_2(j) = \mathbf{X}_2^{(1,j)}$, we obtain

$$\bar{p}_{e,0} \leq \mathbb{E} \left[\min \left\{ 1, (M_0 - 1) \mathbb{E} \left[\min \left\{ 1, M_{11} \mathbb{E} \left[\min \left\{ 1, M_{12} \mathbb{P} \left[\frac{q^n(\bar{\mathbf{X}}, \mathbf{Y})}{q^n(\mathbf{X}, \mathbf{Y})} \geq 1 \mid \bar{\mathbf{X}}_1 \right] \right\} \mid \bar{\mathbf{U}} \right] \right\} \mid \mathbf{U}, \mathbf{X}, \mathbf{Y} \right] \right\} \right]. \quad (5.101)$$

Using (5.49), Markov's inequality, and $\min\{1, \zeta\} \leq \zeta^\rho$ ($\rho \in [0, 1]$), we obtain³

$$\begin{aligned} \bar{p}_{e,0} &\leq (M_0 M_{11}^{\rho_1} M_{12}^{\rho_1 \rho_2})^{\rho_0} \sum_{\mathbf{u}, \mathbf{x}_1, \mathbf{x}_2} P_U(\mathbf{u}) P_{\mathbf{X}_1}(\mathbf{x}_1) P_{\mathbf{X}_2}(\mathbf{x}_2) \sum_{\mathbf{y}} W^n(\mathbf{y} | \Xi(\mathbf{u}, \mathbf{x}_1, \mathbf{x}_2)) \\ &\quad \times \left(\sum_{\bar{\mathbf{u}}} P_U(\bar{\mathbf{u}}) \times \left(\sum_{\bar{\mathbf{x}}_1} P_{\mathbf{X}_1}(\bar{\mathbf{x}}_1) \left(\frac{q^{n_1}(\bar{\mathbf{x}}_1, \mathbf{y}_1(\bar{\mathbf{u}}))}{q^{n_1}(\mathbf{x}_1, \mathbf{y}_1(\mathbf{u}))} \right)^{\rho_2 s} \right)^{\rho_1} \right. \\ &\quad \times \left. \left(\sum_{\bar{\mathbf{x}}_2} P_{\mathbf{X}_2}(\bar{\mathbf{x}}_2) \left(\frac{q^{n_2}(\bar{\mathbf{x}}_2, \mathbf{y}_2(\bar{\mathbf{u}}))}{q^{n_2}(\mathbf{x}_2, \mathbf{y}_2(\mathbf{u}))} \right)^s \right)^{\rho_1 \rho_2} \right)^{\rho_0}, \end{aligned} \quad (5.102)$$

³In the case of continuous alphabets, the summations should be replaced by integrals as necessary.

where $s \geq 0$ and $\rho_1, \rho_2 \in [0, 1]$ are arbitrary. Using the definition of the cost-constrained ensemble in (5.85)–(5.87), we obtain

$$\begin{aligned} \bar{p}_{e,0} &\leq (M_0 M_{11}^{\rho_1} M_{12}^{\rho_1 \rho_2})^{\rho_0} \sum_{\mathbf{u}, \mathbf{x}_1, \mathbf{x}_2} P_U(\mathbf{u}) P_{\mathbf{X}_1}(\mathbf{x}_1) P_{\mathbf{X}_2}(\mathbf{x}_2) \sum_{\mathbf{y}} W^n(\mathbf{y} | \Xi(\mathbf{u}, \mathbf{x}_1, \mathbf{x}_2)) \\ &\quad \times \left(\sum_{\bar{\mathbf{u}}} P_U(\bar{\mathbf{u}}) \left(\sum_{\bar{\mathbf{x}}_1} P_{\mathbf{X}_1}(\bar{\mathbf{x}}_1) \left(\frac{q^{n_1}(\bar{\mathbf{x}}_1, \mathbf{y}_1(\bar{\mathbf{u}}))}{q^{n_1}(\mathbf{x}_1, \mathbf{y}_1(\mathbf{u}))} \right)^{\rho_2 s} \frac{e^{a_1^{n_1}(\bar{\mathbf{x}}_1)}}{e^{a_1^{n_1}(\mathbf{x}_1)}} \right)^{\rho_1} \right. \\ &\quad \times \left. \left(\sum_{\bar{\mathbf{x}}_2} P_{\mathbf{X}_2}(\bar{\mathbf{x}}_2) \left(\frac{q^{n_2}(\bar{\mathbf{x}}_2, \mathbf{y}_2(\bar{\mathbf{u}}))}{q^{n_2}(\mathbf{x}_2, \mathbf{y}_2(\mathbf{u}))} \right)^s \frac{e^{a_2^{n_2}(\bar{\mathbf{x}}_2)}}{e^{a_2^{n_2}(\mathbf{x}_2)}} \right)^{\rho_1 \rho_2} \right)^{\rho_0}, \end{aligned} \quad (5.103)$$

where for $u = 1, 2$, $a_u(\cdot)$ is one of the $L_u = 2$ cost functions in (5.86), and $a_u^{n_u}(\mathbf{x}_u) \triangleq \sum_{i=1}^{n_u} a_u(x_{u,i})$. For each $(\mathbf{u}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{y})$, we write the argument to the summation over $\mathbf{y}$ in (5.103) as a product of two terms, namely

$$\begin{aligned} \mathsf{T}_1 &\triangleq W^n(\mathbf{y} | \Xi(\mathbf{u}, \mathbf{x}_1, \mathbf{x}_2)) \\ &\quad \times \left(q^{n_1}(\mathbf{x}_1, \mathbf{y}_1(\mathbf{u}))^{-\rho_1 \rho_2 s} e^{-\rho_1 a_1^{n_1}(\mathbf{x}_1)} q^{n_2}(\mathbf{x}_2, \mathbf{y}_2(\mathbf{u}))^{-\rho_1 \rho_2 s} e^{-\rho_1 \rho_2 a_2^{n_2}(\mathbf{x}_2)} \right)^{\rho_0} \end{aligned} \quad (5.104)$$

$$\begin{aligned} \mathsf{T}_2 &\triangleq \left(\sum_{\bar{\mathbf{u}}} P_U(\bar{\mathbf{u}}) \left(\sum_{\bar{\mathbf{x}}_1} P_{\mathbf{X}_1}(\bar{\mathbf{x}}_1) q^{n_1}(\bar{\mathbf{x}}_1, \mathbf{y}_1(\bar{\mathbf{u}}))^{\rho_2 s} e^{a_1^{n_1}(\bar{\mathbf{x}}_1)} \right)^{\rho_1} \right. \\ &\quad \times \left. \left(\sum_{\bar{\mathbf{x}}_2} P_{\mathbf{X}_2}(\bar{\mathbf{x}}_2) q^{n_2}(\bar{\mathbf{x}}_2, \mathbf{y}_2(\bar{\mathbf{u}}))^s e^{a_2^{n_2}(\bar{\mathbf{x}}_2)} \right)^{\rho_1 \rho_2} \right)^{\rho_0}. \end{aligned} \quad (5.105)$$

Since $P_{\mathbf{X}_u}(\mathbf{x}_u)$ is upper bounded by a subexponential prefactor times $\prod_{i=1}^n Q_{X|U}(x_{u,i} | u)$ for $u = 1, 2$ (see Proposition 2.1), we have

$$\sum_{\bar{\mathbf{x}}_1} P_{\mathbf{X}_1}(\bar{\mathbf{x}}_1) q^{n_1}(\bar{\mathbf{x}}_1, \mathbf{y}_1(\bar{\mathbf{u}}))^{\rho_2 s} e^{a_1^{n_1}(\bar{\mathbf{x}}_1)} \leq \prod_{i=1}^{n_1} \sum_{\bar{x}_1} Q_{X|U}(\bar{x}_1 | 1) q(\bar{x}_1, y_{1,i}(\bar{\mathbf{u}}))^{\rho_2 s} e^{a_1(\bar{x}_1)} \quad (5.106)$$

$$\sum_{\bar{\mathbf{x}}_2} P_{\mathbf{X}_2}(\bar{\mathbf{x}}_2) q^{n_2}(\bar{\mathbf{x}}_2, \mathbf{y}_2(\bar{\mathbf{u}}))^s e^{a_2^{n_2}(\bar{\mathbf{x}}_2)} \leq \prod_{i=1}^{n_2} \sum_{\bar{x}_2} Q_{X|U}(\bar{x}_2 | 2) q(\bar{x}_2, y_{2,i}(\bar{\mathbf{u}}))^s e^{a_2(\bar{x}_2)}, \quad (5.107)$$

where for $u = 1, 2$, $y_{u,i}(\bar{\mathbf{u}})$ is the i -th entry of $\mathbf{y}_u(\bar{\mathbf{u}})$. Using the definitions in (5.83) along with

$a(u, x) \triangleq a_u(x)$, we thus obtain

$$\left(\sum_{\bar{x}_1} P_{X_1}(\bar{x}_1) q^{n_1}(\bar{x}_1, \mathbf{y}_1(\bar{\mathbf{u}}))^{\rho_2 s} e^{a_1^{n_1}(\bar{x}_1)} \right)^{\rho_1} \left(\sum_{\bar{x}_2} P_{X_2}(\bar{x}_2) q^{n_2}(\bar{x}_2, \mathbf{y}_2(\bar{\mathbf{u}}))^s e^{a_2^{n_2}(\bar{x}_2)} \right)^{\rho_1 \rho_2} \quad (5.108)$$

$$\leq \left(\prod_{i=1}^{n_1} \sum_{\bar{x}_1} Q_{X|U}(\bar{x}_1|1) q(\bar{x}_1, y_{1,i}(\bar{\mathbf{u}}))^{\rho_2 s} e^{a_1(\bar{x}_1)} \right)^{\rho_1} \times \left(\prod_{i=1}^{n_2} \sum_{\bar{x}_2} Q_{X|U}(\bar{x}_2|2) q(\bar{x}_2, y_{2,i}(\bar{\mathbf{u}}))^s e^{a_2(\bar{x}_2)} \right)^{\rho_1 \rho_2} \quad (5.109)$$

$$= \prod_{i=1}^n \left(\sum_{\bar{x}} Q_{X|U}(\bar{x}|\bar{u}_i) q(\bar{x}, y_i)^{s'(\bar{u}_i)} e^{a(\bar{u}_i, \bar{x})} \right)^{\rho'(\bar{u}_i)}. \quad (5.110)$$

Hence, and using the fact that $P_U(\mathbf{u}) \leq Q_U^n(\mathbf{u})$ (see (A.14)), we obtain

$$\tau_2 \leq \prod_{i=1}^n \left(\sum_{\bar{u}} Q_U(\bar{u}) \left(\sum_{\bar{x}} Q_{X|U}(\bar{x}|\bar{u}) q(\bar{x}, y_i)^{s'(\bar{u})} e^{a(\bar{u}, \bar{x})} \right)^{\rho'(\bar{u})} \right)^{\rho_0}. \quad (5.111)$$

A similar argument (without the need for the $\leq$ steps) gives

$$\tau_1 = \prod_{i=1}^n W(y_i|x_i) \left(q(x_i, y_i)^{-\rho'(u_i) s'(u_i)} e^{-\rho'(u_i) a(u_i, x_i)} \right)^{\rho_0}, \quad (5.112)$$

where we have used the fact that $W^n(\mathbf{y}|\Xi(\mathbf{u}, \mathbf{x}_1, \mathbf{x}_2)) = W^{n_1}(\mathbf{y}_1(\mathbf{u})|\mathbf{x}_1) W^{n_2}(\mathbf{y}_2(\mathbf{u})|\mathbf{x}_2)$. Substituting (5.111) and (5.112) into (5.103), we obtain

$$\bar{p}_{e,0} \leq (M_0 M_{11}^{\rho_1} M_{12}^{\rho_1 \rho_2})^{\rho_0} \sum_{\mathbf{u}, \mathbf{x}} P_{UX}(\mathbf{u}, \mathbf{x}) \times \prod_{i=1}^n \sum_y W(y|x_i) \left(\sum_{\bar{u}} Q_U(\bar{u}) \left(\sum_{\bar{x}} Q_{X|U}(\bar{x}|\bar{u}) \left(\frac{q(\bar{x}, y_i)}{q(x_i, y_i)} \right)^{s'(\bar{u}_i)} \frac{e^{a(\bar{u}_i, \bar{x})}}{e^{a(u_i, x_i)}} \right)^{\rho'(\bar{u}_i)} \right)^{\rho_0}, \quad (5.113)$$

where

$$P_{UX}(\mathbf{u}, \mathbf{x}) \triangleq \sum_{\mathbf{x}_1, \mathbf{x}_2} P_U(\mathbf{u}) P_{X_1}(\mathbf{x}_1) P_{X_2}(\mathbf{x}_2) \mathbb{1}\{\mathbf{x} = \Xi(\mathbf{u}, \mathbf{x}_1, \mathbf{x}_2)\}. \quad (5.114)$$

If P_{UX} were i.i.d. on Q_{UX} , then (5.113) would yield an error exponent which is positive when (5.81) holds with strict inequality, by taking $\rho_0 \rightarrow 0$ similarly to Theorem 4.5. The same can be done in the present setting by upper bounding P_{UX} by a subexponential prefactor times

Q_{UX}^n , analogously to (5.106)–(5.107). More precisely, we have

$$P_{UX}(\mathbf{u}, \mathbf{x}) \stackrel{\cdot}{\leq} \sum_{\mathbf{x}_1, \mathbf{x}_2} P_U(\mathbf{u}) \left(\prod_{i=1}^n Q_{X|U}(x_{1,i}|1) \right) \left(\prod_{i=1}^n Q_{X|U}(x_{2,i}|2) \right) \mathbb{1}\{\mathbf{x} = \Xi(\mathbf{u}, \mathbf{x}_1, \mathbf{x}_2)\} \quad (5.115)$$

$$= P_U(\mathbf{u}) Q_{X|U}^n(\mathbf{x}|\mathbf{u}) \quad (5.116)$$

$$\stackrel{\cdot}{\leq} Q_U^n(\mathbf{u}) Q_{X|U}^n(\mathbf{x}|\mathbf{u}) \quad (5.117)$$

$$= Q_{UX}^n(\mathbf{u}, \mathbf{x}). \quad (5.118)$$

Chapter 6

Conclusions and Future Work

As we discussed in Chapter 1, the problem of channel coding with a mismatched decoding rule is relevant to a diverse range of practical communication settings, and has interesting connections with theoretical problems such as zero-error communication. We have presented a variety of information-theoretic performance bounds for the mismatched decoding problem, with a particular focus on the asymptotics of single-user and multi-user random-coding ensembles. A recurring theme, both in this thesis and in the existing literature, is that the performance is improved by certain kinds of structure: Constant-composition and cost-constrained codes outperform i.i.d. codes, and multi-user coding techniques provide benefits over standard single-user coding techniques.

We proceed by presenting an overview of the main contributions of each chapter, and discussing some possibilities for further research.

Motivated by the fact that most of the literature has focused on achievable rates, we presented a more comprehensive study of the random-coding performance in Chapters 2 and 3. In Chapter 2, we introduced a cost-constrained random ensemble with multiple auxiliary costs, and used it to derive achievable asymptotic bounds in the regimes of error exponents (both expurgated and non-expurgated), second-order coding rates, and moderate deviations. This ensemble serves as an attractive alternative to the constant-composition ensemble for improving the performance compared to i.i.d. coding, since it is directly applicable to channels with infinite and continuous alphabets. We showed that two auxiliary costs suffice to obtain the given asymptotic expansions in each of the preceding regimes, and that fewer may suffice if the input distribution or decoding rule is optimized.

We proved that our random-coding error exponents are tight with respect to the ensemble average for any mismatched DMC; determining whether the same holds true for the second-order rates and moderate deviations results remains an open problem. Moreover, even in

the fixed-rate regime, it remains open as to whether the exponents are ensemble-tight for continuous channels. Such an investigation may require more advanced techniques from large deviations theory [139].

In Chapter 3, we provided refined asymptotic results for both the i.i.d. and constant-composition ensembles. For the former, we derived saddlepoint approximations [65] of the random-coding bounds rcu and rcu_s (see Theorem 2.1) that unify the fixed-rate and varying-rate regimes, including several higher-order terms. For the constant-composition ensemble, we focused separately on the fixed-rate prefactor and the fixed-error third-order term, and we provided significant refinements to standard techniques that introduce polynomials depending on the alphabet sizes. In the special case of maximum-likelihood decoding, our bounds were seen to be tight in several cases of interest. In particular, a loss for sum channels [1] was shown to be unavoidable in general.

The technical assumptions considered for i.i.d. and constant-composition coding were related yet different, and it would be of interest to determine whether the cost-constrained ensemble can unify or generalize the two. Our saddlepoint approximation of rcu in Theorem 3.2 was proved in two steps, and only the second was proved to be asymptotically tight. The result would be more powerful still if both steps were proved to be tight; alternatively, a different analysis may yield an asymptotically tight bound of a different form. Saddlepoint approximations for the constant-composition and cost-constrained ensembles would be highly valuable, since these yield higher achievable rates and exponents in general when the decoding rule is suboptimal. Generalizations to continuous channels may also be possible; we believe that the changes would be minimal for the approximation of rcu_s in Section 3.2.2, whereas the analysis of rcu in Section 3.2.3 would require alternative methods that avoid the use of (or use an alternative notion of) types. While we focused on random-coding bounds, similar approximations for non-asymptotic converse bounds in the matched setting (e.g. the meta-converse [61]) would be of great interest. The analysis of [79, 83] suggests that symmetric channels may be significantly easier to handle compared to the general case.

In Chapters 4 and 5, we presented a framework for deriving achievable rates and error exponents for multi-user random-coding ensembles in the presence of mismatch. We derived ensemble-tight error exponents for the mismatched MAC that recover Lapidath's achievable rate region [35], and we proved the ensemble tightness of the exponent of Liu-Hughes [66] for maximum-likelihood decoding. In the point-to-point setting, we studied two types of superposition coding, and showed that a refined version can yield improved rates compared to the standard version, which in turn is at least as high as Lapidath's expurgated parallel coding rate [35]. For each ensemble, alternative expressions of the rates were presented using Lagrange

duality, and cost-constrained random coding was again used to prove their validity for infinite and continuous alphabets.

The multi-user coding techniques that we considered are far from exhaustive, and it is therefore of interest to determine whether the performance can be improved further by considering other coding techniques. We observed that the ensembles yielding the highest rates (e.g. refined SC) can yield worse error exponents and low to moderate rates, thus raising the question of whether there is a single ensemble that achieves the best of both worlds. Moreover, while we demonstrated that refined superposition coding can improve on the standard version for a given input distribution, it is yet to be determined whether the same is true after optimizing the input distribution. A limitation of refined SC is that it relies on the auxiliary alphabet $\mathcal{U}$ being finite; another direction for future research is to seek a variant that can be applied with more general auxiliary alphabets.

In Section 4.3, we presented an achievable rate region for a multi-letter successive decoding rule, and showed that performance gains over the corresponding maximum-metric rule are possible. It would be of interest to quantify the performance gain analytically; for example, by presenting conditions under which a strict improvement is observed. We are also yet to determine whether the region that we presented is ensemble-tight.

In Section 4.4, we presented an inner bound on the set of locally achievable second-order coding rates for the matched MAC. We used a variant of coded time-sharing to obtain non-standard tangent vector terms that enlarge the inner bound, and we showed that constant-composition random coding can provide improved bounds compared to i.i.d. random coding. A direction for further research is to extend the result to continuous channels via cost-constrained random coding; the main hurdle in doing so appears to be finding a suitable alternative to the combinatorial central limit theorem. Perhaps more challenging still is the problem of obtaining second-order asymptotic converse results. While the strong converse has been proved (e.g. see [140]), it appears that significant refinements are needed in order to obtain a non-trivial outer bound on the second-order rate region.

The definition of the mismatched capacity that we considered is the “classical” definition, and has thus far received the most attention in the literature. As discussed by Merhav *et al.* [9, Sec. VII] and Alsan-Telatar [50], there are several variations that lead to rates exceeding the classical mismatched capacity. Further studies of these variations and their consequences may help in understanding the effect of mismatch in a wider variety of settings. Moreover, there are several related problems in information theory where it may be of interest to study the effect of mismatch, including joint source-channel coding [115, Sec. 3.9], channel coding with feedback [115, Ch. 17], and variable-rate channel coding [141]. Once again, there are

several ways to formulate the mismatched versions of these problems, and care should be taken in deciding which of them are of the most interest and relevance.

Finally, a highly challenging open problem is to obtain upper bounds on the mismatched capacity. While a multi-letter characterization has recently been given [39], little is known in the way of single-letter converse results. As discussed in Section 1.1.4, a computable characterization of the mismatched capacity may be formidable in general due to the fact that it yields the zero-error capacity as a special case. However, the zero-error capacity is indeed known for input alphabet sizes up to six [142], so it is conceivable that the mismatched capacity could be found in such cases. Moreover, even if a complete characterization cannot be found, it is of significant interest to obtain single-letter converse results that improve on the trivial converse, i.e. the matched capacity. Another interesting open problem that was raised in [8], and that is yet to be solved, is whether the strong converse holds (i.e. the error probability tends to one for any rate above the mismatched capacity).

Appendix A

The Method of Types

In this appendix, we give an overview of the method of types [27, Ch. 2], and present the properties that are used throughout the thesis. We do not provide proofs, since we only present well-known properties of simple variations thereof (e.g. see [27, Ch. 2], [56], [59]).

Let $\mathcal{X}$ be a finite alphabet. The type of a sequence $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$ is its empirical distribution, defined by

$$\hat{P}_{\mathbf{x}}(x) \triangleq \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{x_i = x\}. \quad (\text{A.1})$$

The set of all types on $\mathcal{X}^n$ is denoted by $\mathcal{P}_n(\mathcal{X})$. For a given $P_X \in \mathcal{P}_n(\mathcal{X})$, the type class $T^n(P_X)$ is the set of all sequences in $\mathcal{X}^n$ with type P_X ,

$$T^n(P_X) \triangleq \{\mathbf{x} \in \mathcal{X}^n : \hat{P}_{\mathbf{x}} = P_X\}. \quad (\text{A.2})$$

For two finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, the quantities $\hat{P}_{\mathbf{xy}}(x, y)$, $\mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ and $T^n(P_{XY})$ are defined analogously. In this case, the type is also referred to as a joint type.

The conditional type of $\mathbf{y}$ given $\mathbf{x}$ is defined by

$$\hat{P}_{\mathbf{y}|\mathbf{x}}(y|x) \triangleq \begin{cases} \frac{\hat{P}_{\mathbf{xy}}(x, y)}{\hat{P}_{\mathbf{x}}(x)} & \hat{P}_{\mathbf{x}}(x) > 0 \\ \frac{1}{|\mathcal{Y}|} & \text{otherwise.} \end{cases} \quad (\text{A.3})$$

The set of all conditional types on $\mathcal{Y}^n$ given $\mathcal{X}^n$ is denoted by $\mathcal{P}_n(\mathcal{Y}|\mathcal{X})$. For any $P_{XY} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ and sequence $\mathbf{x}$, the conditional type class $T_{\mathbf{x}}^n(P_{XY})$ is the set of all sequences $\mathbf{y}$ in $\mathcal{Y}^n$ such that $(\mathbf{x}, \mathbf{y})$ has joint type P_{XY} ,

$$T_{\mathbf{x}}^n(P_{XY}) \triangleq \{\mathbf{y} \in \mathcal{Y}^n : \hat{P}_{\mathbf{xy}} = P_{XY}\}. \quad (\text{A.4})$$

Similarly, given $\mathbf{x} \in T^n(P_X)$, we define

$$T_{\mathbf{x}}^n(P_{Y|X}) \triangleq \{\mathbf{y} \in \mathcal{Y}^n : \hat{P}_{\mathbf{x}\mathbf{y}} = P_X \times P_{Y|X}\}. \quad (\text{A.5})$$

The key property of types permitting an exponentially tight analysis of certain quantities is that the number of (possibly joint or conditional) types grows polynomially in n :

$$|\mathcal{P}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|-1} \quad (\text{A.6})$$

$$|\mathcal{P}_n(\mathcal{Y}|\mathcal{X})| \leq (n+1)^{|\mathcal{X}||\mathcal{Y}|-1}. \quad (\text{A.7})$$

A.1 Properties for the Single-User Setting

Here we present several properties of types that will be used for the single-user setting. We define $(\mathbf{X}', \mathbf{Y}, \bar{\mathbf{X}}') \sim \prod_{i=1}^n Q(x_i)W(y_i|x_i)Q(\bar{x}_i)$; while we consider random-coding ensembles that need not be i.i.d., the properties given here will suffice. Recalling the notations $\doteq$, $\dot{\leq}$, etc. defined in Section 1.4, we have the following:

- There exists a type $Q_n \in \mathcal{P}_n(\mathcal{X})$ such that

$$\mathbb{P}[\mathbf{X} \in T^n(Q_n)] \geq (n+1)^{-(|\mathcal{X}|-1)}. \quad (\text{A.8})$$

- For any joint distribution P_{XY} , we have

$$\mathbb{P}[(\mathbf{X}', \mathbf{Y}) \in T^n(P_{XY})] \doteq e^{-nD(P_{XY} \| Q \times W)}. \quad (\text{A.9})$$

- If $\mathbf{y} \in T^n(P_Y)$, then we have

$$\mathbb{P}[(\bar{\mathbf{X}}', \mathbf{y}) \in T^n(\tilde{P}_{XY})] \doteq e^{-nD(\tilde{P}_{XY} \| Q \times P_Y)} \quad (\text{A.10})$$

for any $\tilde{P}_{XY}$ such that $\tilde{P}_Y = P_Y$.

- For any joint distribution $\tilde{P}_{X\bar{X}}$, we have

$$\mathbb{P}[(\mathbf{X}', \bar{\mathbf{X}}') \in T^n(\tilde{P}_{X\bar{X}})] \doteq e^{-nD(\tilde{P}_{X\bar{X}} \| Q \times Q)}. \quad (\text{A.11})$$

- If $(\mathbf{x}, \bar{\mathbf{x}}) \in T^n(\tilde{P}_{X\bar{X}})$, then we have

$$\mathbb{P}[(\mathbf{x}, \bar{\mathbf{x}}, \mathbf{Y}) \in T^n(P_{X\bar{X}Y}) \mid \mathbf{X} = \mathbf{x}] \doteq e^{-nD(P_{X\bar{X}Y} \| \tilde{P}_{X\bar{X}} \times W)} \quad (\text{A.12})$$

for any $P_{X\bar{X}Y}$ such that $P_{X\bar{X}} = \tilde{P}_{X\bar{X}}$.

A.2 Properties for the MAC

Here we present several properties of types in the context of the multiple-access channel (cf. Chapter 4), though the properties are also used in other settings throughout the thesis. We fix the conditional input distributions $Q_1 \in \mathcal{P}(\mathcal{X}_1)$ and $Q_2 \in \mathcal{P}(\mathcal{X}_2)$. For $\nu = 1, 2$, we let $Q_{\nu,n} \in \mathcal{P}_n(\mathcal{X}_\nu)$ be a type with the same support as Q_ν such that $\max_{x_\nu} |Q_\nu(x_\nu) - Q_{\nu,n}(x_\nu)| \leq \frac{1}{n}$, and we let P_{X_ν} be uniform on $T^n(Q_{\nu,n})$ (see (4.21)). Defining the random variables

$$(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}, \bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2) \sim P_{\mathbf{X}_1}(\mathbf{x}_1)P_{\mathbf{X}_2}(\mathbf{x}_2)W^n(\mathbf{y}|\mathbf{x}_1, \mathbf{x}_2)P_{\mathbf{X}_1}(\bar{\mathbf{x}}_1)P_{\mathbf{X}_2}(\bar{\mathbf{x}}_2), \quad (\text{A.13})$$

we have the following:

- For any sequence $\mathbf{x}_1$, we have

$$\mathbb{P}[\mathbf{X}_1 = \mathbf{x}_1] \leq \prod_{i=1}^n Q_1(x_{1,i}). \quad (\text{A.14})$$

- For any $P_{X_1X_2}$ such that $P_{X_1} = Q_{1,n}$ and $P_{X_2} = Q_{2,n}$, we have

$$\mathbb{P}[(\mathbf{X}_1, \mathbf{X}_2) \in T^n(P_{X_1X_2})] \doteq e^{-nI_P(X_1;X_2)}. \quad (\text{A.15})$$

- For any $P_{X_1X_2Y}$ such that $P_{X_1} = Q_{1,n}$ and $P_{X_2} = Q_{2,n}$, we have

$$\mathbb{P}[(\mathbf{X}_1, \mathbf{X}_2, \mathbf{Y}) \in T^n(P_{X_1X_2Y})] \doteq e^{-nD(P_{X_1X_2Y} \| Q_1 \times Q_2 \times W)}. \quad (\text{A.16})$$

- For $\nu = 1, 2$, if $\mathbf{y} \in T^n(P_Y)$, then

$$\mathbb{P}[(\bar{\mathbf{X}}_\nu, \mathbf{y}) \in T^n(\tilde{P}_{X_\nu Y})] \doteq e^{-nI_{\tilde{P}}(X_\nu;Y)} \quad (\text{A.17})$$

for any $\tilde{P}_{X_\nu Y}$ such that $\tilde{P}_{X_\nu} = Q_{\nu,n}$ and $\tilde{P}_Y = P_Y$.

- If $\mathbf{y} \in T^n(P_Y)$, then

$$\mathbb{P}[(\bar{\mathbf{X}}_1, \bar{\mathbf{X}}_2, \mathbf{y}) \in T^n(\tilde{P}_{X_1X_2Y})] \doteq e^{-nD(\tilde{P}_{X_1X_2Y} \| Q_1 \times Q_2 \times \tilde{P}_Y)} \quad (\text{A.18})$$

for any $\tilde{P}_{X_1X_2Y}$ such that $\tilde{P}_{X_1} = Q_{1,n}$, $\tilde{P}_{X_2} = Q_{2,n}$ and $\tilde{P}_Y = P_Y$.

Appendix B

Properties of the LM Rate and GMI

In this appendix, we provide some properties of the LM rate and GMI to complement the existing properties outlined in Section 1.1.4. The observations and results presented here can easily be extended to the functions $E_0^{\text{iid}}(Q, \rho)$ and $E_0^{\text{cc}}(Q, \rho)$ defined in Section 2.4.2 (see (2.38) and (2.40)) for any fixed $\rho \in (0, 1]$, but we limit our attention to the rates for brevity.

B.1 Suprema over s and $a(\cdot)$

We first show the the suprema in (1.9) and (1.11) are not achieved in general, and then provide a simple sufficient condition under which they are achieved. Consider the binary erasure channel (BEC) with $q(x, y) = 1 + \mathbb{1}\{W(y|x) > 0\}$:

$$\mathbf{W} = \begin{bmatrix} 1 - \epsilon & \epsilon & 0 \\ 0 & \epsilon & 1 - \epsilon \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 2 & 2 \end{bmatrix}. \quad (\text{B.1})$$

Under the uniform distribution $Q = (\frac{1}{2}, \frac{1}{2})$, the objective function in (1.11) equals

$$(1 - \epsilon) \log \frac{2^s}{\frac{1}{2}(2^s + 1)} + \epsilon \log \frac{2^s}{\frac{1}{2}(2^s + 2^s)} = (1 - \epsilon) \left(\log 2 + \log \frac{2^s}{2^s + 1} \right). \quad (\text{B.2})$$

Taking $s \rightarrow \infty$, it follows that $I_{\text{GMI}}(Q) = (1 - \epsilon) \log 2$, which equals the matched capacity. However, no finite value of s achieves this value.

Next, consider the noiseless channel with an additional non-zero value in the metric:

$$\mathbf{W} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}. \quad (\text{B.3})$$

Under the uniform distribution $Q = (\frac{1}{2}, \frac{1}{2})$, the objective function in (1.9) equals

$$\frac{1}{2} \log \frac{e^{a(0)}}{\frac{1}{2}e^{a(0)}} + \frac{1}{2} \log \frac{e^{a(1)}}{\frac{1}{2}(e^{a(0)} + e^{a(1)})} = \frac{1}{2} \log 2 + \frac{1}{2} \left(\log 2 + \log \frac{e^{a(1)}}{e^{a(0)} + e^{a(1)}} \right). \quad (\text{B.4})$$

We obtain $I_{\text{LM}}(Q) = \log 2$, which is the matched capacity, by choosing $a(\cdot)$ such that $a(1) - a(0) \rightarrow \infty$. However, this value is not achieved for any finite values of $a(0)$ and $a(1)$.

The crucial property in the preceding examples is that $q(x, y)$ takes a non-zero value where $W(y|x)$ is zero. If the opposite is true (i.e. there exists a pair (x, y) such that $q(x, y) = 0$ and $W(y|x) > 0$), then the LM rate and GMI are zero when the corresponding input is used (see Section 1.1.4). We will show that in all other cases, the suprema in (1.9) and (1.11) are achieved. Before doing so, we present the following technical lemma.

Lemma B.1. *For any non-negative function $q(x, y)$ on the finite sets $\mathcal{X}$ and $\mathcal{Y}$, the following statements are equivalent:*

1. *There exist functions $g(x)$ and $h(y)$ such that $q(x, y) = g(x)h(y)\mathbb{1}\{q(x, y) > 0\}$.*
2. *(i) For all $(x, y, \bar{x})$ such that $q(x, y)q(\bar{x}, y) > 0$, the ratio $\frac{q(x, y)}{q(\bar{x}, y)}$ is equal to some function $r(x, \bar{x})$ not depending on y ; (ii) Given any positive integer m and any sequence of the form $(x_1, \dots, x_m, x_1)$ such that each consecutive pair $(x, \bar{x})$ satisfies $q(x, y)q(\bar{x}, y) > 0$ for some y , the following holds:*

$$r(x_1, x_2)r(x_2, x_3) \dots r(x_{m-1}, x_m)r(x_m, x_1) = 1. \quad (\text{B.5})$$

Proof. We begin by showing that the first condition implies the second. A direct substitution yields $\frac{q(x, y)}{q(\bar{x}, y)} = \frac{g(x)}{g(\bar{x})}$ whenever both $q(\cdot, \cdot)$ terms are positive; as desired, this ratio does not depend on y . Moreover, the left-hand side of (B.5) is given by

$$\frac{g(x_1)}{g(x_2)} \frac{g(x_2)}{g(x_3)} \dots \frac{g(x_{m-1})}{g(x_m)} \frac{g(x_m)}{g(x_1)}, \quad (\text{B.6})$$

which equals one.

It remains to prove that the second condition implies the first. We let $\mathcal{G}$ denote the bipartite graph with vertices given by the elements of $\mathcal{X}$ and $\mathcal{Y}$ such that there is an edge between x and y if and only if $q(x, y) > 0$. We limit our attention to the case that $\mathcal{G}$ is connected, i.e. there exists a path between any two vertices. More generally, the graph can be split into connected components and the subsequent arguments can be applied separately for each component.

We define the set

$$\mathcal{A} \triangleq \left\{ (x, \bar{x}) : q(x, y)q(\bar{x}, y) > 0 \text{ for some } y \right\}. \quad (\text{B.7})$$

We select an arbitrary symbol $x_0 \in \mathcal{X}$ and set $g(x_0) = 1$. We then repeatedly set $g(x) = r(x, \bar{x})g(\bar{x})$ for all $(x, \bar{x})$ such that $g(\bar{x})$ has been assigned a value, $g(x)$ has not been assigned a value, and $(x, \bar{x}) \in \mathcal{A}$. Since we have assumed that $\mathcal{G}$ is connected, all values $g(\cdot)$ are assigned after finitely many steps.

We wish to show that $q(x, y) = g(x)h(y)\mathbb{1}\{q(x, y) > 0\}$ for some $h(y)$. Let $(x, \bar{x})$ be an arbitrary pair in $\mathcal{A}$. By the above procedure, we have

$$g(x) = r(x_1, x_0)r(x_2, x_1) \dots r(x_m, x_{m-1})r(x, x_m), \quad (\text{B.8})$$

for some integer m and sequence $(x_0, x_1, \dots, x_m)$ such that consecutive elements (x_i, x_{i+1}) are in $\mathcal{A}$ (and similarly for (x_m, x)). An analogous statement to (B.8) holds for $g(\bar{x})$, for some integer $\bar{m}$ and sequence $(x_0, \bar{x}_1, \dots, \bar{x}_{\bar{m}})$. Moreover, since $r(x', x'')r(x'', x') = 1$ whenever $(x', x'') \in \mathcal{A}$ (see the lemma statement), we have

$$\frac{1}{g(\bar{x})} = r(x_0, \bar{x}_1)r(\bar{x}_1, \bar{x}_2) \dots r(\bar{x}_{\bar{m}-1}, \bar{x}_{\bar{m}})r(\bar{x}_{\bar{m}}, \bar{x}). \quad (\text{B.9})$$

Combining (B.8) and (B.9), we obtain

$$\frac{g(x)}{g(\bar{x})} = r(x, x_m) \dots r(x_1, x_0)r(x_0, \bar{x}_1) \dots r(\bar{x}_{\bar{m}}, \bar{x}) \quad (\text{B.10})$$

$$= r(x, x_m) \dots r(x_1, x_0)r(x_0, \bar{x}_1) \dots r(\bar{x}_{\bar{m}}, \bar{x})r(\bar{x}, x)r(x, \bar{x}) \quad (\text{B.11})$$

$$= r(x, \bar{x}), \quad (\text{B.12})$$

where (B.11) follows since $r(\bar{x}, x)r(x, \bar{x}) = 1$, and (B.12) follows from (B.5). Substituting the definition of $r(x, \bar{x})$ into (B.12) and rearranging, we obtain

$$\frac{q(x, y)}{g(x)} = \frac{q(\bar{x}, y)}{g(\bar{x})} \quad (\text{B.13})$$

for all y such that $q(x, y)q(\bar{x}, y) > 0$. Thus, among the pairs (x, y) such that $q(x, y) > 0$, the ratio $\frac{q(x, y)}{g(x)}$ is only a function of y ; the lemma follows by naming this function $h(y)$. $\square$

We now have the following.

Lemma B.2. *For any input distribution Q and pair (W, q) such that*

$$q(x, y) > 0 \iff W(y|x) > 0, \quad (\text{B.14})$$

the supremum in (1.9) is achieved by at least one pair (s, a) having finite values, and the supremum in (1.11) is achieved by at least one finite value of s .

Proof. We assume without loss of generality that $Q(x) > 0$ for all x , and we begin by considering the GMI. We write the objective in (1.11) as

$$I_s \triangleq - \sum_{x,y} Q(x)W(y|x) \log \sum_{\bar{x}} Q(\bar{x}) \left(\frac{q(\bar{x}, y)}{q(x, y)} \right)^s. \quad (\text{B.15})$$

In the case that there exists $(x, y, \bar{x})$ such that $W(y|x)W(y|\bar{x}) > 0$ and $q(\bar{x}, y) > q(x, y)$, we have

$$I_s \leq -Q(x)W(y|x) \log \frac{Q(\bar{x})q(\bar{x}, y)^s}{q(x, y)^s}. \quad (\text{B.16})$$

This bound tends to $-\infty$ as $s \rightarrow \infty$, and hence the supremum cannot be achieved in this limit. Since I_s is continuous in s , we conclude that the supremum is achieved by a finite value. If the above-mentioned condition fails, then we have $q(\bar{x}, y) = q(x, y)$ wherever $W(y|x)W(y|\bar{x}) > 0$. Recalling the condition in (B.14), we see that I_s does not depend on s , since the argument to $(\cdot)^s$ is always zero or one. Thus, any $s > 0$ achieves the supremum.

We prove the claim regarding the LM rate using similar ideas. The objective in (1.9) is denoted by

$$I_{s,a} \triangleq - \sum_{x,y} Q(x)W(y|x) \log \sum_{\bar{x}} Q(\bar{x}) \left(\frac{q(\bar{x}, y)}{q(x, y)} \right)^s \frac{e^{a(\bar{x})}}{e^{a(x)}}. \quad (\text{B.17})$$

We consider sequences of parameters $\{s_k\}_{k \geq 1}$ and $\{a_k\}_{k \geq 1}$ such that some $a_k(\cdot)$ is unbounded, and show that if the LM rate is achieved as $k \rightarrow \infty$, then it is also achieved by different choices that remain bounded. This will prove that $a_k(\cdot)$ can be chosen to be bounded without loss of optimality, from which it follows via the same arguments as those for the GMI that s_k can be assumed bounded. The lemma will then follow from the fact that $I_{s,a}$ is continuous in (s, a) .

Let $\mathcal{A}$ denote the set of triplets $(x, y, \bar{x})$ such that $W(y|x)W(y|\bar{x}) > 0$ (which is equivalent to $q(x, y)q(\bar{x}, y) > 0$ by (B.14)). By the assumption in (B.14), $I_{s,a}$ depends on (s, a) only through the values of

$$\psi_{s,a}(x, y, \bar{x}) \triangleq \left(\frac{q(x, y)}{q(\bar{x}, y)} \right)^s \frac{e^{a(x)}}{e^{a(\bar{x})}} \quad (\text{B.18})$$

for triplets within $\mathcal{A}$.

Let $\{s_k\}_{k \geq 1}$ and $\{a_k\}_{k \geq 1}$ be sequences achieving the supremum in (1.9) in the limit as $k \rightarrow \infty$. For any $(x, y, \bar{x}) \in \mathcal{A}$, $\psi_{s_k, a_k}(x, y, \bar{x})$ must not tend towards ∞ , since this would imply $I_{s_k, a_k} \rightarrow -\infty$; this can be seen by upper bounding (B.17) by replacing each summation by a single value. By reversing the roles of x and $\bar{x}$, we conclude that $\psi_{s_k, a_k}(x, y, \bar{x})$ must not tend towards 0. Thus, $\log \psi_{s_k, a_k}$ is bounded.

Consider the case that the ratio $\frac{e^{a_k(x)}}{e^{a_k(\bar{x})}}$ tends to ∞ for some $(x, \bar{x})$. From (B.18) and the fact that $\psi_{s_k, a_k} \not\rightarrow \infty$, this is only possible if $\frac{q(x, y)}{q(\bar{x}, y)} < 1$ and $s_k \rightarrow \infty$. This, in turn, implies that within $\mathcal{A}$, $\frac{q(x, y)}{q(\bar{x}, y)}$ does not depend on y ; otherwise, the two different values of y would have a different argument to $(\cdot)^s$ in (B.18), and at least one would yield $\log \psi_{s_k, a_k}$ growing unbounded. Moreover, writing $r(x, \bar{x}) \triangleq \frac{q(x, y)}{q(\bar{x}, y)}$, we have the following for any sequence of the form $(x_1, \dots, x_m, x_1)$ such that adjacent symbols share a common output:

$$\psi_{s, a}(x_1, x_2, y) \dots \psi_{s, a}(x_{m-1}, x_m, y) \psi_{s, a}(x_m, x_1, y) = \left(r(x_1, x_2) \dots r(x_{m-1}, x_m) r(x_m, x_1) \right)^s. \quad (\text{B.19})$$

This follows by observing that the $a(\cdot)$ terms in (B.18) cancel. Since we have established that $s_k \rightarrow \infty$ and $\log \psi_{s_k, a_k}$ is bounded, we conclude that the argument to $(\cdot)^s$ in (B.19) equals one. We have thus shown that $q(x, y)$ satisfies the conditions in second item of Lemma B.1, and we conclude that $q(x, y) = g(x)h(y)\mathbb{1}\{q(x, y) > 0\}$ for some functions $g(x)$ and $h(y)$. Given the original parameters s_k and $a_k(\cdot)$, we choose alternative parameters as follows: Set $s' = 1$, and choose $a'_k(\cdot)$ such that

$$\left(\frac{g(x)}{g(\bar{x})} \right)^{s_k} \frac{e^{a_k(x)}}{e^{a_k(\bar{x})}} = \frac{g(x)}{g(\bar{x})} \frac{e^{a'_k(x)}}{e^{a'_k(\bar{x})}}. \quad (\text{B.20})$$

Any such choice yields $\psi_{s', a'_k}(x, y, \bar{x}) = \psi_{s_k, a_k}(x, y, \bar{x})$ within $\mathcal{A}$, and hence $I_{s', a'_k} = I_{s_k, a_k}$. There are infinitely many such choices (e.g. one can always shift $a'_k(\cdot)$ by a constant), and at least one has bounded values, since the logarithm of the left-hand side of (B.20) (i.e. $\log \psi_{s_k, a_k}$) is bounded. $\square$

B.2 Optimization of the LM Rate Over Q

For a given input distribution Q , the LM rate can be computed efficiently using (1.9) or (1.10), since the optimization problems are convex. However, the optimization over Q is difficult, since $I_{\text{LM}}(Q)$ is non-concave in general [11]. It should be noted that, unlike the parameters (s, a) , the maximum over Q is always achieved; this follows from a simple continuity argument and the fact that $\mathcal{P}(\mathcal{X})$ is compact.

Here we provide an alternative form for $I_{\text{LM}}(Q)$ that permits an extension of the Arimoto-

Blahut algorithm [143, 144] that finds a *local* optimum. Interestingly, the additional parameter $a(\cdot)$ in (1.9) compared to (1.11) is seen to be beneficial (rather than making the computation more difficult), at least for the algorithm that we provide.

Maximizing (1.9) with respect to Q yields an overall optimization given by

$$\text{maximize}_{s \geq 0, a(\cdot), Q} \sum_{x,y} Q(x) W(y|x) \log \frac{q(x, y)^s e^{a(x)}}{\sum_{\bar{x}} Q(\bar{x}) q(\bar{x}, y)^s e^{a(\bar{x})}}. \quad (\text{B.21})$$

Suppose for the time being that Q is restricted to satisfy $Q(x) > 0$ for all x . Defining $\tilde{a}(\cdot)$ such that $e^{\tilde{a}(x)} = Q(x) e^{a(x)}$, we can equivalently write (B.21) as

$$\text{maximize}_{s \geq 0, \tilde{a}(\cdot), Q} \sum_{x,y} Q(x) W(y|x) \log \frac{q(x, y)^s e^{\tilde{a}(x)}}{Q(x) \sum_{\bar{x}} q(\bar{x}, y)^s e^{\tilde{a}(\bar{x})}}. \quad (\text{B.22})$$

In this form, the objective is concave in Q for fixed $(s, \tilde{a})$, and vice versa. Thus, we can obtain a sequence of parameters such that the objective is monotonically non-decreasing by alternating between the following two steps:

1. Maximize the objective over the input distribution Q with $(s, \tilde{a})$ fixed. Using the Karush-Kuhn-Tucker conditions [38], it is readily verified that the maximizing Q is given by

$$Q(x) = \frac{\exp \left(\sum_y W(y|x) \log P'_{s, \tilde{a}}(x|y) \right)}{\sum_{\bar{x}} \exp \left(\sum_y W(y|\bar{x}) \log P'_{s, \tilde{a}}(\bar{x}|y) \right)}, \quad (\text{B.23})$$

where

$$P'_{s, \tilde{a}}(x|y) \triangleq \frac{q(x, y)^s e^{\tilde{a}(x)}}{\sum_{\bar{x}} q(\bar{x}, y)^s e^{\tilde{a}(\bar{x})}}. \quad (\text{B.24})$$

2. Maximize the objective over $s \geq 0$ and $\tilde{a}(\cdot)$ with Q fixed. This can be done using efficient convex optimization algorithms [38].

A suitable choice of initial parameters is $s = 1$, $\tilde{a}(x) = 0$ and $Q(x) = \frac{1}{|\mathcal{X}|}$.

We conclude by discussing the possibility that $Q(x) = 0$ for some values of x . As mentioned in Section 1.1.4, if the metric satisfies (1.15), then the LM rate is continuous in Q , and thus the constraint $Q(x) > 0$ is not restrictive. On the other hand, if (1.15) fails, then the LM rate is always zero when the corresponding inputs are used, so one can simply set $Q(x) = 0$ accordingly and repeat the above steps for the channel with the corresponding inputs removed.

Appendix C

Sums of Independent Random Variables

Many of the results of this thesis are proved by studying the behavior of sums of random variables. The goal of this appendix is to provide an introductory overview of the relevant terminology and results. For clarity, and to avoid presenting cumbersome technical conditions, we limit our attention to simple cases, rather than stating the results in their most general form. In particular, we focus on independent and identically distributed (i.i.d.) scalar random variables, even though we will be interested in varying distributions, dependent variables, and multivariate settings throughout the thesis. Comprehensive treatments can be found in the books of Feller [76], Petrov [100], and Dembo-Zeitouni [139].

We let $\{Z_i\}_{i=1}^n$ be an i.i.d. sequence with distribution P_Z , which we assume to have mean zero and variance $\sigma^2 \in (0, \infty)$. The former assumption is without loss of generality, since in the general case the mean can be subtracted. We are interested in characterizing the behavior of the summation $S_n \triangleq \sum_{i=1}^n Z_i$. The following definition will be useful.

Definition C.1. *We say that Z is a lattice random variable with offset γ and span h if its support is a subset of $\{\gamma + ih : i \in \mathbb{Z}\}$, and the same is not true for any such set with a larger value of h . We say that Z is a non-lattice random variable if it is not supported on any set of the form $\{\gamma + ih : i \in \mathbb{Z}\}$.*

C.1 Asymptotic Behavior of the Tail Probability

A basic result of probability theory is the law of large numbers [76, Ch. VII], which can be written as follows:

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{1}{n} S_n \geq z \right] = \begin{cases} 0 & z > 0 \\ 1 & z < 0. \end{cases} \quad (\text{C.1})$$

That is, S_n concentrates about its mean in the limit of large n . A refinement of this statement is the central limit theorem (CLT), which states that [76, Ch. VIII]

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[\frac{S_n}{\sqrt{n\sigma^2}} \geq z \right] = Q(z). \quad (\text{C.2})$$

That is, $\frac{S_n}{\sqrt{n}}$ is asymptotically normal in the limit of large n .

Equation (C.2) is concerned with the tail probability of S_n at values that deviate from the mean by a $\Theta(\sqrt{n})$ term. In many cases, one is interested in the behavior for deviations of a larger order. We proceed by presenting examples of such results. For simplicity, we assume that $\mathbb{E}[e^{\tau Z}] < \infty$ for $Z \sim P_Z$ and any real number τ ; more general cases can be handled [139], but they are omitted here.

Large deviations theory is concerned with probabilities of the form $\mathbb{P}[\frac{1}{n}S_n \geq z]$ for fixed $z > 0$. The asymptotics in this regime can be written in terms of the cumulant transform $\kappa(\tau) \triangleq \log \mathbb{E}[e^{\tau Z}]$ as follows [139, Sec. 2.2.1]:

$$\lim_{n \rightarrow \infty} -\frac{1}{n} \log \mathbb{P} \left[\frac{1}{n} S_n \geq z \right] = E(z), \quad (\text{C.3})$$

where

$$E(z) \triangleq \sup_{\tau \in \mathbb{R}} \tau z - \kappa(\tau). \quad (\text{C.4})$$

That is, the probability in (C.3) decays to zero exponentially fast, and the exponent is given by $E(z)$. A particularly simple non-asymptotic upper bound can be obtained using the Chernoff bound: $\mathbb{P}[\frac{1}{n}S_n \geq z] \leq e^{-nE(z)}$. Studies of the exact asymptotics (or strong large deviations) seek to provide a refined characterization of the prefactor to $e^{-nE(z)}$, and the behavior turns out to vary depending on whether or not Z is a lattice random variable. In the non-lattice case, we have [139, Thm. 3.7.4]

$$\mathbb{P} \left[\frac{1}{n} S_n \geq z \right] \asymp \frac{1}{\hat{\tau} \sqrt{2\pi n \kappa''(\hat{\tau})}} e^{-nE(z)}, \quad (\text{C.5})$$

where $\hat{\tau} \triangleq \arg \max_{\tau} \{\tau z - \kappa(\tau)\}$, and κ'' is the second derivative of κ . A similar result holds in the lattice case, except that the right-hand side of (C.5) is multiplied by an extra $\Theta(1)$ term which, in general, depends on both n and the lattice parameters (span and offset).

Another regime of interest is that of moderate deviations, which is concerned with probabilities of the form $\mathbb{P}[\frac{1}{n}S_n \geq z_n]$ for some positive sequence z_n satisfying $z_n = \omega(\frac{1}{\sqrt{n}})$ and $z_n = o(1)$. In such cases, the tail probability decays to zero, but the rate of decay is no longer

exponential. More precisely, we have [139, Thm. 3.7.1]

$$\lim_{n \rightarrow \infty} -\frac{1}{nz_n^2} \log \mathbb{P}\left[\frac{1}{n}S_n \geq z_n\right] = \frac{1}{2\sigma^2}. \quad (\text{C.6})$$

For example, if $z_n = K\sqrt{\frac{\log n}{n}}$ for some $K > 0$, then the tail probability behaves as $O(n^{-\gamma})$ for some constant $\gamma > 0$ depending on K and σ .

While the preceding results are asymptotic, they are often accompanied by non-asymptotic bounds (e.g. the Chernoff bound stated above). At an intuitive level, large deviations results are more suited to characterizing the behavior of the probability $p_n \triangleq \mathbb{P}[S_n \geq \gamma_n]$ when p_n is small and γ_n is far above $\mathbb{E}[S_n]$. Conversely, results based on the CLT are generally more suited to higher values of p_n , corresponding to values of γ_n closer to $\mathbb{E}[S_n]$. The moderate deviations regime can be thought of as being in between the other two.

C.2 Variations of the Central Limit Theorem

A more precise statement of (C.2) is given by the Berry-Esseen theorem, a simple version of which is as follows [76, Sec. XVI.5]:

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}\left[\frac{S_n}{\sqrt{n}\sigma^2} \geq z\right] - Q(z) \right| \leq \frac{6T}{\sqrt{n}\sigma^{3/2}}, \quad (\text{C.7})$$

where $T \triangleq \mathbb{E}_P[|Z|^3]$. This result strengthens (C.2) by showing that the tail probability differs from the Q-function by $O(\frac{1}{\sqrt{n}})$. It has the notable feature of being non-asymptotic, and depending only on σ and T .

A further refinement of the CLT is given by the Edgeworth expansion, which is stated separately for lattice and non-lattice random variables. Assuming that the third moment $T_0 \triangleq \mathbb{E}_P[Z^3]$ is finite, we have the following in the non-lattice case [76, Sec. XVI.4]:

$$\mathbb{P}\left[\frac{S_n}{\sqrt{n}\sigma^2} \geq z\right] = Q(z) + \frac{T_0}{6\sigma^3\sqrt{n}}(1 - z^2)\phi(z) + o\left(\frac{1}{\sqrt{n}}\right) \quad (\text{C.8})$$

uniformly in z , where $\phi(z) \triangleq \frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$ is the standard normal density. In the lattice case, the expansion in (3.107) holds when z is the midpoint of two consecutive points on the lattice of $\frac{S_n}{\sqrt{n}\sigma^2}$, and a similar expansion can be inferred for general values of z . Under some further assumptions of finite moments, higher-order expansions can be obtained.

The preceding results characterize the behavior of the tail probability of S_n . Another result

of significant interest throughout this thesis is the local limit theorem, which characterizes the behavior of probabilities of the form $\mathbb{P}[S_n \in [z, z + \eta)]$ for some $\eta > 0$. In the case that $Z \sim P_Z$ is a non-lattice random variable, we have the following uniformly in z [99]:

$$\mathbb{P}[S_n \in [z, z + \eta)] = \frac{\eta}{\sqrt{2\pi n\sigma^2}} e^{-\frac{z^2}{2n\sigma^2}} + o\left(\frac{1}{\sqrt{n}}\right). \quad (\text{C.9})$$

One can think of this result as stating that S_n is asymptotically distributed as $N(0, n\sigma^2)$, since the first term approximately equals the integral of the corresponding density over $[z, z + \eta)$.

In the lattice case, a natural choice of η is the span h of the lattice of Z . In this case, S_n is also a lattice variable with span h , and $\mathbb{P}[S_n \in [z, z + \eta)]$ is the probability mass function of S_n evaluated at the unique lattice point in the range $[z, z + \eta)$. Assuming for notational convenience that this point is z itself, we have the following uniformly in z [100, Sec. VII.1]:

$$\mathbb{P}[S_n = z] = \frac{h}{\sqrt{2\pi n\sigma^2}} e^{-\frac{z^2}{2n\sigma^2}} + o\left(\frac{1}{\sqrt{n}}\right). \quad (\text{C.10})$$

Closely related to the local limit theorem is the concentration function (e.g. see [101]), which is defined as follows:

$$G(S_n, \eta) \triangleq \sup_{z \in \mathbb{R}} \mathbb{P}[S_n \in [z, z + \eta)]. \quad (\text{C.11})$$

We immediately obtain from (C.9) that, in the non-lattice case,

$$G(S_n, \eta) = \frac{\eta}{\sqrt{2\pi n\sigma^2}} + o\left(\frac{1}{\sqrt{n}}\right). \quad (\text{C.12})$$

The asymptotic behavior of $G(S_n, \eta)$ can similarly be inferred from (C.10) in the lattice case.

We conclude with some comments on settings beyond the scalar i.i.d. case. Throughout the thesis, we will often be interested in sums of independent variables with varying distributions. In the case that there are finitely many such distributions, each of which is discrete, the distinction between the lattice and non-lattice case turns out to depend on the union of the mass points of the distributions that have a positive variance (e.g. see [100, Ch. VII]).

We will also be interested in the multivariate setting, where one considers sums of d -dimensional vectors of the form $S_n = \sum_{i=1}^n \mathbf{Z}_i$. One may no longer consider the lattice and non-lattice cases alone, as it may be the case that some entries of the vector have a lattice distribution and others do not. Moreover, there are multiple ways to generalize the local limit theorem and concentration function; the two most common ones consider the probability of S_n lying in a d -dimensional cube [74], and a d -dimensional sphere [101].

Appendix D

Bounds on the Probability of a Multiply-Indexed Union

In channel coding settings, bounds on the random-coding error probability are often obtained using the union bound, which states that for any set of events $\{A_i\}_{i=1}^N$,

$$\mathbb{P}\left[\bigcup_i A_i\right] \leq \sum_i \mathbb{P}[A_i]. \quad (\text{D.1})$$

In many cases, the right-hand side of (D.1) is greater than one, and hence a tighter bound is given by the truncated union bound, namely

$$\mathbb{P}\left[\bigcup_i A_i\right] \leq \min\left\{1, \sum_i \mathbb{P}[A_i]\right\}. \quad (\text{D.2})$$

We will also be interested in lower bounds on the probability of a union, which are used to prove ensemble tightness results. In particular, we will make use of de Caen's bound [145]:

$$\mathbb{P}\left[\bigcup_i A_i\right] \geq \sum_i \frac{\mathbb{P}[A_i]^2}{\sum_{i'} \mathbb{P}[A_i \cap A_{i'}]}. \quad (\text{D.3})$$

In the case that the events are pairwise independent and identically distributed, it is easily verified that (D.3) proves the tightness of (D.2) to within a factor of $\frac{1}{2}$ [70]. This tightness result in fact holds even when the events are not identically distributed [73, Lemma A.2].

In this appendix, we provide a number of upper and lower bounds on the probability of a multiply-indexed union. In several cases of interest, the upper and lower bounds coincide to within a constant factor, and generalize the above-mentioned tightness result of [70] to certain settings where pairwise independence need not hold.

Lemma D.1. Let $\{Z_1(i)\}_{i=1}^{N_1}$ and $\{Z_2(j)\}_{j=1}^{N_2}$ be independent sequences of identically distributed random variables on the alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ respectively, with $Z_1(i) \sim P_{Z_1}$ and $Z_2(j) \sim P_{Z_2}$. For any set $\mathcal{A} \subseteq \mathcal{Z}_1 \times \mathcal{Z}_2$, we have

$$\mathbb{P}\left[\bigcup_{i,j}\{(Z_1(i), Z_2(j)) \in \mathcal{A}\}\right] \leq \min \left\{ 1, N_1 \mathbb{E} \left[\min \left\{ 1, N_2 \mathbb{P}[(Z_1, Z_2) \in \mathcal{A} \mid Z_1] \right\} \right], \right. \\ \left. N_2 \mathbb{E} \left[\min \left\{ 1, N_1 \mathbb{P}[(Z_1, Z_2) \in \mathcal{A} \mid Z_2] \right\} \right] \right\}, \quad (\text{D.4})$$

where $(Z_1, Z_2) \sim P_{Z_1} \times P_{Z_2}$.

Proof. Applying the union bound to the union over i , we obtain

$$\mathbb{P}\left[\bigcup_{i,j}\{(Z_1(i), Z_2(j)) \in \mathcal{A}\}\right] \leq N_1 \mathbb{P}\left[\bigcup_j \{(Z_1, Z_2(j)) \in \mathcal{A}\}\right] \quad (\text{D.5})$$

$$= N_1 \mathbb{E} \left[\mathbb{P}\left[\bigcup_j \{(Z_1, Z_2(j)) \in \mathcal{A}\} \mid Z_1\right] \right]. \quad (\text{D.6})$$

Applying the truncated union bound to the union over j , we recover the second term in the outer minimization in (D.4). The third term is obtained similarly by applying the union bounds in the opposite order, and the first term is trivial. $\square$

Lemma D.2. Let $\{Z_1(i)\}_{i=1}^{N_1}$ and $\{Z_2(j)\}_{j=1}^{N_2}$ be independent sequences of pairwise independent and identically distributed random variables on the alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ respectively, with $Z_1(i) \sim P_{Z_1}$ and $Z_2(j) \sim P_{Z_2}$. For any set $\mathcal{A} \subseteq \mathcal{Z}_1 \times \mathcal{Z}_2$, we have

$$\mathbb{P}\left[\bigcup_{i,j}\{(Z_1(i), Z_2(j)) \in \mathcal{A}\}\right] \geq \frac{1}{4} \min \left\{ 1, N_1 \frac{\mathbb{P}[(Z_1, Z_2) \in \mathcal{A}]^2}{\mathbb{P}[(Z_1, Z_2) \in \mathcal{A} \cap (Z_1, Z'_2) \in \mathcal{A}]}, \right. \\ \left. N_2 \frac{\mathbb{P}[(Z_1, Z_2) \in \mathcal{A}]^2}{\mathbb{P}[(Z_1, Z_2) \in \mathcal{A} \cap (Z'_1, Z_2) \in \mathcal{A}]}, N_1 N_2 \mathbb{P}[(Z_1, Z_2) \in \mathcal{A}] \right\}, \quad (\text{D.7})$$

where $(Z_1, Z_2, Z'_1, Z'_2) \sim P_{Z_1}(z_1)P_{Z_2}(z_2)P_{Z_1}(z'_1)P_{Z_2}(z'_2)$.

Proof. We make use of de Caen's bound in (D.3). Noting by symmetry that each term in the outer summation is equal, and splitting the inner summation according to which of the (i, j)

indices coincide with (i', j') , we obtain

$$\begin{aligned} & \mathbb{P}\left[\bigcup_{i,j}\left\{(Z_1(i), Z_2(j)) \in \mathcal{A}\right\}\right] \\ & \geq N_1 N_2 \mathbb{P}[(Z_1, Z_2) \in \mathcal{A}]^2 \left(\mathbb{P}[(Z_1, Z_2) \in \mathcal{A}] + (N_1 - 1)(N_2 - 1) \mathbb{P}[(Z_1, Z_2) \in \mathcal{A}]^2 \right. \\ & \quad \left. + (N_2 - 1) \mathbb{P}[(Z_1, Z_2) \in \mathcal{A} \cap (Z_1, Z'_1) \in \mathcal{A}] + (N_1 - 1) \mathbb{P}[(Z_1, Z_2) \in \mathcal{A} \cap (Z'_1, Z_2) \in \mathcal{A}] \right)^{-1} \end{aligned} \quad (\text{D.8})$$

$$\begin{aligned} & \geq N_1 N_2 \mathbb{P}[(Z_1, Z_2) \in \mathcal{A}]^2 \left(4 \max \left\{ \mathbb{P}[(Z_1, Z_2) \in \mathcal{A}], N_1 N_2 \mathbb{P}[(Z_1, Z_2) \in \mathcal{A}]^2, \right. \right. \\ & \quad \left. \left. N_2 \mathbb{P}[(Z_1, Z_2) \in \mathcal{A} \cap (Z_1, Z'_1) \in \mathcal{A}], N_1 \mathbb{P}[(Z_1, Z_2) \in \mathcal{A} \cap (Z'_1, Z_2) \in \mathcal{A}] \right\} \right)^{-1}, \end{aligned} \quad (\text{D.9})$$

from which the lemma follows. $\square$

The following lemma gives conditions under which a weakened version of (D.4) coincides with (D.7) to within a factor of four.

Lemma D.3. *Let $\{Z_1(i)\}_{i=1}^{N_1}$ and $\{Z_2(j)\}_{j=1}^{N_2}$ be independent sequences of identically distributed random variables on the alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ respectively, with $Z_1(i) \sim P_{Z_1}$ and $Z_2(j) \sim P_{Z_2}$. Fix a set $\mathcal{A} \subseteq \mathcal{Z}_1 \times \mathcal{Z}_2$, and define*

$$\mathcal{A}_1 \triangleq \left\{ z_1 \in \mathcal{Z}_1 : (z_1, z_2) \in \mathcal{A} \text{ for some } z_2 \in \mathcal{Z}_2 \right\} \quad (\text{D.10})$$

$$\mathcal{A}_2 \triangleq \left\{ z_2 \in \mathcal{Z}_2 : (z_1, z_2) \in \mathcal{A} \text{ for some } z_1 \in \mathcal{Z}_1 \right\}. \quad (\text{D.11})$$

1. A general upper bound is given by

$$\begin{aligned} \mathbb{P}\left[\bigcup_{i,j}\left\{(Z_1(i), Z_2(j)) \in \mathcal{A}\right\}\right] & \leq \min \left\{ 1, N_1 \mathbb{P}[Z_1 \in \mathcal{A}_1], \right. \\ & \quad \left. N_2 \mathbb{P}[Z_2 \in \mathcal{A}_2], N_1 N_2 \mathbb{P}[(Z_1, Z_2) \in \mathcal{A}] \right\}, \end{aligned} \quad (\text{D.12})$$

where $(Z_1, Z_2) \sim P_{Z_1} \times P_{Z_2}$.

2. If (i) $\{Z_1(i)\}_{i=1}^{N_1}$ are pairwise independent, (ii) $\{Z_2(j)\}_{j=1}^{N_2}$ are pairwise independent, (iii) $\mathbb{P}[(z_1, Z_2) \in \mathcal{A}]$ is the same for all $z_1 \in \mathcal{A}_1$, and (iv) $\mathbb{P}[(Z_1, z_2) \in \mathcal{A}]$ is the same for all

$z_2 \in \mathcal{A}_2$, then

$$\mathbb{P}\left[\bigcup_{i,j}\left\{(Z_1(i), Z_2(j)) \in \mathcal{A}\right\}\right] \geq \frac{1}{4} \min\left\{1, N_1\mathbb{P}[Z_1 \in \mathcal{A}_1], N_2\mathbb{P}[Z_2 \in \mathcal{A}_2], N_1N_2\mathbb{P}[(Z_1, Z_2) \in \mathcal{A}]\right\}. \quad (\text{D.13})$$

Proof. We obtain (D.12) by weakening (D.4) in multiple ways. The second term in (D.12) follows since the inner probability in the second term of (D.4) is zero whenever $\mathbb{P}[Z_1 \notin \mathcal{A}]$, and since $\min\{1, \zeta\} \leq 1$. The third term in (D.12) is obtained similarly, and the fourth term follows from the fact that $\min\{1, \zeta\} \leq \zeta$.

The lower bound in (D.13) follows from (D.7), and since the additional assumptions in the second part of the lemma statement imply

$$\frac{\mathbb{P}[(Z_1, Z_2) \in \mathcal{A}]^2}{\mathbb{P}[(Z_1, Z_2) \in \mathcal{A} \cap (Z_1, Z'_2) \in \mathcal{A}]} = \frac{\mathbb{P}[Z_1 \in \mathcal{A}_1]^2 \mathbb{P}[(z_1, Z_2) \in \mathcal{A}]^2}{\mathbb{P}[Z_1 \in \mathcal{A}_1] \mathbb{P}[(z_1, Z_2) \in \mathcal{A}]^2} \quad (\text{D.14})$$

$$= \mathbb{P}[Z_1 \in \mathcal{A}_1], \quad (\text{D.15})$$

where z_1 is an arbitrary element of $\mathcal{A}_1$. The third term in the minimization in (D.7) can be handled similarly. $\square$

While we will primarily be interested in doubly-indexed unions, we also give the following generalization of Lemma D.3 to the probability of a union indexed by K values, which is stated without proof. For a given subset $\mathcal{K} = \{k_1, \dots, k_{|\mathcal{K}|}\}$ of $\{1, \dots, K\}$, we write $\mathbf{Z}_{\mathcal{K}}$ as a shorthand for $(Z_{k_1}, \dots, Z_{k_{|\mathcal{K}|}})$, and similarly for $\mathbf{Z}_{\mathcal{K}^c}$. The corresponding realizations are written by $\mathbf{z}_{\mathcal{K}} = (z_{k_1}, \dots, z_{k_{|\mathcal{K}|}})$, and similarly for $\mathbf{z}_{\mathcal{K}^c}$.

Lemma D.4. *Let $\{Z_1(i_1)\}_{i_1=1}^{N_1}, \dots, \{Z_K(i_K)\}_{i_K=1}^{N_K}$ be independent sequences of identically distributed random variables on the alphabets $\mathcal{Z}_1, \dots, \mathcal{Z}_K$ respectively, with $Z_k(i_k) \sim P_{Z_k}$ for $k = 1, \dots, K$. Fix $\mathcal{A} \subseteq \mathcal{Z}_1 \times \dots \times \mathcal{Z}_K$, and for each $\mathcal{K} = \{k_1, \dots, k_{|\mathcal{K}|}\} \subseteq \{1, \dots, K\}$, define*

$$\mathcal{A}_{\mathcal{K}} \triangleq \left\{ \mathbf{z}_{\mathcal{K}} : (z_1, \dots, z_K) \in \mathcal{A} \text{ for some } \mathbf{z}_{\mathcal{K}^c} \right\}. \quad (\text{D.16})$$

1. A general upper bound is given by

$$\mathbb{P}\left[\bigcup_{i_1, \dots, i_K} \left\{ (Z_1(i_1), \dots, Z_K(i_K)) \in \mathcal{A} \right\}\right] \leq \min \left\{ 1, \min_{\mathcal{K} \subseteq \{1, \dots, K\}, \mathcal{K} \neq \emptyset} \left(\prod_{k \in \mathcal{K}} N_k \right) \mathbb{P}[\mathbf{Z}_{\mathcal{K}} \in \mathcal{A}_{\mathcal{K}}] \right\}, \quad (\text{D.17})$$

where $(Z_1, \dots, Z_K) \sim \prod_{k=1}^K P_{Z_k}$.

2. If (i) $\{Z_k(i)\}_{i=1}^{N_k}$ are pairwise independent for any given k , and (ii) for any given $\mathcal{K} \subseteq \{1, \dots, K\}$ ($\mathcal{K} \neq \emptyset$), $\mathbb{P}[(Z_1, \dots, Z_K) \in \mathcal{A} | \mathbf{Z}_{\mathcal{K}} = \mathbf{z}_{\mathcal{K}}]$ is the same for all $\mathbf{z}_{\mathcal{K}} \in \mathcal{A}_{\mathcal{K}}$, then

$$\mathbb{P}\left[\bigcup_{i_1, \dots, i_K} \left\{ (Z_1(i_1), \dots, Z_K(i_K)) \in \mathcal{A} \right\}\right] \geq 2^{-K} \min \left\{ 1, \min_{\mathcal{K} \subseteq \{1, \dots, K\}, \mathcal{K} \neq \emptyset} \left(\prod_{k \in \mathcal{K}} N_k \right) \mathbb{P}[\mathbf{Z}_{\mathcal{K}} \in \mathcal{A}_{\mathcal{K}}] \right\}, \quad (\text{D.18})$$

where $(Z_1, \dots, Z_K) \sim \prod_{k=1}^K P_{Z_k}$.

Appendix E

Convex Optimization and Lagrange Duality

In this appendix, we provide a number of general lemmas based on Lagrange duality [38] that will be used throughout the thesis. The following lemma generalizes the result that (1.9) and (1.10) are equivalent, which was proved in [9].

Lemma E.1. *Fix the finite alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$, the non-negative functions $f(z_1, z_2)$ and $g(z_1, z_2)$, the distributions $P_{Z_1} \in \mathcal{P}(\mathcal{Z}_1)$ and $P_{Z_2} \in \mathcal{P}(\mathcal{Z}_2)$, and a constant β . The optimization*

$$\min_{\substack{\tilde{P}_{Z_1 Z_2} : \tilde{P}_{Z_1} = P_{Z_1}, \tilde{P}_{Z_2} = P_{Z_2}, \\ \mathbb{E}_{\tilde{P}}[\log f(Z_1, Z_2)] \geq \beta}} I_{\tilde{P}}(Z_1; Z_2) - \mathbb{E}_{\tilde{P}}[\log g(Z_1, Z_2)] \quad (\text{E.1})$$

has the same value as

$$\sup_{\lambda \geq 0, \mu_1(\cdot)} - \sum_{z_2} P_{Z_2}(z_2) \log \sum_{z_1} P_{Z_1}(z_1) f(z_1, z_2)^\lambda g(z_1, z_2) e^{\mu_1(z_1)} + \sum_{z_1} P_{Z_1}(z_1) \mu_1(z_1) + \lambda \beta, \quad (\text{E.2})$$

where the supremum over $\mu_1(\cdot)$ is taken over all real-valued functions on $\mathcal{Z}_1$.

Proof. See Section E.1. □

When using Lemma E.1, we will typically be interested the case that either $g(\cdot, \cdot)$ is absent from (E.1), or the constraint $\mathbb{E}_{\tilde{P}}[\log f(Z_1, Z_2)] \geq \beta$ is absent. In the former case, the dual expression is given by (E.2) with $g(\cdot, \cdot) = 1$, and in the latter case, the dual expression is given by (E.2) with $\lambda = 0$. By the symmetry of (E.1) in Z_1 and Z_2 , the roles of the two can also be swapped in (E.2).

Lemma E.1 allows us to obtain dual expressions for the rates and exponents for constant-composition random coding. The following lemma provides variations that will be useful for cost-constrained random coding (see Section 2.3).

Lemma E.2. *Fix the finite alphabets $\mathcal{Z}_1$ and $\mathcal{Z}_2$ and the distribution $P_{Z_1} \in \mathcal{P}(\mathcal{Z}_1)$. For $l = 1, \dots, L$, let $a_l(z_1)$ be a function with mean $\phi_l \triangleq \sum_{z_1} P_{Z_1}(z_1) a_l(z_1)$, and define the set*

$$\mathcal{A} \triangleq \left\{ \tilde{P}_{Z_1} \in \mathcal{P}(\mathcal{Z}_1) : \mathbb{E}_{\tilde{P}}[a_l(Z_1)] = \phi_l, l = 1, \dots, L \right\}. \quad (\text{E.3})$$

(i) *For any non-negative function $f(z_1, z_2)$, distribution $P_{Z_2} \in \mathcal{P}(\mathcal{Z}_2)$, and constant β , the optimization*

$$\min_{\substack{\tilde{P}_{Z_1 Z_2} : \tilde{P}_{Z_1} \in \mathcal{A}, \tilde{P}_{Z_2} = P_{Z_2}, \\ \mathbb{E}_{\tilde{P}}[\log f(Z_1, Z_2)] \geq \beta}} D(\tilde{P}_{Z_1 Z_2} \| P_{Z_1} \times P_{Z_2}) \quad (\text{E.4})$$

has the same value as

$$\sup_{\lambda \geq 0, \{\theta_l\}} - \sum_{z_2} P_{Z_2}(z_2) \log \sum_{z_1} P_{Z_1}(z_1) f(z_1, z_2)^\lambda e^{\sum_{l=1}^L \theta_l (a_l(z_1) - \phi_l)} + \lambda \beta, \quad (\text{E.5})$$

where the supremum over each θ_l is over all real numbers.

(ii) *For any non-negative function $g(z_1, z_2)$ and conditional distribution $P_{Z_2|Z_1} \in \mathcal{P}(\mathcal{Z}_2|\mathcal{Z}_1)$, the optimization*

$$\min_{\tilde{P}_{Z_1 Z_2} : \tilde{P}_{Z_1} \in \mathcal{A}} D(\tilde{P}_{Z_1 Z_2} \| P_{Z_1} \times P_{Z_2|Z_1}) - \mathbb{E}_{\tilde{P}}[\log g(Z_1, Z_2)] \quad (\text{E.6})$$

has the same value as

$$\sup_{\{\theta_l\}} - \log \sum_{z_1, z_2} P_{Z_1}(z_1) P_{Z_2|Z_1}(z_2|z_1) g(z_1, z_2) e^{\sum_{l=1}^L \theta_l (a_l(z_1) - \phi_l)}. \quad (\text{E.7})$$

(iii) *For any non-negative function $f(z_1, z_2)$, distribution $P_{Z_1} \in \mathcal{P}(\mathcal{Z}_1)$, conditional distribution $P_{Z_2|Z_1} \in \mathcal{P}(\mathcal{Z}_2|\mathcal{Z}_1)$, and constant β , the optimization*

$$\min_{\substack{\tilde{P}_{Z_1 Z_2} : \tilde{P}_{Z_1} = P_{Z_1}, \\ \mathbb{E}_{\tilde{P}}[\log f(Z_1, Z_2)] \geq \beta}} D(\tilde{P}_{Z_1 Z_2} \| P_{Z_1} \times P_{Z_2|Z_1}) \quad (\text{E.8})$$

has the same value as

$$\sup_{\lambda \geq 0} - \sum_{z_1} P_{Z_1}(z_1) \log \sum_{z_2} P_{Z_2|Z_1}(z_2|z_1) f(z_1, z_2)^\lambda + \lambda \beta. \quad (\text{E.9})$$

(iv) For any non-negative function $g(z_1, \bar{z}_1)$, the optimization

$$\min_{\substack{\tilde{P}_{Z_1 \bar{Z}_1} : \tilde{P}_{Z_1} \in \mathcal{A}, \tilde{P}_{\bar{Z}_1} \in \mathcal{A} \\ D(\tilde{P}_{Z_1 \bar{Z}_1} \| P_{Z_1} \times P_{\bar{Z}_1}) \leq R}} -\mathbb{E}_{\tilde{P}}[\log g(Z_1, \bar{Z}_1)] + D(\tilde{P}_{Z_1 \bar{Z}_1} \| P_{Z_1} \times P_{\bar{Z}_1}) - R \quad (\text{E.10})$$

has the same value as

$$\sup_{\lambda \geq 0, \{\theta_l\}, \{\bar{\theta}_l\}} -(1 + \lambda) \log \sum_{z_1, \bar{z}_1} P_{Z_1}(z_1) P_{\bar{Z}_1}(\bar{z}_1) \left(g(z_1, \bar{z}_1) e^{\sum_l \theta_l (a_l(z_1) - \phi_l)} e^{\sum_l \bar{\theta}_l (a_l(\bar{z}_1) - \phi_l)} \right)^{\frac{1}{1+\lambda}} - (1 + \lambda)R. \quad (\text{E.11})$$

Proof. The proof is similar to that of Lemma E.1, and is thus omitted. $\square$

The following lemma will allow certain convex optimization problems to be expressed in a form where, after some simple manipulations, Lemma E.1 can be applied.

Lemma E.3. Fix a positive integer d , and let $\mathcal{A}$ be a convex subset of $\mathbb{R}^d$. Let $f(z)$, $g(z)$, $g_1(z)$ and $g_2(z)$ be convex functions mapping $\mathbb{R}^d$ to $\mathbb{R}$ such that

$$g_1(z) + g_2(z) \leq g(z) \quad (\text{E.12})$$

for all $z \in \mathcal{A}$. Then the optimization

$$\min_{z \in \mathcal{A}} f(z) + \left[\max \{g_1(z), g_2(z), g(z)\} \right]^+ \quad (\text{E.13})$$

has the same value as

$$\max \left\{ \min_{z \in \mathcal{A}} f(z) + \left[\max \{g_1(z), g(z)\} \right]^+, \min_{z \in \mathcal{A}} f(z) + \left[\max \{g_2(z), g(z)\} \right]^+ \right\}. \quad (\text{E.14})$$

Proof. See Section E.2. $\square$

E.1 Proof of Lemma E.1

The Lagrangian [38, Sec. 5.1.1] of the optimization problem in (E.1) is given by

$$\begin{aligned} \mathcal{L} = \sum_{z_1, z_2} \tilde{P}_{Z_1 Z_2}(z_1, z_2) & \left(\log \frac{\tilde{P}_{Z_1 Z_2}(z_1, z_2)}{P_{Z_1}(z_1)P_{Z_2}(z_2)} - \log g(z_1, z_2) - \lambda \log f(z_1, z_2) \right) \\ & + \sum_{z_1} \mu_1(z_1)(P_{Z_1}(z_1) - \tilde{P}_{Z_1}(z_1)) + \sum_{z_2} \mu_2(z_2)(P_{Z_2}(z_2) - \tilde{P}_{Z_2}(z_2)) + \lambda\beta, \end{aligned} \quad (\text{E.15})$$

where $\lambda \geq 0$, $\mu_1(\cdot)$ and $\mu_2(\cdot)$ are Lagrange multipliers. Since the objective in (E.1) is convex and the constraints are affine, the optimal value equals $\mathcal{L}$ for some choices of $\tilde{P}_{Z_1 Z_2}$ and the Lagrange multipliers satisfying the Karush-Kuhn-Tucker (KKT) conditions [38, Sec. 5.5.3].

We proceed to simplify (E.15) using the KKT conditions. Setting $\frac{\partial \mathcal{L}}{\partial \tilde{P}_{Z_1 Z_2}(z_1, z_2)} = 0$ yields

$$1 + \log \frac{\tilde{P}_{Z_1 Z_2}(z_1, z_2)}{P_{Z_1}(z_1)P_{Z_2}(z_2)f(z_1, z_2)^\lambda g(z_1, z_2)} - \mu_1(z_1) - \mu_2(z_2) = 0, \quad (\text{E.16})$$

and hence

$$\tilde{P}_{Z_1 Z_2}(z_1, z_2) = P_{Z_1}(z_1)P_{Z_2}(z_2)f(z_1, z_2)^\lambda g(z_1, z_2)e^{-1+\mu_1(z_1)+\mu_2(z_2)}. \quad (\text{E.17})$$

Applying the constraint $\tilde{P}_{Z_2} = P_{Z_2}$ yields

$$\sum_{\bar{z}_1} P_{Z_1}(\bar{z}_1)f(\bar{z}_1, z_2)^\lambda g(\bar{z}_1, z_2)e^{\mu_1(\bar{z}_1)} = e^{1-\mu_2(z_2)}, \quad (\text{E.18})$$

and hence

$$\mu_2(z_2) = 1 - \log \sum_{\bar{z}_1} P_{Z_1}(\bar{z}_1)f(\bar{z}_1, z_2)^\lambda g(\bar{z}_1, z_2)e^{\mu_1(\bar{z}_1)}. \quad (\text{E.19})$$

Substituting (E.16) into (E.15) yields

$$\mathcal{L} = -1 + \sum_{z_1} \mu_1(z_1)P_{Z_1}(z_1) + \sum_{z_2} \mu_2(z_2)P_{Z_2}(z_2) + \lambda\beta. \quad (\text{E.20})$$

Applying (E.19) and taking the supremum over $\lambda \geq 0$ and $\mu_1(\cdot)$, we obtain (E.2).

Since we took the supremum over $\lambda \geq 0$ and $\mu_1(\cdot)$ without verifying that the resulting parameters satisfy the KKT conditions, we have only proved that (E.2) is an upper bound to (E.1). To obtain a matching lower bound, we follow [9, Appendix A] and make use of the

log-sum inequality [3, Thm. 2.7.1]: For non-negative sequences $\{\alpha_i\}_{i=1}^n$ and $\{\beta_i\}_{i=1}^n$,

$$\sum_{i=1}^n \alpha_i \log \frac{\alpha_i}{\beta_i} \geq \left(\sum_{i=1}^n \alpha_i \right) \log \frac{\sum_{i=1}^n \alpha_i}{\sum_{i=1}^n \beta_i}. \quad (\text{E.21})$$

For any $\tilde{P}_{Z_1 Z_2}$ satisfying the constraints in (E.1), we can lower bound the objective in (E.1) as follows:

$$\sum_{z_1, z_2} \tilde{P}_{Z_1 Z_2}(z_1, z_2) \log \frac{\tilde{P}_{Z_1 Z_2}(z_1, z_2)}{P_{Z_1}(z_1) P_{Z_2}(z_2) g(z_1, z_2)} \quad (\text{E.22})$$

$$\geq \sum_{z_1, z_2} \tilde{P}_{Z_1 Z_2}(z_1, z_2) \log \frac{\tilde{P}_{Z_1 Z_2}(z_1, z_2)}{P_{Z_1}(z_1) P_{Z_2}(z_2) f(z_1, z_2)^\lambda g(z_1, z_2)} + \lambda \beta \quad (\text{E.23})$$

$$= \sum_{z_1, z_2} \tilde{P}_{Z_1 Z_2}(z_1, z_2) \log \frac{\tilde{P}_{Z_1 Z_2}(z_1, z_2)}{P_{Z_1}(z_1) P_{Z_2}(z_2) f(z_1, z_2)^\lambda g(z_1, z_2) e^{\mu_1(\bar{z}_1)}} + \sum_{z_1} P_{Z_1}(z_1) \mu_1(z_1) + \lambda \beta, \quad (\text{E.24})$$

where (E.23) holds for any $\lambda \geq 0$ due to the constraint $\mathbb{E}_{\tilde{P}}[\log f(Z_1, Z_2)] \geq \beta$, and (E.24) holds for any $\mu_1(\cdot)$ by a simple expansion of the logarithm. Applying (E.21) to the summation over z_1 in (E.24) and using the constraint $\tilde{P}_{Z_2} = P_{Z_2}$, we can lower bound (E.24) by the objective in (E.2). Since $\lambda \geq 0$ and $\mu_1(\cdot)$ are arbitrary, the proof is complete.

E.2 Proof of Lemma E.3

We define the functions

$$\Phi_0(z) \triangleq f(z) + [g(z)]^+ \quad (\text{E.25})$$

$$\Phi_1(z) \triangleq f(z) + \left[\max \{g_1(z), g(z)\} \right]^+ \quad (\text{E.26})$$

$$\Phi_2(z) \triangleq f(z) + \left[\max \{g_2(z), g(z)\} \right]^+. \quad (\text{E.27})$$

Since $f(\cdot)$, $g(\cdot)$, $g_1(\cdot)$ and $g_2(\cdot)$ are convex by assumption, it follows from the composition rules in [38, Sec. 3.2.4] that $\Phi_0(\cdot)$, $\Phi_1(\cdot)$ and $\Phi_2(\cdot)$ are also convex.

We wish to show that

$$\min_{z \in \mathcal{A}} \max \{ \Phi_1(z), \Phi_2(z) \} = \max \left\{ \min_{z \in \mathcal{A}} \Phi_1(z), \min_{z \in \mathcal{A}} \Phi_2(z) \right\}. \quad (\text{E.28})$$

We define the regions

$$\mathcal{R}_1 = \{z : \Phi_1(z) > \Phi_0(z)\} \quad (\text{E.29})$$

$$\mathcal{R}_2 = \{z : \Phi_2(z) > \Phi_0(z)\}. \quad (\text{E.30})$$

The key observation is that $\mathcal{R}_1$ and $\mathcal{R}_2$ are disjoint. To see this, we observe from (E.25)–(E.27) that any $z \in \mathcal{R}_1 \cap \mathcal{R}_2$ must satisfy $g_1(z) > g(z)$ and $g_2(z) > g(z)$. When combined with (E.12), these imply $g_1(z) < 0$ and $g_2(z) < 0$, and it follows from (E.25)–(E.27) that $\Phi_0(z) = \Phi_1(z) = \Phi_2(z) = f(z)$, in contradiction with the assumption that $z \in \mathcal{R}_1 \cap \mathcal{R}_2$. Thus, $\mathcal{R}_1 \cap \mathcal{R}_2 = \emptyset$, and hence $g_1(z)$ and $g_2(z)$ cannot simultaneously be the unique maximizers in (E.26) and (E.27) respectively. Combining this observation with (E.25), we obtain

$$\Phi_0(z) = \min \{\Phi_1(z), \Phi_2(z)\}. \quad (\text{E.31})$$

To prove (E.28), we use a proof by contradiction. Let the left-hand side and right-hand side be denoted by f^* and $\tilde{f}^*$ respectively. It is easily seen that $f^* \geq \tilde{f}^*$ always holds, so we assume that $f^* > \tilde{f}^*$. Let z_1^* and z_2^* minimize Φ_1 and Φ_2 respectively on the right-hand side of (E.28), so that

$$\tilde{f}^* = \max \{\Phi_1(z_1^*), \Phi_2(z_2^*)\}. \quad (\text{E.32})$$

The assumption $f^* > \tilde{f}^*$ implies that

$$\Phi_2(z_1^*) > \Phi_1(z_1^*) \quad (\text{E.33})$$

$$\Phi_1(z_2^*) > \Phi_2(z_2^*). \quad (\text{E.34})$$

Next, we define

$$\hat{\Phi}_0(\lambda) \triangleq \Phi_0(\lambda z_1^* + (1 - \lambda)z_2^*) \quad (\text{E.35})$$

$$\hat{\Phi}_1(\lambda) \triangleq \Phi_1(\lambda z_1^* + (1 - \lambda)z_2^*) \quad (\text{E.36})$$

$$\hat{\Phi}_2(\lambda) \triangleq \Phi_2(\lambda z_1^* + (1 - \lambda)z_2^*) \quad (\text{E.37})$$

for $\lambda \in [0, 1]$. Since any convex function is also convex when restricted to a straight line [38, Section 3.1.1], it follows that $\hat{\Phi}_0$, $\hat{\Phi}_1$ and $\hat{\Phi}_2$ are convex in λ . From (E.33)–(E.34), we have

$$\hat{\Phi}_2(1) > \hat{\Phi}_1(1) \quad (\text{E.38})$$

$$\hat{\Phi}_1(0) > \hat{\Phi}_2(0). \quad (\text{E.39})$$

Since $\hat{\Phi}_1$ and $\hat{\Phi}_2$ are convex, they are also continuous (at least in the region that they are finite), and it follows that the two must intersect somewhere in $(0, 1)$, say at λ^* . Therefore,

$$\hat{\Phi}_0(\lambda^*) = \min \{ \hat{\Phi}_1(\lambda^*), \hat{\Phi}_2(\lambda^*) \} \quad (\text{E.40})$$

$$= \max \{ \hat{\Phi}_1(\lambda^*), \hat{\Phi}_2(\lambda^*) \} \quad (\text{E.41})$$

$$\geq \min_{z \in \mathcal{A}} \max \{ \Phi_1(z), \Phi_2(z) \} \quad (\text{E.42})$$

$$= f^*, \quad (\text{E.43})$$

where (E.40) follows from (E.31). Finally, we have the following contradiction:

- Combining (E.43) with the assumption that $f^* > \tilde{f}^*$, we have

$$\hat{\Phi}_0(\lambda^*) > \tilde{f}^* = \max \{ \hat{\Phi}_1(1), \hat{\Phi}_2(0) \}, \quad (\text{E.44})$$

where the equality follows from (E.32).

- From (E.31), we have $\hat{\Phi}_0(\lambda) = \min \{ \hat{\Phi}_1(\lambda), \hat{\Phi}_2(\lambda) \}$, and it follows from (E.38)–(E.39) that $\hat{\Phi}_0(1) = \hat{\Phi}_1(1)$ and $\hat{\Phi}_0(0) = \hat{\Phi}_2(0)$. Hence, using the convexity of $\hat{\Phi}_0$ and Jensen's inequality, we have

$$\hat{\Phi}_0(\lambda^*) \leq \lambda^* \hat{\Phi}_1(1) + (1 - \lambda^*) \hat{\Phi}_2(0) \quad (\text{E.45})$$

$$\leq \max \{ \hat{\Phi}_1(1), \hat{\Phi}_2(0) \}. \quad (\text{E.46})$$

Bibliography

- [1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. Journal*, vol. 27, pp. 379–423, July and Oct. 1948.
- [2] R. Gallager, *Information Theory and Reliable Communication*. John Wiley & Sons, 1968.
- [3] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 2001.
- [4] A. Feinstein, "A new basic theorem of information theory," *IRE Prof. Group. on Inf. Theory*, vol. 4, no. 4, pp. 2–22, Sept. 1954.
- [5] N. Alon and J. Spencer, *The Probabilistic Method*. John Wiley & Sons, Inc., 2011.
- [6] I. Csiszár and J. Körner, "Graph decomposition: A new key to coding theorems," *IEEE Trans. Inf. Theory*, vol. 27, no. 1, pp. 5–12, Jan. 1981.
- [7] J. Hui, "Fundamental issues of multiple accessing," Ph.D. dissertation, MIT, 1983.
- [8] I. Csiszár and P. Narayan, "Channel capacity for a given decoding metric," *IEEE Trans. Inf. Theory*, vol. 45, no. 1, pp. 35–43, Jan. 1995.
- [9] N. Merhav, G. Kaplan, A. Lapidoth, and S. Shamai, "On information rates for mismatched decoders," *IEEE Trans. Inf. Theory*, vol. 40, no. 6, pp. 1953–1967, Nov. 1994.
- [10] V. Balakirsky, "Coding theorem for discrete memoryless channels with given decision rule," in *Algebraic Coding*. Springer Berlin / Heidelberg, 1992, vol. 573, pp. 142–150.
- [11] A. Ganti, A. Lapidoth, and E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.
- [12] P. Elias, "Coding for two noisy channels," in *Third London Symp. Inf. Theory*, 1955.
- [13] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [14] A. Lapidoth and S. Shamai, "Fading channels: how perfect need "perfect side information" be?" *IEEE Trans. Inf. Theory*, vol. 48, no. 5, pp. 1118–1134, May 2002.

- [15] H. Weingarten, Y. Steinberg, and S. Shamai, "Gaussian codes and weighted nearest neighbor decoding in fading multiple-antenna channels," *IEEE Trans. Inf. Theory*, vol. 50, no. 8, pp. 1665–1686, Aug. 2004.
- [16] A. Lapidoth, "Nearest neighbor decoding for additive non-Gaussian noise channels," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1520–1529, Sept. 1996.
- [17] J. Salz and E. Zehavi, "Decoding under integer metrics constraints," *IEEE Trans. Comms.*, vol. 43, no. 234, pp. 307–317, Feb. 1995.
- [18] N. Binshatok and S. Shamai, "Integer metrics for binary input symmetric output memoryless channels," *IEEE Trans. Comms.*, vol. 47, no. 11, pp. 1636–1645, 1999.
- [19] A. Guillén i Fàbregas, A. Martinez, and G. Caire, "Bit-interleaved coded modulation," *Found. Trends Comms. Inf. Theory*, vol. 5, no. 1-2, pp. 1–153, Jan. 2008.
- [20] A. Martinez, A. Guillén i Fàbregas, G. Caire, and F. Willems, "Bit-interleaved coded modulation revisited: A mismatched decoding perspective," *IEEE Trans. Inf. Theory*, vol. 55, no. 6, pp. 2756–2765, June 2009.
- [21] J. Jaldén, P. Fertl, and G. Matz, "On the generalized mutual information of BICM systems with approximate demodulation," in *IEEE Inf. Theory. Workshop*, Jan. 2010.
- [22] T. T. Nguyen and L. Lampe, "Bit-interleaved coded modulation with mismatched decoding metrics," *IEEE Trans. Comms.*, vol. 59, no. 2, pp. 437–447, Feb. 2011.
- [23] L. Peng, A. Guillén i Fàbregas, and A. Martinez, "Mismatched shaping schemes for bit-interleaved coded modulation," in *IEEE Int. Symp. Inf. Theory*, Cambridge, MA, 2012.
- [24] L. Peng, A. Fabregas, and A. Martinez, "Improved exponents and rates for bit-interleaved coded modulation," in *IEEE Int. Symp. Inf. Theory*, July 2013.
- [25] M. Oizumi, T. Ishii, K. Ishibashi, T. Hosoya, and M. Okada, "Mismatched decoding in the brain," *J. Neuroscience*, vol. 30, no. 13, pp. 4815–4826, March 2010.
- [26] R. A. Ince, R. Senatore, E. Arabzadeh, F. Montani, M. E. Diamond, and S. Panzeri, "Information-theoretic methods for studying population codes," *Neural Networks*, vol. 23, no. 6, pp. 713–727, 2010.
- [27] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.
- [28] I. Csiszár and J. Körner, "Many coding theorems follow from an elementary combinatorial lemma," in *Proc. 3rd Czech.-Sov.-Hung. Sem. Inf. Theory*, 1980.
- [29] M. Fossorier, M. Mihaljević, and H. Imai, "Reduced complexity iterative decoding of low-density parity check codes based on belief propagation," *IEEE Trans. Comms.*, vol. 47, no. 5, pp. 673–680, May 1999.
- [30] I. Stiglitz, "Coding for a class of unknown channels," *IEEE Trans. Inf. Theory*, vol. 12, no. 2, pp. 189–195, April 1966.

- [31] D. Kazakos, "Upper and lower bounds for noisy channel coding under mismatch," in *Proc. Conf. Info. Sci. Sys.*, John Hopkins Univ., 1981, pp. 37–42.
- [32] D. Divsalar, "Performance of mismatched receivers on bandlimited channels," Ph.D. dissertation, Univ. California L.A., 1978.
- [33] T. Fischer, "Some remarks on the role of inaccuracy in Shannon's theory of information transmission," in *Trans. 8th Prague Conf. on Inf. Theory*, 1971, pp. 211–226.
- [34] V. Balakirsky, "A converse coding theorem for mismatched decoding at the output of binary-input memoryless channels," *IEEE Trans. Inf. Theory*, vol. 41, no. 6, pp. 1889–1902, Nov. 1995.
- [35] A. Lapidoth, "Mismatched decoding and the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1439–1452, Sept. 1996.
- [36] G. Kaplan and S. Shamai, "Information rates and error exponents of compound channels with application to antipodal signaling in a fading environment," *Arch. Elek. Über.*, vol. 47, no. 4, pp. 228–239, 1993.
- [37] S. Shamai and I. Sason, "Variations on the Gallager bounds, connections, and applications," *IEEE Trans. Inf. Theory*, vol. 48, no. 12, pp. 3029–3051, Dec. 2002.
- [38] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [39] A. Somekh-Baruch, "A general formula for the mismatch capacity," <http://arxiv.org/abs/1309.7964>.
- [40] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, July 1994.
- [41] E. Abbe, M. Medard, S. Meyn, and L. Zheng, "Finding the best mismatched detector for channel coding and hypothesis testing," in *Inf. Theory. Apps. Workshop*, Feb. 2007.
- [42] A. Somekh-Baruch, "On achievable rates for channels with mismatched decoding," 2013, <http://arxiv.org/abs/1305.0547>.
- [43] C. Hucher and P. Sadeghi, "Mismatched decoding for the relay channel," in *Proc. Int. Zurich Sem. on Comms.*, 2010.
- [44] A. Lapidoth, "On the role of mismatch in rate distortion theory," *IEEE Trans. Inf. Theory*, vol. 43, no. 1, pp. 38–47, Jan. 1997.
- [45] J. Chen, D. He, and A. Jagmohan, "On the duality between Slepian-Wolf coding and channel coding under mismatched decoding," *IEEE Trans. Inf. Theory*, vol. 55, no. 9, pp. 4006–4018, Sept. 2009.
- [46] N. Merhav, "Universal decoding for arbitrary channels relative to a given class of decoding metrics," *IEEE Trans. Inf. Theory*, vol. 59, no. 9, pp. 5566–5576, Sept. 2013.

- [47] A. Asyhari and A. Guillén i Fàbregas, “Nearest neighbor decoding in MIMO block-fading channels with imperfect CSIR,” *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1483–1517, March 2012.
- [48] A. Lapidoth and S. Shamai, “A lower bound on the bit-error-rate resulting from mismatched Viterbi decoding,” *European Trans. Telecom.*, vol. 9, no. 6, pp. 473–482, 1998.
- [49] M. Alsan, “A lower bound on achievable rates by polar codes with mismatch polar decoding,” in *IEEE Inf. Theory Workshop*, Seville, 2013.
- [50] M. Alsan and E. Telatar, “Polarization as a novel architecture to boost the classical mismatched capacity of B-DMCs,” Jan. 2014, <http://arxiv.org/abs/1401.6097>.
- [51] R. Fano, *Transmission of information: A statistical theory of communications*. MIT Press, 1961.
- [52] R. Gallager, “A simple derivation of the coding theorem and some applications,” *IEEE Trans. Inf. Theory*, vol. 11, no. 1, pp. 3–18, Jan. 1965.
- [53] C. E. Shannon, R. Gallager, and E. Berlekamp, “Lower bounds to error probability for coding on discrete memoryless channels. I,” *Information and Control*, vol. 10, no. 1, pp. 65–103, 1967.
- [54] —, “Lower bounds to error probability for coding on discrete memoryless channels. II,” *Information and Control*, vol. 10, no. 5, pp. 522–552, 1967.
- [55] I. Csiszár, J. Körner, and K. Marton, “A new look at the error exponent of discrete memoryless channels,” in *IEEE Int. Symp. Inf. Theory*, Ithaca, NY, 1977.
- [56] R. Gallager, “Fixed composition arguments and lower bounds to error probability,” <http://web.mit.edu/gallager/www/notes/notes5.pdf>.
- [57] —, “The random coding bound is tight for the average code,” *IEEE Trans. Inf. Theory*, vol. 19, no. 2, pp. 244–246, March 1973.
- [58] A. G. D’yachkov, “Bounds on the average error probability for a code ensemble with fixed composition,” *Prob. Inf. Transm.*, vol. 16, no. 4, pp. 3–8, 1980.
- [59] G. Poltyrev, “Random coding bounds for discrete memoryless channels,” *Prob. Inf. Transm.*, vol. 18, no. 1, pp. 9–21, 1982.
- [60] V. Strassen, “Asymptotische Abschätzungen in Shannon’s Informationstheorie,” in *Trans. 3rd Prague Conf. on Inf. Theory*, 1962, pp. 689–723, [English Translation: <http://www.math.wustl.edu/~luthy/strassen.pdf>].
- [61] Y. Polyanskiy, H. V. Poor, and S. Verdú, “Channel coding rate in the finite blocklength regime,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [62] M. Hayashi, “Information spectrum approach to second-order coding rate in channel coding,” *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.
- [63] Y. Altuğ and A. B. Wagner, “Moderate deviations in channel coding,” 2012, <http://arxiv.org/abs/1208.1924>.

- [64] Y. Polyanskiy and S. Verdú, “Channel dispersion and moderate deviations limits for memoryless channels,” in *Allerton Conf. on Comm., Control and Comp.*, 2010.
- [65] J. L. Jensen, *Saddlepoint Approximations*. Oxford University Press, 1995.
- [66] Y. Liu and B. Hughes, “A new universal random coding bound for the multiple-access channel,” *IEEE Trans. Inf. Theory*, vol. 42, no. 2, pp. 376–386, March 1996.
- [67] R. Nomura and T. S. Han, “Second-order Slepian-Wolf coding theorems for non-mixed and mixed sources,” in *IEEE Int. Symp. Inf. Theory*, Istanbul, 2013.
- [68] 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.
- [69] J. Scarlett, L. Peng, N. Merhav, A. Martinez, and A. Guillén i Fàbregas, “Expurgated random-coding ensembles: Exponents, refinements and connections,” 2014, *IEEE Trans. Inf. Theory*, to appear [Online: <http://arxiv.org/abs/1307.6679>].
- [70] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, “Ensemble-tight error exponents for mismatched decoders,” in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, Oct. 2012.
- [71] —, “Cost-constrained random coding and applications,” in *Inf. Theory and Apps. Workshop*, San Diego, CA, Feb. 2013.
- [72] J. Scarlett, L. Peng, N. Merhav, A. Martinez, and A. Guillén i Fàbregas, “Expurgated random-coding ensembles: Exponents, refinements and connections,” in *Int. Zurich Sem. on Comms.*, Feb. 2014.
- [73] N. Shulman, “Communication over an unknown channel via common broadcasting,” Ph.D. dissertation, Tel Aviv University, 2003.
- [74] C. Stone, “On local and ratio limit theorems,” in *Proc. Fifth Berkeley Symp. Math. Stat. Prob.*, 1965, pp. 217–224.
- [75] H. Rubin and J. Sethuraman, “Probabilities of moderate deviations,” *Indian Journal of Stats.*, vol. 27, no. 2, pp. 325–346, Dec. 1965.
- [76] W. Feller, *An introduction to probability theory and its applications*, 2nd ed. John Wiley & Sons, 1971, vol. 2.
- [77] J. Löfberg, “YALMIP : A toolbox for modeling and optimization in MATLAB,” in *Proc. CACSD Conf.*, Taipei, 2004.
- [78] K. Fan, “Minimax theorems,” *Proc. Nat. Acad. Sci.*, vol. 39, pp. 42–47, 1953.
- [79] R. L. Dobrushin, “Asymptotic estimates of the probability of error for transmission of messages over a discrete memoryless communication channel with a symmetric transition probability matrix,” *Theory Prob. Apps.*, vol. 7, no. 3, pp. 270–300, 1962.
- [80] C. E. Shannon, “Probability of error for optimal codes in a gaussian channel,” *Bell Syst. Tech. Journal*, vol. 38, no. 3, pp. 611–656, 1959.

- [81] Y. Altuğ and A. B. Wagner, “Refinement of the random coding bound,” 2014, <http://arxiv.org/abs/1312.6875>.
- [82] —, “Refinement of the sphere-packing bound: Asymmetric channels,” *IEEE Trans. Inf. Theory*, vol. 60, no. 3, pp. 1592–1614, March 2013.
- [83] —, “Refinement of the sphere packing bound for symmetric channels,” in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, 2011.
- [84] Y. Altuğ, “Moderate deviations and exact asymptotics in channel coding,” Ph.D. dissertation, Cornell University, 2013.
- [85] Y. Polyanskiy, “Channel coding: Non-asymptotic fundamental limits,” Ph.D. dissertation, Princeton University, 2010.
- [86] P. Moulin, “The log-volume of optimal codes for memoryless channels, within a few nats,” 2013, <http://arxiv.org/abs/1311.0181>.
- [87] —, “The log-volume of optimal constant-composition codes for memoryless channels, within $\mathcal{O}(1)$ bits,” in *IEEE Int. Symp. Inf. Theory*, Boston, MA, 2012.
- [88] Y. Altuğ and A. B. Wagner, “The third-order term in the normal approximation for singular channels,” 2013, <http://arxiv.org/abs/1309.5126>.
- [89] M. Tomamichel and V. Tan, “A tight upper bound for the third-order asymptotics for most discrete memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7041–7051, Nov 2013.
- [90] V. Y. F. Tan and M. Tomamichel, “The third-order term in the normal approximation for the AWGN channel,” 2013, <http://arxiv.org/abs/1311.2337>.
- [91] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, “A derivation of the asymptotic random-coding prefactor,” in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, 2013.
- [92] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, “The saddlepoint approximation: Unified random coding asymptotics for fixed and varying rates,” 2014, accepted to *IEEE Int. Symp. Inf. Theory*.
- [93] J. Scarlett, A. Martinez, and A. G. i Fàbregas, “Refinements of the third-order term in the fixed error asymptotics of constant-composition codes,” 2014, submitted to *IEEE Inf. Theory Workshop*.
- [94] A. Martinez and A. Guillén i Fàbregas, “Saddlepoint approximation of random-coding bounds,” in *Inf. Theory Apps. Workshop*, San Diego, CA, 2011.
- [95] R. Bahadur and R. Ranga Rao, “On deviations of the sample mean,” *Annals Math. Stats.*, vol. 31, pp. 1015–1027, Dec. 1960.
- [96] Y. Altuğ and A. B. Wagner, “A refinement of the random coding bound,” in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, 2012.

- [97] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 881–903, Feb. 2014.
- [98] V. Kostina and S. Verdú, "Channels with cost constraints: Strong converse and dispersion," in *IEEE Int. Symp. Inf. Theory*, Istanbul, 2013.
- [99] J. Mineka and S. Silverman, "A local limit theorem and recurrence conditions for sums of independent non-lattice random variables," *Annals Math. Stats.*, vol. 41, no. 2, pp. 592–600, April 1970.
- [100] V. V. Petrov, *Sums of Independent Random Variables*. Springer-Verlag, 1975.
- [101] C. Esseen, "On the concentration function of a sum of independent random variables," *Zeit. für Wahr. und Verwandte Gebiete*, vol. 9, no. 4, pp. 290–308, 1968.
- [102] R. Ahlswede, "Multi-way communication channels," in *Int. Symp. Inf. Theory*, Tsaghkadzor, 1971.
- [103] H. Liao, "Multiple-access channels," Ph.D. dissertation, Dept. Elec. Eng. Univ. Hawaii, Honolulu, 1972.
- [104] A. Dyachkov, "Random constant composition codes for multiple access channels," *Prob. Contr. Inf. Theory*, vol. 13, pp. 357–369, 1984.
- [105] R. Gallager, "A perspective on multiaccess channels," *IEEE Trans. Inf. Theory*, vol. 31, no. 2, pp. 124–142, March 1985.
- [106] J. Pokorný and H. Wallmeier, "Random coding bound and codes produced by permutations for the multiple-access channel," *IEEE Trans. Inf. Theory*, vol. 31, no. 6, pp. 741–750, Nov. 1985.
- [107] A. Nazari, A. Anastasopoulos, and S. S. Pradhan, "Error exponent for multiple-access channels: Lower bounds," 2010, <http://arxiv.org/abs/1010.1303>.
- [108] W. Hoeffding, "A combinatorial central limit theorem," *Annals Math. Stats.*, vol. 22, no. 4, pp. 558–566, 1951.
- [109] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Multiuser coding techniques for mismatched decoding," 2013, submitted to *IEEE Trans. Inf. Theory* [Online: <http://arxiv.org/abs/1311.6635>].
- [110] J. Scarlett and A. Guillén i Fàbregas, "An achievable error exponent for the mismatched multiple-access channel," in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, Oct. 2012.
- [111] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "The mismatched multiple-access channel: General alphabets," in *IEEE Int. Symp. Inf. Theory*, Istanbul, 2013.
- [112] —, "Mismatched multi-letter successive decoding for the multiple-access channel," 2014, accepted to *IEEE Int. Symp. Inf. Theory*.

- [113] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, “Second-order rate region of constant-composition codes for the multiple-access channel,” 2013, <http://arxiv.org/abs/1303.6167>.
- [114] H. Yagi, “Finite blocklength bounds for multiple access channels with correlated sources,” in *Int. Symp. Inf. Theory and Apps.*, 2012, pp. 377–381.
- [115] A. El Gamal and Y. H. Kim, *Network Information Theory*. Cambridge University Press, 2011.
- [116] G. Poltyrev, “Random coding bounds for some broadcast channels,” *Prob. Peredachi Inf.*, vol. 19, no. 1, pp. 9–20, 1983.
- [117] M. Bierbaum and H. Wallmeier, “A note on the capacity region of the multiple-access channel,” *IEEE Trans. Inf. Theory*, vol. 25, no. 4, pp. 484–484, July 1979.
- [118] R. Etkin, N. Merhav, and E. Ordentlich, “Error exponents of optimum decoding for the interference channel,” *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 40–56, Jan. 2010.
- [119] Y. Huang and P. Moulin, “Finite blocklength coding for multiple access channels,” in *IEEE Int. Symp. Inf. Theory*, Boston, MA, July 2012.
- [120] E. MolavianJazi and J. N. Laneman, “Simpler achievable rate regions for multiaccess with finite blocklength,” in *IEEE Int. Symp. Inf. Theory*, Boston, MA, July 2012.
- [121] J. Scarlett and V. Y. F. Tan, “Second-order asymptotics for the Gaussian MAC with degraded message sets,” 2013, submitted to *IEEE Trans. Inf. Theory* [Online: <http://arxiv.org/abs/1310.1197>].
- [122] S.-Q. Le, V. Y. F. Tan, and M. Motani, “A case where interference does not affect the channel dispersion,” <http://arxiv.org/abs/1404.0255>.
- [123] E. Haim, Y. Kochman, and U. Erez, “A note on the dispersion of network problems,” in *IEEE Conv. Elec. Eng. in Israel*, 2012.
- [124] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, “Second-order rate region of constant-composition codes for the multiple-access channel,” in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, 2013.
- [125] V. Y. F. Tan, “Achievable second-order coding rates for the wiretap channel,” in *IEEE Int. Conf. Comm. Sys.*, Singapore, 2012, pp. 65–69.
- [126] E. MolavianJazi and J. N. Laneman, “A finite-blocklength perspective on Gaussian multi-access channels,” <http://arxiv.org/abs/1309.2343>.
- [127] A. Grant, B. Rimoldi, R. Urbanke, and P. Whiting, “Rate-splitting multiple access for discrete memoryless channels,” *IEEE Trans. Inf. Theory*, vol. 47, no. 3, pp. 873–890, 2001.
- [128] N. Merhav, “Error exponents of erasure/list decoding revisited via moments of distance enumerators,” *IEEE Trans. Inf. Theory*, vol. 54, no. 10, pp. 4439–4447, Oct. 2008.

- [129] A. Somekh-Baruch and N. Merhav, "Exact random coding exponents for erasure decoding," *IEEE Trans. Inf. Theory*, vol. 57, no. 10, pp. 6444–6454, Oct. 2011.
- [130] W. Loh, "On Latin hypercube sampling," *Annals of Stats.*, vol. 24, no. 5, pp. 2058–2080, 1996.
- [131] E. Bolthausen, "An estimate of the remainder in a combinatorial central limit theorem," *Prob. Theory and Rel. Fields*, vol. 66, pp. 379–386, 1984.
- [132] E. Bolthausen and F. Götze, "The rate of convergence for multivariate sampling statistics," *Annals of Stats.*, vol. 21, no. 4, pp. 1692–1710, 1993.
- [133] Y. Wu and S. Verdú, "The impact of constellation cardinality on Gaussian channel capacity," in *Allerton Conf. on Comm., Control and Comp.*, Monticello, IL, 2010.
- [134] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 23, no. 1, pp. 60–64, Jan. 1977.
- [135] R. Gallager, "Capacity and coding for degraded broadcast channels," *Prob. Peredachi Inf.*, vol. 10, no. 3, pp. 3–14, 1974.
- [136] J. Scarlett, A. Martinez, and A. Guillén i Fàbregas, "Superposition codes for mismatched decoding," in *IEEE Int. Symp. Inf. Theory*, Istanbul, 2013.
- [137] J. Körner and A. Sgarro, "Universally attainable error exponents for broadcast channels with degraded message sets," *IEEE Trans. Inf. Theory*, vol. 26, no. 6, pp. 670–679, Nov. 1980.
- [138] R. Ahlswede, N. Cai, and Z. Zhang, "Erasure, list, and detection zero-error capacities for low noise and a relation to identification," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 55–62, Jan. 1996.
- [139] A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*. Boston, MA: Jones and Bartlett, 1992.
- [140] R. Ahlswede, "An elementary proof of the strong converse theorem for the multiple-access channel," *Journal Comb. Inf. and Sys. Sci.*, vol. 7, no. 3, pp. 216–230, 1982.
- [141] S. Verdú and S. Shamai, "Variable-rate channel capacity," *IEEE Trans. Inf. Theory*, vol. 56, no. 6, pp. 2651–2667, June 2010.
- [142] J. Körner and A. Orlitsky, "Zero-error information theory," *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2207–2229, Oct. 1998.
- [143] S. Arimoto, "An algorithm for computing the capacity of arbitrary discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 18, no. 1, pp. 14–20, Jan. 1972.
- [144] R. Blahut, "Computation of channel capacity and rate-distortion functions," *IEEE Trans. Inf. Theory*, vol. 18, no. 4, pp. 460–473, July 1972.
- [145] D. de Caen, "A lower bound on the probability of a union," *Discrete Math.*, vol. 169, pp. 217–220, 1997.