Network Source Coding in Scenarios with Uncertainty at the Decoder

Thesis by

Siming Sun

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2026
Defended April 8, 2026

Siming Sun
ORCID: 0000-0003-1314-7485

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to an EAS-Institute Fellowship, funding from the Oringer Fellowship Fund, funding from the Yunni and Maxine Pao Graduate Fellowship Fund, a grant from the Caltech Center for Sensing to Intelligence, and funding from the California Institute of Technology for the financial support that made this research possible.

I would like to thank my advisor, Professor Effros, for her insightful guidance and unwavering support throughout my journey at Caltech. I am profoundly grateful for the intellectual rigor she brought to our discussions. From her, I have learned the importance of structural logic and the craft of academic writing. I would also like to thank Professor Kostina, Professor Hassibi and Professor Low for serving on my committee and for their valuable insights.

Furthermore, I wish to thank Professor Effros, Professor Kostina, and Professor Vaidyanathan for the opportunity to serve as a teaching assistant under their mentorship, which allowed me to develop a deeper appreciation for the field through teaching.

I also want to thank my labmates Dr. Yavas, Yuxin, Zihao, Pranav and Rosa.

A very special thanks goes to my friends from the neighbouring research group - Mingshu, Siyuan, Oumeng, Zhenyu, Haowen, Max, Steven, Ninghe, Shi and Siying.

Finally, I want to express my deepest gratitude to my parents. Your unconditional love, patience, and constant belief in me have been my greatest source of strength throughout this journey.

ABSTRACT

In this thesis, I investigate four network source coding problems in scenarios with uncertainty at the decoder.

In network communications, uncertainty may arise from both extrinsic forces such as a noisy environment and intrinsic forces such as the choices of participating agents. For example, both imperfect coordination capabilities and deliberate removal of synchronization requirements lead to uncertainty about the relative timing of messages from different communicators. Similarly, errors in source code reconstructions can arise either due to channel code failure or because compression algorithms are designed to tolerate a certain amount of error for the sake of improving rate or latency. Since distributed source codes have the potential to benefit from source dependencies but these dependencies may be damaged or destroyed by prior source coding decisions, the existence of reconstruction errors is another potential source of uncertainty in network communications. My thesis investigates the performance of communication systems under uncertainties that include the examples described above.

I first propose an asynchronous random access source code (ARASC) and present bounds on its performance. In traditional random access source codes (RASCs), the source blocks from different terminals are aligned, meaning that the codeword transmissions begin and end at the same time. ARASC removes the synchronization constraint, allowing each encoder to encode and start a codeword transmission any time it has available information to describe and the channel is idle. In this work, I show that ARASC can take advantage of source dependence without enforcing synchrony; I propose an ARASC algorithm, bound the second-order rate of the proposed strategy, and show simulation results that explore the performance tradeoffs between RASC and ARASC.

I then introduce the problem of source coding with unreliable side information. This problem models the scenario where a decoder is tasked with reconstructing a source with the help of side information that is not necessarily reliable. I propose a code that uses a 2-stage decoding strategy that prevents the uncertainty in the side information from damaging the reliability of the source reconstruction while taking advantage of the side information.

I next present bounds on the performance of an almost lossless point-to-point source

code in scenarios where the description from the given code is later used to aid the description of another dependent source. I demonstrate that, even under strict efficiency and reliability constraints, point-to-point source codes are not equally powerful when the decoder outputs are used as side information to help a downstream user decode a dependent source, and the difference is big enough to affect the downstream user in a significant way. I show the full range of possible performances and prove that the designer of the point-to-point code can control where on the spectrum, from the least to the most helpful, the reconstruction is expected to be. Furthermore, I show that, if the designer wishes to provide more information rather than less about a dependent source to a downstream user through the source reconstruction, they can do so without observing the joint distribution with the later source. As a step towards practicality, I also show that it is possible to generate a linear point-to-point code that is helpful to later users if the designer accepts a small rate penalty.

Finally, I present the performance of a random binning source code when the source reconstruction from that code is used as an input to a distributed hypothesis test. I show that random binning generates a source code whose codeword serves simultaneously as a source description and as an input to a distributed hypothesis test that seeks to determine whether the coded source is independent of or dependent on side information available to the decoder. The given random binning code can serve these dual roles without significantly worsening the source coding rate. When the joint distribution between sources is uncertain, such a code offers a way to distinguish between possible joint distributions and to decode only if the source description rate and hypothesis regarding dependency together render decoding appropriate. I also discuss the implications of such a hypothesis test on the statistics of the random binning code output, analyze the statistics through an alternative route that involves direct analysis of the random binning code, and compare the results.

PUBLISHED CONTENT AND CONTRIBUTIONS

[1] S. Sun and M. Effros, "Distributed hypothesis testing and the output statistics of random binning," in *Proceedings of the 61st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Available at IDEALS: https://hdl.handle.net/2142/130280, Champaign, IL, USA, 2025,
Siming Sun participated in the conception of the project and prepared the manuscript.

[2] S. Sun and M. Effros, "Code design in almost lossless onion peeling data compression," in *2025 Data Compression Conference (DCC)*, 2025, pp. 293–302. DOI: 10.1109/DCC62719.2025.00037,
Siming Sun participated in the conception of the project and prepared the manuscript.

[3] S. Sun and M. Effros, "Almost lossless onion peeling data compression," in *2024 Data Compression Conference (DCC)*, 2024, pp. 472–481. DOI: 10.1109/DCC58796.2024.00055,
Siming Sun participated in the conception of the project and prepared the manuscript.

[4] S. Sun and M. Effros, "Asynchronous random access data compression," in *2024 Data Compression Conference (DCC)*, 2024, pp. 482–491. DOI: 10.1109/DCC58796.2024.00056,
Siming Sun participated in the conception of the project and prepared the manuscript.

[5] S. Sun and M. Effros, "Source coding with unreliable side information in the finite blocklength regime," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 222–227. DOI: 10.1109/ISIT50566.2022.9834490,
Siming Sun participated in the conception of the project and prepared the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Chapter 1

INTRODUCTION

1.1 Source Coding for Network Communication

Shannon's 1948 paper [1] considers the problem of describing a single source with a short description that allows for the reconstruction of that source. When the probability that the reconstruction differs from the described source converges to 0 as the number of source observations described in one sitting grows without bound, the description-generating and reconstruction-building algorithms (known as the encoder and decoder) are together called *lossless* data compression. A fixed-length scenario is illustrated in Figure 1.1(a).

When there are more than two nodes that communicate with one another, the nodes and communication links between them form a communication *network*. Each node may observe and encode one or many sources, transmit encoded or unencoded descriptions through communication links, receive such descriptions, and attempt reconstruction.

One simple example of a communication network comprises three nodes, with node $\mathcal{A}$ and node $\mathcal{B}$ each observing and compressing a source simultaneously, and node C receiving compressed descriptions from $\mathcal{A}$ and $\mathcal{B}$ and reconstructing both sources. A straightforward way to operate this 3-node communication network is to use the technique for compressing a single source. Specifically, the coding requirement can be satisfied by designing a single-source compression algorithm for each source and having C perform the reconstructions of the sources separately. However, this is not the most efficient solution in general.

In practice, sources observed at two nodes in a shared environment are often dependent. For example, two sensors in close spatial proximity may collect dependent measurements of their shared environment. If the dependence is fully exploited, the descriptions needed by C to reconstruct the sources can be shorter than those needed using a pair of single-source algorithms. In the ideal case, the redundancy between the two sources does not need to be described repeatedly.

The 1973 paper by Slepian and Wolf [2] shows that if two encoders independently compress two dependent sources and a single decoder jointly reconstructs both

sources (see Figure 1.1(c)), there can be significant performance gain compared to using a point-to-point code for each source (Figure 1.1(b)). The best possible performance is equivalent to what would be achieved if the two sources are encoded jointly (Figure 1.1(d)). The problems investigated in this thesis are variations of the Slepian-Wolf scenario.

1.2 Slepian-Wolf Code and its Variations

Let $\mathcal{A}$ and $\mathcal{B}$ observe a pair of discrete sources $(X_1, Y_1), (X_2, Y_2), \ldots$. When the sources are stationary and memoryless, [2] proves that asymptotically lossless compression is achievable as long as the compression rate pair (R_X, R_Y) satisfies

$$R_X > H(X|Y) \tag{1.1}$$

$$R_Y > H(Y|X) \tag{1.2}$$

$$R_X + R_Y > H(X,Y) \tag{1.3}$$

where $H(X|Y)$, $H(Y|X)$ and $H(X,Y)$ are conditional and joint entropies of the two sources.

This result assumes that the two sources are synchronized on both the *symbol* level and the *block* level. This means that nodes $\mathcal{A}$ and $\mathcal{B}$ share the same clock and always observe and encode source samples simultaneously, making source block start- and end-times agree exactly. However, synchronization is not always possible, and even when it is, it may not be the best choice in terms of rate and latency.

In [3], Willems investigates a 2-encoder, 1-decoder scenario without the assumption of block synchronization in the asymptotic regime. The notion of asynchrony in [3] is similar to one studied in the channel coding literature [4]. Under Willems' asynchrony model, the first source blocks from the two transmitters are offset by a fixed amount and, since each transmitter is assumed to be sending an unending stream of back-to-back blocks, that fixed offset remains unchanged from one block to the next. This fixed asynchrony model is also used in [5] and [6], where the decoder can use a finite number of each transmitter's codewords from prior and upcoming blocks in decoding the current block. To avoid the long decoding delay that results from using future blocks in decoding, Matsuta and Uyematsu [7] and Merhav [8] consider, in the error exponent regime, only one asynchronous block from each of the system's two transmitters and show that, in this framework, it is no longer possible to achieve the Slepian-Wolf region in general.

Aside from the problem of asynchrony, in some communication networks, there

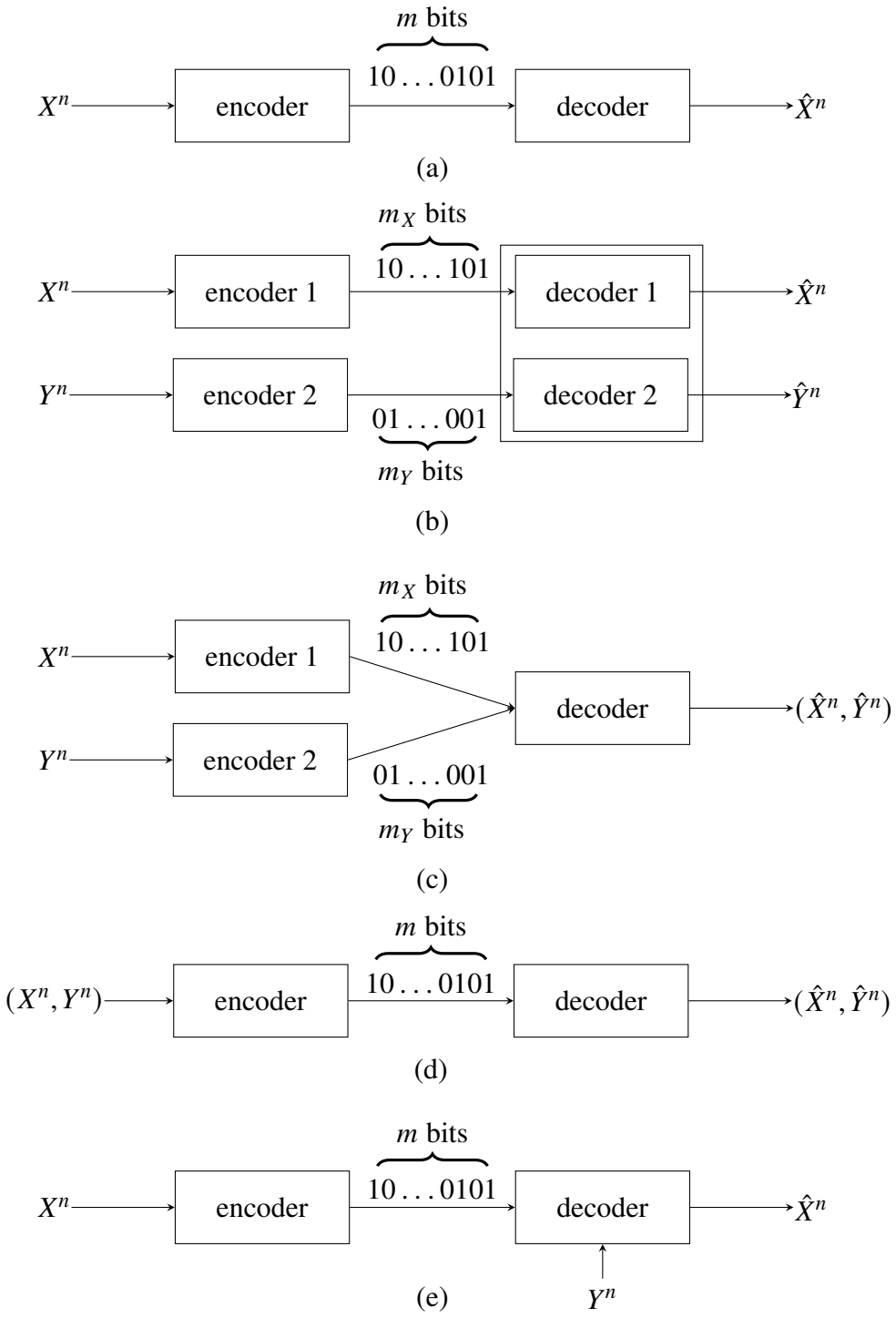


Figure 1.1: Block diagrams of (a) fixed-length point-to-point block source code, (b) a pair of independent codes, (c) fixed-length Slepian-Wolf (multiple access) block source code, (d) a single joint code, and (e) fixed-length block source code with side information.

may be occasions when $\mathcal{A}$ has the need and ability to observe its source, compress it, and send the description to the receiver C while $\mathcal{B}$ does not; similarly, it may sometimes occur that $\mathcal{B}$ can participate but $\mathcal{A}$ cannot or neither can participate. This scenario may arise due to a failure at one or both sensors or the fact that one or both channels to receiver C is occupied by other transmission tasks. To give encoding nodes the opportunity to decide whether to encode and transmit or to remain silent, [9] proposes a variation of the Slepian-Wolf code called a random access source code (RASC). The RASC allows each encoding node to choose whether to encode or not at the start of each coding period, but still assumes that the coding periods are synchronized. That is, $\mathcal{A}$ and $\mathcal{B}$ need to make the decision to encode at the same time. [1]

Another variation of the Slepian-Wolf problem is a special case of it. As an extreme case of the Slepian-Wolf scenario, the decoder node C has access to the source Y. It may use Y as a helper to reduce the rate when reconstructing X from its encoded description. This problem is referred to as source coding with side information (see Figure 1.1(e)). Variations of this problem include having rate-limited side information [14], side information available at both the encoder and the decoder [15], etc. There is also a rich literature on the role of side information in lossy data compression (e.g. [16], [17], [18]). This thesis focuses on lossless data compression.

Even when both sources in a multi-encoder, single-decoder communication network need to be decoded, they can have inherent differences in their nature and therefore making the need to decode them asymmetric. For example, suppose that $\mathcal{A}$ and $\mathcal{B}$ are close enough, with $\mathcal{A}$ observing the temperature and $\mathcal{B}$ observing the level of solar radiation. Dependence exists due to the spatial proximity and the nature of observed data streams. Temperature is a core variable in nearly all meteorological models and forecasts, but the need for solar radiation data can be less frequent [19]. Users may want to be able to reconstruct the temperature from its compressed description more efficiently without having to consult the compressed description of the solar radiation. For Slepian-Wolf code, in general, the decoding node C needs to receive encoded descriptions from both $\mathcal{A}$ and $\mathcal{B}$ to reconstruct any of the two sources, but the process can be made sequential by choosing to operate at a corner point of the Slepian-Wolf rate region. The decoder at node C reconstructs the

[1] In this thesis, the term "random access" refers to the encoder's autonomy to decide whether to encode and transmit an encoded description. It is similar to the random access channel coding problem in [10] but distinct from the meaning of "random access" in privacy and access control context (e.g., [11], [12], [13]), which refers to retrieving pieces of information without accessing the entire compressed file.

source observed at $\mathcal{A}$ solely from the description received from $\mathcal{A}$. To reconstruct the source observed at $\mathcal{B}$, it consults both the description received from $\mathcal{B}$ and the available reconstruction of the other source. The practice of using the decoder output of one code as a helper to another code is often referred to as onion peeling, rate slicing, or stripping.

In the rest of Chapter 1, I provide an overview of the four problems studied in this thesis, all of which stem from the variations of the Slepian-Wolf source code discussed above. Furthermore, common to each problem is the concept of *almost lossless* source coding, a setting where the source code's error probability is bounded by a constant. [2]

1.3 Notation and Definitions

Below, I introduce notation used broadly through this dissertation. Notation that is restricted to specific contexts is defined at the point of use.

Unless stated otherwise, I employ the following conventions. I use upper case letters for random variables and lower case letters for realizations. Sans-serif letters represent functions and operators (e.g., deterministic function $\mathsf{f}(\cdot)$, random function $\mathsf{F}(\cdot)$). Times new roman letters specify sources (e.g., source realization x, random source X) and constants (e.g., constant C). Vectors are expressed using either superscripts to emphasize the dimension (e.g., x^n and X^n) or bold font to simplify notation (e.g., $\mathbf{x}$ and $\mathbf{X}$). Blackboard bold letters capture expectations and measures (e.g., $\mathbb{E}$, $\mathbb{P}$, $\mathbb{U}$). Subscripts clarify the randomness associated with an expectation or measure when the information is not clear from context (e.g., $\mathbb{E}_{P_X}$). Calligraphic letters describe sets (e.g., $\mathcal{A}$) and events. Whether a set is random or deterministic is either stated in the text or clear from context. The complement of set $\mathcal{A}$ is denoted by $\mathcal{A}_{\mathsf{c}}$.

Let P and Q be two probability mass functions (p.m.f.s) on support $\mathcal{X}$. I use $\|P - Q\|_{\mathrm{TV}}$ to denote the total variation distance between the distributions, $\|P - Q\|_p$, $p > 0$ to denote the L^p-distance, and $D(P||Q)$ to denote the KL-divergence between

[2] The study of almost lossless source codes—where the error probability is bounded by a constant—is a component of the finite blocklength regime. Within this framework, research focuses on characterizing the coding rate as a function of blocklength, often by isolating the dispersion term. This line of work originated with Strassen [20] and includes foundational and recent contributions such as [9], [10], [21], [22], [23], [24], [25], [26], [27], [28], [29].

P and Q, giving

$$\|P-Q\|_{\mathrm{TV}} \triangleq \max_{\mathcal{A}\subseteq\mathcal{X}} |P(\mathcal{A})-Q(\mathcal{A})| = \frac{1}{2}\sum_{x\in\mathcal{X}} |P(x)-Q(x)|$$

$$\|P-Q\|_p \triangleq \left(\sum_{x\in\mathcal{X}} |P(x)-Q(x)|^p\right)^{\frac{1}{p}}$$

$$D(P||Q) \triangleq \mathbb{E}_P\left[\log\frac{P(X)}{Q(X)}\right] = \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)}.$$

Let $P_{X|Y}, Q_{X|Y}$ be two conditional p.m.f.s on alphabet $\mathcal{X}\times\mathcal{Y}$, and let W_Y be a p.m.f. on $\mathcal{Y}$. I use $D(P_{X|Y}||Q_{X|Y}|P_Y)$ to denote the conditional KL-divergence, giving

$$\begin{aligned} D(P_{X|Y}||Q_{X|Y}|W_Y) &\triangleq \mathbb{E}_{P_{X|Y}W_Y}\left[\log\frac{P_{X|Y}(X|Y)}{Q_{X|Y}(X|Y)}\right] \\ &= \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} P_{X|Y}(x|y)W_Y(y)\log\frac{P_{X|Y}(x|y)}{Q_{X|Y}(x|y)}. \end{aligned}$$

All instances of "log" assume base 2 while "ln" uses base e. The base of "exp" can be inferred from context.

Suppose a pair of discrete random variables, $(X,Y)\in\mathcal{X}\times\mathcal{Y}$, are distributed according to P_{XY}. Then, P_X and P_Y denote the marginal distributions on $\mathcal{X}$ and $\mathcal{Y}$, respectively, $P_{X|Y}$ and $P_{Y|X}$ denote the conditional distributions of X given Y and Y given X, respectively. The following notation defines the information $\imath$, entropy H, varentropy V, and third centered moment of information T in single-variate, conditional, and multi-variate forms, and defines the information density i and mutual information I.

$$\begin{aligned}
&\imath(x) \triangleq \log\frac{1}{P_X(x)} && H(X) \triangleq \mathbb{E}[\imath(X)] \\
&V(X) \triangleq \mathrm{Var}[\imath(X)] && T(X) \triangleq \mathbb{E}[|\imath(X)-H(X)|^3] \\
&\imath(x|y) \triangleq \log\frac{1}{P_{X|Y}(x|y)} && H(X|Y) \triangleq \mathbb{E}[\imath(X|Y)] \\
&V(X|Y) \triangleq \mathrm{Var}[\imath(X|Y)] && T(X|Y) \triangleq \mathbb{E}[|\imath(x|y)-H(X|Y)|^3] \\
&\imath(x,y) \triangleq \log\frac{1}{P_{XY}(x,y)} && H(X,Y) \triangleq \mathbb{E}[\imath(X,Y)] \\
&V(X,Y) \triangleq \mathrm{Var}[\imath(X,Y)] && T(X,Y) \triangleq \mathbb{E}[|\imath(X,Y)-H(X,Y)|^3] \\
&i(x;y) \triangleq \log\frac{P_{XY}(x,y)}{P_X(x)P_Y(y)} && I(X;Y) \triangleq \mathbb{E}[i(X;Y)].
\end{aligned}$$

For positive integer k, let $[k] \triangleq \{1, 2, \ldots, k\}$ and $(k] \triangleq \{0, 1, \ldots, k\}$. For vector $\mathbf{x}$, let $\lfloor \mathbf{x} \rfloor_k$ denote the first k dimensions of $\mathbf{x}$.

Notation $\doteq$ stands for equality to the first order in the exponent. For two pairs (a, b) and (c, d), $(a, b) \prec (c, d)$ means that (a, b) is lexicographically smaller than (c, d).

1.4 Asynchronous Random Access Source Code

Chapter 2 investigates the problem of asynchronous random access source coding (ARASC). It arises from the uncertainty in synchronization patterns of source blocks that distributed encoders encode.

A *synchronous* random access source code (RASC) [9] is a 1-decoder source code with an unknown number of active encoders. Synchrony among encoders is preserved, but time is divided into *epochs* whose starting times are the only opportunities for encoders to decide whether they want to encode. Should the code be used over time, each epoch begins in the time step after the end of the prior epoch. In each epoch, the decoder decodes using only the codeword symbols it receives in that epoch.

Under the RASC model, an encoder that is inactive at the start of one epoch must wait for the current epoch to end before it can begin describing a new source block. The resulting delay can be as much as one time step shorter than the time needed for decoding under the worst-case (independent-decoding) scenario. In delay-sensitive scenarios, one may want to avoid such waiting time. Therefore, in Chapter 2, I propose a new ARASC framework and a code design, and bound the performance of the design.

In the ARASC scenario, as in the Slepian-Wolf and RASC settings, multiple encoders independently encode their respective source observations and send their source descriptions to a common decoder. An encoder becomes active when an opportunity to send a codeword to the decoder arises (e.g., when the device gains access to the communication channel to the decoder). When the opportunity arises, the encoder encodes the latest observed block (a source sequence of length n) and begins sending codeword symbols to the decoder. As in the RASC in [9], the codewords are assumed to be sent through a noiseless channel, and a single bit of feedback is used by the decoder to tell each transmitter when to stop transmitting its current description, yielding a rateless code. In decoding each subsequent source block, the receiver may employ both the as-yet undecoded source descriptions and any past source reconstructions that overlap in time with these source block(s).

To bound the performance achievable under the (block) ARASC framework proposed in this work, I design a random code for a pair of discrete, stationary memoryless sources $\{(X_{1,i}, X_{2,i})\}_{i=1}^{\infty}$. For simplicity, I assume a symmetrical joint source distribution and equal-capacity noiseless channels from each transmitter to the single receiver. Under mild assumptions (see Chapter 2), the proposed code achieves an average rate across blocks

$$\bar{R} \leqslant \bar{H} + \sqrt{\frac{C}{n}} Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right), \tag{1.4}$$

with ϵ being the incremental error allowed per decoding and

$$C = \max(V(X_1), V(X_1|X_2)).$$

The first-order term, $\bar{H}$, is the weighted average of $H(X_1)$ (same as $H(X_2)$ due to the symmetrical setting) and $H(X_1, X_2)$, with weights approaching the fractions of time in which one and both sources are encoded; that is, the code asymptotically does as well as a Slepian-Wolf code when multiple sources are described and as well as a point-to-point code when only one source is described, and this performance is achieved despite the *a priori* uncertainty around the number of active transmitters.

I also discuss the performance tradeoffs between RASC and ARASC based on simulation results.

1.5 Source Coding with Unreliable Side Information at the Decoder

In Chapter 3, I study the problem of source coding with unreliable side information at the decoder.

A decoder receives a compressed description of a source block X^n and a side information block $\hat{Y}^n$ that is an unreliable copy of the true side information Y^n which is dependent on X^n with a known joint distribution. The decoder knows that $\hat{Y}^n$ equals Y^n with at least a certain probability. Despite the imperfection of $\hat{Y}^n$, the decoder still hopes to take advantage of the information it contains about X^n to aid in the reconstruction of X^n.

The setup of this problem shares some similarity with the Kaspi problem in lossy source coding [17], [30].

The uncertainty in the side information may arise from a sensor failure, a synchronization mistake, or a decoding error of an upstream code. In Chapter 3, I propose a 2-stage decoding strategy that allows the decoder to decode at a rate close to $H(X|Y)$

when it is confident, and request for more description bits about X^n when it isn't. I analyze the performance of such a 2-stage strategy under two different models of the joint distribution of X^n, Y^n and $\hat{Y}^n$. One model captures the worst case under the assumption that the distribution is known to the decoder. The other captures the general case where the decoder is unaware of the distribution.

1.6 Almost Lossless Onion Peeling Code

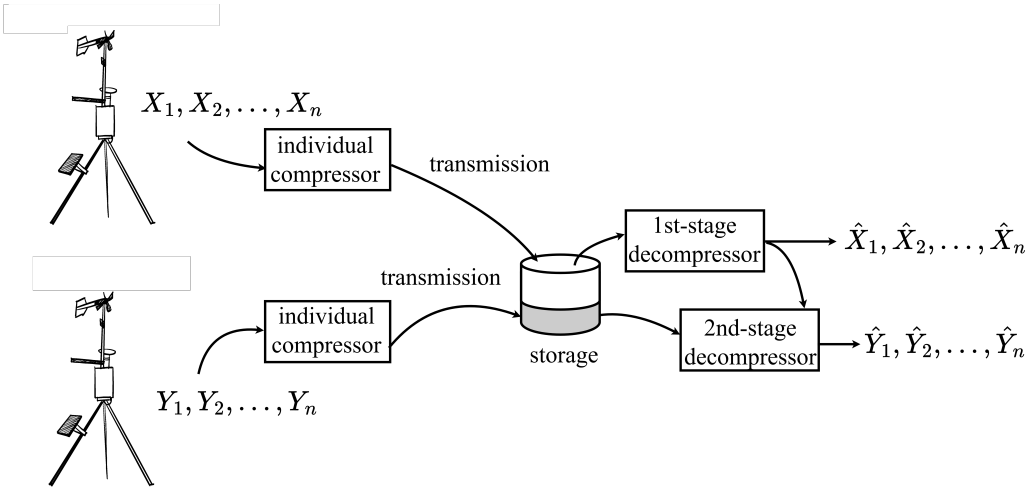


Figure 1.2: Illustration of an onion peeling code for a pair of dependent sources from weather stations.

While Chapter 3 focuses on what a decoder can do when the side information that it can access is an imperfect copy, Chapter 4 looks at the problem from the perspective of a possible party providing such unreliable side information: an upstream code that can error and thereby produce uncertainty for a downstream user.

For a fixed-length point-to-point source code that compresses source $X^n \sim P_X^n$ almost losslessly with error probability at most ϵ, $\epsilon \in (0, 1)$, [28] and [9] show that the best achievable rate is

$$R_n^* = H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} \pm O\left(\frac{1}{n}\right). \tag{1.5}$$

Codes that satisfy (1.5) are not unique. If satisfying the rate constraint and the error probability bound are the only requirements, then one could choose to use any of these codes. However, the codes may have vastly different impact on downstream users who rely on the reconstruction for other tasks.

Consider, for example, a downstream user whose goal is to decode its own source of interest, Y^n, with the help of $\hat{X}^n$, the reconstruction of X^n by an almost lossless

point-to-point code. Figure 1.2 illustrates the scenario. In such a case, it is natural for the uncertainty from the almost-losslessness to pass to downstream users. The downstream user may either accept an error propagation or use a strategy to prevent it (for example, the source code with unreliable side information in Chapter 3). However, it is not always necessary for the downstream user to suffer significantly from the loss of information due to the imperfect reconstruction of X^n.

In Chapter 4, I show that the code design of the almost lossless point-to-point source code can significantly affect the amount of information $\hat{X}^n$ provides about a dependent source Y^n. I prove that, for two sequences of codes for X^n that both satisfy (1.5) and $\mathbb{P}[\hat{X}^n \neq X^n] \leqslant \epsilon$, their respective conditional entropy rate $\lim_{n\longrightarrow\infty} \frac{1}{n} H(Y^n|\hat{X}^n)$ can differ by as much as $\epsilon I(X;Y)$. Their respective conditional spectral sup-entropy rate [31]

$$\inf\left\{\alpha \left| \lim_{n\longrightarrow\infty} \mathbb{P}\left[\frac{1}{n}\log\frac{1}{P_{Y^n|\hat{X}^n}\left(Y^n|\hat{X}^n\right)} > \alpha\right] = 0\right.\right\} \tag{1.6}$$

can differ by as much as $I(X;Y)$. Furthermore, if the code designer knows that the source reconstruction would be used by a downstream user as a side information, they have the authority to design the code in a way that provides exactly the amount of useful information to that downstream user as they wish.

I further prove that it is possible to design a sequence of code such that both the conditional entropy rate and the conditional spectral sup-entropy rate lie on the better side of their respective intervals for universal $P_{Y|X}$. This means that the code designer is able to design the code knowing that the reconstruction would serve well as a helper to the coding of another source but without knowing the exact $P_{Y|X}$.

I also show that, if the designer accepts a small rate penalty, they can design a sequence of linear code that operates at the better side of the two intervals, potentially reducing the coding complexity.

1.7 Distributed Hypothesis Testing and Source Coding

Chapter 5 studies the problem of distributed hypothesis testing and source coding. It is a problem that arises from a decoder being uncertain of whether there is dependence between the source block described by a received codeword and a source block that the decoder itself observes locally.

When the decoder node C has a perfect copy of one of the sources and only needs to decode the other, the copy can be used as a side information to improve the

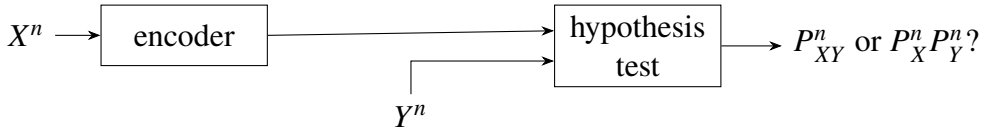
efficiency of the code of the source of interest. However, note that this system can be part of a larger communication network, and communication links can carry other information, for example, the encoded description of a source that is independent of the copy held by C. One possibility is that the encoded description of a source observed by a sufficiently far sensor uses this communication link as a hop on its route to C. As another possibility, C has a perfect copy of the observed temperature from $\mathcal{A}$ and receives an encoded description of the solar radiation data from $\mathcal{B}$, but due to timing issues, it is uncertain whether the compressed solar radiation data is in synchrony with the temperature data or so temporally distant from it that no dependence exists. In this case, C needs to decide whether the received description contains dependent information. If it does, C should decompress the description reliably; if not, C can act accordingly, either requesting more information or passing the description to other nodes in the network that might be able to decode. This problem, investigated in Chapter 5, is the last topic of this thesis.

Before providing an overview of the problem, I briefly describe a closely related problem: distributed hypothesis testing. When the pair of random vectors (X^n, Y^n) can be distributed according to P_{XY}^n or $P_X^n P_Y^n$, a hypothesis test can be used to distinguish between the two distributions. However, the information flowing through communication links is often compressed. To understand the dependence, one needs to decompress the information first or perform hypothesis tests based on the compressed descriptions. In the simplest form of this problem, the source X^n is compressed by an encoder $\mathsf{f}_n : \mathcal{X}^n \mapsto [M_n]$. The hypothesis test $\mathsf{t}_n : [M_n] \times \mathcal{Y}^n \mapsto \{0, 1\}$ observes the compressed description $\mathsf{f}_n(X^n)$ and an uncompressed source Y^n, and outputs a boolean that indicates whether it believes $(X^n, Y^n) \sim P_{XY}^n$ or $(X^n, Y^n) \sim P_X^n P_Y^n$. This problem, often referred to as distributed hypothesis testing or hypothesis testing with communication constraint, is illustrated in Figure 1.3(a).

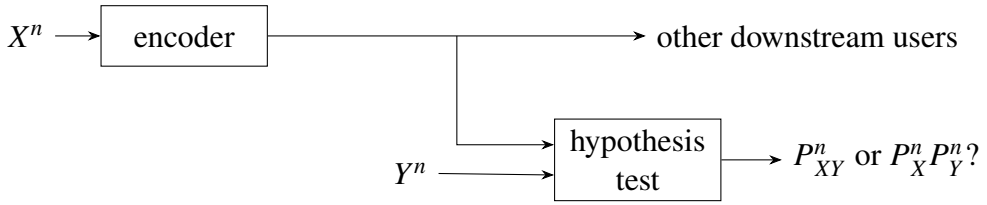
Studies of distributed hypothesis testing include [32], [33], [34], [35], [36]. The strategy of the distributed hypothesis test typically involves generating the compressed description by finding a random variable U that captures some shared information between the two sources, and compressing U^n instead of X^n. While it is possible to set $U = X$, [33] and [35] suggest that this may not yield the best performance. Therefore, in general, if distinguishing between distributions is the only goal, U is chosen to be different from X.

However, in many cases, the hypothesis test is not the only user of the compressed description of X^n. For example, downstream users may use the compressed de-

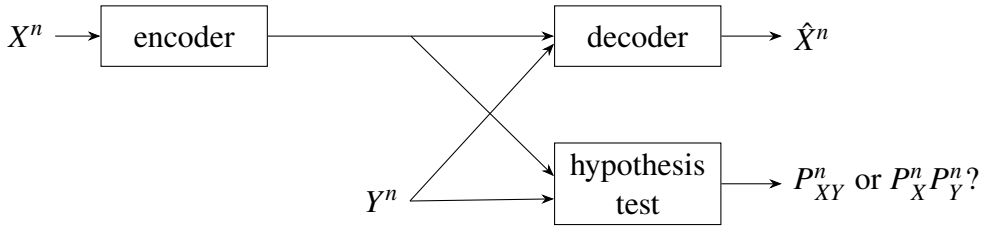
scriptions to perform their own hypothesis tests, help decode another source of their own interest, or attempt to reconstruct X^n itself as described at the beginning of this section. A special case that also involves Y^n is Figure 1.3(c). The decoder has access to Y^n and would like to reliably reconstruct X^n when there is, indeed, a dependency between the two sources. In this case, choosing $U = X$ enables such reconstruction.



(a)



(b)



(c)

Figure 1.3: Block diagrams of (a) a distributed hypothesis test, (b) a distributed hypothesis test where the encoded description of X^n has other downstream users, and (c) a special case of (b) where the downstream user hopes to reconstruct X^n with the help of Y^n if dependence exists.

In Chapter 5, I study the problem of distributed hypothesis testing when the compressed description is also used to reconstruct X^n. I focus on the performance of random binning codebook generation. Random binning is a strategy widely used to prove results in source coding. For example, the achievability bound of the rate

region of the Slepian-Wolf code uses random binning [2]. This codebook generation method picks codewords for each source realization independently and uniformly at random. Figure 1.4 describes this generation process.

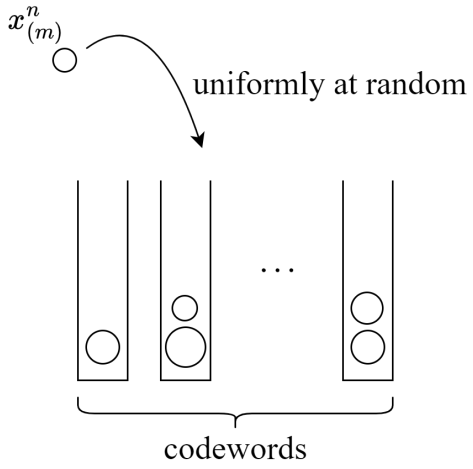


Figure 1.4: Illustration of codeword generation by uniform random binning.

I show that when X^n is compressed using a random binning code $\mathsf{F}_n : \mathcal{X}^n \mapsto [M_n]$ to a rate $R_n \triangleq \frac{\log M_n}{n}$ that is $O(1/\sqrt{n})$ above $H(X|Y)$, the encoded description, combined with Y^n, can be used to perform a hypothesis test against independence. The type-II error of this test decays to zero subexponentially with the blocklength. I explore the implications of this result on the statistics of source descriptions of random binning codes, showing that the expected KL-divergence between $P_{\mathsf{F}_n(X^n)Y^n}$ and $P_{\mathsf{F}_n(X^n)}P_Y^n$ is asymptotically bounded below by $\sqrt{n}$. I extend the result to testing against an alternative joint distribution with identical marginals and higher conditional entropy (less dependence). Finally, I show that, within the given framework, with high probability, no two distributions with identical marginals and conditional entropies Y exceeding the rate are distinguishable.

Chapter 2

ASYNCHRONOUS RANDOM ACCESS SOURCE CODE

The materials in this chapter are published in part as [37].

2.1 Background

The Slepian-Wolf code [2], or multiple access source code (MASC), involves multiple transmitters independently encoding their respective sources and transmitting the encoded descriptions to a common decoder that reconstructs both sources. The transmitters are assumed to be synchronized. That is, they start and stop transmitting their encoded descriptions to the decoder simultaneously. In Section 2.2, I briefly introduce the finite blocklength results on the 2-user MASC by [9], [26]. I provide a refinement of the results that ensures that the encoders are identical. This feature is useful in practice because it removes the need to employ different software/hardware on different devices.

In [9], Chen *et al.* propose a synchronous random access source code (RASC). RASC differs from MASC in that the number of transmitters is unknown *a priori* and may vary over time. The Chen *et al.* RASC divides time into epochs, and each transmitter can choose whether to encode or not at the start of each epoch. Once the decision is made, it cannot be changed until the start of a new epoch. The rate, and therefore length, of an epoch depends on the number of participating transmitters. Figure 2.1 illustrates how the RASC works when there are two transmitters in the network and the codewords are binary; the figure shows the decoder's view of the codeword transmission channel. In the first 5 time slots, neither transmitter decides to encode, resulting in five epochs of length 1. Then, one transmitter decides to participate, resulting in a rate that is close to the best rate achievable for a point-to-point code. After another length-1 epoch resulting from no transmitter deciding to encode, both transmitters decide to encode in the next epoch. The resulting rate is close to the best rate achievable for a MASC. Afterwards, for 3 time steps, no transmitter encodes, and so on.

The RASC removes the need for the number of active transmitters to be fixed and known *a priori*. However, since an ongoing epoch blocks any inactive transmitter from encoding, an inactive encoder can experience a significant delay in the transi-

: epoch with 0 active encoder, length = 1

: epoch with 1 active encoder,
length $\approx nH(X_1) + \sqrt{nV(X_1)}Q^{-1}(\epsilon)$

: epoch with 2 active encoders,
length $\approx \frac{1}{2}\left(nH(X_1, X_2) + \sqrt{nV(X_1, X_2)}Q^{-1}(\epsilon)\right)$

Figure 2.1: Behavior of a 2-user synchronous random access code (RASC) [9] with binary codewords over multiple epochs with different numbers of active transmitters.

tion from inactive to active status. The maximum delay is 1 time slot shorter than the length of the longest active epoch.

In order to reduce such delay, I propose an asynchronous RASC (ARASC) in Section 2.3 of this chapter. The ARASC allows each transmitter to encode whenever it has collected enough source observations for the code blocklength and the channel to the decoder is idle. In Section 2.3, I analyze the proposed code's performance in the finite blocklength regime.

Throughout the remainder of this chapter, I study the canonical two-encoder case, and focus on the case where the joint distribution of the two sources is symmetrical and the encoders are identical.

2.2 Refinement of MASC Result

Finite-blocklength studies on the MASC [26][9] employ the following definition.

Definition 1 (MASC)**.** *An* (M_1, M_2, ϵ) *MASC for a pair of random variables* (X_1, X_2) *defined on discrete alphabet* $\mathcal{X}_1 \times \mathcal{X}_2$ *comprises two separate encoding functions,* $\mathsf{f}_1 : \mathcal{X}_1 \mapsto [M_1]$ *and* $\mathsf{f}_2 : \mathcal{X}_2 \mapsto [M_2]$*, and a decoding function* $\mathsf{g} : [M_1] \times [M_2] \mapsto \mathcal{X}_1 \times \mathcal{X}_2$ *such that the error probability satisfies*

$$\mathbb{P}\left[\mathsf{g}(\mathsf{f}_1(X_1), \mathsf{f}_2(X_2)) \neq (X_1, X_2)\right] \leqslant \epsilon.$$

Definition 2 (Block MASC)**.** *An* (n, M_1, M_2, ϵ) *block MASC is an* (M_1, M_2, ϵ) *MASC for random vectors* $(X_1^n, X_2^n) \in \mathcal{X}_1^n \times \mathcal{X}_2^n$.

For a pair of sources $\{X_1\}$ and $\{X_2\}$, the (n, ϵ) MASC rate region, $\mathscr{R}^*(n, \epsilon)$, is the closure of all (R_1, R_2) pairs such that there exists an (n, M_1, M_2, ϵ) block MASC with

$$\frac{\log M_1}{n} \leqslant R_1, \quad \frac{\log M_2}{n} \leqslant R_2.$$

The best known characterization of $\mathscr{R}^*(n, \epsilon)$ comes from [9].

Theorem 1 (MASC Rate Region - Achievability part of [9, Theorem 20])**.** *For a pair of stationary, memoryless sources with single-letter joint distribution* $P_{X_1X_2}$ *on discrete alphabet* $\mathcal{X}^2$ *satisfying*

$$V(X_1|X_2) > 0, V(X_2|X_1) > 0, V(X_1, X_2) > 0 \tag{2.1}$$

$$T(X_1|X_2) < \infty, T(X_2|X_1) < \infty, T(X_1, X_2) < \infty \tag{2.2}$$

$$\mathbb{E}[|\imath(X_1|X_2) - \mathbb{E}[\imath(X_1|X_2)|X_2]|^3|X_2] < \infty \tag{2.3}$$

$$\mathbb{E}[|\imath(X_2|X_1) - \mathbb{E}[\imath(X_2|X_1)|X_1]|^3|X_1] < \infty, \tag{2.4}$$

the MASC rate region $\mathscr{R}^*(n, \epsilon) \supseteq \mathscr{R}_{\text{in}}^*(n, \epsilon)$*, where*

$$\mathscr{R}_{\text{in}}^*(n, \epsilon) \triangleq \left\{ \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} : \begin{bmatrix} R_1 \\ R_2 \\ R_1 + R_2 \end{bmatrix} \in \overline{\mathscr{R}}^*(n, \epsilon) + O\left(\frac{1}{n}\right)\mathbf{1} \right\}, \tag{2.5}$$

$$\overline{\mathscr{R}}^*(n, \epsilon) \triangleq \begin{bmatrix} H(X_1|X_2) \\ H(X_2|X_1) \\ H(X_1, X_2) \end{bmatrix} + \frac{\mathcal{Q}_{\text{inv}}(\mathsf{V}, \epsilon)}{\sqrt{n}} - \frac{\log n}{2n}\mathbf{1},$$

$\mathcal{Q}_{\text{inv}}$ *is the set-valued multivariate inverse Q-function, and* V *is the covariance matrix of the random vector* $[\imath(X_1|X_2) \;\; \imath(X_2|X_1) \;\; \imath(X_1, X_2)]'$.

A key auxiliary result used to derive this characterization is the random coding union (RCU) bound.

Theorem 2 (MASC RCU Bound [9, Theorem 18])**.** *Given discrete random variables* $(X_1, X_2) \in \mathcal{X}_1 \times \mathcal{X}_2$*, there exists an* (M_1, M_2, ϵ) *MASC that satisfies*

$$\epsilon \leqslant \mathbb{E}[\min\{1, A_1 + A_2 + A_{12}\}],$$

where

$$A_1 \triangleq \frac{1}{M_1}\mathbb{E}\left[\exp(\imath(\bar{X}_1'|X_2))\mathbf{1}\{\imath(\bar{X}_1'|X_2) \leqslant \imath(X_1|X_2)\}|X_1, X_2\right]$$
$$A_2 \triangleq \frac{1}{M_2}\mathbb{E}\left[\exp(\imath(\bar{X}_2'|X_1))\mathbf{1}\{\imath(\bar{X}_2'|X_1) \leqslant \imath(X_2|X_1)\}|X_1, X_2\right]$$
$$A_{12} \triangleq \frac{1}{M_1 M_2}\mathbb{E}\left[\exp(\imath(\bar{X}_1, \bar{X}_2))\mathbf{1}\{\imath(\bar{X}_1, \bar{X}_2) \leqslant \imath(X_1, X_2)\}|X_1, X_2\right]$$

$$P_{X_1X_2\bar{X}_1\bar{X}_2\bar{X}_1'\bar{X}_2'}\left(a, b, \bar{a}, \bar{b}, \bar{a}', \bar{b}'\right) = P_{X_1X_2}(a, b)P_{X_1X_2}\left(\bar{a}, \bar{b}\right)P_{X_1|X_2}\left(\bar{a}'|b\right)P_{X_2|X_1}\left(\bar{b}'|a\right)$$

for all $a, b, \bar{a}, \bar{b}, \bar{a}', \bar{b}'$.

Refinement: The identical encoder case

In scenarios where the observed sources have the same alphabet (for example, scattered sensors measuring the same type of data in a shared environment), one can modify the proof of Theorem 1 to achieve a similar result under the following "identical encoders" framework.

Definition 3 (MASC with identical encoders)**.** *For random variables* (X_1, X_2) *on the same discrete alphabets* $\mathcal{X}_1 = \mathcal{X}_2 = \mathcal{X}$*, an* $(|C|^{m_1}, |C|^{m_2}, \epsilon)$ *MASC with identical encoders and codeword symbol alphabet* C *comprises a single function* $\mathsf{f} : \mathcal{X} \mapsto C^{\max(m_1,m_2)}$ *and a decoding function* $\mathsf{g} : C^{m_1} \times C^{m_2} \mapsto \mathcal{X}^2$ *such that*

$$\mathbb{P}\left[\mathsf{g}\left(\lfloor \mathsf{f}(X_1)\rfloor_{m_1}, \lfloor \mathsf{f}(X_2)\rfloor_{m_2}\right) \neq (X_1, X_2)\right] \leqslant \epsilon.$$

Definition 4 (Block MASC with identical encoders)**.** *An* $(n, |C|^{m_1}, |C|^{m_2}, \epsilon)$ *block MASC with identical encoders and binary codeword symbols is an* $(|C|^{m_1}, |C|^{m_2}, \epsilon)$ *MASC with identical encoders and binary codeword symbols for random vectors* $(X_1^n, X_2^n) \in \mathcal{X}^n \times \mathcal{X}^n$.

The identical encoders framework from Definition 4 allows both the possibility that the encoders operate at the same rate and the possibility that they operate at different rates; the rate at which each encoder communicates is controlled by allowing transmitter j, $j \in [2]$, to transmit only the first m_j symbols of the output of the single encoding function f for any $m_j \leqslant \max(m_1, m_2)$. Having a single f ensures that the transmitters can use the same hardware or software to encode their respective source observations. I use $|C|^{m_1}$ and $|C|^{m_2}$ to emphasize that, in practice, codewords are strings of codeword symbols.

Theorem 3 (MASC rate region with identical encoders)**.** *For a pair of stationary, memoryless sources with single-letter joint distribution* $P_{X_1X_2}$ *on discrete alphabet*

$\mathcal{X}^2$ *satisfying the conditions* (2.1)-(2.4)*, for sufficiently large* n *and any* $0 < \epsilon < 1$*, there exists a constant* $0 < \delta < \infty$ *such that, for any* R_1, R_2 *satisfying*

$$\begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \in \mathscr{R}_{\text{in}}^*(n, \epsilon) + \mathbf{1}\frac{\delta}{n}, \tag{2.6}$$

there exists an $(n, |C|^{m_1}, |C|^{m_2}, \epsilon)$ *MASC code with identical encoders and codeword symbol alphabet* C *such that*

$$\frac{m_1 \log |C|}{n} \leqslant R_1, \quad \frac{m_2 \log |C|}{n} \leqslant R_2.$$

Proof. See Section 2.8. □

The proof uses an updated RCU bound of two abstract sources. This modified RCU bound incorporates the identical encoder constraint.

Theorem 4 (MASC RCU bound with identical encoders)**.** *Given discrete random variables* $(X_1, X_2) \in \mathcal{X}_1 \times \mathcal{X}_2$*, there exists an* $(|C|^{m_1}, |C|^{m_2}, \epsilon)$ *MASC with identical encoders and codeword symbol alphabet* C *with*

$$\epsilon \leqslant \mathbb{E}[\min\{1, A_1 + A_2 + A_{12}\}] + \mathbb{P}[X_1 = X_2],$$

where A_1*,* A_2 *and* A_{12} *are defined as in Theorem 2 with* $M_1 = |C|^{m_1}, M_2 = |C|^{m_2}$*.*

Proof. See Section 2.8. □

As shown in [9, Theorem 20] and reproduced in Theorem 1 above, the achievable rate region for the MASC is known to the third-order accuracy ($O(\log n/n)$). Theorem 3 shows that enforcing the use of identical encoders across transmitters does not damage the performance; that is, the performance bounds in Theorem 3 are identical to those in Theorem 1 without the identical encoder constraint. Since Theorem 1 is tight to within $O(1/n)$ [9], Theorem 3 is tight to within $O(1/n)$ as well.

This result extends to the scenario of $J > 2$ sources satisfying the conditions necessary for finite-blocklength analyses [9]. In that case, the event $\left\{X_1^n = X_2^n\right\}$ in the error probability bound becomes $\left\{\exists j, j' \in [J], j \neq j', X_j^n = X_{j'}^n\right\}$. For constant J, the probability of this event decays exponentially with n.

The RASC result in [9] can be refined in a similar way.

2.3 Asynchronous Random Access Source Code

I hereby state the definition of an ARASC for block codes. In this chapter, I focus on the 2-transmitter case with symmetrical source distribution and codeword symbol alphabet.

A J-transmitter ARASC comprises J independent transmitters and a single receiver. The random variable $X_{j,t}$ is observed at transmitter $j \in [J]$ at time $t \in \mathbb{Z}_+$. There is no direct communication between the transmitters. If an encoding opportunity arises at transmitter j at time t, the transmitter encodes $X_{j,t}^{-n} \triangleq (X_{j,t-n+1}, \ldots, X_{j,t})$ by mapping it to a length $m_{\max}$ string of symbols from code alphabet C_j and begins sending its description to the receiver through a channel that carries $\log |C_j|$ bits per channel use. I assume that encoding opportunities are independent of the source observations. I denote the source block encoded at the kth opportunity at transmitter j by $\mathbf{B}_{j,k}$, and the time that this opportunity arises, called the *encoding time*, by $T_{j,k}$. The transmission stops when the transmitter receives stop feedback from the receiver; denote the time of the stop feedback by $T_{j,k}^f$. Under this framework, from time $T_{j,k}$ to time $T_{j,k}^f$, the transmitter sends the first $m_{j,k} \triangleq T_{j,k}^f - T_{j,k} + 1$ symbols of the encoded description of $\mathbf{B}_{j,k}$ to the receiver. This corresponds to a source coding rate of $R_{j,k} = m_{j,k} \log |C_j|/n$ for source block $\mathbf{B}_{j,k}$. An encoding opportunity can only arise when i) the last transmission has ended (that is, $T_{j,k} > T_{j,k-1}^f$ for $k \geqslant 2$), and ii) there are enough new observations to make a block (that is, $T_{j,k} \geqslant T_{j,k-1} + n$ for $k \geqslant 2$ and $T_{j,1} \geqslant n$).

At the receiver, the decoder observes the encoded source descriptions until it is ready to decode one or more of the received descriptions. When the decoder is ready to decode $\mathbf{B}_{j,k}$ at time $T_{j,k}^f$, it sends a stop feedback to transmitter j and outputs a reconstruction $\hat{\mathbf{B}}_{j,k}$ of $\mathbf{B}_{j,k}$.

When the decoder decodes $\mathbf{B}_{j,k}$ at time $T_{j,k}^f$, it can use the received codeword symbols of $\mathbf{B}_{j,k}$, namely, $\left\lfloor \mathsf{F}\left(\mathbf{B}_{j,k}\right) \right\rfloor_{m_{j,k}}$. For (j', k') such that $\mathbf{B}_{j,k}$ has time overlap with $\mathbf{B}_{j',k'}$ and $(T_{j',k'}, j') \prec (T_{j,k}, j)$, the reconstruction of $\mathbf{B}_{j',k'}$, $\hat{\mathbf{B}}_{j',k'}$, is also available for the decoding of $\mathbf{B}_{j,k}$. The lexicographical comparison ensures that, when observations from different transmitters are encoded at the same time, the observation from a transmitter with the lower index is considered the earlier by convention. For (j', k') such that $\mathbf{B}_{j,k}$ has time overlap with $\mathbf{B}_{j',k'}$ and $(T_{j,k}, j) \prec (T_{j',k'}, j')$, and $T_{j',k'} \leqslant T_{j,k}^f$, the as-yet available codeword symbols, $\left\lfloor \mathsf{f}\left(\mathbf{B}_{j',k'}\right) \right\rfloor_{T_{j,k}^f - T_{j',k'}+1}$, can also be used by the decoder. Intuitively speaking, to reconstruct $\mathbf{B}_{j,k}$, the decoder may use the as-yet transmitted codeword symbols of $\mathbf{B}_{j,k}$, any reconstruction of

earlier dependent observations, and any as-yet transmitted codeword symbols of later observations that have arrived by $T_{j,k}^{f}$.

In the rest of the section, I formally define an asynchronous random access source code tailored for the following scenario (the ***symmetrical scenario***).

Let $J = 2$. Consider dependent source observations $\{(X_{1,t}, X_{2,t})\}, t = 1, 2, \ldots$ distributed i.i.d. on discrete alphabet $\mathcal{X}_1 \times \mathcal{X}_2 = \mathcal{X}^2$ with single-letter joint distribution $P_{X_1X_2}$ that is symmetrical in X_1 and X_2 (i.e., $P_{X_1X_2}(x_1, x_2) = P_{X_1X_2}(x_2, x_1)$ for all $x_1, x_2 \in \mathcal{X}$). Assume that

$$V(X_1) > 0, V(X_1|X_2) > 0, V(X_1, X_2) > 0 \tag{2.7}$$

$$T(X_1) < \infty, T(X_1|X_2) < \infty, T(X_1, X_2) < \infty \tag{2.8}$$

These conditions are similar to (2.1)-(2.4) and are common in finite-blocklength analysis (see, e.g [9], [26], [28]). Fix code alphabets $C_1 = C_2 = C$.

The following definition is useful to the construction of the ARASC code.

Definition 5 (Codebook)**.** *An* $(n, |C|^{m_{\max}})$ *deterministic codebook for a pair of sources* $\{X_1\} = (X_{1,1}, X_{1,2}, \ldots)$ *and* $\{X_2\} = (X_{2,1}, X_{2,2}, \ldots)$ *with identical encoders is a mapping* $\mathsf{f} : \mathcal{X}^n \mapsto C^{m_{\max}}$ *from length-n source sequences from alphabet* $\mathcal{X}$ *to length-*$m_{\max}$ *codewords from alphabet* C*. An* $(n, |C|^{m_{\max}})$ *random codebook is a random variable taking values on the set of* $(n, |C|^{m_{\max}})$ *code.*

Before stating the main result on the performance of ARASC, I will describe the code design that I propose. The code design, in essence, employs a greedy decoder that reconstructs a source block as soon as it has enough information to do so. The information that the decoder may use includes the decoder's latest prior decoding outcome and the latest encoded descriptions from the two transmitters.

The code is designed as follows.

First generate an $(n, |C|^{m_{\max}})$ random codebook by mapping each sequence $x^n \in \mathcal{X}^n$ independently and uniformly at random to a codeword sequence in $C^{m_{\max}}$. I employ the resulting random mapping F as the encoder for both $\{X_1\}$ and $\{X_2\}$. For the k-th source block from the j-th source, the decoder outputs the reconstruction using a decoding function

$$\mathsf{G} : C^{m_{j,k}} \times C^{\max(0, m_{j,k} - d_{j,k})} \times \mathcal{X}^{n - d_{j,k}^{p}} \mapsto \mathcal{X}^n, \tag{2.9}$$

where $m_{j,k}$ is the codelength associated with the block, $d_{j,k}$ and $d^p_{j,k}$ are delays between blocks; I define delays $d_{j,k}$ and $d^p_{j,k}$ in the detailed description below.

The codelength $m_{j,k}$ for source block $\mathbf{B}_{j,k}$ is determined by the encoding time and any blocks from the other transmitter that overlap with $\mathbf{B}_{j,k}$ in time. I next describe these components.

1. Notation $T_{j,k}$ describes the encoding time.
2. Denote the most recent preceding encoded source block from the other transmitter by $\mathbf{B}^p_{j,k} \triangleq \mathbf{B}_{j^c,k^p}$, where $j^c \triangleq 1 \cdot \mathbf{1}\{j = 2\} + 2 \cdot \mathbf{1}\{j = 1\}$ and $k^p \triangleq \max\{k' : (T_{j^c,k'}, j^c) \prec (T_{j,k}, j)\}$. I call $\mathbf{B}^p_{j,k}$ the *predecessor* of $\mathbf{B}_{j,k}$. The comparison of j to j^c in the definition of k^p establishes the transmitter with the lower index to be the earlier transmitter in the case of a tie in transmission times. By convention, if the maximum does not exist, $\mathbf{B}^p_{j,k} = \mathbf{B}_{j^c,-1}$ and $T_{j,-1} = -1$ for $j \in [2]$; that is, define two dummy blocks with index and encoding time -1 for convenience. Denote the encoding time of $\mathbf{B}^p_{j,k}$ by $T^p_{j,k}$.
3. Denote the most recent succeeding encoded source block from the other transmitter by $\mathbf{B}^s_{j,k} \triangleq \mathbf{B}_{j^c,k^s}$, where $k^s \triangleq \min\{k' : (T_{j,k}, j) \prec (T_{j^c,k'}, j^c)\}$. I call $\mathbf{B}^s_{j,k}$ the *successor* of $\mathbf{B}_{j,k}$. By convention, if the minimum does not exist, $\mathbf{B}^s_{j,k} = \mathbf{B}_{j^c,\infty}$, and, for $j \in [2]$, $T_{j,\infty} = \infty$; these dummy blocks with index and encoding time ∞ are defined for convenience. I denote the encoding time of $\mathbf{B}^s_{j,k}$ by $T^s_{j,k}$.

Figure 2.2 depicts one example of the block of interest $\mathbf{B}_{j,k}$, its predecessor $\mathbf{B}^p_{j,k}$, and successor $B^s_{j,k}$ from the encoder j^c. It also illustrates the delay between $\mathbf{B}^p_{j,k}$ and $\mathbf{B}_{j,k}$, $d^p_{j,k}$, and the delay between $\mathbf{B}_{j,k}$ and $\mathbf{B}^s_{j,k}$, $d_{j,k}$.

The decoder uses the following rules to determine the decoding time from its inputs.

1. FIFO: If $(T_{j^c,k'}, j^c) \prec (T_{j,k}, j)$, then $T^f_{j,k} > T^f_{j^c,k'}$.
2. Sufficient Symbols: $m_{j,k} \geqslant \mathsf{m}(d^p_{j,k}, d_{j,k})$, where $d^p_{j,k} \triangleq \min(n, T_{j,k} - T^p_{j,k})$ and $d_{j,k} \triangleq \min(n, T^s_{j,k} - T_{j,k})$.

The function $\mathsf{m} : (n] \times (n] \mapsto [m_{\max}]$ specifies the target codelength as a function of the potential availability of a dependent predecessor and the codeword symbols for a dependent successor.

$$\mathsf{m}(d^p, d) \triangleq \min\left(m_{\mathrm{A}}(d^p), m_{\mathrm{B}}(d^p, d)\right), \tag{2.10}$$

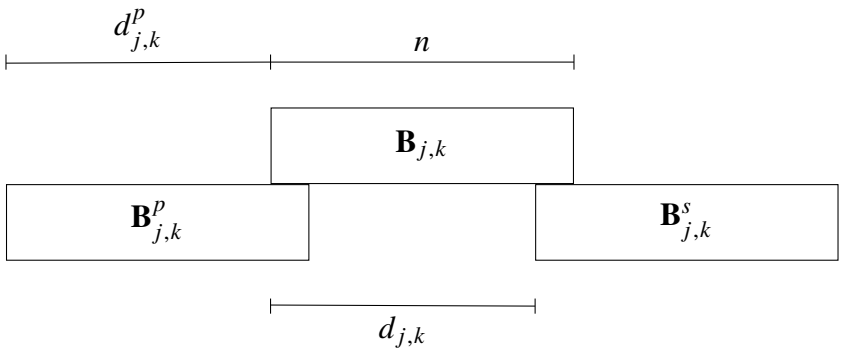


Figure 2.2: Illustration of the predecessor and successor of an encoded source block $\mathbf{B}_{j,k}$ in the asynchronous random access code (ARASC) and the delays between the three source blocks.

where

$$m_{\mathrm{A}}(d^p) \triangleq \min\left\{m \in \mathbb{Z}_+ : m \geqslant \frac{1}{\log|C|}\left(nH_{d^p} + \sqrt{nV_{d^p}}Q^{-1}\left(\epsilon - \frac{C_1}{\sqrt{n}}\right)\right)\right\} \tag{2.11}$$

$$m_{\mathrm{B}}(d^p, d) \triangleq \min\left\{m \in \mathbb{Z}_+ : m \geqslant \frac{1}{2\log|C|}\left(nH'_{d^p,d} + \sqrt{nV'_{d^p,d}}Q^{-1}\left(\epsilon - \frac{C_2}{\sqrt{n}}\right) + d\log|C|\right)\right\}.$$

Here, $0 < \epsilon < 1/2$, C_1 and C_2 are constants, and

$$H_{d^p} \triangleq \frac{n-d^p}{n}H(X_1|X_2) + \frac{d^p}{n}H(X_1) \tag{2.12}$$

$$V_{d^p} \triangleq \frac{n-d^p}{n}V(X_1|X_2) + \frac{d^p}{n}V(X_1) \tag{2.13}$$

$$H'_{d^p,d} \triangleq \frac{n-d^p}{n}H(X_1|X_2) + \frac{2d-d^p}{n}H(X_1) + \frac{n-d}{n}H(X_1,X_2) \tag{2.14}$$

$$V'_{d^p,d} \triangleq \frac{n-d^p}{n}V(X_1|X_2) + \frac{2d-d^p}{n}V(X_1) + \frac{n-d}{n}V(X_1,X_2) \tag{2.15}$$

are the weighted sum of conditional, unconditional, and joint entropies and varentropies.

The decoder relies on its memory of at most one prior decoding outcome in making its next decoding decision. Its knowledge about this prior outcome includes the source reconstruction, the encoding time, and the transmitter from which it originates. Thus, when $\mathbf{B}_{j,k}$ is decoded at $T^f_{j,k}$, the decoder may only use the information from transmitter j^c's prior outcome and the encoded descriptions from transmitter j and j^c since their respective last stop feedback signals.

At time $T_{j,k}^f$, let $\hat{\mathbf{B}}_{j,k}^p$ be transmitter j^c's prior outcome. It is worth noting that $\hat{\mathbf{B}}_{j,k}^p$ may not always be available, as the decoder only stores one prior decoding outcome, and that might be a decoding outcome for transmitter j. The decoder only uses $\hat{\mathbf{B}}_{j,k}^p$ when it is accessible.

The decoder decodes $\mathbf{B}_{j,k}$ using a decoding function introduced in (2.9) with codeword input $\lfloor \mathsf{F}(\mathbf{B}_{j,k}) \rfloor_{m_{j,k}} \in \mathcal{C}^{m_{j,k}}$ describing the transmitted codeword symbols, codeword input $\lfloor \mathsf{F}(\mathbf{B}_{j,k}^s) \rfloor_{m_{j,k}-d_{j,k}} \in \mathcal{C}^{m_{j,k}-d_{j,k}}$ describing the subsequent transmitted codeword symbols from the other decoder if $m_{j,k} > d_{j,k}$ (or a null input otherwise), and the dependent symbols $(\hat{X}_{j^c,T_{j,k}-n+1}, \ldots, \hat{X}_{j^c,T_{j,k}-d_{j,k}^p}) \in \mathcal{X}^{n-d_{j,k}^p}$ from the prior decoder output (or a null input if there is no overlapping prior decoding output). The decoder output $\hat{X}_{j,T_{j,k}}^{-n} \in \mathcal{X}^n$ is the reconstruction of block $\mathbf{B}_{j,k}$. The decoding function behaves as follows, with all ties broken equiprobably at random.

1. If $m_{j,k} \geqslant m_{\mathrm{A}}(d_{j,k}^p)$, output $\hat{X}^n$ is the sequence x^n that has the highest probability

$$P_{X_1|X_2}^{n-d_{j,k}^p}\left(x^n\left[1:n-d_{j,k}^p\right]\middle|\hat{\mathbf{B}}_{j,k}^p\left[d_{j,k}^p+1:n\right]\right) \times P^{d_{j,k}^p}\left(x^n\left[n-d_{j,k}^p+1:n\right]\right)$$

   among all sequences in $\mathcal{X}^n$ that encode to a codeword that matches the received codeword sequence from encoder j (that is, $\lfloor \mathsf{F}(x^n) \rfloor_{m_{j,k}} = \lfloor \mathsf{F}(\mathbf{B}_{j,k}) \rfloor_{m_{j,k}}$).

2. If $m_{j,k} < m_{\mathrm{A}}(d_{j,k}^p)$, the decoder finds the pair $(x_j^n, x_{j^c}^n)$ that has the highest probability

$$\begin{aligned} &P_{X_1|X_2}^{n-d_{j,k}^p}\left(x_j^n[1:n-d_{j,k}^p]\middle|\hat{\mathbf{B}}_{j,k}^p\left[d_{j,k}^p+1:n\right]\right) \\ &\quad\times P^{d_{j,k}^p+d_{j,k}-n}\left(x_j^n\left[n-d_{j,k}^p+1:d_{j,k}\right]\right) \\ &\quad\times P_{X_1X_2}^{n-d_{j,k}}\left(x_j^n\left[d_{j,k}+1:n\right], x_{j^c}^n\left[1:n-d_{j,k}\right]\right) \\ &\quad\times P^{d_{j,k}}\left(x_{j^c}^n\left[n-d_{j,k}+1:n\right]\right) \end{aligned}$$

   among all pairs from $\mathcal{X}^n \times \mathcal{X}^n$ that encode to codewords satisfying

$$\left\lfloor \mathsf{F}\left(x_j^n\right)\right\rfloor_{m_{j,k}} = \left\lfloor \mathsf{F}\left(\mathbf{B}_{j,k}\right)\right\rfloor_{m_{j,k}} \tag{2.16}$$

   and

$$\left\lfloor \mathsf{F}\left(x_{j^c}^n\right)\right\rfloor_{m_{j,k}-d_{j,k}} = \left\lfloor \mathsf{F}\left(\mathbf{B}_{j,k}^s\right)\right\rfloor_{m_{j,k}-d_{j,k}},$$

respectively. The decoder output is x_j^n, the first sequence in the pair. The probability maximization ensures that $(x_j^n, x_{j^c}^n)$ is the pair with maximum likelihood given the prior decoding output.

Lemma 1 and Lemma 2 constitute the main results of this chapter, bounding the performance of the proposed ARASC in terms of reliability and efficiency.

Lemma 1, stated below, bounds the *per use* error probability given that any dependent side information from previous blocks is available for conditioning. Due to the FIFO rule, if the block to be decoded depends on a previously decoded block, the reconstruction of that previous block should be stored in the receiver's buffer. Since this reconstruction may itself contain error, Lemma 1 can be viewed as an upper bound on the incremental error.

Lemma 1. *For a block $\mathbf{B}_{j,k}$ of observations encoded using the above-described code and for any $m \geqslant \mathsf{m}(d_{j,k}^p, d_{j,k})$,*

$$\mathbb{P}\left[\mathsf{G}\left(\left\lfloor \mathsf{F}\left(\mathbf{B}_{j,k}\right)\right\rfloor_m, \left\lfloor \mathsf{F}\left(\mathbf{B}_{j,k}^s\right)\right\rfloor_{\max(0,m-d_{j,k})}, \mathbf{B}_{j,k}^p\left[d_{j,k}^p+1:n\right]\right) \neq \mathbf{B}_{j,k}\right] \leqslant \epsilon.$$

Proof. See Section 2.8. □

Unlike [9] which proves the existence of a deterministic code that simultaneously achieves a set of error probability constraints for $O(1)$ possible combinations of active encoders, I only characterize the expected error probability. The difficulty of finding a deterministic code lies in the number of different synchronization patterns, which grows polynomially with the blocklength.

Lemma 2 bounds the average rate of encoded source blocks up to each individually decoded block. The system can be used without a time limit, but if a time limit exists, the average rate of all blocks can be found by applying Lemma 2 to the last block. *Decoding individually* here means that the block of interest is decoded due to hitting m_{A}. Meanwhile, *decoding jointly* means that the block of interest is decoded due to hitting m_{B}. It does not mean that the decoder outputs two reconstructions simultaneously.

Lemma 2. *For a block $\mathbf{B}_{j,k}$ that is decoded at $m_{j,k} < m_{\mathrm{B}}(d_{j,k}^p, d_{j,k})$ or $m_{j,k} = m_{\mathrm{A}}(d_{j,k}) = m_{\mathrm{B}}(d_{j,k}^p, d_{j,k})$, let $k^c = \min\{k' : (T_{j^c,k'}, j^c) \prec (T_{j,k}, j)\}$. Then, the*

average rate up to block $\mathbf{B}_{j,k}$ *is*

$$\bar{R}_{j,k} \triangleq \frac{1}{k+k^c}\left(\sum_{k'=1}^{k} R_{j,k'} + \sum_{k'=1}^{k^c} R_{j^c,k'}\right),$$

which satisfies

$$\bar{R}_{j,k} \leqslant \bar{H}_{j,k} + \sqrt{\frac{V_m}{n}}Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right). \tag{2.17}$$

Here, $R_{j,k} \triangleq m_{j,k}\log|C|/n$ *is the effective rate of block* $\mathbf{B}_{j,k}$*,*

$$V_m \triangleq \max\left(V(X_1), V(X_1|X_2)\right) \tag{2.18}$$

is the larger of conditional and individual varentropies, and $\bar{H}_{j,k}$ *is the weighted average*

$$\bar{H}_{j,k} \triangleq \frac{\left(N_{j,k}^{j} + N_{j,k}^{j^c}\right)H(X_1) + N_{j,k}^{j,j^c}H(X_1,X_2)}{N_j + N_{j,k}^{j^c} + 2N_{j,k}^{j,j^c}}$$

where $N_{j,k}^{j}$ *is the number of time steps that only* $X_{j,t}$ *is encoded,* $N_{j,k}^{j^c}$ *is the number of time steps that only* $X_{j^c,t}$ *is encoded, and* $N_{j,k}^{j,j^c}$ *is the number of time steps that both sources are encoded. Specifically,*

$$\begin{aligned} N_{j,k}^{j} = &\sum_{t=1}^{T_{j,k}}\left(\sum_{i=1}^{k}\mathbf{1}\{T_{j,i}-n+1\leqslant t\leqslant T_{j,i}\}\right) \\ &\times\left(1-\sum_{i=1}^{k^c}\mathbf{1}\{T_{j^c,i}-n+1\leqslant t\leqslant T_{j^c,i}\}\right) \end{aligned} \tag{2.19}$$

$$\begin{aligned} N_{j,k}^{j^c} = &\sum_{t=1}^{T_{j^c,k^c}}\left(\sum_{i=1}^{k^c}\mathbf{1}\{T_{j^c,i}-n+1\leqslant t\leqslant T_{j^c,i}\}\right) \\ &\times\left(1-\sum_{i=1}^{k}\mathbf{1}\{T_{j,i}-n+1\leqslant t\leqslant T_{j,i}\}\right) \end{aligned} \tag{2.20}$$

$$\begin{aligned} N_{j,k}^{j,j^c} = &\sum_{t=1}^{T_{j^c,k^c}}\left(\sum_{i=1}^{k^c}\mathbf{1}\{T_{j^c,i}-n+1\leqslant t\leqslant T_{j^c,i}\}\right) \\ &\times\left(1-\sum_{i=1}^{k}\mathbf{1}\{T_{j,i}-n+1\leqslant t\leqslant T_{j,i}\}\right). \end{aligned} \tag{2.21}$$

Remark 1. *While Lemma 2 is not tight in general, it provides a simple upper bound on the rate of convergence of the average rate of encoded source blocks. When* $V(X_1) \geqslant V(X_1|X_2)$*, ARASC achieves an improvement in its first-order rate relative to point-to-point source coding by taking advantage of source dependence.*

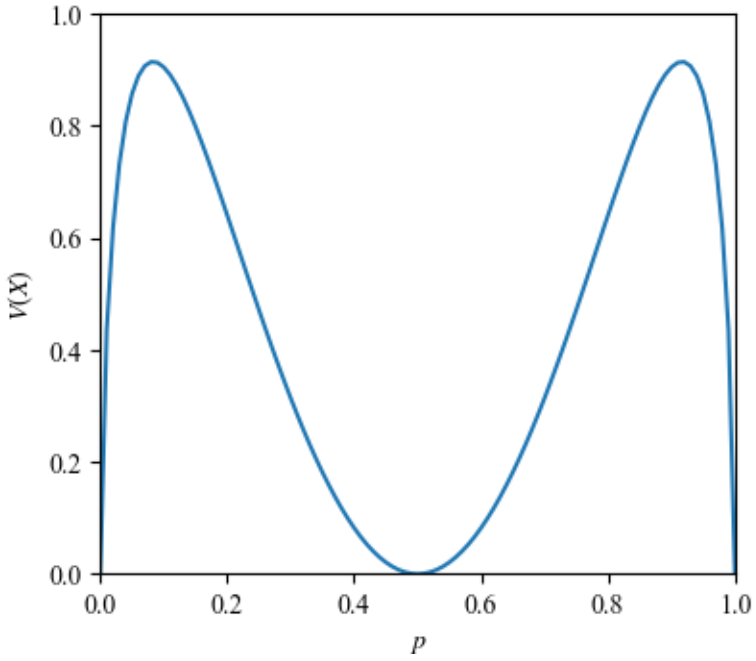


Figure 2.3: Varentropy of a Bernoulli source.

It accomplishes this benefit while maintaining a rate of convergence that matches the optimal rate of convergence of a point-to-point source code [28][9].

The varentropy does not have simple convexity properties. For example, Figure 2.3 plots the varentropy of a single Bernoulli-p source with respect to parameter p. As a result, determining when $V(X_1)$ *is larger than* $V(X_1|X_2)$ *may be a complicated function of the source parameters. Below, I include two examples of symmetrical joint source distributions* $P_{X_1X_2}$ *such that example* P_1 *satisfies* $V(X_1) \geqslant V(X_1|X_2)$ *while example* P_2 *satisfies* $V(X_1) < V(X_1|X_2)$.

$$\mathsf{P}_1 = \begin{bmatrix} 0.59 & 0.2 \\ 0.2 & 0.01 \end{bmatrix}, \qquad \mathsf{P}_2 = \begin{bmatrix} 0.5 & 0.2 \\ 0.2 & 0.1 \end{bmatrix}. \tag{2.22}$$

For symmetrical binary sources, fixing $P_{X_1X_2}(0,1) = P_{X_1X_2}(1,0) = p < 0.33$, $V(X_1) > V(X_1|X_2)$ *when either* $P_{X_1X_2}(0,0)$ *or* $P_{X_1X_2}(1,1)$ *is close to* 0. *Figure 2.4 illustrates how* $V(X_1)$ *and* $V(X_1|X_2)$ *change with* $P_{X_1X_2}(0,0)$ *when p is set to* 0.01, 0.2, *and* 0.4.

Lemma 2 is proved by examining each encoded source block up to $\mathbf{B}_{j,k}$. The proof is in Section 2.8. Specifically, the proof repeatedly use the following auxiliary results.

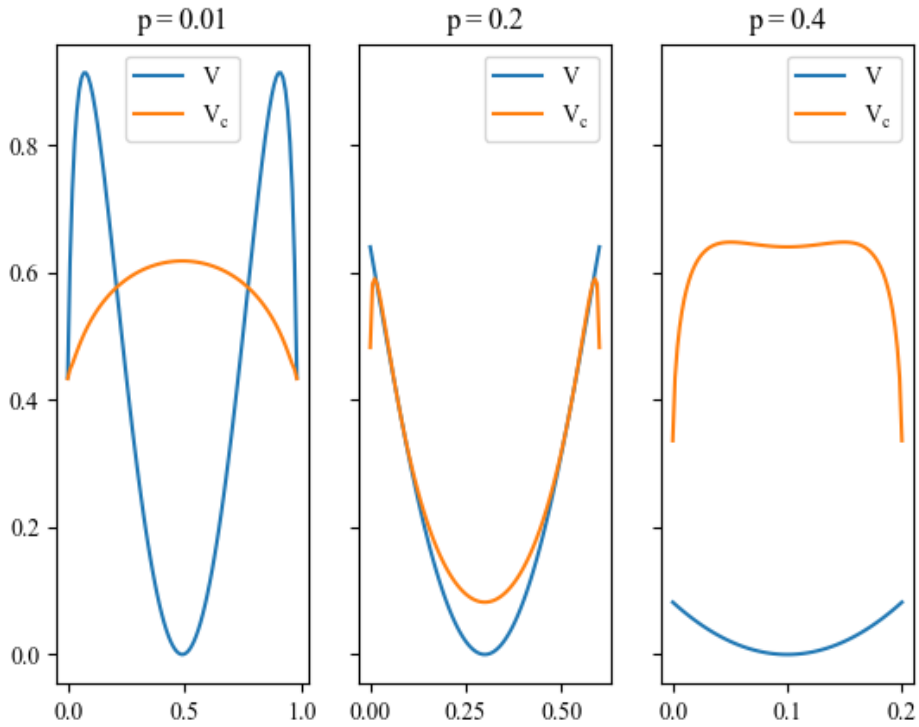


Figure 2.4: Illustration of varentropy $V = V(X_1)$ vs. conditional varentropy $V_c = V(X_1|X_2)$ for three different distributions.

Lemma 3. *For a block* $\mathbf{B}_{j,k}$ *that is decoded at codelength* $m_{j,k} = m_{\mathrm{B}}(d^p_{j,k}, d_{j,k}) < m_{\mathrm{A}}(d^p_{j,k})$,

$$R_{j,k} + R^s_{j,k} \leqslant H'_{d^p_{j,k},d_{j,k}} + 2\sqrt{\frac{V_m}{n}}Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right),$$

where V_m *is defined in* (2.18), $R^s_{j,k}$ *is the effective rate of source block* $\mathbf{B}^s_{j,k}$, *and* $H'_{d^p_{j,k},d_{j,k}}$ *is defined in* (2.14).

Proof. See Section 2.8. □

Lemma 3 provides an upper bound on the sum of the rate $R_{j,k}$ of a source block $\mathbf{B}_{j,k}$ decoded at $m_{\mathrm{B}}(d^p_{j,k}, d_{j,k})$ and the rate $R^s_{j,k}$ of its successor $\mathbf{B}^s_{j,k}$. It follows from the fact that the successor of a jointly decoded block must be decoded individually (stated in Lemma 4), and that $\sqrt{V(X_1, X_2)} \leqslant \sqrt{V(X_1)} + \sqrt{V(X_1|X_2)}$.

Lemma 4. *For sufficiently large* n, *if a block* $\mathbf{B}_{j,k}$ *is decoded at codelength* $m_{j,k} = m_{\mathrm{B}}(d^p_{j,k}, d_{j,k}) < m_{\mathrm{A}}(d^p_{j,k})$, *then its successor* $\mathbf{B}^s_{j,k} = \mathbf{B}_{j^c,k'}$, *if such a suc-*

cessor exists, satisfies $m_{j^c,k'} < m_{\mathrm{B}}(d_{j,k}, d^s_{j,k})$*, where* $d^s_{j,k} = \min(n, T_{j,k+1} - T_{j^c,k'})$ *describes the delay between* $\mathbf{B}^s_{j,k}$ *and its own potential successor.*

Proof. See Section 2.8. □

Intuitively, if $\mathbf{B}_{j,k}$ is jointly decoded, then a significant number of codeword symbols for $\mathbf{B}^s_{j,k}$ are transmitted before $T^f_{j,k}$. Hence, very few additional symbols, if any, are needed for the decoder to individually decode $\mathbf{B}^s_{j,k}$. The dependence between $\mathbf{B}^s_{j,k}$ and $\mathbf{B}_{j,k+1}$ cannot be significant enough for $\mathbf{B}^s_{j,k}$ to be decoded jointly with its own successor.

The next lemma helps characterize the sum rate across multiple blocks.

Lemma 5. *For sufficiently large n, if a block* $\mathbf{B}_{j,k}$ *is decoded at codelength* $m_{j,k} < m_{\mathrm{B}}(d^p_{j,k}, d_{j,k})$ *or* $m_{j,k} = m_{\mathrm{A}}(d_{j,k}) = m_{\mathrm{B}}(d^p_{j,k}, d_{j,k})$*, then its successor* $\mathbf{B}^s_{j,k} = \mathbf{B}_{j^c,k'}$*, if it exists, is decoded at* $m_{j^c,k'} = \mathsf{m}(d_{j,k}, d^s_{j,k}) + O(1)$*, where* $d^s_{j,k} = \min(n, T_{j,k+1} - T_{j^c,k'})$*. Furthermore,* $m_{j^c,k'} < m_{\mathrm{B}}(d_{j,k}, d^s_{j,k})$*.*

Proof: See Section 2.8.

This lemma indicates that if $\mathbf{B}_{j,k}$ is decoded individually, then its successor $\mathbf{B}^s_{j,k}$ is guaranteed to be decoded at no more than $O(1)$ time steps away from $\mathsf{m}(d_{j,k}, d^s_{j,k})$. At $T^f_{j,k}$, the decoder would not have significantly more than enough codeword symbols to decode $\mathbf{B}^s_{j,k}$ by conditioning on the newly obtained reconstruction of $\mathbf{B}_{j,k}$. Therefore, the problem of the decoder having significantly more than enough codeword symbols with the help of the latest decoder output would not arise.

Lemma 2 can be proved by examining the encoded source blocks in ascending order of transmission times. If a block is decoded individually, its rate is determined by (2.11). If a block is decoded jointly, compute its rate together with the rate of its successor by applying Lemma 3. Because of Lemma 5, in this process, one would never mis-compute the rate of an individually decoded block more than $O(1/n)$.

2.4 Simulation Results

While ARASC frees the encoders from synchronization constraints and potential waiting times for a new epoch to start, whether it provides performance gain depends on the scenario. In general, when encoding occurs in a relatively sparse manner, ARASC yields a performance gain. When encoding occurs frequently, enforcing synchronization (using RASC [9]) has an advantage over ARASC. This claim is

backed-up by simulation results. Below is a discussion of the performance of the code under Markovian arrival process (MAP) of encoding requests.

Numerical Performance under MAP

To evaluate the performance of the proposed ARASC code against existing results, I simulate the behavior of ARASC and RASC [9] code under the assumption that encoding requests arrive at the two transmitters as a pair of independent MAPs where inter-arrival times are governed by geometric distributions with identical parameter λ_a. For ARASC, an encoding opportunity arises at transmitter j when j's last transmission has ended, there are enough observations, and there is an awaiting request. For RASC, a transmitter becomes active in an epoch if there is an awaiting request at the time that the epoch starts. Both systems resemble a pair of M/G/1 queues [38], although the processing times of requests are dependent. I compute the codelengths with accuracy up to the second order. I use w_{avg} to denote the average time elapsed between the initiation of an encoding request and the time that the request is processed. I use m_{avg} to denote the average codelength. The sum $w_{\text{avg}} + m_{\text{avg}}$ is the average time that an encoding request spends in the system.

Figure 2.5 shows the performance of the two codes under a parameter setting that satisfies $\frac{H(X_1,X_2)}{2} < \log|C| < H(X_1)$. This setting ensures that the channels have sufficient capacity to transmit encoded descriptions of all observations if taking full advantage of the source dependence. Simulation results show that ARASC outperforms RASC when encoding requests are relatively infrequent. It can be observed that, as the arrivals become more frequent (but not so frequent that the queues are likely to go unstable), RASC starts to see a lower average codeword length, yet the average delay of ARASC continues to outperform that of RASC due to the smaller waiting time. When λ_a is large, the average codelength is smaller than the observation blocklength. This results in the performance (both average time in the system and average codelength) of ARASC and RASC to agree in the first order. The small difference in m_{avg} results from the ARASC system being stuck in a state where one transmitter always operates at a longer codelength while the other always operates at a shorter one. Figure 2.6 shows how this occurs.

Longer codelengths change the performance comparison. When $\frac{H(X_1)}{2} < \log|C| < \frac{H(X_1,X_2)}{2}$, codewords are, in general, longer than the blocklength n but not significantly so. In this case, ARASC outperforms RASC at low λ_a (See Figure 2.7). However, if the codelength is much larger than n, although ARASC still outper-

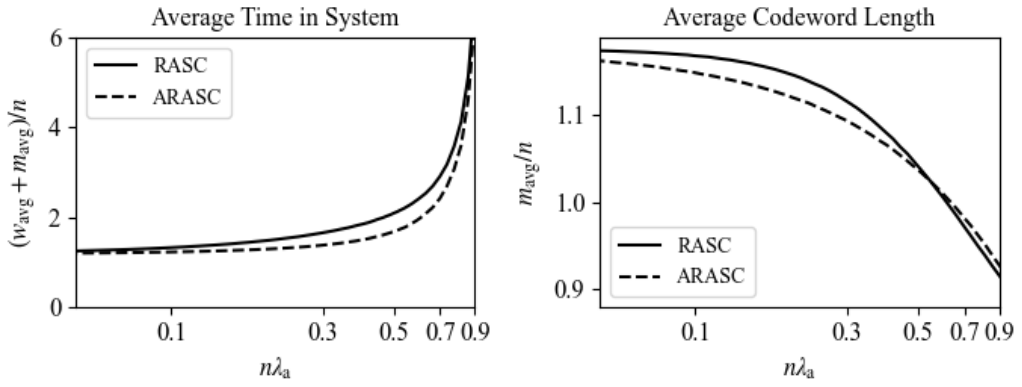


Figure 2.5: ARASC and RASC performance in terms of $w_{\text{avg}} + m_{\text{avg}}$ and m_{avg} when $H(X_1, X_2)/2 < \log|C| < H(X_1)$.

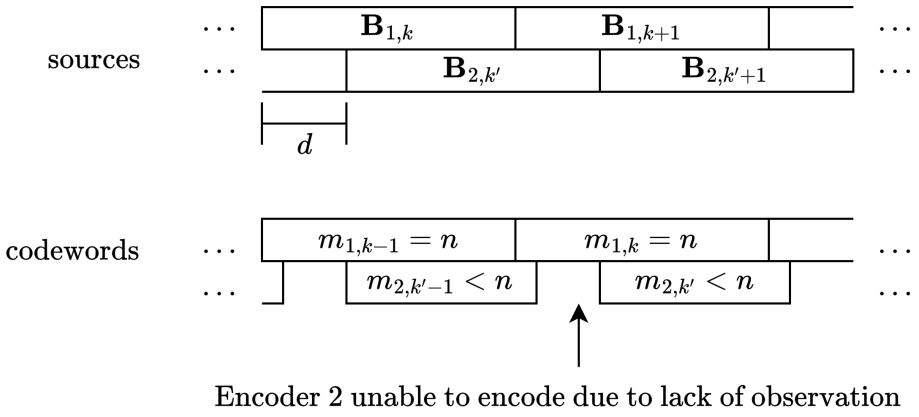


Figure 2.6: An example of the effect of dependence when $H(X_1, X_2)/2 < \log|C| < H(X_1)$ and λ_{a} is large.

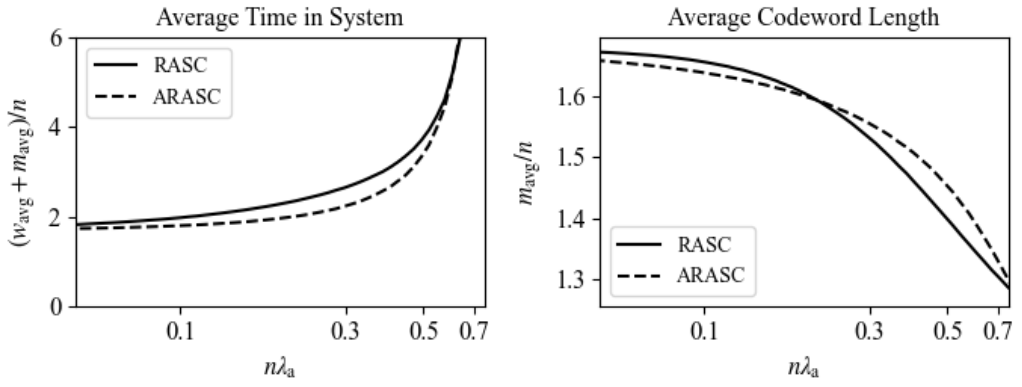


Figure 2.7: ARASC and RASC performance in terms of $w_{\text{avg}} + m_{\text{avg}}$ and m_{avg} when $H(X_1)/2 < \log|C| < H(X_1, X_2)/2$.

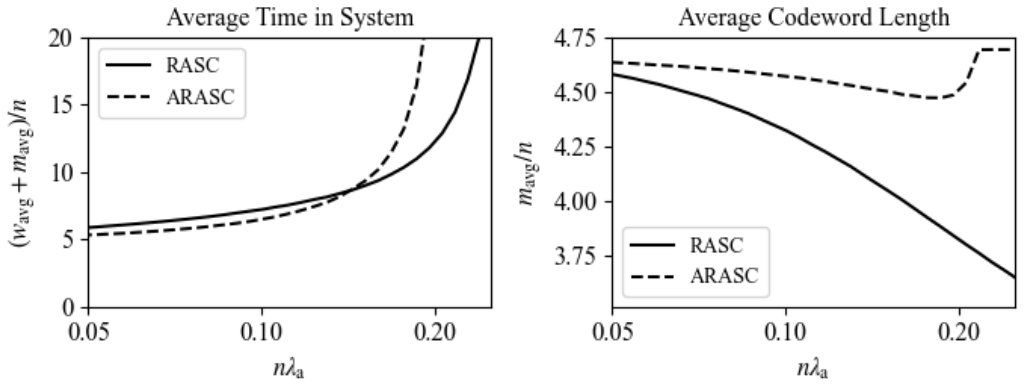


Figure 2.8: ARASC and RASC performance in terms of $w_{\text{avg}} + m_{\text{avg}}$ and m_{avg} when $\log|C| < H(X_1)/2$.

forms RASC when arrival rate is low, the average codelength begins to suffer from an undesirable effect as arrival rate increases (See Figure 2.8). ARASC runs the risk of being trapped in a state where every block is decoded at $\mathsf{m}(n,n)$. Figure 2.9 illustrates this phenomenon.

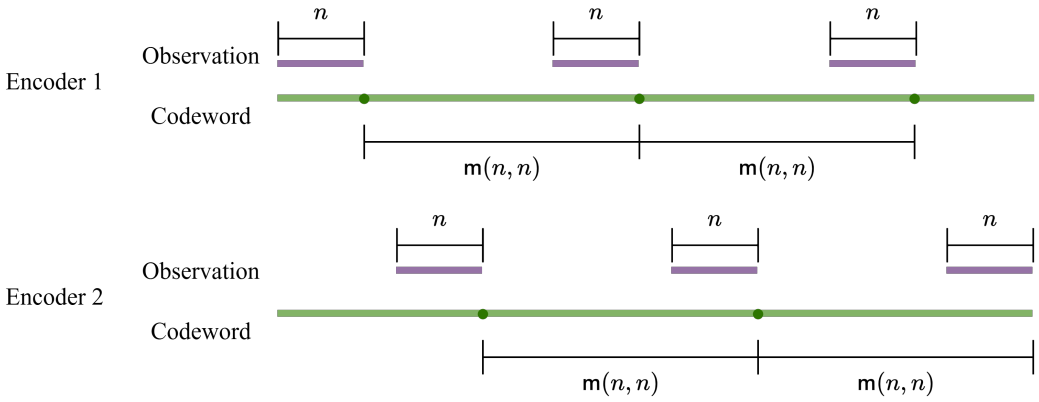


Figure 2.9: Behavior of ARASC for Large λ_a

Remark 2. *In addition to efficiency considerations, there are two other factors that also make the scenario with low λ_a better for ARASC than the scenario with high λ_a. Firstly, the ARASC proposed in this chapter uses its own prior decoding results to make future decoding decisions, which makes it prone to the effect of error propagation. When λ_a is small, it is expected that decoding errors do not propagate for multiple hops. However, when λ_a is large and $\log|C| > H(X_1, X_2)/2$, errors can continue to propagate. An extreme example is the scenario in Figure 2.6. Secondly, the low λ_a scenario gives ARASC an advantage over RASC in terms of the number*

of time steps that an encoder needs to listen for feedback from the decoder. ARASC encoders have to listen for stop feedback when they are transmitting but do not have to listen when they are not transmitting. On the other hand, during the period when an RASC encoder u transmits a codeword, the number of occasions that it has to listen for stop feedback is $O(1)$ [9], but when a RASC encoder is not transmitting, it has to continuously listen for feedback from the decoder to know whether it can transmit if there is an encoding request.

2.5 Mitigation of Error Propagation

As discussed in Remark 2, the ARASC is potentially vulnerable to error propagation. This issue is common in sequential decision systems (see, for example, [39][40][41]). When encoding requests are sparse, a decoding error is not expected to propagate for many hops. However, if the decoder wishes to stop error propagation at any time, it can easily do so by flushing its own prior decoding results and making the next decoding decision as if no side information is available. An alternative simple strategy that is more rate-efficient is to use a an approach similar to an unreliable side information code, which I present in the next chapter of this thesis.

2.6 Summary

Synchronous RASC allows transmitters to decide whether to encode or not at the start of each epoch, thereby granting more freedom compared to the MASC scenario. However, it prevents a transmitter from encoding during an epoch that has already started without its participation. The ARASC proposed in this chapter allows a transmitter to encode and transmit the codeword as long as source observations and channel resource are available, without depending on the behavior of the other transmitter. Analytically, ARASC yields an optimal first-order sum rate, and the rate of convergence is bounded.

2.7 Some Useful Auxiliary Results and Their Proofs

The following auxiliary results are widgets that are useful in proving the results in this chapter.

Lemma 6. *For $a, b, c, \ell, m \geqslant 0$, if*

$$\sqrt{a} + \sqrt{b} \geqslant \sqrt{c} \tag{2.23}$$

then

$$\sqrt{a+\ell} + \sqrt{b+m} \geqslant \sqrt{c+\ell+m}.$$

Proof. Inequality (2.23) implies that

$$a + b \geqslant c - 2\sqrt{ab}.$$

Therefore,

$$\begin{aligned}(\sqrt{a+\ell} + \sqrt{b+m})^2 &= a + \ell + b + m + 2\sqrt{(a+\ell)(b+m)} \\ &\geqslant (c + \ell + m) - 2\sqrt{ab} + 2\sqrt{(a+\ell)(b+m)} \\ &\geqslant c + \ell + m \\ &= \left(\sqrt{c+\ell+m}\right)^2.\end{aligned}$$

□

Lemma 7. *Let (X_1, X_2) be a pair of discrete random variables distributed according to $P_{X_1X_2}$. Then*

$$\sqrt{V(X_1)} + \sqrt{V(X_2|X_1)} \geqslant \sqrt{V(X_1, X_2)}. \tag{2.24}$$

Proof.

$$\begin{aligned}V(X_1) + V(X_2|X_1) &= \mathrm{Var}[\imath(X_1)] + \mathrm{Var}[\imath(X_2|X_1)] \\ &= \mathrm{Var}[\imath(X_1) + \imath(X_2|X_1)] - 2\mathrm{Cov}[\imath(X_1), \imath(X_2|X_1)] \\ &= V(X_1, X_2) - 2\mathrm{Cov}[\imath(X_1), \imath(X_2|X_1)].\end{aligned}$$

Then,

$$\begin{aligned}&\left(\sqrt{V(X_1)} + \sqrt{V(X_2|X_1)}\right)^2 \\ &\quad= V(X_1) + V(X_2|X_1) + 2\sqrt{V(X_1)V(X_2|X_1)} \\ &\quad= V(X_1, X_2) - 2\mathrm{Cov}[\imath(X_1), \imath(X_2|X_1)] + 2\sqrt{V(X_1)V(X_2|X_1)} \\ &\quad= V(X_1, X_2) + 2\left(\sqrt{V(X_1)V(X_2|X_1)} - \mathrm{Cov}[\imath(X_1), \imath(X_2|X_1)]\right) \\ &\quad= V(X_1, X_2) + 2(1 - r)\sqrt{V(X_1)V(X_2|X_1)} \\ &\quad\geqslant \left(\sqrt{V(X_1, X_2)}\right)^2,\end{aligned}$$

where $r \in [-1, 1]$ is the correlation coefficient. Since

$$V(X_1), V(X_2|X_1), V(X_1, X_2) > 0,$$

the lemma is proved. □

Lemma 8. *For a pair of discrete random variables $(X_1, X_2) \sim P_{X_1X_2}$ such that $P_{X_1X_2}(x_1, x_2) = P_{X_1X_2}(x_2, x_1)$ for all $x_1, x_2 \in \mathcal{X}$, let $P_{X_1X_2}$ satisfy the conditions specified for the symmetrical scenario. For all $(d^p, d) \in (n]^2$,*

$$\sqrt{V'_{d^p,d}} \leqslant 2\sqrt{V_m}, \tag{2.25}$$

where $V_m \triangleq \max\{V(X_1), V(X_2|X_1)\}$.

Proof.

$$\begin{aligned}
\sqrt{nV'_{d^p,d}} &\leqslant \sqrt{2dV_m + (n-d)V(X_1, X_2)} \\
&= \sqrt{(n-d)V(X_1, X_2) + dV_m + dV_m} \\
&\leqslant \sqrt{(n-d)V(X_1) + dV_m} + \sqrt{(n-d)V(X_2|X_1) + dV_m} \qquad (2.26) \\
&\leqslant 2\sqrt{nV_m},
\end{aligned}$$

where (2.26) follows from Lemma 6 and Lemma 7. □

The lemma below is an RCU bound that is useful in the proof of Lemma 1. It bounds the error probability of decoding an abstract source from its compressed description with the help of a side information and a compressed description of a dependent source.

Lemma 9 (ARASC RCU Bounds)**.** *Let X, Y, Z be sources on the same alphabet $\mathcal{X}$. The uniform random binning code F, which independently and uniformly maps source realizations to codewords in $\mathcal{C}^{m_{\max}}$, satisfies the following properties.*

1. *Probability*

$$\mathbb{P}\left[\mathsf{G}_{\mathrm{ML}}\left(\lfloor \mathsf{F}(X) \rfloor_{m_{\mathrm{A}}}, Z\right) \neq X\right] \leqslant \mathbb{E}[\min\{1, A\}], \tag{2.27}$$

where G_{ML} is the maximum likelihood decoder

$$\begin{aligned}
\mathsf{G}_{\mathrm{ML}}(c, z) &\triangleq \max_{x \in \mathcal{X}_{m_{\mathrm{A}}}(c)} P_{X|Z}(x|z) \\
\mathcal{X}_{m_{\mathrm{A}}}(c) &\triangleq \left\{x \in \mathcal{X} : \lfloor \mathsf{F}(x) \rfloor_{m_{\mathrm{A}}} = c\right\},
\end{aligned}$$

and A is defined as

$$\begin{aligned}
A &\triangleq \frac{1}{|\mathcal{C}|^{m_{\mathrm{A}}}} \mathbb{E}\left[\exp\left(\imath\left(\bar{X}|Z\right)\right) \mathbf{1}\left\{\imath\left(\bar{X}|Z\right) \leqslant \imath\left(X|Z\right)\right\} \middle| X, Z\right] \\
\imath(x|z) &\triangleq \log \frac{1}{P_{X|Z}(x|z)}.
\end{aligned}$$

2. *Probability*

$$\mathbb{P}\left[\hat{X}^* \neq X\right] \leqslant \mathbb{E}[\min\{1, A'\}] + \mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{I}], \tag{2.28}$$

where $m_{\mathrm{B}}, m_{\mathrm{S}} \in \mathbb{Z}_+$ *and* $m_{\mathrm{B}} > m_{\mathrm{S}}$,

$$\begin{aligned}
\mathcal{E}_1 &\triangleq \big\{\exists \bar{x} \in \mathcal{X}\backslash\{X\} : \\
&\qquad P_{XY|Z}(\bar{x}, Y|Z) \geqslant P_{XY|Z}(XY|Z), \\
&\qquad \lfloor \mathsf{F}(\bar{x}) \rfloor_{m_{\mathrm{B}}} = \lfloor \mathsf{F}(X) \rfloor_{m_{\mathrm{B}}} \big\} \\
\mathcal{I} &\triangleq \{X = Y\} \\
A' &\triangleq \frac{1}{|C|^{m_{\mathrm{B}}+m_{\mathrm{S}}}} \mathbb{E}\left[\exp\left(\imath\left(\bar{X}, \bar{Y}\middle| Z\right)\right)\right. \\
&\qquad \left. \cdot \mathbf{1}\left\{\imath\left(\bar{X}, \bar{Y}\middle| Z\right) \leqslant \imath(X, Y|Z)\right\}\middle| X, Y, Z\right] \\
\imath(x, y|z) &\triangleq \log \frac{1}{P_{XY|Z}(x, y|z)},
\end{aligned}$$

$\mathsf{G}'_{\mathrm{ML}}$ *is the joint maximum likelihood decoder*

$$\begin{aligned}
\mathsf{G}'_{\mathrm{ML}}(c_1, c_2, z) &\triangleq \max_{\substack{(x,y)\in \\ \mathcal{X}_{m_{\mathrm{B}}}(c_1)\times\mathcal{X}_{m_{\mathrm{S}}}(c_2)}} P_{XY|Z}(x, y|z) \\
\mathcal{X}_{m_{\mathrm{B}}}(c_1) &\triangleq \left\{x \in \mathcal{X} : \lfloor \mathsf{F}(x) \rfloor_{m_{\mathrm{B}}} = c_1\right\} \\
\mathcal{X}_{m_{\mathrm{S}}}(c_2) &\triangleq \left\{x \in \mathcal{X} : \lfloor \mathsf{F}(x) \rfloor_{m_{\mathrm{S}}} = c_2\right\},
\end{aligned}$$

and

$$\left(\hat{X}^*, \hat{Y}^*\right) = \mathsf{G}'_{\mathrm{ML}}\left(\lfloor \mathsf{F}(X) \rfloor_{m_{\mathrm{B}}}, \lfloor \mathsf{F}(Y) \rfloor_{m_{\mathrm{S}}}, Z\right).$$

Here,

$$\left(X, Y, \bar{X}, \bar{Y}\right) \sim P_{XY} P_{XY}.$$

Proof. For every $x \in \mathcal{X}$, $\mathsf{F}(x)$ is drawn i.i.d. uniformly at random from $C^{m_{\max}}$. Then, the first m_{A} symbols of $\mathsf{F}(x)$, $\lfloor \mathsf{F}(x) \rfloor_{m_{\mathrm{A}}}$, is uniformly distributed on $C^{m_{\mathrm{A}}}$.

Due to the use of maximum likelihood decoder G_{ML}, the error probability can be bounded by

$$\mathbb{P}[\mathsf{G}_{\mathrm{ML}}(\lfloor \mathsf{F}(X) \rfloor_{m_{\mathrm{A}}}, Z) \neq X] \leqslant \mathbb{P}[\mathcal{E}],$$

where

$$\mathcal{E} \triangleq \{\exists \bar{x} \in \mathcal{X}\backslash\{X\} \text{ s.t. } \imath(\bar{x}|Z) \leqslant \imath(X|Z), \lfloor \mathsf{F}(\bar{x}) \rfloor_{m_{\mathrm{A}}} = \lfloor \mathsf{F}(X) \rfloor_{m_{\mathrm{A}}}\}.$$

Applying the union bound,

$$\begin{aligned}
\mathbb{P}[\mathcal{E}] &\leqslant \mathbb{P}\left[\bigcup_{\bar{x}\in\mathcal{B}\setminus\{X\}} \{\imath(\bar{x}|Z) \leqslant \imath(X|Z), \lfloor \mathsf{F}(\bar{x})\rfloor_{m_\mathrm{A}} = \lfloor \mathsf{F}(X)\rfloor_{m_\mathrm{A}}\}\right] \\
&= \mathbb{E}\left[\mathbb{P}\left[\bigcup_{\bar{x}\in\mathcal{X}\setminus\{X\}} \{\imath(\bar{x}|Z) \leqslant \imath(X|Z), \lfloor \mathsf{F}(\bar{x})\rfloor_{m_\mathrm{A}} = \lfloor \mathsf{F}(X)\rfloor_{m_\mathrm{A}}\}\middle| X, Z\right]\right] \\
&\leqslant \mathbb{E}\left[\min\left\{1, \sum_{\bar{x}\in\mathcal{B}\setminus\{X\}} \mathbb{P}[\imath(\bar{x}|Z) \leqslant \imath(X|Z), \lfloor \mathsf{F}(\bar{x})\rfloor_{m_\mathrm{A}} = \lfloor \mathsf{F}(X)\rfloor_{m_\mathrm{A}}|X, Z]\right\}\right] \\
&\leqslant \mathbb{E}\left[\min\left\{1, \frac{1}{|\mathcal{C}|^{m_\mathrm{A}}} \sum_{\bar{x}\in\mathcal{X}} \mathbf{1}\{\imath(\bar{x}|Z) \leqslant \imath(X|Z)\}\right\}\right] \\
&= \mathbb{E}\left[\min\left\{1, \frac{1}{|\mathcal{C}|^{m_\mathrm{A}}} \mathbb{E}\left[\frac{1}{P_{X|Z}(\bar{X}|Z)} \mathbf{1}\{\imath(\bar{X}|Z) \leqslant \imath(X|Z)\}\middle| X, Z\right]\right\}\right] \\
&= \mathbb{E}[\min\{1, A\}].
\end{aligned}$$

Next, I bound $\mathbb{P}\left[\hat{X}^* \neq X\right]$.

Since I only compare the first dimension of the output of $\mathsf{G}'_{\mathrm{ML}}$ against X, the error probability is bounded by the union of two events, $\mathcal{E}_1$ and

$$\begin{aligned}
\mathcal{E}_2 \triangleq \{\exists \bar{x} \in \mathcal{X}\setminus\{X\}, \bar{y} \in \mathcal{X}\setminus\{Y\} : \imath(\bar{x}, \bar{y}|Z) \leqslant \imath(X, Y|Z), \\
\lfloor \mathsf{F}(\bar{x})\rfloor_{m_\mathrm{B}} = \lfloor \mathsf{F}(X)\rfloor_{m_\mathrm{B}}, \lfloor \mathsf{F}(\bar{y})\rfloor_{m_\mathrm{S}} = \lfloor \mathsf{F}(Y)\rfloor_{m_\mathrm{S}}\}. \qquad (2.29)
\end{aligned}$$

By the union bound, $\mathbb{P}[\mathcal{E}_1 \cup \mathcal{E}_2] \leqslant \mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{E}_2]$. Therefore, one only needs to show that

$$\mathbb{P}[\mathcal{E}_2] \leqslant \mathbb{E}[\min\{1, A'\}] + \mathbb{P}[\mathcal{I}]. \qquad (2.30)$$

The analysis of $\mathbb{P}[\mathcal{E}_2]$ below follows the same reasoning as the proof of Theorem 4.

$$\begin{aligned}
\mathbb{P}[\mathcal{E}_2] &\leqslant \mathbb{P}\Bigg[\bigcup_{\substack{\bar{x}\in\mathcal{X}\setminus\{X\}\\ \bar{y}\in\mathcal{X}\setminus\{Y\}}} \{\imath(\bar{x},\bar{y}|Z) \leqslant \imath(X,Y|Z), \\
&\qquad\qquad \lfloor \mathsf{F}(\bar{x})\rfloor_{m_\mathrm{B}} = \lfloor \mathsf{F}(X)\rfloor_{m_\mathrm{B}}, \lfloor \mathsf{F}(\bar{y})\rfloor_{m_\mathrm{S}} = \lfloor \mathsf{F}(Y)\rfloor_{m_\mathrm{S}}\}\Bigg] \\
&\leqslant \mathbb{E}\Bigg[\min\Bigg\{1, \sum_{\substack{\bar{x}\in\mathcal{X}\setminus\{X\}\\ \bar{y}\in\mathcal{X}\setminus\{Y\}}} \mathbb{P}[\imath(\bar{x},\bar{y}|Z) \leqslant \imath(X,Y|Z), \\
&\qquad\qquad \lfloor \mathsf{F}(\bar{x})\rfloor_{m_\mathrm{B}} = \lfloor \mathsf{F}(X)\rfloor_{m_\mathrm{B}}, \lfloor \mathsf{F}(y)\rfloor_{m_\mathrm{S}} = \lfloor \mathsf{F}(Y)\rfloor_{m_\mathrm{S}}]\Bigg\}\Bigg] \\
&\leqslant \mathbb{E}\left[\min\left\{1, \frac{1}{|C|^{m_\mathrm{B}+m_\mathrm{S}}} \sum_{(\bar{x},\bar{y})\in\mathcal{X}^2} \mathbf{1}\{\imath(\bar{x},\bar{y}|Z) \leqslant \imath(X,Y|Z)\}\right\}\right] + \mathbb{P}[\mathcal{I}] \\
&= \mathbb{E}\left[\min\left\{1, \frac{1}{|C|^{m_\mathrm{B}+m_\mathrm{S}}} \mathbb{E}\left[\frac{\mathbf{1}\{\imath(\bar{X},\bar{Y}|Z) \leqslant \imath(X,Y|Z)\}}{P_{XY|Z}(\bar{X},\bar{Y}|Z)}\middle| X,Y,Z\right]\right\}\right] + \mathbb{P}[\mathcal{I}] \\
&= \mathbb{E}[\min\{1, A'\}] + \mathbb{P}[\mathcal{I}].
\end{aligned}$$

□

2.8 Proofs of Theorems and Lemmas

Proof of Theorem 3

Proof. Like the achievability proof of [9, Theorem 20], the proof of Theorem 3 applies the RCU bound to a pair of stationary, memoryless sources $\left(X_1^n, X_2^n\right)$. The only difference is the existence of the term $\mathbb{P}\left[X_1^n = X_2^n\right]$. The probability of this event decays exponentially with n. Specifically, $\mathbb{P}\left[X_1^n = X_2^n\right] = p_{\mathrm{eq}}^n$, where

$$p_{\mathrm{eq}} \triangleq \mathbb{P}[X_{1,1} = X_{2,1}] = \sum_{x\in\mathcal{X}} P_{X_1X_2}(x,x) < 1.$$

Set a constant $C_{\mathrm{eq}} > 0$. For sufficiently large n, $p_{\mathrm{eq}}^n \leqslant C_{\mathrm{eq}}/\sqrt{n}$. Therefore, picking any pair of rates (R_1, R_2) satisfying

$$\bar{\mathbf{R}} \in \mathscr{R}_{\mathrm{in}}^*\left(n, \epsilon - \frac{C_{\mathrm{eq}}}{\sqrt{n}}\right)$$

gives an error probability bounded by ϵ. Applying [9, Lemma 13-1] completes the proof. □

When P_{X_1} and P_{X_2} are different, pre-processing the source observation from one source (let it be source 1 without loss of generality) with a bijective mapping may reduce p_{eq}, which in turn affects the achievable rate region in the forth-order term. In this sense, the best choice of the bijective mapping $\mathsf{p}^* : \mathcal{X} \mapsto \mathcal{X}$ is

$$\mathsf{p}^* = \min_{\mathsf{p}:\mathcal{X}\mapsto\mathcal{X}} \sum_{x\in\mathcal{X}} P_{X_1X_2}\left(\mathsf{p}^{-1}(x), x\right).$$

Whereas such a pre-processing can be important in the error exponent regime, it does not affect the rate region up to the third order in the rate-of-convergence regime.

Proof of Theorem 4

Proof. For every $x \in \mathcal{X}$, draw $\mathsf{F}(x)$ i.i.d. uniformly at random from $\{0,1\}^{\max(m_1,m_2)}$, and let $\mathsf{F}_1(x) = \lfloor \mathsf{F}(x) \rfloor_{m_1}$ and $\mathsf{F}_2(x) = \lfloor \mathsf{F}(x) \rfloor_{m_2}$ for all $x \in \mathcal{X}$. The maximum likelihood decoder is defined for each $(c_1, c_2) \in \{0,1\}^{m_1} \times \{0,1\}^{m_2}$ by

$$\mathsf{g}(c_1, c_2) = \arg \min_{\substack{(x_1,x_2)\in\mathcal{X}^2: \\ \mathsf{F}_1(x_1)=c_1, \mathsf{F}_2(x_2)=c_2}} \imath(x_1, x_2),$$

where ties are broken lexicographically. Consider the following three events (these definitions are local to this proof and not the same as the definitions in the proof of Lemma 9)

$$\begin{aligned}
\mathcal{E}_1 &\triangleq \{\exists \bar{x}_1 \in \mathcal{X}\backslash\{X_1\} : \imath(\bar{x}_1|X_2) \leqslant \imath(X_1|X_2), \mathsf{F}_1(\bar{x}_1) = \mathsf{F}_1(X_1)\} \\
\mathcal{E}_2 &\triangleq \{\exists \bar{x}_2 \in \mathcal{X}\backslash\{X_2\} : \imath(\bar{x}_2|X_1) \leqslant \imath(X_2|X_1), \mathsf{F}_2(\bar{x}_2) = \mathsf{F}_2(X_2)\} \\
\mathcal{E}_{12} &\triangleq \{\exists \bar{x}_1 \in \mathcal{X}\backslash\{X_1\}, \bar{x}_2 \in \mathcal{X}\backslash\{X_2\} : \imath(\bar{x}_1, \bar{x}_2) \leqslant \imath(X_1, X_2), \\
&\quad \mathsf{F}_1(\bar{x}_1) = \mathsf{F}_1(X_1), \mathsf{F}_2(\bar{x}_2) = \mathsf{F}_2(X_2)\}.
\end{aligned}$$

Since the lexicographical tie breaker gives lower error probability than a "malicious gene" tie breaker that always chooses the wrong reconstruction, these event definitions capture the case where X_1 is reconstructed incorrectly and X_2 correctly, X_2 is reconstructed incorrectly and X_1 correctly, and X_1, X_2 are both reconstructed incorrectly. The random code's expected error probability is bounded by the probability of the union of these events. I further define events

$$\mathcal{I} \triangleq \{X_1 = X_2\}, \quad \mathcal{I}^{\mathsf{c}} \triangleq \{X_1 \neq X_2\}.$$

Then, the error probability is bounded by

$$\begin{aligned}
&\mathbb{P}[\mathsf{g}(\mathsf{f}(X_1), \mathsf{f}(X_2)) \neq (X_1, X_2)] \\
&\leqslant \mathbb{P}[\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_{12}] \\
&= \mathbb{P}[(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_{12}) \cap \mathcal{I}] + \mathbb{P}[(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_{12}) \cap \mathcal{I}^{\mathsf{c}}] \\
&\leqslant \mathbb{P}[\mathcal{I}] + \mathbb{P}[(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_{12}) \cap \mathcal{I}^{\mathsf{c}}]. \qquad (2.31)
\end{aligned}$$

The law of iterated expectations gives

$$\begin{aligned}
\mathbb{P}[(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_{12}) \cap \mathcal{I}^{\mathsf{c}}] &\leqslant \mathbb{P}[\mathcal{E}_1 \cup \mathcal{E}_2 \cup (\mathcal{E}_{12} \cap \mathcal{I}^{\mathsf{c}})] \\
&= \mathbb{E}\left[\mathbb{P}\left[\mathcal{E}_1' \cup \mathcal{E}_2' \cup \mathcal{E}_{12}' | X_1, X_2\right]\right],
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{E}_1' &\triangleq \bigcup_{\bar{x}_1 \in \mathcal{X} \setminus \{X_1\}} \{\imath(\bar{x}_1|X_2) \leqslant \imath(X_1|X_2), \mathsf{F}_1(\bar{x}_1') = \mathsf{F}_1(X_1)\} \\
\mathcal{E}_2' &\triangleq \bigcup_{\bar{x}_2 \in \mathcal{X} \setminus \{X_2\}} \{\imath(\bar{x}_2|X_1) \leqslant \imath(X_2|X_1), \mathsf{F}_2(\bar{x}_2') = \mathsf{F}_2(X_2)\} \\
\mathcal{E}_{12}' &\triangleq \bigcup_{\substack{\bar{x}_1 \in \mathcal{X} \setminus \{X_1\}, \\ \bar{x}_2 \in \mathcal{X} \setminus \{X_2\}}} \{\imath(\bar{x}_1, \bar{x}_2) \leqslant \imath(X_1, X_2), \mathsf{F}_1(\bar{x}_1) = \mathsf{F}_1(X_1), \mathsf{F}_2(\bar{x}_2) = \mathsf{F}_2(X_2), X_1 \neq X_2\}
\end{aligned}$$

are the unions of error events corresponding to $\mathcal{E}_1$, $\mathcal{E}_2$ and $\mathcal{E}_{12} \cap \mathcal{I}^{\mathsf{c}}$. Applying the union bound and the maximal probability bound,

$$\mathbb{P}\left[(\mathcal{E}_1 \cup \mathcal{E}_2 \cup \mathcal{E}_{12}) \cap \mathcal{I}^{\mathsf{c}}\right] \leqslant \mathbb{E}\left[\min\{1, A_1' + A_2' + A_{12}'\}\right],$$

where

$$\begin{aligned}
A_1' &\triangleq \sum_{\bar{x}_1 \in \mathcal{X} \setminus \{X_1\}} \mathbb{P}[\imath(\bar{x}_1|X_2) \leqslant \imath(X_1|X_2), \mathsf{F}_1(\bar{x}_1) = \mathsf{F}_1(X_1)|X_1, X_2], \\
A_2' &\triangleq \sum_{\bar{x}_2 \in \mathcal{X} \setminus \{X_2\}} \mathbb{P}[\imath(\bar{x}_2|X_1) \leqslant \imath(X_2|X_1), \mathsf{F}_2(\bar{x}_2) = \mathsf{F}_2(X_2)|X_1, X_2], \\
A_{12}' &\triangleq \sum_{\substack{\bar{x}_1 \in \mathcal{X} \setminus \{X_1\}, \\ \bar{x}_2 \in \mathcal{X} \setminus \{X_2\}}} \mathbb{P}[\imath(\bar{x}_1, \bar{x}_2) \leqslant \imath(X_1, X_2), \mathsf{F}_1(\bar{x}_1) = \mathsf{F}_1(X_1), \\
&\qquad\qquad \mathsf{F}_2(\bar{x}_2) = \mathsf{F}_2(X_2), X_1 \neq X_2 | X_1, X_2].
\end{aligned}$$

Since the encoder outputs are drawn i.i.d. uniformly at random,

$$\begin{aligned}
&\mathbb{E}[\min\{1, A_1' + A_2' + A_{12}'\}] \\
&\leqslant \mathbb{E}\left[\min\left\{1, \frac{1}{|C|^{m_1}} \sum_{\bar{x}_1 \in \mathcal{X}} \mathbf{1}\left\{\imath(\bar{x}_1|X_2) \leqslant \imath(X_1|X_2)\right\}\right.\right. \\
&\qquad\qquad \left.\left. + \frac{1}{|C|^{m_2}} \sum_{\bar{x}_2 \in \mathcal{X}} \mathbf{1}\left\{\imath(\bar{x}_2|X_1) \leqslant \imath(X_2|X_1)\right\} + A_{12}'\right\}\right] \\
&\leqslant \mathbb{E}\left[\min\left\{1, \frac{1}{|C|^{m_1}} \sum_{\bar{x}_1 \in \mathcal{X}} \mathbf{1}\left\{\imath(\bar{x}_1|X_2) \leqslant \imath(X_1|X_2)\right\}\right.\right. \\
&\qquad\qquad + \frac{1}{|C|^{m_2}} \sum_{\bar{x}_2 \in \mathcal{X}} \mathbf{1}\left\{\imath(\bar{x}_2|X_1) \leqslant \imath(X_2|X_1)\right\} \\
&\qquad\qquad \left.\left. + \frac{1}{|C|^{m_1+m_2}} \sum_{\bar{x}_1 \in \mathcal{X}, \bar{x}_2 \in \mathcal{X}} \mathbf{1}\left\{\imath(\bar{x}_1, \bar{x}_2) \leqslant \imath(X_1, X_2)\right\}\right\}\right].
\end{aligned}$$

The manipulation of A_{12}' follows from the fact that $\mathsf{F}_1(x_1)$ and $\mathsf{F}_2(x_2)$ are independently drawn for $x_1 \neq x_2$. That is, when $x_1 \neq x_2$,

$$\Pr\left(\mathsf{F}(\bar{x}_1) = \mathsf{F}(x_1), \mathsf{F}(\bar{x}_2) = \mathsf{F}(x_2), x_1 \neq x_2\right) = \frac{1}{|C|^{m_1+m_2}}.$$

Otherwise, this probability is 0, as $x_1 \neq x_2$ is deterministically impossible. Combining with (2.31) completes the proof. □

Proof of Lemma 1

Proof. The proof uses Lemma 9. The two bounds in Lemma 9 help bound the per-use error probability of the block ARASC when decoding a block without and with the help of the description of its successor, respectively.

The RCU bound for individual decoding, (2.27), is used to prove that the error probability is bounded when decoding is done because of codeword length reaching m_{A}.

Without loss of generality, let $j = 1$. Denote

$$\begin{aligned}
I_n &\triangleq \imath\left(\mathbf{B}_{j,k}[n] \,\middle|\, \mathbf{B}_{j,k}^{p}[n]\right) \\
&= \sum_{t=T_{j,k}-n+1}^{T_{j,k}-d_{j,k}^{p}} \imath(X_{1,t}|X_{2,t}) + \sum_{t=T_{j,k}-d_{j,k}^{p}+1}^{T_{j,k}} \imath(X_{1,t})
\end{aligned}$$

and

$$\begin{aligned}\bar{I}_n &\triangleq \imath\left(\bar{X}_{T_{j,k}}^{-n}\middle|\mathbf{B}_{j,k}^{p}[n]\right)\\ &= \sum_{t=T_{j,k}-n+1}^{T_{j,k}-d_{j,k}^{p}} \imath(\bar{X}_{1,t}|X_{2,t}) + \sum_{t=T_{j,k}-d_{j,k}^{p}+1}^{T_{j,k}} \imath(\bar{X}_{1,t}),\end{aligned}$$

where $\bar{X}_{T_{j,k}}^{-n}$ is distributed with respect to P_X^n and is independent of the block of interest ($\mathbf{B}_{j,k}$) or its predecessor ($\mathbf{B}_{j,k}^{p}$).

Both I_n and $\bar{I}_n$ are sums of n independent (but not necessarily identical) random variables. Applying (2.27), the individual RCU bound, when decoded at $m \geqslant m_{\mathrm{A}}(d_{j,k}^{p})$, the ARASC code satisfies

$$\epsilon' \leqslant \mathbb{E}\left[\min\left\{1, \frac{1}{M}\mathbb{E}\left[\exp(\bar{I}_n)\mathbf{1}\{\bar{I}_n \leqslant I_n\}\middle|X_{1,T_{j,k}}^{-n}, X_{2,T_{j,k}-d_{j,k}^{p}}^{-\left(n-d_{j,k}^{p}\right)}\right]\right\}\right]$$

where $M = |C|^m$. Using [22, Lemma 47],

$$\mathbb{E}\left[\exp(\bar{I}_n)\mathbf{1}\{\bar{I}_n \leqslant I_n\}\middle|X_{1,T_{j,k}}^{-n}, X_{2,T_{j,k}-d_{j,k}^{p}}^{-\left(n-d_{j,k}^{p}\right)}\right] \leqslant \frac{C'}{\sqrt{n}}\exp(I_n). \tag{2.32}$$

Here,

$$C' \triangleq 2\left(\frac{\log 2}{\sqrt{2\pi V_{d_{j,k}^{p}}}} + 2B'\right) \tag{2.33}$$

with

$$B' = C_0\frac{T_{d_{j,k}^{p}}}{\left(V_{j,k}^{p}\right)^{3/2}}, \tag{2.34}$$

where

$$T_{d^p} \triangleq \frac{n-d^p}{n}T(X_1|X_2) + \frac{d^p}{n}T(X_1). \tag{2.35}$$

The distribution assumptions (2.7) and (2.8) guarantee that C' is a finite positive constant. Therefore.

$$\epsilon' \leqslant \mathbb{E}\left[\min\{1, \frac{C'}{M\sqrt{n}}\exp(I_n)\}\right] \tag{2.36}$$

$$= \mathbb{P}\left[I_n > \log\frac{M\sqrt{n}}{C'}\right] + \frac{C'}{M\sqrt{n}}\mathbb{E}\left[\exp(I_n)\mathbf{1}\left\{I_n \leqslant \log\frac{M\sqrt{n}}{C'}\right\}\right] \tag{2.37}$$

$$\leqslant \mathbb{P}[I_n > \log M + \frac{1}{2}\log n - \log C'] + \frac{C'}{\sqrt{n}}. \tag{2.38}$$

Since I choose $m_{\mathrm{A}}(d^p)$ to satisfy

$$m_{\mathrm{A}}\left(d^p\right)\log|C| = \log M \geqslant nH_{d^p} + \sqrt{nV_{d^p}}Q^{-1}\left(\epsilon - \frac{B_m + C_m}{\sqrt{n}}\right) \tag{2.39}$$

in (2.11), when n is sufficiently large,

$$\log M \geqslant nH_{d^p_{j,k}} + \sqrt{nV_{d^p_{j,k}}}Q^{-1}\left(\epsilon - \frac{B_m + C_m}{\sqrt{n}}\right) - \frac{1}{2}\log n + \log C', \tag{2.40}$$

hence $\epsilon' \leqslant \epsilon$ by the Berry-Esseen bound.

Note that our choice of C_1 in (2.11) is

$$C_1 \triangleq C_m + B_m \tag{2.41}$$

$$= \max_{d^p}\{C'\} + \max_{d^p}\{B'\}. \tag{2.42}$$

Although choosing different C_1 for different d^p yields better performance (a rate gain on the forth order), I choose to make it fixed for simplicity.

The part of Lemma 9 for joint decoding is used to prove that the error probability is bounded when decoding is done because of codeword length reaching m_{B}. For any pair of $(d^p, d) \in (n]^2$, for the joint decoder to activate,

$$m_{\mathrm{A}}(d^p) > m_{\mathrm{B}}(d^p, d). \tag{2.43}$$

By the definitions of m_{A} and m_{B},

$$\frac{1}{2\log|C|}\left(nH'_{d^p,d} + \sqrt{nV'_{d^p,d}}Q\left(\epsilon - \frac{C_2}{\sqrt{n}}\right) + d\log|C|\right) \tag{2.44}$$

$$< \frac{1}{\log|C|}\left(nH_{d^p} + \sqrt{nV_{d^p}}Q^{-1}\left(\epsilon - \frac{C_1}{\sqrt{n}}\right)\right) + 1. \tag{2.45}$$

For any $\delta_1 > 0$, when n is sufficiently large,

$$\frac{1}{2}H'_{d^p,d} - H_{d^p} + \frac{d}{2n}\log|C| < \delta_1, \tag{2.46}$$

which is equivalent to

$$\frac{n-d}{n} > \frac{\log|C| - 2\delta_1}{H(X_1) - H(X_1|X_2) + \log|C|}. \tag{2.47}$$

Denote

$$H^c_{d^p,d} \triangleq H'_{d^p,d} - H(X_1). \tag{2.48}$$

Note that

$$H^c_{d^p,d} = \frac{(n-d^p)+(n-d)}{n} H(X_1|X_2) + \frac{d-n+d^p}{n} H(X_1) \tag{2.49}$$

is the horizontal coordinate of the corner point of the following asymptotic rate region boundary. Due to (2.47), there exists $\delta_2 > 0$ such that $H^c_{d^p,d} < \frac{1}{2} H'_{d^p,d} - \delta_2$.

Denote by

$$R^*_{\mathbf{B}_{j,k}} \triangleq \frac{\log |C|}{n} m \tag{2.50}$$

the "rate" of block $\mathbf{B}_{j,k}$ when m symbols of its codeword have been received by the decoder. As a result of the FIFO rule,

$$R^*_{\mathbf{B}_{j,k}} \geqslant \frac{1}{2} H'_{d^p,d}. \tag{2.51}$$

Therefore,

$$R^*_{\mathbf{B}_{j,k}} > H^c_{d^p,d} + \delta_2. \tag{2.52}$$

Consequently, $\mathbb{P}[\mathcal{E}_1]$ decays exponentially with n, where

$$\begin{aligned}\mathcal{E}_1 \triangleq \Big\{ & \exists \bar{x}^n \in \mathcal{X}^n \backslash \left\{ \mathbf{B}_{j,k}[n] \right\} : \\ & P_{\mathbf{B}_{j,k}\mathbf{B}^s_{j,k}\left|\mathbf{B}^p_{j,k}\right.} \left(\bar{x}^n, \mathbf{B}^s_{j,k}[n] \middle| \mathbf{B}^p_{j,k}[n] \right) \geqslant P_{\mathbf{B}_{j,k}\mathbf{B}^s_{j,k}\left|\mathbf{B}^p_{j,k}\right.} \left(\mathbf{B}_{j,k}[n], \mathbf{B}^s_{j,k}[n] \middle| \mathbf{B}^d_{j,k}[n] \right), \\ & \lfloor \mathsf{F}(\bar{x}^n) \rfloor_m = \left\lfloor \mathsf{F}\left(\mathbf{B}_{j,k}[n] \right) \right\rfloor_m \Big\}\end{aligned}$$

is defined in the same essence as in Lemma 9. The probability $\mathbb{P}[\mathcal{I}] = \mathbb{P}[\mathbf{B}_{j,k}[n] = \mathbf{B}^s_{j,k}[n]]$ also decays exponentially. For n sufficiently large,

$$\mathbb{P}[\mathcal{E}_1] + \mathbb{P}[\mathcal{I}] \leqslant \frac{1}{\sqrt{n}}. \tag{2.53}$$

The rest of the proof resembles the m_{A} case. Denote

$$\begin{aligned} J_n &\triangleq \imath \left(\mathbf{B}_{j,k}[n], \mathbf{B}^s_{j,k}[n] \middle| \mathbf{B}^p_{j,k}[n] \right) \\ &= \sum_{t=T_{j,k}-n+1}^{T_{j,k}-d^p_{j,k}} \imath(X_{1,t}|X_{2,t}) + \sum_{t=T_{j,k}-d^p_{j,k}+1}^{T_{j,k}} \imath(X_{1,t}, X_{2,t}) + \sum_{t=T_{j,k}+1}^{T_{j,k}+d_{j,k}} \imath(X_{2,t}) \end{aligned}$$

and

$$\begin{aligned} \bar{J}_n &\triangleq \imath \left(\bar{X}^{-n}_{1,T_{j,k}}, \bar{X}^{-n}_{2,T_{j,k}+d} \middle| \mathbf{B}^p_{j,k}[n] \right) \\ &= \sum_{t=T_{j,k}-n+1}^{T_{j,k}-d^p_{j,k}} \imath(\bar{X}_{1,t}|X_{2,t}) + \sum_{t=T_{j,k}-d^p_{j,k}+1}^{T_{j,k}} \imath(\bar{X}_{1,t}, \bar{X}_{2,t}) + \sum_{t=T_{j,k}+1}^{T_{j,k}+d_{j,k}} \imath(\bar{X}_{2,t}) \end{aligned}$$

where $(\bar{X}^{-n}_{1,T_{j,k}}, \bar{X}^{-n}_{2,T_{j,k}+d_{j,k}})$ are distributed with respect to $P^n_X P^n_X$ and is independent of $\mathbf{B}_{j,k}$ or $\mathbf{B}^p_{j,k}$. Both J_n and $\bar{J}_n$ are sums of $n + d_{j,k}$ independent (but not necessarily identical) random variables. Applying the part of Lemma 9 for joint decoding, (2.28), when decoded at $m \geqslant m_{\mathrm{B}}(d^p_{j,k}, d_{j,k})$ but $m \leqslant m_{\mathrm{A}}(d^p_{j,k})$, the ARASC code satisfies

$$\epsilon' \leqslant \frac{1}{\sqrt{n}} + \mathbb{E}\left[\min\left\{1, \frac{1}{M_1 M_2}\bar{A}\right\}\right] \tag{2.54}$$

where $\log M_1 = nR_{\mathbf{B}_{j,k}} = m \log |C|$, $\log M_2 = (m - d_{j,k}) \log |C|$, and

$$\bar{A} = \mathbb{E}\left[\exp(\bar{J}_n)\mathbf{1}\{\bar{J}_n \leqslant J_n\}\middle| X^{-n}_{1,T_{j,k}}, X^{-n}_{1,T_{j,k}+d}, X^{-\left(b-d^p_{j,k}\right)}_{2,T_{j,k}-d^p_{j,k}}\right].$$

Using [22, Lemma 47],

$$\bar{A} \leqslant \frac{C''}{\sqrt{n+d}} \exp(J_n). \tag{2.55}$$

Here,

$$C'' \triangleq 2\left(\frac{\log 2}{\sqrt{2\pi \frac{nV'_{d^p_{j,k},d_{j,k}}}{n+d}}} + 2B''\right) \tag{2.56}$$

with

$$B'' = C_0 \frac{\frac{nT'_{d^p_{j,k},d_{j,k}}}{n+d}}{\left(\frac{nV'_{d^p_{j,k},d_{j,k}}}{n+d}\right)^{3/2}} \tag{2.57}$$

where

$$T'_{d^p,d} \triangleq \frac{n-d^p}{n} T(X_1|X_2) + \frac{2d-d^p}{n} T(X_1) + \frac{n-d}{n} T(X_1, X_2). \tag{2.58}$$

The assumptions (2.7) and (2.8), together with the fact that $d_{j,k} \in (n]$, guarantees that C'' is a finite positive constant. Therefore,

$$\epsilon' \leqslant \mathbb{P}[J_n \geqslant \log M_1 M_2 + \frac{1}{2}\log(n+d) - \log C''] + \frac{C''}{\sqrt{n+d}} + \frac{1}{\sqrt{n}}.$$

Since I chose $m_{\mathrm{B}}(d^p, d)$ to satisfy

$$(m_{\mathrm{B}}(d^p, d) + (m_{\mathrm{B}}(d^p, d) - d)) \log |C| \tag{2.59}$$

$$= \log M_1 M_2 \tag{2.60}$$

$$\geqslant nH'_{d^p,d} + \sqrt{nV'_{d^p,d}} Q^{-1}\left(\epsilon - \frac{B'_m + C'_m + 1}{\sqrt{n}}\right), \tag{2.61}$$

when n is sufficiently large,

$$\log M_1 M_2 \geqslant nH'_{d^p_{j,k},d_{j,k}} + \sqrt{nV'_{d^p_{j,k},d_{j,k}}}Q^{-1}\left(\epsilon - \frac{B'_m + C'_m + 1}{\sqrt{n}}\right) - \frac{1}{2}\log(n + d_{j,k}) + \log C'', \tag{2.62}$$

hence $\epsilon' \leqslant \epsilon$ by the Berry-Esseen bound. Note that our choice of C_2 is

$$C_2 \triangleq C'_m + B'_m + 1 \tag{2.63}$$

where

$$C'_m \triangleq \max_{d^p,d} \sqrt{\frac{n}{n+d}} C'' \tag{2.64}$$

$$B'_m \triangleq \max_{d^p,d} \sqrt{\frac{n}{n+d}} B''. \tag{2.65}$$

Again, these constants are chosen for simplicity.

From the above two cases, it can be concluded that Lemma 1 is true. □

Proof of Lemma 2

Proof. Number the blocks $\mathbf{B}_{j',k'}$ such that $(T_{j',k'}, j') \preceq (T_{j,k}, j)$ as $\mathbf{B}_1, \ldots, \mathbf{B}_{k+k^c}$, sorted in ascending order of encoding time and transmitter index. Refer to their encoding times by $T_1, \ldots, T_{k+k^c}$, their codelengths by $m_1, \ldots, m_{k+k^c}$, and their $\left(d^p_{j,k}, d_{j,k}, d^s_{j,k}\right)$ triples by

$$\left(d^p_1, d_1, d^s_1\right), \ldots, \left(d^p_{k+k^c}, d_{k+k^c}, d^s_{k+k^c}\right).$$

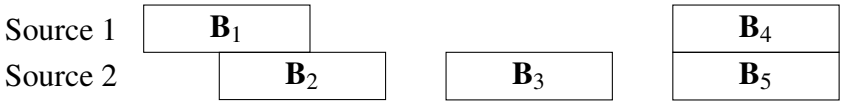


Figure 2.10: Numbering of blocks in ARASC.

The proof is by induction. I restate (2.17) here since it is used repeatedly:

$$\bar{R}_{j,k} \leqslant \bar{H}_{j,k} + \sqrt{\frac{V_m}{n}}Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right). \tag{2.66}$$

Base: There are two base cases. The first one is $m_1 = m_{\mathrm{A}}(n) < m_{\mathrm{B}}(n, d_1)$. In this case,

$$R_{\mathbf{B}_1} = H(X_1) + \sqrt{\frac{V(X_1)}{n}}Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right). \tag{2.67}$$

$\bar{R} = R_{\mathbf{B}_1}$ satisfies (2.66) as $N_j = n$, $N_{j^c} = N_{j,j^c} = 0$.

The second case is $m_1 = m_{\mathrm{B}}(n, d_1^p)$, and $m_2 < m_{\mathrm{B}}(d_2^p, d_2)$. $\bar{R}$ satisfies (2.66) by Lemma 3.

Hypothesis: Let $\mathcal{A}$ denote the set of blocks that satisfy the condition in Lemma 2. For $k + k^c > 1$ under the first base case or $k + k^c > 2$ under the second, let

$$k' \triangleq \min\{k'' < k + k^c, \mathbf{B}_{k'} \in \mathcal{A}\} \tag{2.68}$$

be the last block that is decoded because of hitting m_{A}. By Lemma 3, $k+k^c-k' \leqslant 2$. Hypothesize that the average rate up to $\mathbf{B}_{k'}$ satisfies (2.66).

Induction: If $k + k^c - k' = 1$, by the design of m_{A} and Lemma 5,

$$R_{\mathbf{B}_{k'}} = H_{d_{k'}^p} + \sqrt{\frac{V_{d_{k'}^p}}{n}} Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right). \tag{2.69}$$

Combining definitions of $H_{d^p}, H'_{d^p,d}, N^j_{j,k}, N^{j^c}_{j,k}$ and $N^{j,j^c}_{j,k}$ ((2.12), (2.14), (2.19)-(2.21)) and the fact that $V_{d_{k'}^p} \leqslant V_m$, $\mathbf{B}_{k+k^c}$ satisfies (2.66).

If $k + k^c - k' = 2$, $\mathbf{B}_{k'+1}$ is decoded at $m_{k'+1} = m_{\mathrm{B}}(d^p_{k'+1}, d_{k'+1})$. By Lemma 3,

$$\begin{aligned} &R_{\mathbf{B}_{k'+1}} + R_{\mathbf{B}_{k+k^c}} \\ &\quad\leqslant H'_{d^p_{k'+1},d_{k+k^c}} + 2\sqrt{\frac{V_m}{n}} Q^{-1}(\epsilon) + O\left(\frac{1}{n}\right). \end{aligned} \tag{2.70}$$

Combining the above equation with definitions for weighted entropies, (2.12)-(2.14), and definitions for time of one or two sources being encoded, (2.19)-(2.21), $\mathbf{B}_{k+k^c}$ satisfies (2.66). □

Proof of Lemma 3

Proof. Let $m' \triangleq m_{j,k} - d_{j,k}$ be the number of codeword symbols of successor $\mathbf{B}^s_{j,k}$ received by the decoder at time $T^f_{j,k}$. The case $m_{j,k} = m_{\mathrm{B}}(d^p_{j,k}, d_{j,k})$ can be divided into two categories. In the first case, $m' \geqslant m_{\mathrm{A}}(d_{j,k})$. In this case, $\mathbf{B}^s_{j,k}$ is decoded at $T^f_{j,k} + 1$, giving

$$\begin{aligned} &(m + (m - d_{j,k})) \log |C| \\ &\quad= nH'_{d^p_{j,k},d_{j,k}} + \sqrt{nV'_{d^p_{j,k},d_{j,k}}} Q^{-1}\left(\epsilon - \frac{C_2}{\sqrt{n}}\right) + O(1). \end{aligned} \tag{2.71}$$

Therefore,

$$R_{j,k} + R^s_{j,k} \\ = H'_{d^p_{j,k},d_{j,k}} + \sqrt{\frac{V'_{d^p_{j,k},d_{j,k}}}{n}} Q^{-1}\left(\epsilon - \frac{C_2}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right). \tag{2.72}$$

Applying Lemma 8 yields

$$R_{j,k} + R^s_{j,k} \\ \leqslant H'_{d^p_{j,k},d_{j,k}} + 2\sqrt{\frac{V_m}{n}} Q^{-1}\left(\epsilon - \frac{C_2}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right). \tag{2.73}$$

On the other hand, when $m' < m_{\mathrm{A}}(d_{j,k})$, by Lemma 4,

$$m^s_{j,k} \log |C| = nH_{d_{j,k}} + \sqrt{nV_{d_{j,k}}} Q^{-1}\left(\epsilon - \frac{C_1}{\sqrt{n}}\right) + O(1). \tag{2.74}$$

At the same time, since $m_{j,k} < m_{\mathrm{A}}(d^p_{j,k})$,

$$m_{j,k} \log |C| < nH_{d^p_{j,k}} + \sqrt{nV_{d^p_{j,k}}} Q^{-1}\left(\epsilon - \frac{C_1}{\sqrt{n}}\right). \tag{2.75}$$

Adding (2.74) and (2.75) gives

$$R_{j,k} + R^s_{j,k} \tag{2.76}$$

$$\leqslant H'_{d^p_{j,k},d_{j,k}} + \left(\sqrt{\frac{V_{d^p_{j,k}}}{n}} + \sqrt{\frac{V_{d_{j,k}}}{n}}\right) Q^{-1}\left(\epsilon - \frac{C_1}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right) \tag{2.77}$$

$$\leqslant H'_{d^p_{j,k},d_{j,k}} + 2\sqrt{\frac{V_m}{n}} Q^{-1}\left(\epsilon - \frac{C_2}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right). \tag{2.78}$$

Therefore, no matter $m' \geqslant m_{\mathrm{A}}(d_{j,k})$ or $m' < m_{\mathrm{A}}(d_{j,k})$, the second order term of the sum rate is always bounded from above by $2\sqrt{V_m/n}Q^{-1}(\epsilon)$, completing the proof. □

Proof of Lemma 4

Proof. Since $m_{j,k} = m_{\mathrm{B}}(d^p_{j,k}, d_{j,k})$, the codelength $m_{j,k}$ satisfies

$$m_{j,k} \log |C| + (m_{j,k} - d_{j,k}) \log |C| \\ \geqslant nH'_{d^p_{j,k},d_{j,k}} + \sqrt{nV'_{d^p_{j,k},d_{j,k}}} Q^{-1}\left(\epsilon - \frac{C_2}{\sqrt{n}}\right). \tag{2.79}$$

Note that, at time $T^f_{j,k}$, the number of codeword symbols reaching the decoder that belongs to the successor $\mathbf{B}^s_{j,k}$ is

$$m' = m_{j,k} - d_{j,k}. \tag{2.80}$$

Meanwhile, $m_{j,k} < m_{\mathrm{A}}(d^p_{j,k})$ implies that

$$m_{j,k} \log |C| < nH_{d^p_{j,k}} + \sqrt{nV_{d^p_{j,k}}} Q^{-1}\left(\epsilon - \frac{C_1}{\sqrt{n}}\right). \tag{2.81}$$

Combining (2.79), (2.80) and (2.81),

$$m' \log |C| > nH_{d_{j,k}} - O(\sqrt{n}). \tag{2.82}$$

Therefore,

$$m_{\mathrm{A}}(d_{j,k}) - m' \leqslant O(\sqrt{n}). \tag{2.83}$$

As encoder j cannot encode again until $T^f_{j,k} + 1$,

$$d^s_{j,k} \geqslant m', \tag{2.84}$$

hence

$$m_{\mathrm{B}}(d_{j,k}, d^s_{j,k}) - m' = O(n). \tag{2.85}$$

Therefore, for sufficiently large n, $\mathbf{B}^s_{j,k}$ is decoded because of hitting $m_{\mathrm{A}}(d_{j,k})$, instead of $m_{\mathrm{B}}(d_{j,k}, d^s_{j,k})$. □

Proof of Lemma 5

Proof. Consider $m' = T^f_{j,k} - T_{j',k'} + 1$, the number of codeword symbols for $\mathbf{B}^s_{j,k}$ that the decoder has already received by the time $\mathbf{B}_{j,k}$ is decoded. Since $m_{\mathrm{A}}(d_{j,k}) \leqslant m_{j,k} \leqslant m_{\mathrm{B}}(d^p_{j,k}, d_{j,k})$,

$$\begin{aligned}
m' \log |C| &= (m_{j,k} - d_{j,k}) \log |C| \\
&\leqslant nH'_{d^p_{j,k}, d_{j,k}} + \sqrt{nV'_{d^p_{j,k}, d_{j,k}}} Q^{-1}\left(\epsilon - \frac{C_2}{\sqrt{n}}\right) \\
&\quad - nH_{d^p_{j,k}} - \sqrt{nV_{d^p_{j,k}}} Q^{-1}\left(\epsilon - \frac{C_1}{\sqrt{n}}\right) \\
&\leqslant nH_{d_{j,k}} + \left(\sqrt{nV'_{d^p_{j,k}, d_{j,k}}} - \sqrt{nV_{d^p_{j,k}}}\right) Q^{-1}(\epsilon) + O(1) \\
&\leqslant nH_{d_{j,k}} + \sqrt{nV_{d_{j,k}}} Q^{-1}(\epsilon) + O(1).
\end{aligned}$$

Therefore, for sufficiently large n, $m' < m_{\mathrm{A}}(d_{j,k}) + O(1)$. The last inequality holds due to Lemma 6 and Lemma 7. Meanwhile, by the same reasoning as in the proof of Lemma 4, $m_{\mathrm{B}}(d_{j,k}, d^s_{j,k}) - m' = O(n)$. When n is sufficiently large, $m' < m_{\mathrm{B}}(d_{j,k}, d^s_{j,k})$, completing the proof. □

Chapter 3

SOURCE CODING WITH UNRELIABLE SIDE INFORMATION AT THE DECODER

The materials in this chapter are published in part as [42].

3.1 Introduction

In source coding, the availability of dependent side information at the decoder improves the efficiency of lossless source sequence description. In this chapter, I consider the impact of side information uncertainty caused by failures such as faulty sensors or noisy data collection. Given unreliable decoder side information, the generalized Slepian-Wolf code should take advantage of the side information observation when it is useful, ignore it when it is not, and in both cases ensure a bound on the error probability. This is challenging because encoders typically don't have access to side information available to the decoder, and neither the encoder nor the decoder knows whether the side information observation is good, though the decoder may be able to learn something about that from the side information itself.

In this chapter, I introduce two models of unreliable side information, the *worst-case model* and the *unknown model*. In the worst-case model, a Bernoulli-p random variable that is unknown to the encoder and the decoder and independent of the source and side information determines whether the decoder sees the side information sequence or an independent random sequence with the same marginal. I call this the worst-case model because any source coding advantage of decoder side information must arise from some sort of statistical dependence between the source and side information; here there is no dependence available. In the unknown model, the observed side information again equals the intended side information with at least probability p; in this case, however, when the side information does not equal the intended side information, the dependence between the decoder's observation and the source-side information sequence pair is unknown. To derive achievability bounds for both cases, I propose a source code using the classical random binning encoder design and stop feedback from the decoder. Under the worst-case model, the resulting achievability bound matches the converse up to the second order. Under the unknown model, the code achieves first order Slepian-Wolf optimal rate when the side information observation equals the true side information and first order

optimal point-to-point rate when the side information is faulty, despite the fact that neither the encoder nor the decoder receives side information to distinguish between these scenarios. I discuss the performance of the code under some example joint distributions.

The code that I propose in this chapter focuses only on reconstructing the source of interest with the help of decoder side information. It does not attempt to also determine the joint distribution of the source and side information. The study of this alternative problem under a scenario similar to the worst-case model appears in Chapter 5 of this thesis.

Since the unknown model imposes an upper bound on the probability that the side information is wrong but places no other constraints on the joint probability, it can be employed in a range of scenarios with unreliable side information. One of them is the case where the side information is the decoder output of some other source code, for example, the output of one decoding action in the ARASC described in Chapter 2. Therefore, the code I propose under the unknown model in this chapter can be used as a strategy to stop error propagation whenever the receiver deems appropriate. In fact, though not boasting optimal rate, the proposed technique is a simple strategy to bound the probability of output error in the wake of an unreliable input. It is useful beyond this source coding problem and can be adapted to stop errors from propagating in successive interference cancelation [43], [44], [45], [46].

3.2 Problem Formulation

The Worst-Case Model

Consider a pair of stationary, memoryless, dependent random processes $\{X\}_{i=1}^{\infty}$ and $\{Y\}_{i=1}^{\infty}$ with single-letter joint distribution P_{XY} on discrete alphabet $\mathcal{X} \times \mathcal{Y}$. Random processes $\{X\}_{i=1}^{\infty}$ and $\{Y\}_{i=1}^{\infty}$ represent the source and side information, respectively. Since they are dependent by assumption,

$$H(X|Y) < H(X).$$

For any $\epsilon_0 \in [0, 1]$, an ϵ_0-unreliable observation $\{\hat{Y}\}_{i=1}^{\infty}$ of side information process $\{Y\}_{i=1}^{\infty}$ is defined by

$$\hat{Y}_i = W \cdot Y_i + (1 - W) \cdot Z_i, \tag{3.1}$$

where W is a single Bernoulli random variable (rather than a random process) with $\mathbb{P}[W = 0] = \epsilon_0$, $\{Z\}_{i=1}^{\infty}$ is a stationary, memoryless random process with single-letter distribution P_Y on $\mathcal{Y}$. Thus, $\{Z\}_{i=1}^{\infty}$ is, on its own, indistinguishable from

$\{Y\}_{i=1}^{\infty}$, and (X, Y, Z, W) has single-letter distribution

$$P_{XYZW}(x, y, z, w) = P_{XY}(x, y)P_Y(z)P_W(w)$$

for all $(x, y, z, w) \in \mathcal{X} \times \mathcal{Y}^2 \times \{0, 1\}$.

The Unknown Model

Under the unknown model, joint random process $\{(X, Y)\}_{i=1}^{\infty}$ is unchanged and (3.1) continues to hold with $\mathbb{P}[W = 0] = \epsilon_0$, but no further assumptions are made about P_{XYZW}. Specifically, neither the encoder, nor the decoder, nor the code designer knows $P_{ZW|XY}$, and $\{Z\}_{i=1}^{\infty}$ need not be memoryless. As noted earlier, the flexibility in the unknown model ensures that codes designed for the unknown model work in a wide variety of scenarios where the true statistics may be difficult to capture.

While the worst-case model statistics are the worst possible statistics that can occur under the unknown model, performance under the unknown model may be inferior to performance under the worst-case model since not knowing the true statistics complicates code design.

Problem Setup

For simplicity, I assume that the codeword symbol alphabet is binary. Define 2-dimensional rate vector

$$\mathbf{R} \triangleq [R_D \quad R_I]'$$

with $R_D < R_I$.

Definition 6. *An* $(n, \mathbf{R})$ *unreliable side-information source code (USSC)* $(\mathsf{f}, \mathsf{g}_D, \mathsf{g}_I)$ *for source random process* $\{X\}_{i=1}^{\infty}$*, true side-information random process* $\{Y\}_{i=1}^{\infty}$ *and observed side-information random process* $\{\hat{Y}\}_{i=1}^{\infty}$ *with discrete alphabets* $\mathcal{X}$*,* $\mathcal{Y}$*, and* $\mathcal{Y}$ *comprises an encoding function* $\mathsf{f} : \mathcal{X}^n \to \{0, 1\}^{nR_I}$ *and a pair of decoding functions,* $\mathsf{g}_D : \{0, 1\}^{nR_D} \times \mathcal{Y}^n \to \mathcal{X}^n \cup \{\mathsf{e}\}$ *for some symbol* $\mathsf{e} \notin \mathcal{X}^n$ *and* $\mathsf{g}_I : \{0, 1\}^{nR_I} \to \mathcal{X}^n$*. The encoder describes the source to the decoder in two phases. The first* nR_D *bits of* $\mathsf{f}(X^n)$*, denoted by* $\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}$*, constitute the first phase. The decoder* g_D *maps these bits either to a reconstruction from* $\mathcal{X}^n$ *or to a symbol* $\mathsf{e} \notin \mathcal{X}^n$ *indicating that it is not yet able to decode. Following the tradition of [47][9], a single bit of feedback to the encoder specifies whether (*$\mathsf{g}_D(\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}, \hat{Y}^n) = \mathsf{e}$*) or not (*$\mathsf{g}_D(\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}, \hat{Y}^n) \neq \mathsf{e}$*) the remaining* $R_I - R_D$ *bits of* $\mathsf{f}(X^n)$ *should be sent. An error occurs either if the decoder* g_D *fails to decode correctly and promptly when the side information observation is good (*$W = 1$*) or if source reproduction differs*

from X^n when the side information observation fails ($W = 0$), giving

$$P_{\mathsf{e}}^{(n)}(1) = \mathbb{P}[\mathsf{g}_D(\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}, \hat{Y}^n) \neq X^n | W = 1] \tag{3.2}$$

and

$$P_{\mathsf{e}}^{(n)}(0) = \mathbb{P}\left[\{\mathsf{g}_D(\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}, \hat{Y}^n) \in \mathcal{X}^n \backslash \{X^n\}\right. \tag{3.3}$$

$$\left. \cup \{\{\mathsf{g}_D(\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}, \hat{Y}^n) = \mathsf{e}\} \cap \{\mathsf{g}_I(\mathsf{f}(X^n)) \neq X^n\}\} \middle| W = 0\right]. \tag{3.4}$$

The discussion that follows considers the rate achievable by a random ensemble of $(n, \mathbf{R})$ codes. While strategies for using random codes to prove the existence of deterministic codes that meet close performance bounds are common in literature, attempts to remove the randomness from our USSC arguments to date result in penalties in the 3rd and 2rd order terms in Theorem 5 and Theorem 6, respectively. I will not include detailed proofs about deterministic USSC codes, but arguments of a similar style can be found in Chapter 4 and Chapter 5 of this thesis. Definitions relevant to the random USSC scenario follow.

Definition 7. *An $(n, \mathbf{R})$ USSC random code ensemble is a random variable on the set of all $(n, \mathbf{R})$ USSC codes.*

Definition 8. *Define 2-dimensional error vector*

$$\boldsymbol{\epsilon} \triangleq [\epsilon_D \quad \epsilon_I]' \in (0, 1)^2.$$

Rate pair $\mathbf{R} = [R_D \quad R_I]'$ is $(n, \boldsymbol{\epsilon})$-achievable if there exists an $(n, [R_D' \quad R_I']')$ USSC random code ensemble $(\mathsf{F}, \mathsf{G}_D, \mathsf{G}_I)$ such that $R_D' \leqslant R_D$, $R_I' \leqslant R_I$, $\mathbb{E}[P_{\mathsf{e}}^{(n)}(1)] \leqslant \epsilon_D$ and $\mathbb{E}[P_{\mathsf{e}}^{(n)}(0)] \leqslant \epsilon_I$. Rate region $\mathscr{R}_{US}^*(n, \boldsymbol{\epsilon})$ is the closure of the set of $(n, \boldsymbol{\epsilon})$-achievable rate pairs.

Remark 3. *While the case $\epsilon_D = \epsilon_I = \epsilon$ may be most useful for many applications, differing ϵ_D and ϵ_I may be useful, for example, when an impending deadline changes the user's priorities regarding acceptable error probability.*

3.3 Main Results for the Worst-Case Model

Consider the third-order optimal Slepian-Wolf rate

$$R_D^* \triangleq H(X|Y) + \sqrt{\frac{V(X|Y)}{n}} Q^{-1}(\epsilon_D) - \frac{\log n}{2n} \tag{3.5}$$

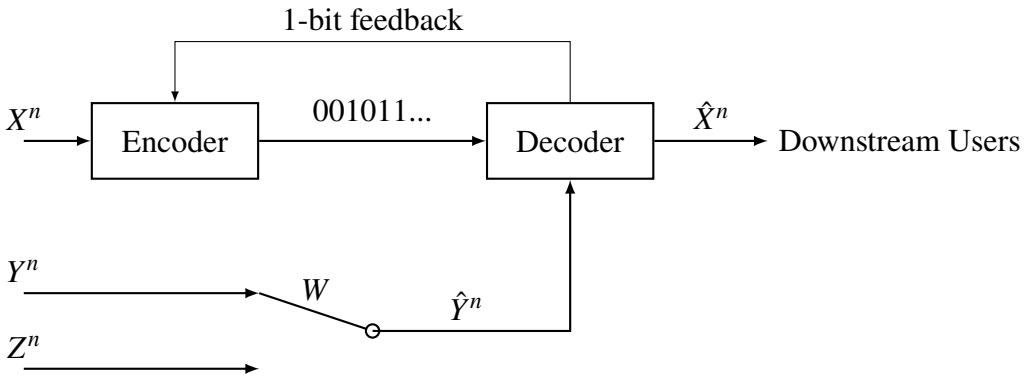


Figure 3.1: Block diagram of the unreliable side information source code (USSC). Binary random variable W controls whether $\hat{Y}^n = Y^n$ or $\hat{Y}^n = Z^n$.

from [15] and the third-order optimal point-to-point rate

$$R_I^* \triangleq H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon_I) - \frac{\log n}{2n} \tag{3.6}$$

from [28][1]. Define inner and outer bounding rate regions

$$\mathscr{R}_{\text{in}}(n, \boldsymbol{\epsilon}) \triangleq \left\{ \mathbf{R} : R_D \geqslant R_D^* + \frac{\log n}{2n} + O\left(\frac{1}{n}\right), R_I \geqslant R_I^* + O\left(\frac{1}{n}\right) \right\} \tag{3.7}$$

$$\mathscr{R}_{\text{out}}(n, \boldsymbol{\epsilon}) \triangleq \left\{ \mathbf{R} : R_D \geqslant R_D^* - O\left(\frac{1}{n}\right), R_I \geqslant R_I^* - O\left(\frac{1}{n}\right) \right\}. \tag{3.8}$$

Theorem 5. *Under the worst-case model, if*

$$V(X) > 0,\ V(X|Y) > 0,\ T(X) < \infty,\ T(X|Y) < \infty, \tag{3.9}$$

then

$$\mathscr{R}_{\text{in}}(n, \boldsymbol{\epsilon}) \subseteq \mathscr{R}_{US}^*(n, \boldsymbol{\epsilon}) \subseteq \mathscr{R}_{\text{out}}(n, \boldsymbol{\epsilon}).$$

Remark 4. *This code design can be extended to more than two possible joint distributions $P_{XY}^{(1)}, P_{XY}^{(2)}, \ldots$ by increasing the number of stages, allowing $R^{(1)}, R^{(2)}, \ldots$ in place of R_D and R_I. The result is a universal Slepian-Wolf code for a finite family of possible side information distributions.*

[1]The second- and third-order optimal rate regions for the Slepian-Wolf scenario where side information is available directly to the decoder are due to [26, Eqn. 29a] and [15, Theorem 7, 8], respectively. The third-order gap between the R_D boundary in our result and the rate region for a code with true side information by Gavalakis and Kontonyiannis [15] is due to the inability of the decoder to distinguish Y^n from Z^n. The gap can be closed if the marginals are different.

Proof of Achievability

While it is possible, under the given framework, to first send a subsequence of $\{X\}$ losslessly and then use that subsequence to learn W at the decoder, I here employ instead a code design where the decoder determines W and the encoded message X^n through a single decoding operation. This approach is preferable when W can change across many uses of the code.

Codebook Generation

For each $x^n \in \mathcal{X}^n$, randomly and independently draw encoder output $\mathsf{f}(x^n)$ from the uniform distribution on $[2^{nR_I}]$. Each index in $[2^{nR_I}]$ represents a unique binary string of length nR_I.

Decoder

The decoder g_D performs threshold decoding using a constant threshold $\log \gamma$ defined below, giving

$$\mathsf{g}_D(c_D, \hat{y}^n) = \begin{cases} \hat{x}^n & \text{if } \lfloor \mathsf{f}(\hat{x}^n) \rfloor_{nR_D} = c_D,\ \imath(\hat{x}^n|\hat{y}^n) \leqslant \log \gamma \\ \mathsf{e} & \text{otherwise.} \end{cases} \tag{3.10}$$

When more than one source vector $\hat{x}^n$ satisfies the threshold test in (3.10), choose one that minimizes $\imath(\hat{x}^n|\hat{y}^n)$. The decoder g_I performs maximum likelihood decoding, giving

$$\mathsf{g}_I(c_I) = \arg \min_{x^n \in \mathcal{X}^n : \mathsf{f}(x^n) = c_I} \imath(x^n).$$

Ties are broken uniformly at random.

Error Analysis

When $W = 1$, the dependence of the side information makes it useful for source coding. As defined in Definition 6, $P_{\mathsf{e}}^{(n)}(1)$, the probability of error or late reconstruction when $W = 1$, includes any case where the decoder cannot decode correctly after receiving nR_D bits, giving

$$\begin{aligned} \mathbb{E}\left[P_{\mathsf{e}}^{(n)}(1)\right] &= \mathbb{P}[\mathsf{G}_D(\lfloor \mathsf{F}(X^n) \rfloor_{nR_D}, \hat{Y}^n) \neq X^n | W = 1] \\ &\leqslant \mathbb{P}[\{\exists \bar{x}^n \in \mathcal{X}^n \backslash \{X^n\} : \lfloor \mathsf{F}(\bar{x}^n) \rfloor_{nR_D} = \lfloor \mathsf{F}(X^n) \rfloor_{nR_D}, \imath(\bar{x}^n|\hat{Y}^n) \leqslant \log \gamma\} \\ &\qquad \cup \{\imath(X^n|\hat{Y}^n) > \log \gamma\} | W = 1] \\ &\leqslant \mathbb{P}\left[\exists \bar{x}^n \in \mathcal{X}^n \backslash \{X^n\} : \lfloor \mathsf{F}(\bar{x}^n) \rfloor_{nR_D} = \lfloor \mathsf{F}(X^n) \rfloor_{nR_D}, \imath(\bar{x}^n|\hat{Y}^n) \leqslant \log \gamma | W = 1\right] \\ &\qquad + \mathbb{P}\left[\imath(X^n|\hat{Y}^n) > \log \gamma | W = 1\right] \end{aligned} \tag{3.11}$$

by the union bound. For the first term in (3.11),

$$\begin{aligned}
&\mathbb{P}\left[\exists \bar{x}^n \in \mathcal{X}^n \backslash \{X^n\} : \lfloor \mathsf{F}(\bar{x}^n) \rfloor_{nR_D} = \lfloor \mathsf{F}(X^n) \rfloor_{nR_D}, \imath(\bar{x}^n|\hat{Y}^n) \leqslant \log\gamma | W = 1\right] \\
=&\mathbb{P}\left[\exists \bar{x}^n \in \mathcal{X}^n \backslash \{X^n\} : \lfloor \mathsf{F}(\bar{x}^n) \rfloor_{nR_D} = \lfloor \mathsf{F}(X^n) \rfloor_{nR_D}, \imath(\bar{x}^n|Y^n) \leqslant \log\gamma\right] && (3.12)\\
\leqslant&\frac{1}{2^{nR_D}} \sum_{\bar{x}^n \in \mathcal{X}^n,\, y^n \in \mathcal{Y}^n} P_Y^n(y^n) \mathbf{1}\{\imath(\bar{x}^n|y^n) \leqslant \log\gamma\} && (3.13)\\
=&\frac{1}{2^{nR_D}} \mathbb{E}\left[\frac{1}{P_{X|Y}^n(X^n|Y^n)} \mathbf{1}\{\imath(X^n|Y^n) \leqslant \log\gamma\}\right], && (3.14)
\end{aligned}$$

where the equality (3.12) follows since $Y^n = \hat{Y}^n$ when $W = 1$ and Y^n, Z^n and W are independent under the worst-case model assumption.

Let

$$\gamma = 2^{nR_D} \tag{3.15}$$

$$nR_D = nH(X|Y) + \sqrt{nV(X|Y)}Q^{-1}\left(\epsilon_D - \frac{B+C}{\sqrt{n}}\right). \tag{3.16}$$

Given the theorem's distribution assumptions from (3.9), applying [21, Lemma 47] gives

$$\mathbb{E}\left[\frac{1}{P_{X|Y}^n(X^n|Y^n)} \mathbf{1}\{\imath(X^n|Y^n) \leqslant \log\gamma\}\right] \leqslant \frac{C}{\sqrt{n}} 2^{nR_D}$$

for some constant C. Therefore,

$$\mathbb{P}\left[\exists \bar{x}^n \in \mathcal{X}^n \backslash \{X^n\} : \lfloor \mathsf{F}(\bar{x}^n) \rfloor_{nR_D} = \lfloor \mathsf{F}(X^n) \rfloor_{nR_D}, \imath(\bar{x}^n|\hat{Y}^n) \leqslant \log\gamma | W = 1\right] \leqslant \frac{C}{\sqrt{n}}. \tag{3.17}$$

For the second term in (3.11),

$$\begin{aligned}
\mathbb{P}\left[\imath(X^n|\hat{Y}^n) > \log\gamma | W = 1\right] &= \mathbb{P}\left[\imath(X^n|Y^n) > \log\gamma\right] && (3.18)\\
&\leqslant Q\left(Q^{-1}\left(\epsilon_D - \frac{B+C}{\sqrt{n}}\right)\right) + \frac{B}{\sqrt{n}} && (3.19)\\
&= \epsilon_D - \frac{C}{\sqrt{n}}, && (3.20)
\end{aligned}$$

for some constant B, where (3.18) follows again from $\hat{Y}^n = Y^n$ when $W = 1$ and the independence among W, (X^n, Y^n) and Z^n, and (3.19) follows from the Berry Esseen bound.

Thus, combining (3.17) and (3.20), $\mathbb{E}[P_{\mathsf{e}}^{(n)}(1)] \leqslant \epsilon_D$ under the setting of R_D in (3.16). Here

$$R_D = H(X|Y) + \sqrt{\frac{V(X|Y)}{n}} Q^{-1}(\epsilon_D) + O\left(\frac{1}{n}\right), \tag{3.21}$$

where

$$\frac{1}{n}\sqrt{nV(X|Y)}Q^{-1}\left(\epsilon_D - \frac{B+C}{\sqrt{n}}\right) = \sqrt{\frac{V(X|Y)}{n}}Q^{-1}(\epsilon_D) + O\left(\frac{1}{n}\right) \tag{3.22}$$

follows from the differentiability and Taylor expansion of $Q^{-1}(\cdot)$.

To bound the expectation of error probability $P_{\mathsf{e}}^{(n)}(0)$ for the case where $W = 0$, recall that two scenarios – incorrect decoding after nR_D bits and incorrect decoding later – yield errors when the side information is independent of the source. Therefore, by the union bound,

$$\begin{aligned}\mathbb{E}\left[P_{\mathsf{e}}^{(n)}(0)\right] \leqslant \mathbb{P}\left[\mathsf{G}_D(\lfloor \mathsf{F}(X^n)\rfloor_{nR_D}, \hat{Y}^n) \in \mathcal{X}^n\backslash\{X^n\}|W=0\right] \\ + \mathbb{P}\left[\mathsf{G}_I(\mathsf{F}(X^n)) \neq X^n|W=0\right],\end{aligned}$$

The probability of decoding to an incorrect source vector $\bar{x}^n$ after nR_D bits is identical whether $W = 0$ or $W = 1$ since in both cases the incorrect codeword is independent of the observed side information. Therefore,

$$\begin{aligned}&\mathbb{P}[\mathsf{G}_D(\lfloor \mathsf{F}(X^n)\rfloor_{nR_D}, \hat{Y}^n) \in \mathcal{X}^n\backslash\{X^n\}|W=0] \\ &\quad\leqslant \mathbb{P}[\exists \bar{x}^n \in \mathcal{X}^n\backslash\{X^n\} : \lfloor \mathsf{F}(\bar{x}^n)\rfloor_{nR_D} = \lfloor \mathsf{F}(X^n)\rfloor_{nR_D}, \imath(\bar{x}^n|Z^n) \leqslant \log\gamma] \\ &\quad\leqslant \frac{1}{2^{nR_D}} \sum_{\bar{x}^n\in\mathcal{X}^n, z^n\in\mathcal{Y}^n} P_Y^n(z^n)\mathbf{1}\{\imath(\bar{x}^n|z^n) \leqslant \log\gamma\} \\ &\quad= \frac{1}{2^{nR_D}}\mathbb{E}\left[\frac{1}{P_{X|Y}^n(X^n|Y^n)}\mathbf{1}\{\imath(X^n|Y^n) \leqslant \log\gamma\}\right] \\ &\quad\leqslant \frac{C}{\sqrt{n}}\end{aligned}$$

under the same parameter choices, identical to (3.17). For the scenario where the error occurs at rate R_I, $W = 0$ implies $\hat{Y}^n$ is independent of X^n and therefore $\mathbb{P}[\mathsf{G}_I(\mathsf{F}(X^n)) \neq X^n|W = 0]$ is analogous to the error probability for a maximum likelihood decoder that has no useful side information. From [9, Proof of Theorem 5],

$$\mathbb{P}[\mathsf{G}_I(\mathsf{F}(X^n)) \neq X^n|W=0] \leqslant \mathbb{P}[\imath(X^n) > nR_I + \frac{1}{2}\log n - \log C'] + \frac{C'}{\sqrt{n}}.$$

Choosing

$$R_I = H(X) + \sqrt{\frac{V(X)}{n}}Q^{-1}\left(\epsilon_I - \frac{B'+C'+C}{\sqrt{n}}\right) - \frac{\log n}{2n} + \frac{\log C'}{n}$$

(for large enough n) keeps the total expected error probability bounded as $\mathbb{E}[P_{\mathsf{e}}^{(n)}(0)] \leqslant \epsilon_I$. Again relying on the differentiability of $Q^{-1}(\cdot)$ and Taylor's expansion,

$$R_I = H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon_I) - \frac{\log n}{2n} + O\left(\frac{1}{n}\right). \tag{3.23}$$

Combining (3.21) and (3.23) shows that

$$\mathscr{R}_{\text{in}}(n, \boldsymbol{\epsilon}) \subseteq \mathscr{R}_{US}^{*}(n, \boldsymbol{\epsilon}).$$

Proof of Converse

When $W = 1$, the decoder output yields an error if $\mathsf{g}_D(\lfloor \mathsf{f}(X^n) \rfloor_{nR_D}, \hat{Y}^n) = \mathsf{e}$, no matter whether $\mathsf{g}_I(\mathsf{f}(X^n)) = X^n$ or not. Therefore, when $W = 1$, the decoder must decode at a fixed rate R_D, before the encoder receives any feedback. Applying earlier bounds on fixed-rate lossless source coding with side information and without feedback from [26], [15] (as in footnote 1) gives

$$R_D \geqslant H(X|Y) + \sqrt{\frac{V(X|Y)}{n}} Q^{-1}(\epsilon_D) - \frac{\log n}{2n} - O\left(\frac{1}{n}\right). \tag{3.24}$$

In contrast, when $W = 0$, feedback plays an active role.

Since the channel in this source coding problem is noiseless and the decoder is deterministic by assumption,[2] the stop feedback is a deterministic function of the encoder's output and the side information sequence observed at the decoder. Therefore, when $W = 0$, the system can perform no better than a 2-rate system without feedback for which an independent side information sequence is available at both the encoder and the decoder. Since the presence of independent side information does not increase the source coding rate region, and since I here constrain the source code to operate at the same rates R_D and R_I, applying Han's converse [31, Lemma 1.3.2] with $\gamma = \frac{\log n}{2}$ gives

$$\log\left(2^{nR_D} + 2^{nR_I}\right) \geqslant nH(X) + \sqrt{nV(X)} Q^{-1}(\epsilon_I) - \frac{1}{2}\log n - O(1). \tag{3.25}$$

Since $R_I \geqslant R_D$, the left-hand side of (3.25) is no larger than $nR_I + 1$. Therefore,

$$R_I \geqslant H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon_I) - \frac{\log n}{2n} - O\left(\frac{1}{n}\right). \tag{3.26}$$

[2] Allowing a random decoder complicates the argument but does not change the outcome. Every random decoder can be expressed as a deterministic decoder with an additional random input that is independent of the source. Making this random input available to the encoder does not enhance its performance.

Combining (3.24) with (3.26) yields

$$\mathscr{R}^*_{US}(n,\boldsymbol{\epsilon}) \subseteq \mathscr{R}_{\text{out}}(n,\boldsymbol{\epsilon}).$$

□

3.4 Main Results for the Unknown Model

Under the unknown model, both the encoder and the decoder know $\mathbb{P}[W=0]=\epsilon_0$, but neither knows how the side information has been corrupted, here captured by the conditional distribution of $\hat{Y}^n$ given X^n and Y^n when $W=0$. In the proof of the theorem that follows, I use the 2-stage code from Theorem 5 but modify the value of γ to bound the achievable rate in this scenario.

Theorem 6. *Under an unknown model that satisfies the source and side information distribution constraints in* (3.9)*, for any* $0<\epsilon_D<1$ *and* $0<\epsilon_I<1$*, the* $(n,\boldsymbol{\epsilon})$*-rate region* $\mathscr{R}^*_{USW}(n,\boldsymbol{\epsilon})$ *satisfies*

$$\mathscr{R}^*_{USW}(n,\boldsymbol{\epsilon}) \supseteq \mathscr{R}'_{\text{in}}(n,\boldsymbol{\epsilon}),$$

where

$$\mathscr{R}'_{\text{in}}(n,\boldsymbol{\epsilon}) \triangleq \left\{\mathbf{R} : R_D \geqslant R^*_{DW} + \frac{\log n}{n} + O\left(\frac{1}{n}\right), R_I \geqslant R^*_{IW} + O\left(\frac{1}{n}\right)\right\}$$

$$R^*_{DW} \triangleq H(X|Y) + \sqrt{\frac{V(X|Y)}{n}}Q^{-1}(\epsilon_D(1-\epsilon_0)) - \frac{\log n}{2n}$$
$$R^*_{IW} \triangleq H(X) + \sqrt{\frac{V(X)}{n}}Q^{-1}(\epsilon_I\epsilon_0) - \frac{\log n}{2n}.$$

The theorem implies that, up to the first order, there exists a code that can perform as well as a code with true side information when the side information is uncorrupted and as well as a point-to-point code when the side information is corrupted.

Proof of Theorem 6

I use the encoder and decoder from the proof of Theorem 5 but modify threshold parameter γ. Since $W=1$ implied $\hat{Y}^n = Y^n$, the expected conditional error probability given $W=1$, $\mathbb{E}[P_{\mathsf{e}}^{(n)}(1)]$, can be calculated as

$$\begin{aligned}
\mathbb{E}\left[P_{\mathsf{e}}^{(n)}(1)\right] &= \frac{1}{\Pr[W=1]}\mathbb{P}[\mathsf{G}_D(\lfloor \mathsf{F}(X^n)\rfloor_{nR_D}, \hat{Y}^n) \neq X^n, W=1] \\
&= \frac{1}{\Pr[W=1]}\mathbb{P}[\mathsf{G}_D(\lfloor \mathsf{F}(X^n)\rfloor_{nR_D}, Y^n) \neq X^n, W=1] \\
&\leqslant \frac{1}{1-\epsilon_0}\mathbb{P}[\mathsf{G}_D(\lfloor \mathsf{F}(X^n)\rfloor_{nR_D}, Y^n) \neq X^n] \\
&\leqslant \frac{1}{1-\epsilon_0}\mathbb{P}[\imath(X^n|Y^n) > \log\gamma] \\
&\quad + \frac{1}{(1-\epsilon_0)2^{nR_D}} \sum_{\bar{x}^n \in \mathcal{X}^n,\, y^n \in \mathcal{Y}^n} P_Y^n(y^n)\mathbf{1}\{\imath(\bar{x}^n|y^n) \leqslant \log\gamma\} \\
&\leqslant \frac{1}{1-\epsilon_0}\mathbb{P}[\imath(X^n|Y^n) > \log\gamma] + \frac{1}{(1-\epsilon_0)2^{nR_D}}\mathbb{U}[\imath(\bar{X}^n|Y^n) \leqslant \log\gamma],
\end{aligned}$$

where $\mathbb{U}$ is a finite measure on $\mathcal{X}^n \times \mathcal{Y}^n$ for which $U_{X^nY^n}(x^n, y^n) = P_Y^n(y^n)$ for all $(x^n, y^n) \in \mathcal{X}^n \times \mathcal{Y}^n$. Extending the Chernoff Bound from probability measures to a finite measure $\mathbb{Q}$ by applying Markov's Inequality for finite measures gives

$$\mathbb{Q}[X \geqslant a] \leqslant \frac{\sum_{x\in\mathcal{X}} \mathbb{Q}[X=x]\exp\{x\}}{\exp\{a\}}.$$

Therefore,

$$\begin{aligned}
\mathbb{U}[\imath(\bar{X}^n|Y^n) \leqslant \log\gamma] &= \mathbb{U}[-\imath(\bar{X}^n|Y^n) \geqslant -\log\gamma] \\
&\leqslant \frac{\sum_{\bar{x}^n, y^n} -P_Y^n(y^n)\exp\{-\imath(\bar{x}^n|y^n)\}}{\exp\{-\log\gamma\}} \\
&= \gamma \sum_{\bar{x}^n, y^n} P_Y^n(y^n)P_{X|Y}^n(\bar{x}^n|y^n) \\
&= \gamma.
\end{aligned}$$

Setting $\gamma = \frac{2^{nR_D}}{\sqrt{n}}$ gives

$$\frac{1}{2^{nR_D}}\mathbb{U}[\imath(\bar{X}^n|Y^n) \leqslant \log\gamma] \leqslant \frac{1}{\sqrt{n}}, \tag{3.27}$$

and setting

$$R_D = H(X|Y) + \sqrt{\frac{V(X|Y)}{n}}Q^{-1}\left(\epsilon_D(1-\epsilon_0) - \frac{B+1}{\sqrt{n}}\right) + \frac{\log n}{2n}$$

gives

$$\log\gamma = nH(X|Y) + \sqrt{nV(X|Y)}Q^{-1}\left(\epsilon_D(1-\epsilon_0) - \frac{B+1}{\sqrt{n}}\right).$$

Again recalling that P_{XY} satisfies (3.9), apply the Berry Esseen bound to give

$$\begin{aligned}\mathbb{P}[\imath(X^n|Y^n) > \log\gamma] &\leqslant Q\left(Q^{-1}\left(\epsilon_D(1-\epsilon_0) - \frac{B+1}{\sqrt{n}}\right)\right) + \frac{B}{\sqrt{n}} \\ &= \epsilon_D(1-\epsilon_0) - \frac{1}{\sqrt{n}} \end{aligned} \tag{3.28}$$

for some constant B. Combining (3.27) and (3.28) gives $\mathbb{E}[P_{\mathsf{e}}^{(n)}(1)] \leqslant \epsilon_D$ under the given parameter settings. Again employing the Taylor's series expansion of $Q^{-1}(\cdot)$, this argument proves the achievability of rate

$$R_D = H(X|Y) + \sqrt{\frac{V(X|Y)}{n}}Q^{-1}(\epsilon_D(1-\epsilon_0)) + \frac{\log n}{2n} + O\left(\frac{1}{n}\right)$$

with expected conditional error probability $\mathbb{E}[P_{\mathsf{e}}^{(n)}(1)] \leqslant \epsilon_D$. Similarly,

$$\begin{aligned}\mathbb{E}\left[P_{\mathsf{e}}^{(n)}(0)\right] \leqslant \frac{1}{\epsilon_0}\big(&\mathbb{P}[\mathsf{G}_D(\lfloor \mathsf{F}(X^n)\rfloor_{nR_D}, \hat{Y}^n) \in \mathcal{X}^n\backslash\{X^n\}, W=0] \\ &+ \mathbb{P}[\mathsf{G}_I(\mathsf{F}(X^n)) \neq X^n, W=0]\big).\end{aligned} \tag{3.29}$$

Define finite measure $\mathbb{V}$ on $\mathcal{X}^n \times \mathcal{Y}^n$ by $V_{X^nY^n}(x^n, y^n) = P_{\hat{Y}^n}(y^n)$. Note that measure $\mathbb{V}$ is unknown to the encoder and decoder and is used only for the purpose of analysis. Then

$$\begin{aligned}&\mathbb{P}\left[\mathsf{G}_D(\lfloor \mathsf{F}(X^n)\rfloor_{nR_D}\hat{Y}^n) \in \mathcal{X}^n\backslash\{X^n\}, W=0\right] \\ &\leqslant \mathbb{P}[\exists \bar{x}^n \in \mathcal{X}\backslash\{X^n\}, \lfloor \mathsf{f}(\bar{x}^n)\rfloor_{nR_D} = \lfloor \mathsf{f}(X^n)\rfloor_{nR_D}, \imath(\bar{x}^n|\hat{Y}^n) \leqslant \log\gamma] \\ &\leqslant \frac{1}{2^{nR_D}}\mathbb{V}[\imath(\bar{X}^n|\hat{Y}^n) \leqslant \log\gamma]\end{aligned}$$

and

$$\begin{aligned}\mathbb{V}[\imath(\bar{X}^n|\hat{Y}^n) \leqslant \log\gamma] &= \gamma \sum_{\bar{x}^n, \hat{y}^n} P_{\hat{Y}^n}(\hat{y}^n) P_{X|Y}^n(\bar{x}^n|\hat{y}^n) \\ &= \gamma,\end{aligned}$$

where the last equality follows since $P_{X|Y}^n P_{\hat{Y}^n}$ is *some* p.m.f. on $\mathcal{X}^n \times \mathcal{Y}^n$.

Hence $\gamma = \frac{2^{nR_D}}{\sqrt{n}}$ implies

$$\mathbb{P}\left[\mathsf{G}_D(\lfloor \mathsf{F}(X^n)\rfloor_{nR_D}\hat{Y}^n) \in \mathcal{X}^n\backslash\{X^n\}, W=0\right] \leqslant \frac{1}{\sqrt{n}}. \tag{3.30}$$

Using the same analysis for the maximum likelihood decoder in (3.3) and choosing

$$R_I = H(X) + \sqrt{\frac{V(X)}{n}}Q^{-1}\left(\epsilon_I\epsilon_0 - \frac{B'+C'+1}{\sqrt{n}}\right) - \frac{\log n}{2n} + \frac{\log C'}{n}$$

gives

$$\mathbb{P}[\mathsf{G}_D(\lfloor \mathsf{F}(X^n)\rfloor_{nR_D}, \hat{Y}^n) \in \mathcal{X}^n \backslash \{X^n\}, W = 0] + \mathbb{P}[\mathsf{G}_I(\mathsf{F}(X^n)) \neq X^n, W = 0] \leqslant \epsilon_I \epsilon_0,$$

which implies that $\mathbb{E}[P_{\mathsf{e}}^{(n)}(0)] \leqslant \epsilon_I$. Therefore, $\mathscr{R}'_{\text{in}}(n, \boldsymbol{\epsilon}) \subseteq \mathscr{R}^*_{USW}(n, \boldsymbol{\epsilon})$.

□

Remark 5. *If the decoder for the unknown model is used when the distribution of the source and observation follows the worst-case model, the code does as well as the achievability part of Theorem 5 up to the first order. If the decoder for the unknown model is used when the true side information is available at the decoder, with probability* $(1 - \epsilon_D)$*, the source sequence is reconstructed correctly at a rate that asymptotically approaches* $H(X|Y)$.

Remark 6. *Due to the unknown nature of* $P_{ZW|XY}$*, to date, the best known outer bound such that* $\mathscr{R}^*_{USW}(n, \boldsymbol{\epsilon}) \subseteq \mathscr{R}'_{\text{out}}(n, \boldsymbol{\epsilon})$ *is*

$$\mathscr{R}'_{\text{out}}(n, \boldsymbol{\epsilon}) \triangleq \left\{ \mathbf{R} : R_D \geqslant R_D^* - O\left(\frac{1}{n}\right), R_I \geqslant R_D^* - O\left(\frac{1}{n}\right) \right\}.$$

Discussion

The achievability result for the unknown model (Theorem 6) matches the achievability bound for the worst-case model (Theorem 5) only up to the first order. The second order gap and the dependence of the dispersion term on ϵ_0 come from the possible but unknown dependence of the side information reliability random variable, W, the corrupted side information, Z^n, the source, X^n, and the side information, Y^n. The converse of the unknown model does not match the converse of the worst-case model (Theorem 5) due to the uncertainty about the true joint distribution of the source and the side information. The main contribution of our code under the unknown model is that it allows us to take advantage of dependent side information roughly fraction $1 - \epsilon_0$ of the time without allowing flaws in the side information (e.g., a sensor that is faulty fraction-ϵ_0 of the time) to cause propagating errors in the source reconstruction.

3.5 Summary

This chapter introduces the problem of source coding with asymptotically unreliable side information at the decoder and analyzes the performance of a two-stage code under the cases of known and unknown joint distributions of the unreliable side information. The strategy can be used to stop error propagation when the

side information is unreliable due to decoding errors in prior codes. Under the unknown model, the decoder does not have to have knowledge of, or control over, the distribution of the side information beyond its error probability constraint.

Alternatively, if one places oneself in the position of the designer of not the USSC code, but the prior code that is prone to decoding error and therefore can affect downstream codes looking to use the decoder output as side information, one could wonder if, under the same reliability and efficiency constraints, there are ways to design the code such that later stages benefit more from its decoding result. I investigate this alternative problem in the next chapter of this thesis.

Chapter 4

ALMOST LOSSLESS ONION PEELING CODE

The materials in this chapter are published in part as [48], [49].

4.1 Introduction

In a 2-user Slepian-Wolf code, two dependent sources $(\mathbf{X}, \mathbf{Y})$ are encoded independently and decoded jointly. Assume that the sources are discrete, memoryless, and stationary. While Slepian-Wolf codes achieve a savings in sum-rate by allowing the description of each source to be used in reconstructing both sources, the downside is that the decoder often needs to consult the full encoded descriptions of both sources even when it is only interested in reconstructing one. This can be especially inefficient when there are asymmetric needs to decode the two sources - for example, when usually only the reconstruction of $\mathbf{X}$ is desired or when significant delays in transmitting the codeword for $\mathbf{Y}$ regularly impede the reconstruction of $\mathbf{X}$. Operating near the corner point $(H(X), H(Y|X))$ of the Slepian-Wolf rate region [2]

$$\left\{ \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} : R_1 > H(X|Y), \quad R_2 > H(Y|X), \quad R_1 + R_2 > H(X,Y) \right\} \tag{4.1}$$

allows $\mathbf{X}$ to be reconstructed without the description of $\mathbf{Y}$. The scenario in which the decoder reconstructs $\mathbf{X}$ from only the coded description of $\mathbf{X}$, and reconstructs $\mathbf{Y}$ from the coded description of $\mathbf{Y}$ and the reconstruction of $\mathbf{X}$, is called an onion peeling code (See Figure 4.1).

A fixed-rate point-to-point source code, a side information code, and an onion peeling code are defined as follows.

Definition 9 (Point-to-point Code, [9])**.** *For any positive integers n and M, an (n, M) point-to-point code for a source from finite alphabet $\mathcal{X}$ comprises an encoder $\mathsf{f}^{(n)} : \mathcal{X}^n \mapsto [M]$ and a decoder $\mathsf{g}^{(n)} : [M] \mapsto \mathcal{X}^n$. The error probability in coding X^n from probability mass function (p.m.f.) P_{X^n} is $P_{\mathsf{e}}^{(n)} \triangleq \mathbb{P}\left[\mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right) \neq X^n\right]$. The rate of the code is $R \triangleq \frac{\log M}{n}$.*

Definition 10 (Side Information Code, [50])**.** *For any positive integers n and M, an (n, M) side information code for a source from finite alphabet $\mathcal{Y}$ using side information $A^{(n)}$ from a finite alphabet $\mathcal{A}^{(n)}$ comprises an encoder $\mathsf{f}^{(n)} : \mathcal{Y}^n \mapsto [M]$*

and a decoder $\mathsf{g}^{(n)} : [M] \times \mathcal{A}^{(n)} \mapsto \mathcal{Y}^n$. *The error probability in coding* X^n *from probability mass function (p.m.f.)* P_{X^n} *is* $P_{\mathsf{e}}^{(n)} \triangleq \mathbb{P}\left[\mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(Y^n), A^{(n)}\right) \neq Y^n\right]$. *The rate of the code is* $R \triangleq \frac{\log M}{n}$.

Definition 11 (Onion Peeling Code)**.** *For any positive integers n,* M_1*, and* M_2*, an* (n, M_1, M_2) *onion peeling code for sources on finite alphabets* $\mathcal{X}$ *and* $\mathcal{Y}$ *comprises an* (n, M_1) *point-to-point code* $\left(\mathsf{f}_1^{(n)}, \mathsf{g}_1^{(n)}\right)$ *(the first-stage code) for a source with alphabet* $\mathcal{X}$ *and an* (n, M_2) *side information code* $\left(\mathsf{f}_2^{(n)}, \mathsf{g}_2^{(n)}\right)$ *(the second-stage code) for a source with alphabet* $\mathcal{Y}$ *and side information from alphabet* $\mathcal{X}^n$. *The error probabilities in coding random vectors* (X^n, Y^n) *from p.m.f.* $P_{X^nY^n}$ *are* $P_{\mathsf{e},1}^{(n)} \triangleq \mathbb{P}\left[X^n \neq \mathsf{g}_1^{(n)}\left(\mathsf{f}_1^{(n)}(X^n)\right)\right]$ *and* $P_{\mathsf{e},2}^{(n)} \triangleq \mathbb{P}\left[Y^n \neq \mathsf{g}_2^{(n)}\left(\mathsf{f}_2^{(n)}(Y^n), \mathsf{g}_1^{(n)}\left(\mathsf{f}_1^{(n)}(X^n)\right)\right)\right]$. *The rates of the code are* $R_1 \triangleq \frac{\log M_1}{n}$ *and* $R_2 \triangleq \frac{\log M_2}{n}$.

In this chapter, I consider how the reconstruction of $\mathbf{Y}$ can be affected by the algorithm used to reconstruct $\mathbf{X}$. Specifically, I consider potential performance of the reconstruction of $\mathbf{Y}$ by a decoder with access to the reconstruction of $\mathbf{X}$, as in the onion peeling code defined above, under the assumption that $\mathbb{P}[\hat{\mathbf{X}} \neq \mathbf{X}] > 0$. Positive error probability is common in Slepian-Wolf codes that rely on fixed-rate coding strategies. I here characterize the potential uncertainty that arises from an almost lossless reconstruction of $\mathbf{X}$ and the effects of that uncertainty on the reconstruction of $\mathbf{Y}$.

It should be noted that an intuitive and straightforward way to use $\hat{\mathbf{X}}$ as a helper is to treat it as if $\hat{\mathbf{X}}$ equals $\mathbf{X}$. Another strategy is to use the USSC described in Chapter 3. An alternative approach is to use the compression algorithm for $\mathbf{X}$ to derive $P_{\mathbf{X}\hat{\mathbf{X}}}$, allowing for an alternative design of the decoder for $\mathbf{Y}$.

4.2 Range of Behaviors

This section demonstrates the importance of carefully designing the first stage of an onion peeling code. I use two proxies to study the impact of the first-stage code design on the potential performance of the second-stage code. The conditional entropy rate

$$H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) \triangleq \lim_{n\to\infty} \frac{1}{n} H\left(Y^n|\hat{X}^n\right) \tag{4.2}$$

characterizes the limit of the expected rate of an optimal variable-rate second-stage code. The conditional spectral sup-entropy rate [31]

$$\mathcal{H}\left(\mathcal{Y}|\hat{\mathcal{X}}\right) \triangleq \text{p-}\limsup_{n\to\infty} \frac{1}{n}\imath\left(Y^n|\hat{X}^n\right) = \inf\left\{\eta \middle| \lim_{n\to\infty} \Pr\left[\frac{1}{n}\imath\left(Y^n|\hat{X}^n\right) > \eta\right] = 0\right\} \tag{4.3}$$

characterizes the limit of the rate of an optimal fixed-rate second-stage code.

When studying the information that an almost lossless first-stage code provides about the second source, I only consider first-stage codes that are efficient as point-to-point codes for $\mathbf{X}$. The best rate that a point-to-point code can achieve for i.i.d. source $X^n \sim P_X^n$ under the reliability constraint $\mathbb{P}[\hat{X}^n \neq X^n] \leqslant \epsilon$ is characterized by [28] and [9] as

$$R^* = H(X) + \sqrt{\frac{V(X)}{n}}Q^{-1}(\epsilon) - \frac{\log n}{2n} \pm O\left(\frac{1}{n}\right); \tag{4.4}$$

the achievability and converse results given there agree up to the third order ($O(\log n/n)$) term. I therefore impose the rate constraint in (4.4) on the first stage of the onion peeling codes studied in this chapter unless specified otherwise.

In this chapter, I focus on the following family of p.m.f.s.

Definition 12. *Set $\mathcal{P}_{XY}$ is the set of all p.m.f.s on finite alphabet $\mathcal{X} \times \mathcal{Y}$ that satisfy*

$$V(X) > 0, \quad V(X|Y) > 0, \quad V(Y|X) > 0, \quad V(X,Y) > 0. \tag{4.5}$$

The following two theorems show that there is a first-order difference between the optimal second-stage performance under the best and worst first-stage codes that meet this constraint. Theorem 7 discusses the conditional entropy rate $H(\mathcal{Y}|\hat{\mathcal{X}})$, while Theorem 8 discusses the conditional spectral sup-entropy rate $\mathcal{H}(\mathcal{Y}|\hat{\mathcal{X}})$.

Theorem 7. *Consider a pair of stationary, memoryless sources $(X_1, Y_1), (X_2, Y_2), \ldots$ drawn i.i.d. according to a p.m.f. $P_{XY} \in \mathcal{P}_{XY}$. For each blocklength n, let*

$$\mathcal{D}_\epsilon^{(n)}(C) \triangleq \left\{\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right) : P_\mathsf{e}^{(n)} \leqslant \epsilon, \quad R \leqslant H(X) + \sqrt{\frac{V(X)}{n}}Q^{-1}(\epsilon) - \frac{\log n}{2n} + \frac{C}{n}\right\} \tag{4.6}$$

be the set of all point-to-point codes that meet certain error and rate constraints.

Let $\hat{X}^n \triangleq \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right)$ be the reconstruction of X^n by point-to-point code $\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right)$. There exists a sequence of codes $\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right) \in \mathcal{D}_\epsilon^{(n)}(C), n = 1, 2, \ldots$ for some $C > 0$

such that the conditional entropy rate

$$H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = H$$

if and only if

$$H \in \left[H(Y|X), \epsilon H(Y) + (1-\epsilon)H(Y|X)\right]. \tag{4.7}$$

Proof. See Section 4.6. □

In this thesis, I will not focus on determining the constant C. Therefore, in the rest of this chapter, I abbreviate $\mathcal{D}_\epsilon^{(n)}(C)$ to $\mathcal{D}_\epsilon^{(n)}$.

Theorem 8. *Consider a pair of stationary, memoryless sources* $(X_1, Y_1), (X_2, Y_2), \ldots$ *drawn i.i.d. according to a p.m.f.* $P_{XY} \in \mathcal{P}_{XY}$. *There exists a sequence of codes* $\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right) \in \mathcal{D}_\epsilon^{(n)}, n = 1, 2, \ldots$ *such that the conditional spectral sup-entropy rate*

$$\mathcal{H}\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = H$$

if and only if

$$H \in \left[H(Y|X), H(Y)\right]. \tag{4.8}$$

Proof. See Section 4.6. □

The two theorems show that, asymptotically, point-to-point codes with identical error probability and rate that is equal up to the third-order term are not equally powerful from the perspective of providing information about a dependent source. Therefore, point-to-point codes are not equally helpful when employed as first-stage codes in an onion peeling code. The designer of the point-to-point code may position themselves anywhere along a spectrum ranging from benevolent to adversarial towards downstream users looking to use the decoder output to assist their own coding task.

I provide some brief intuition on the two theorems above aside from the proofs in Section 4.6. If all realizations that are reconstructed incorrectly by the point-to-point almost lossless source code are mapped to a single codeword (or very few codewords), then the error event loses the opportunity to provide helpful information about Y^n, giving the right (worse) side of both the interval for conditional entropy rate and the interval for conditional spectral sup-entropy rate. However, if the realizations that are reconstructed incorrectly by the point-to-point almost lossless source code

are mapped to codewords in an even way, some useful information about Y^n can be retrieved even if an error occurs. I use two different time-sharing techniques in the proofs to show that both the conditional entropy rate and the conditional spectral sup-entropy rate can vary smoothly within their respective intervals.

The proof of Theorem 7 shows that, if a sequence of codes $\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right) \in \mathcal{D}_\epsilon^{(n)}$ satisfies $H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = H(Y|X)$, then, there exists a sequence of codes $\left(\mathsf{f}_*^{(n)}, \mathsf{g}_*^{(n)}\right) \in \mathcal{D}_\epsilon^{(n)}$ such that all realizations $x^n \in \mathcal{X}^n$ reconstructed correctly by $\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right) \in \mathcal{D}_\epsilon^{(n)}$ are mapped to the same codewords and reconstructed correctly by $\left(\mathsf{f}_*^{(n)}, \mathsf{g}_*^{(n)}\right) \in \mathcal{D}_\epsilon^{(n)}$ while all realizations that are reconstructed in error by $\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right)$ are mapped by $\mathsf{f}_*^{(n)}$ to a single codeword. This $\left(\mathsf{f}_*^{(n)}, \mathsf{g}_*^{(n)}\right)$ makes it possible for the decoder to almost always know whether it is reconstructing correctly (unless the decoder receives that single codeword), and yields $H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = (1-\epsilon)H(Y|X) + \epsilon H(Y)$; this is the worst-case from the bound in Theorem 7. This observation suggests that the performance of a sequence of codes in terms of providing help to the reconstruction of a dependent source may be determined not by how the correctly reconstructed realizations are coded but instead by how the incorrectly reconstructed realizations are coded.

4.3 Universally Beneficial Codes

The preceding section establishes the importance of carefully designing the first-stage code in an onion peeling scenario. However, the designer may not be aware of $P_{Y|X}$ when designing the code for $\mathbf{X}$. If possible, we would like to design the code without knowledge of $P_{Y|X}$ and still achieve

$$H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = \mathcal{H}\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = H(Y|X) \tag{4.9}$$

regardless of the $P_{Y|X}$ later revealed. The following theorem states that this is possible.

Theorem 9 (Universally Beneficial Code)**.** *Consider a stationary, memoryless source drawn i.i.d. according to a p.m.f. P_X on finite alphabet $\mathcal{X}$ satisfying $V(X) > 0$. Let $\mathcal{Y}$ be a finite alphabet. There exists a sequence of codes $(\mathsf{f}_n, \mathsf{g}_n) \in \mathcal{D}_\epsilon^{(n)}, n = 1, 2, \ldots$ such that* (4.9) *holds for all $P_{Y|X} : \mathcal{X} \times \mathcal{Y} \mapsto [0, 1]$ that satisfies $V(Y|X) > 0$.*

Proof. See Section 4.6. □

4.4 Linear Codes

Theorem 7, Theorem 8 and Theorem 9 all rely on the random binning argument directly or indirectly. The argument is helpful to demonstrate the existence of a sequence of codes for which reconstruction $\{\hat{X}^n\}_{n=1}^{\infty}$ of $\{X^n\}_{n=1}^{\infty}$ provides as much of information about $\{Y^n\}_{n=1}^{\infty}$ as $\{X^n\}_{n=1}^{\infty}$ itself. Unfortunately, these theorems do not give insight into whether structured codes can achieve this performance. Structured codes are of particular interest because they are easier to implement from both the software and the hardware perspectives, but enforcing restrictions on code structure without loosening reliability constraints can result in rate penalties. In this section, I pursue a linear, almost lossless onion-peeling first-stage code that gives good performance for both $H(\mathcal{Y}|\hat{\mathcal{X}})$ and $\mathcal{H}(\mathcal{Y}|\hat{\mathcal{X}})$. Since most data is stored in a binary format, the following definition assumes code alphabet $GF(2)$. A non-binary source can be represented in binary before compression using a binary linear code.

Definition 13 (Point-to-point Linear Block Source Code)**.** *A point-to-point linear block source code is a point-to-point block source code described by a matrix* Φ *such that* $\mathsf{f}_n : \{0,1\}^n \mapsto \{0,1\}^m$ *satisfies* $\mathsf{f}_n(x^n) = \Phi x^n$ *for all* $x^n \in \{0,1\}^n$*. The rate of the code is* $R = m/n$*.*

Define

$$\mathcal{D}_{*,\epsilon}^{(n)}(C) \triangleq \left\{ \left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right) : P_{\mathsf{e}}^{(n)} \leqslant \epsilon, \quad R \leqslant H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon) + \frac{C \log n}{n} \right\}. \tag{4.10}$$

For sufficiently large n, this set is larger than $\mathcal{D}_{\epsilon}^{(n)}(C')$ for any C' as it relaxes the requirement on the third order term in the rate bound. This difference reflects the rate penalty paid for requiring a more computationally efficient code.

Theorem 10 (Linear Code)**.** *Consider a pair of stationary, memoryless sources* $(X_1, Y_1), (X_2, Y_2), \ldots$ *drawn i.i.d. according to a p.m.f.* $P_{XY} \in \mathcal{P}_{XY}$*. There exists a sequence of linear codes* $(\mathsf{f}_n, \mathsf{g}_n) \in \mathcal{D}_{*,\epsilon}^{(n)}(C)$*,* $n = 1, 2, \ldots$ *for some* $C > 0$ *such that*

$$H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = \mathcal{H}\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = H(Y|X). \tag{4.11}$$

Proof. See Section 4.6. □

Similar to $\mathcal{D}_{\epsilon}^{(n)}$, in the rest of this chapter, I will abbreviate $\mathcal{D}_{*,\epsilon}^{(n)}(C)$ to $\mathcal{D}_{*,\epsilon}^{(n)}$.

The proof of Theorem 10 uses an LDPC code ensemble constructed by uniform random permutation of Tanner graph edges. While the resulting argument suffers a small rate penalty, the strategy generates linear, low density[1] first-stage codes with reconstructions satisfying $\frac{1}{n}H(Y^n|\hat{X}^n) \approx H(Y|X)$.

4.5 Summary

In almost lossless onion peeling data compression, if the first-stage compression reconstructs the first source with a non-vanishing error probability, then the resulting imperfect reconstruction may be of limited use in reconstructing the second source. The exact limitation varies with the type of compressor employed for the first source; this remains true even comparing first-stage codes with almost identical rates and error probabilities. The gap between the conditional entropy rate/conditional spectral sup-entropy rate of the second source given the first-stage source reconstruction varies significantly between the best and worst case first-stage codes. Since achieving truly lossless coding in a distributed code is expensive from a rate perspective [51], this result suggests the benefit of keeping an eye on the future stages of an onion peeling code when designing the earlier stages. If the designer would like the reconstruction to provide more information to later users, they do not have to know the joint distribution of the sources in advance. Furthermore, one can design an almost lossless linear code whose reconstruction is helpful to later users by accepting a small rate penalty.

4.6 Proofs of Theorems

Proof of Theorem 7

Before providing the proof of Theorem 7, I present some auxiliary results useful for the proof. The rate of convergence result (Lemma 10), stated below, follows primarily from [9, Theorem 25], which characterizes the rate region of a deterministic RASC. The RASC allows both individual and joint compression, and is briefly introduced in Chapter 2 of this thesis. In this work, [9, Theorem 25] helps characterize the rate region of an onion peeling code in Lemma 10, which, in turn, helps show that there exists $\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right) \in \mathcal{D}_{\epsilon}^{(n)}$ such that $\mathcal{H}\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = H(Y|X)$.

Definition 14 (($n, \epsilon_1, \epsilon_2$)-onion peeling rate region)**.** *Rate pair* $\mathbf{R} = (R_1, R_2)$ *is* $(n, \epsilon_1, \epsilon_2)$*-achievable by an onion peeling code if there exists an* (n, M_1, M_2) *onion*

[1] The penalty can be reduced and potentially restrained to higher orders if Φ is dense. However, A dense Φ does not provide the desirable savings in hardware and software complexity that a sparse Φ offers. Therefore, I leave the third-order penalty in the definition of $\mathcal{D}_{*,\epsilon}^{(n)}$.

peeling code with $\frac{1}{n}\log M_1 \leqslant R_1$, $\frac{1}{n}\log M_2 \leqslant R_2$, $P_{\mathsf{e},1}^{(n)} \leqslant \epsilon_1$, *and* $P_{\mathsf{e},2}^{(n)} \leqslant \epsilon_2$. *The* $(n, \epsilon_1, \epsilon_2)$*-rate region* $\mathscr{R}^*(n, \epsilon_1, \epsilon_2)$ *is the closure of the set of* $(n, \epsilon_1, \epsilon_2)$*-achievable rate pairs.*

Lemma 10 employs the inner and outer bounds on the point-to-point rate region at blocklength n and error probability ϵ_1 for a finite-alphabet, stationary, memoryless source with positive varentropy, here denoted by $\mathscr{R}_{\text{pp,in}}(n, \epsilon_1)$ and $\mathscr{R}_{\text{pp,out}}(n, \epsilon_1)$, where

$$\mathscr{R}_{\text{pp,in}}(n, \epsilon) \triangleq \left\{ R : R \geqslant H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} + O\left(\frac{1}{n}\right) \right\} \tag{4.12}$$

$$\mathscr{R}_{\text{pp,out}}(n, \epsilon) \triangleq \left\{ R : R \geqslant H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} - O\left(\frac{1}{n}\right) \right\}. \tag{4.13}$$

Lemma 10 further employs the inner bound on the Slepian-Wolf rate region and outer bound on the rate region for describing $\{Y\}$ at blocklength n and error probability ϵ_2 to a decoder with side information $\{X\}$, denoted by $\mathscr{R}_{\text{SW,in}}(n, \epsilon_2)$ and $\mathscr{R}_{\text{si,out}}$, respectively, where

$$\mathscr{R}_{\text{SW,in}}(n, \epsilon) \triangleq \left\{ \mathbf{R} : \begin{bmatrix} R_1 \\ R_2 \\ R_1 + R_2 \end{bmatrix} \in \begin{bmatrix} H(X|Y) \\ H(Y|X) \\ H(X,Y) \end{bmatrix} + \frac{\mathcal{Q}_{\text{inv}}(\mathsf{V}, \epsilon)}{\sqrt{n}} - \frac{\log n}{2n}\mathbf{1} + O\left(\frac{1}{n}\right)\mathbf{1} \right\} \tag{4.14}$$

$$\mathscr{R}_{\text{si,out}}(n, \epsilon) \triangleq \left\{ R : R \geqslant H(Y|X) + \sqrt{\frac{V(Y|X)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} - O\left(\frac{1}{n}\right) \right\}. \tag{4.15}$$

Here $\mathcal{Q}_{\text{inv}}$ is the set-valued multivariate inverse Q-function and V is the covariance matrix of

$$[\imath(X|Y) \;\; \imath(Y|X) \;\; \imath(X,Y)]'.$$

The third-order characterization in (4.12)-(4.13) is derived for the optimal code in [28]. The same inner bounding region can be derived using a random binning argument [9]. Throughout, $\{(X_i, Y_i)\}$ describes a pair of finite-alphabet, stationary, memoryless sources with $P_{XY} \in \mathcal{P}_{XY}$. Region $\mathscr{R}_{\text{SW,in}}(n, \epsilon_2)$ is characterized up to the second-order term in [26, Theorem 1] and the third-order term in [9, Theorem 20], giving (4.14). The derivation of $\mathscr{R}_{\text{si,out}}(n, \epsilon_2)$ in [15, Theorem 8] yields the characterization in (4.15).

Lemma 10. *Consider stationary, memoryless sources* $(X_1, Y_1), (X_2, Y_2), \ldots$ *drawn i.i.d. according to the p.m.f.* $P_{XY} \in \mathcal{P}_{XY}$*. Then*

$$\mathscr{R}_{\text{in}}(n, \epsilon_1, \epsilon_2) \subseteq \mathscr{R}^*(n, \epsilon_1, \epsilon_2) \subseteq \mathscr{R}_{\text{out}}(n, \epsilon_1, \epsilon_2),$$

where

$$\mathscr{R}_{\text{in}}(n, \epsilon_1, \epsilon_2) \triangleq \mathscr{R}_{\text{SW,in}}(n, \epsilon_2) \cap \{\mathbf{R} : R_1 \in \mathscr{R}_{\text{pp,in}}(n, \epsilon_1)\} \tag{4.16}$$

$$\mathscr{R}_{\text{out}}(n, \epsilon_1, \epsilon_2) \triangleq \{\mathbf{R} : R_1 \in \mathscr{R}_{\text{pp,out}}(n, \epsilon_1), R_2 \in \mathscr{R}_{\text{si,out}}(n, \epsilon_2)\}. \tag{4.17}$$

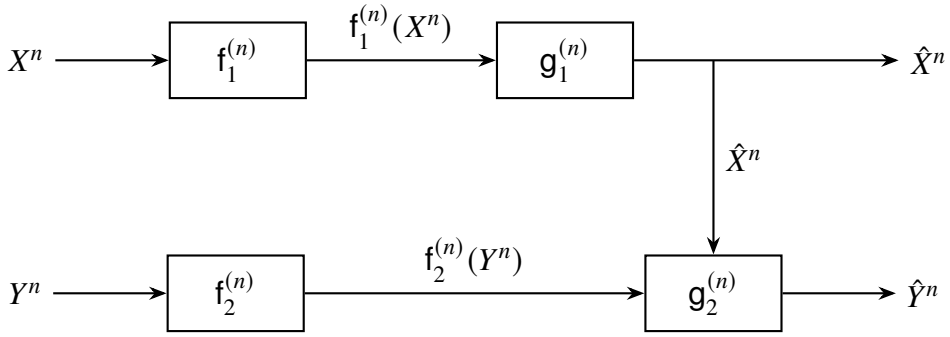


Figure 4.1: Block diagram of the onion peeling problem.

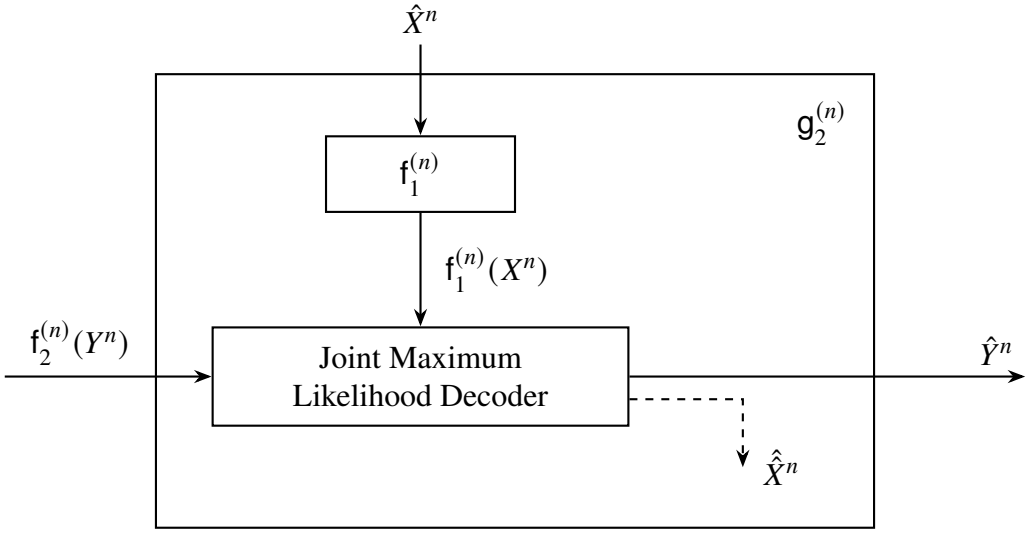


Figure 4.2: Illustration of the decoder used in proofs of Lemma 10, Lemma 16 and Lemma 18.

Proof. For the achievability part, I use a random binning code with a maximum likelihood decoder. With this code, knowing $\hat{X}^n = \mathsf{g}_1^{(n)}(\mathsf{f}_1(X^n))$ implies exact

knowledge of $\mathsf{f}_1(X^n)$. Therefore, the best decoder for Y^n is no worse than the optimal joint Slepian-Wolf decoder for (X^n, Y^n). (Figure 4.2 shows the decoder that maps $\hat{X}^n$ back to $\mathsf{f}_1(X^n)$ and employs the joint decoder to reconstruct Y^n.) Since the best decoder for Y^n is no worse than the joint Slepian-Wolf decoder for (X^n, Y^n), any rate pair in $\mathscr{R}_{\text{in}}(n, \epsilon_1, \epsilon_2)$ would suffice to ensure $P_{\mathsf{e},1}^{(n)} \leqslant \epsilon_1$ and $P_{\mathsf{e},2}^{(n)} \leqslant \epsilon_2$, which means that the error of reconstructing the first source is bounded by ϵ_1, and the error of reconstructing the second source is bounded by ϵ_2. The achievability result of [9, Theorem 24], which shows that there exists a deterministic code from the random binning code generation that simultaneously achieves third-order optimal rate for different number of encoders, is used to demonstrate the existence of a deterministic code from the random binning generation that simultaneously satisfies $P_{\mathsf{e},1}^{(n)} \leqslant \epsilon_1$ and $P_{\mathsf{e},2}^{(n)} \leqslant \epsilon_2$, since $P_{\mathsf{e},1}^{(n)}$ is the error probability when only X^n is encoded, and $P_{\mathsf{e},2}^{(n)}$ is bounded by the joint error probability when both sources are encoded. The converse holds since $Y^n \to X^n \to \hat{X}^n$ forms a Markov chain, which implies that X^n captures all information about Y^n that $\hat{X}^n$ may contain, and therefore the achievable rate region for describing Y^n to a decoder that knows $\hat{X}^n$ must be in the achievable rate region for describing Y^n to a decoder that knows X^n. □

The inner and outer bounds from Lemma 10 are depicted in Figure 4.3(a) and 4.3(b), respectively. Dashed and solid lines show the first- and second-order characterizations. The inner and outer bounds differ in their second-order terms, where the inner bound is determined by the joint Slepian-Wolf rate region and the outer bound by the rate region of side information code. As n increases, the second-order term approaches 0. Figure 4.3(c) provides an enlarged view of the difference between the inner and outer bounds, indicated with hash lines. The size of the gap on the left boundary can be analyzed using the local dispersion result of [26, Theorem 2] and solving for the colored triangle (the second-order term of $\mathscr{R}_{\text{pp,in}}(n, \epsilon_1)$ and $\mathscr{R}_{\text{pp,out}}(n, \epsilon_1)$ gives the length of the horizontal side, [26, Theorem 2] can be used to solve for the vertical side, subtracting the second-order term of the side information code rate region $\mathscr{R}_{\text{si,out}}(n.\epsilon_2)$ gives the gap). Whether either of the second-order terms in (4.16) and (4.17) is tight remains an open problem for future work.

Remark 7. *Lemma 10 uses a random binning encoder and a maximum likelihood decoder to prove the existence of a blocklength-n onion peeling code for* $\{(X, Y)\}$ *with error probabilities* ϵ_1 *and* ϵ_2 *and rates close to* $H(X)$ *and* $H(Y|X)$ *in the first and second stages. The result holds even if* $\epsilon_2 < \epsilon_1$*. Although the proof of Lemma 10*

employs a joint decoder, it does not rule out the possibility that there exists a code that is less computationally expensive.

The following lemma is the spectral converse of the source coding theorem with decoder side information. While Lemma 11 is stated in a slightly different way than either the spectral converse of the point-to-point source code [31, Theorem 1.3.1] or the spectral converse of the Slepian-Wolf code [31, Theorem 7.4.1], it is in some sense a special case of the Slepian-Wolf result and can be proven with an argument almost identical to the ones used in [31, Theorem 1.3.1] and [31, Theorem 7.4.1]. I therefore omit the proof.

Lemma 11 (Modified from converse of the Source Coding Theorem with Decoder Side Information)**.** *Let* $(X^1, Y^1), (X^2, Y^2), \ldots$ *be a sequence of pair of discrete source blocks on finite alphabets* $\mathcal{X} \times \mathcal{Y}$*. If, for all* $\delta > 0$*, there exists a sequence of codes with decoder side information*

$$\mathsf{f}^{(n)} : \mathcal{Y}^n \mapsto \left[2^{nR}\right] \tag{4.18}$$

$$\mathsf{g}^{(n)} : \left[2^{nR}\right] \times \mathcal{X}^n \mapsto \mathcal{Y}^n \tag{4.19}$$

such that $R \leqslant L + \delta$ *and*

$$\lim_{n\longrightarrow\infty} \mathbb{P}\left[\mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(Y^n), X^n\right) \neq Y^n\right] = 0,$$

then

$$H(\mathcal{Y}|\mathcal{X}) \leqslant \mathcal{H}(\mathcal{Y}|\mathcal{X}) \leqslant L. \tag{4.20}$$

Evaluating Lemma 10 with $\epsilon_2 \to 0$ and then applying Lemma 11 shows that there exists a sequence of codes $\left\{\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right)\right\}_{n=1}^{\infty}$, $\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right) \in \mathcal{D}_\epsilon^{(n)}$ such that $\mathcal{H}\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = H(Y|X)$.

The following lemma states that, in the long run, knowing that an event happens does not reduce the uncertainty in Y^n as long as the event's probability does not vanish as $n \to \infty$. In fact, the property holds as long as the probability does not decay exponentially, but the current form is sufficient to serve the purpose of the discussion in this thesis.

Lemma 12. *Let* $\{Y_i\}_{i=1}^{\infty}$ *be a stationary, memoryless sources on a finite alphabet* $\mathcal{Y}$*. For a sequence of events* $\mathcal{E}^{(1)}, \mathcal{E}^{(2)}, \ldots$ *that satisfies* $\lim_{n\to\infty} \mathbb{P}\left[\mathcal{E}^{(n)}\right] > 0$*,*

$$\lim_{n\longrightarrow\infty} \frac{1}{n} H\left(Y^n|\mathcal{E}^{(n)}\right) = H(Y). \tag{4.21}$$

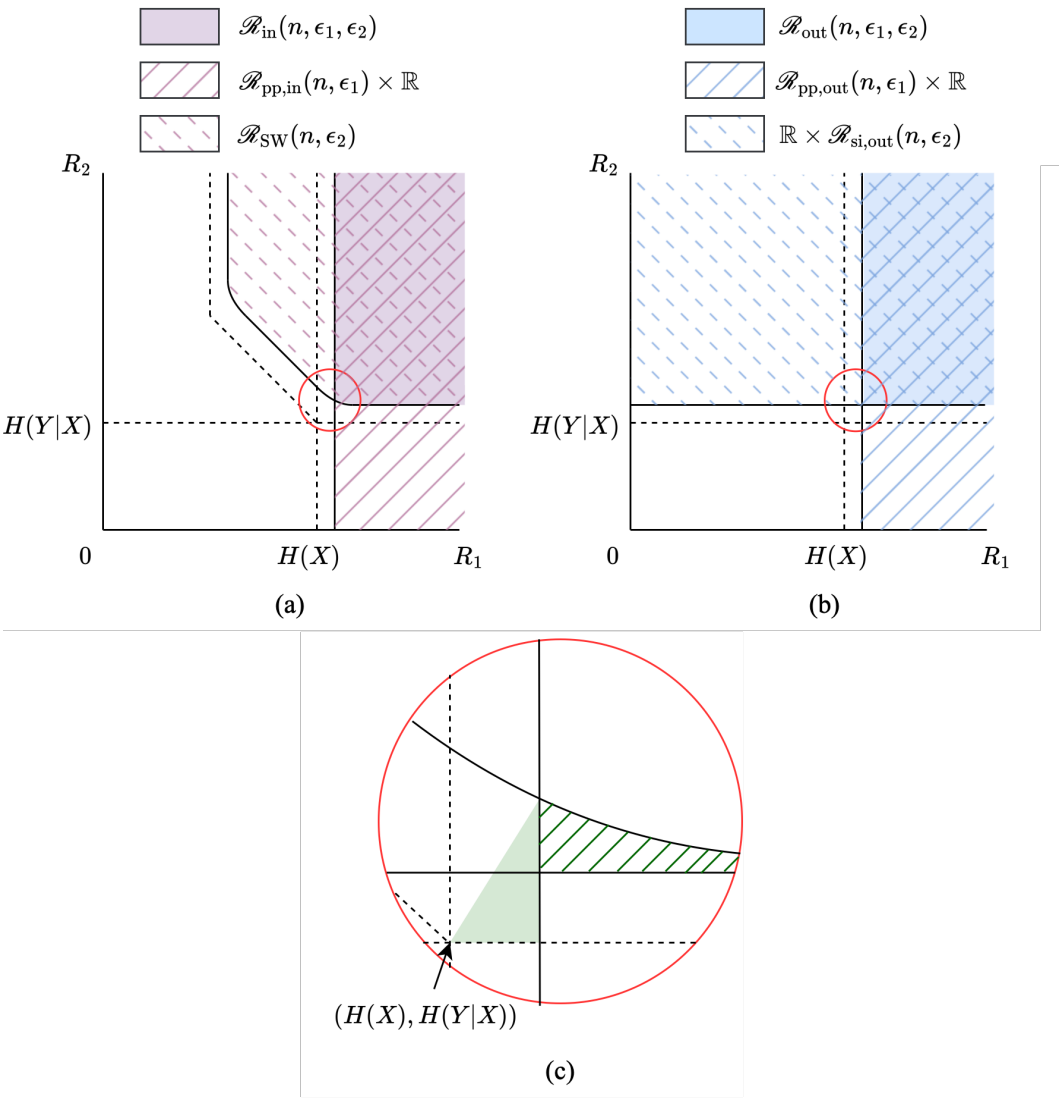


Figure 4.3: (a) Inner bound of onion peeling rate region, $\mathscr{R}_{\text{in}}(n,\epsilon_1,\epsilon_2)$. (b) Outer bound of onion peeling rate region, $\mathscr{R}_{\text{out}}(n,\epsilon_1,\epsilon_2)$. (c) Enlarged view of the second-order gap between the inner and outer bound (hashed) and the triangle to be solved in order to compute the gap size (colored) using local dispersion methods in [26].

Proof. See Section 4.7. □

The following lemma characterizes the behavior of a function of X^n.

Lemma 13. *Let $A_1, A_2, \ldots$ be a stationary, memoryless random process on finite alphabet $\mathcal{A}$. Let $\mathcal{S}^{(n)} \subseteq \mathcal{A}^n, n = 1, 2, \ldots$ be a sequence of sets such that $\lim_{n\to\infty} \mathbb{P}\left[A^n \in \mathcal{S}^{(n)}\right] = p$ where $p \in (0,1)$ is a constant. Let $f^{(n)}(\cdot), n = 1, 2, \ldots$*

be a sequence of functions on $\mathcal{A}^n$. If

$$\lim_{n\longrightarrow\infty}\frac{1}{n}H\left(f^{(n)}(A^n)\right)=H(A),$$

then

$$\lim_{n\longrightarrow\infty}\frac{1}{n}H\left(f^{(n)}(A^n)\middle|A^n\in\mathcal{S}^{(n)}\right)=H(A). \tag{4.22}$$

Proof. See Section 4.7. □

I then proceed to prove Theorem 7.

Proof of Theorem 7. I wish to prove that, if and only if $H\in[H(Y|X),(1-\epsilon)H(Y|X)+\epsilon H(Y)]$, there exists a sequence of point-to-point codes $(\mathsf{f}^{(n)},\mathsf{g}^{(n)})\in\mathcal{D}_\epsilon^{(n)}$, $n=1,2,\ldots$ for X^n such that the conditional entropy rate of Y^n given the reconstruction $\hat{X}^n=\mathsf{g}^{(n)}(\mathsf{f}^{(n)}(X^n))$ satisfies $H(\mathcal{Y}|\hat{\mathcal{X}})=H$.

I first prove that, if $H\notin[H(Y|X),(1-\epsilon)H(Y|X)+\epsilon H(Y)]$, then there is no sequence of codes $(\mathsf{f}^{(n)},\mathsf{g}^{(n)})\in\mathcal{D}_\epsilon^{(n)}$ for X^n such that $H\left(\mathcal{Y}|\hat{\mathcal{X}}\right)=H$.

For any $\left(\mathsf{f}^{(n)},\mathsf{g}^{(n)}\right)\in\mathcal{D}_\epsilon^{(n)}$, $Y^n\to X^n\to\hat{X}^n$ forms a Markov chain, which implies $H\left(Y^n|\hat{X}^n\right)\geqslant H\left(Y^n|X^n\right)$. Therefore, by the definition of $H(\mathcal{Y}|\hat{\mathcal{X}})$, $H(\mathcal{Y}|\hat{\mathcal{X}})\geqslant H(Y|X)$.

Define random variable $E=\mathbf{1}\left\{\hat{X}^n\neq X^n\right\}$. Since E is binary,

$$H\left(Y^n|\hat{X}^n,E\right)\leqslant H\left(Y^n|\hat{X}^n\right)\leqslant H(Y^n,E|\hat{X}^n)\leqslant H\left(Y^n|\hat{X}^n,E\right)+1.$$

Furthermore, if we define $\mathcal{S}_*^{(n)}\triangleq\left\{x^n\in\mathcal{X}^n:\mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}\left(x^n\right)\right)\neq x^n\right\}$,

$$\begin{aligned}
H\left(Y^n|\hat{X}^n,E\right)&=\sum_{\hat{x}^n\in\hat{\mathcal{X}}^n}\sum_{e\in\{0,1\}}P_{\hat{X}^nE}(\hat{x}^n,e)H\left(Y^n|\hat{X}^n=\hat{x}^n,E=e\right)\\
&=\sum_{\hat{x}^n\in\hat{\mathcal{X}}^n}P_{\hat{X}^nE}(\hat{x}^n,0)H(Y^n|\hat{X}^n=\hat{x}^n,E=0)\\
&\qquad+\sum_{\hat{x}^n\in\hat{\mathcal{X}}^n}P_{\hat{X}^nE}(\hat{x}^n,1)H(Y^n|\hat{X}^n=\hat{x}^n,E=1)\\
&=\sum_{x^n\notin\mathcal{S}_*^{(n)}}P_X^n(x^n)H(Y^n|X^n=x^n)+P_{\mathsf{e},1}^{(n)}H(Y^n|\hat{X}^n,E=1)\\
&\leqslant\sum_{x^n\notin\mathcal{S}_*^{(n)}}P_X^n(x^n)H(Y^n|X^n=x^n)+P_{\mathsf{e},1}^{(n)}H(Y^n|E=1)\\
&=\sum_{x^n\notin\mathcal{S}_*^{(n)}}P_X^n(x^n)H(Y^n|X^n=x^n)+P_{\mathsf{e},1}^{(n)}H\left(Y^n|X^n\in\mathcal{S}_*^{(n)}\right).
\end{aligned}$$

By the converse result of finite-blocklength point-to-point code [28], for all first-stage codes in $\mathcal{D}_\epsilon^{(n)}$, $P_{\mathsf{e},1}^{(n)} \to \epsilon$ as $n \to \infty$. Therefore, Lemma 12 shows that

$$\lim_{n\longrightarrow\infty} \frac{1}{n} H\left(Y^n | X^n \in \mathcal{S}_*^{(n)}\right) = H(Y).$$

Note that

$$\begin{aligned}
&\left| \frac{1}{n} \sum_{x^n \notin \mathcal{S}_*^{(n)}} P_X^n(x^n) H(Y^n|X^n = x^n) - (1 - \mathbb{P}[\mathcal{S}_*^{(n)}]) H(Y|X) \right| \\
&= \left| \mathbb{E}_{\bar{X}^n \sim P_X^n} \left[\left(\frac{1}{n} H(Y^n|X^n = \bar{X}^n) - H(Y|X) \right) \mathbf{1}\left\{ X^n \notin \mathcal{S}_*^{(n)} \right\} \right] \right| \\
&\leqslant \mathbb{E}_{\bar{X}^n \sim P_X^n} \left[\left| \frac{1}{n} H(Y^n|X^n = \bar{X}^n) - H(Y|X) \right| \left| \mathbf{1}\left\{ X^n \notin \mathcal{S}_*^{(n)} \right\} \right| \right] \\
&\leqslant \mathbb{E}_{\bar{X}^n \sim P_X^n} \left[\left| \frac{1}{n} H(Y^n|X^n = \bar{X}^n) - H(Y|X) \right| \right] \\
&\to 0,
\end{aligned}$$

implying

$$\lim_{n\longrightarrow\infty} \frac{1}{n} \sum_{x^n \notin \mathcal{S}_*^{(n)}} P_X^n(x^n) H(Y^n|X^n = x^n) = (1 - \epsilon) H(Y|X).$$

Therefore, for any sequence of point-to-point codes $\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right) \in \mathcal{D}_\epsilon^{(n)}$,

$$H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = \lim_{n\longrightarrow\infty} \frac{1}{n} H\left(Y^n|\hat{X}^n, E\right) \leqslant (1 - \epsilon) H(Y|X) + \epsilon H(Y).$$

To prove that a sequence of codes exists if $H \in [H(Y|X), (1-\epsilon)H(Y|X) + \epsilon H(Y)]$, note that invoking Lemma 10 with $\epsilon_2 \to 0$, combined with Lemma 11, shows that there exists a sequence of point-to-point codes $\left(\mathsf{f}_{\mathrm{L}}^{(n)}, \mathsf{g}_{\mathrm{L}}^{(n)}\right) \in \mathcal{D}_\epsilon^{(n)}, n = 1, 2, \ldots$ such that $H(\mathcal{Y}|\hat{\mathcal{X}}) = H(Y|X)$.

Let $\mathcal{S}_{\mathrm{L}}^{(n)} = \left\{ x^n \in \mathcal{X}^n : \mathsf{g}_{\mathrm{L}}^{(n)}\left(\mathsf{f}_{\mathrm{L}}^{(n)}(x^n)\right) = x^n \right\}$ denote the set of realizations $x^n \in \mathcal{X}^n$ that are correctly decoded by $\left(\mathsf{f}_{\mathrm{L}}^{(n)}, \mathsf{g}_{\mathrm{L}}^{(n)}\right)$. Let $\left(\mathsf{f}_{\mathrm{H}}^{(n)}, \mathsf{g}_{\mathrm{H}}^{(n)}\right) \in \mathcal{D}_\epsilon^{(n)}, n = 1, 2, \ldots$ be a sequence of code that is constructed from $\left(\mathsf{f}_{\mathrm{L}}^{(n)}, \mathsf{g}_{\mathrm{L}}^{(n)}\right), n = 1, 2, \ldots$ using the following method.

$$\mathsf{f}_{\mathrm{H}}^{(n)}(x^n) = \begin{cases} \mathsf{f}_{\mathrm{L}}^{(n)}(x^n), & x^n \in \mathcal{S}_{\mathrm{L}}^{(n)} \\ c^*, & x^n \notin \mathcal{S}_{\mathrm{L}}^{(n)} \end{cases} \tag{4.23}$$

where c^* is the (lexicographically) last codeword in the codeword alphabet, and $\mathsf{g}_{\mathrm{H}}^{(n)}(c) = \mathsf{g}_{\mathrm{L}}^{(n)}(c)$. The codes $\left(\mathsf{f}_{\mathrm{L}}^{(n)}, \mathsf{g}_{\mathrm{L}}^{(n)}\right)$ and $\left(\mathsf{f}_{\mathrm{H}}^{(n)}, \mathsf{g}_{\mathrm{H}}^{(n)}\right)$ differ in the encoding of realizations that $\left(\mathsf{f}_{\mathrm{L}}^{(n)}, \mathsf{g}_{\mathrm{L}}^{(n)}\right)$ fails to correctly reconstruct, but are identical in encoding the realizations that $\left(\mathsf{f}_{\mathrm{L}}^{(n)}, \mathsf{g}_{\mathrm{L}}^{(n)}\right)$ successfully reconstructs.

Next, for each n, define a set $\mathcal{V}^{(n)} \subseteq \mathcal{X}^n \backslash \mathcal{S}_{\mathrm{L}}^{(n)}$ by sequentially placing realizations $x^n \notin \mathcal{S}_{\mathrm{L}}^{(n)}$ into $\mathcal{V}^{(n)}$ until the first time that $\mathbb{P}[\mathcal{V}^{(n)}] \geqslant \min(\epsilon p, 1 - \mathbb{P}[\mathcal{S}_{\mathrm{L}}^{(n)}])$ for $p \in (0, 1]$.

Finally, define $\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right)$, the code design for which I want to show that $H(\mathcal{Y}|\hat{\mathcal{X}}) = H$, as follows. For a realization $x^n \in \mathcal{X}^n$, if $x^n \in \mathcal{V}^{(n)}$, then $\mathsf{f}^{(n)}(x^n) = \mathsf{f}_{\mathrm{H}}^{(n)}(x^n)$. If $x^n \notin \mathcal{V}^{(n)}$, then $\mathsf{f}^{(n)}(x^n) = \mathsf{f}_{\mathrm{L}}^{(n)}(x^n)$. The decoder $\mathsf{g}^{(n)}(c) = \mathsf{g}_{\mathrm{L}}^{(n)}(c)$.

The given strategy yields three code sequences, $\left\{\left(\mathsf{f}_{\mathrm{L}}^{(n)}, \mathsf{g}_{\mathrm{L}}^{(n)}\right)\right\}_{n=1}^{\infty}$, $\left\{\left(\mathsf{f}_{\mathrm{H}}^{(n)}, \mathsf{g}_{\mathrm{H}}^{(n)}\right)\right\}_{n=1}^{\infty}$, and $\left\{\left(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}\right)\right\}_{n=1}^{\infty}$, that all fall into $\mathcal{D}_{\epsilon}^{(n)}$ and all share the same error set $\mathcal{S}_{\mathrm{L}}^{(n)}$. By the converse of the finite-blocklength source coding theorem [28][9],

$$\lim_{n \to \infty} \mathbb{P}\left[X^n \in \mathcal{S}_{\mathrm{L}}^{(n)}\right] = \epsilon.$$

The construction of $\mathcal{V}^{(n)}$ ensures that $\mathbb{P}[X^n \in \mathcal{V}^{(n)}] \to p\epsilon$.

Since

$$H\left(Y^n \left| \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right)\right.\right) \geqslant H\left(Y^n \left| \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right), \mathbf{1}\left\{X^n \in \mathcal{V}^{(n)}\right\}\right.\right)$$

and

$$\begin{aligned} H\left(Y^n \left| \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right)\right.\right) &\leqslant H\left(Y^n \left| \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right), \mathbf{1}\left\{X^n \in \mathcal{V}^{(n)}\right\}\right.\right) \\ &\quad + H\left(\mathbf{1}\left\{X^n \in \mathcal{V}^{(n)}\right\} \left| \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right)\right.\right) \\ &\leqslant H\left(Y^n \left| \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right), \mathbf{1}\left\{X^n \in \mathcal{V}^{(n)}\right\}\right.\right) + 1, \end{aligned}$$

it follows that

$$H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = \lim_{n \to \infty} \frac{1}{n} H\left(Y^n \left| \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right), \mathbf{1}\left\{X^n \in \mathcal{V}^{(n)}\right\}\right.\right). \tag{4.24}$$

To analyze $H\left(Y^n \middle| \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right), \mathbf{1}\left\{X^n \in \mathcal{V}^{(n)}\right\}\right)$, note that

$$\begin{aligned}
&H\left(Y^n \middle| \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right), \mathbf{1}\left\{X^n \in \mathcal{V}^{(n)}\right\}\right) \\
&\quad = H\left(Y^n \middle| \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right), X^n \in \mathcal{V}^{(n)}\right) \mathbb{P}\left[X^n \in \mathcal{V}^{(n)}\right] \\
&\qquad + H\left(Y^n \middle| \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}(X^n)\right), X^n \notin \mathcal{V}^{(n)}\right) \mathbb{P}\left[X^n \notin \mathcal{V}^{(n)}\right] \\
&\quad = H\left(Y^n \middle| \mathsf{g}_{\mathrm{H}}^{(n)}\left(\mathsf{f}_{\mathrm{H}}^{(n)}(X^n)\right), X^n \in \mathcal{V}^{(n)}\right) \mathbb{P}\left[X^n \in \mathcal{V}^{(n)}\right] \\
&\qquad + H\left(Y^n \middle| \mathsf{g}_{\mathrm{L}}^{(n)}\left(\mathsf{f}_{\mathrm{L}}^{(n)}(X^n)\right), X^n \notin \mathcal{V}^{(n)}\right) \mathbb{P}\left[X^n \notin \mathcal{V}^{(n)}\right]. \qquad (4.25)
\end{aligned}$$

Due to the construction of $\mathcal{V}^{(n)}$ and $\left(\mathsf{f}_{\mathrm{H}}^{(n)}, \mathsf{g}_{\mathrm{H}}^{(n)}\right)$, $X^n \in \mathcal{V}^{(n)}$ ensures that $\mathsf{g}_{\mathrm{H}}^{(n)}\left(\mathsf{f}_{\mathrm{H}}^{(n)}(X^n)\right)$ is a constant. Therefore,

$$\lim_{n\to\infty} \frac{1}{n} H\left(Y^n \middle| \mathsf{g}_{\mathrm{H}}^{(n)}\left(\mathsf{f}_{\mathrm{H}}^{(n)}(X^n)\right), X^n \in \mathcal{V}^{(n)}\right) = \lim_{n\to\infty} \frac{1}{n} H\left(Y^n \middle| X^n \in \mathcal{V}^{(n)}\right) = H(Y), \qquad (4.26)$$

where the second equality follows from Lemma 12.

Notice that, when deriving and invoking Lemma 10, I used a strategy in which the second stage decoder reconstructs both X^n and Y^n, even though only Y^n is used as the decoder output. Therefore, the error probability bound also applies to the pair. An interesting consequence is that, not only is it true that the reconstruction $\hat{X}^n$ reduces the conditional uncertainty of Y^n by the same amount as X^n asymptotically, giving

$$\lim_{n\to\infty} \frac{1}{n} H\left(Y^n \middle| \mathsf{g}_{\mathrm{L}}^{(n)}\left(\mathsf{f}_{\mathrm{L}}^{(n)}(X^n)\right)\right) = H(Y|X), \qquad (4.27)$$

but also it is true that $\hat{X}^n$ reduces the conditional uncertainty in (X^n, Y^n) by the same amount as X^n asymptotically, giving

$$\lim_{n\to\infty} \frac{1}{n} H\left(X^n, Y^n \middle| \mathsf{g}_{\mathrm{L}}^{(n)}\left(\mathsf{f}_{\mathrm{L}}^{(n)}(X^n)\right)\right) = H(Y|X). \qquad (4.28)$$

Together, these observations imply

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} H\left(\mathsf{g}_{\mathrm{L}}^{(n)}\left(\mathsf{f}_{\mathrm{L}}^{(n)}(X^n)\right)\right) &= H(X) - \lim_{n\to\infty} \frac{1}{n} H\left(X^n \middle| \mathsf{g}_{\mathrm{L}}^{(n)}\left(\mathsf{f}_{\mathrm{L}}^{(n)}(X^n)\right)\right) && (4.29)\\
&= H(X) - 0 && (4.30)\\
&= H(X) && (4.31)
\end{aligned}$$

and

$$\lim_{n\to\infty} \frac{1}{n} H\left(\mathsf{g}_{\mathrm{L}}^{(n)}\left(\mathsf{f}_{\mathrm{L}}^{(n)}(X^n)\right), Y^n\right) = H(X,Y). \qquad (4.32)$$

Recall that $\mathbb{P}\left[X^n \notin \mathcal{V}^{(n)}\right] \to 1 - p\epsilon \in (0,1)$. Therefore, applying Lemma 13 with $A^n = X^n$, $f^{(n)}(x^n) = \mathsf{g}_{\mathrm{L}}^{(n)}\left(\mathsf{f}_{\mathrm{L}}^{(n)}(x^n)\right)$, we can conclude that

$$\lim_{n\to\infty} \frac{1}{n} H\left(\mathsf{g}_{\mathrm{L}}^{(n)}\left(\mathsf{f}_{\mathrm{L}}^{(n)}(X^n)\right) \middle| X^n \notin \mathcal{V}^{(n)}\right) = H(X); \tag{4.33}$$

similarly, setting $A^n = (X^n, Y^n)$, $f^{(n)}(x^n, y^n) = \left(\mathsf{g}_{\mathrm{L}}^{(n)}\left(\mathsf{f}_{\mathrm{L}}^{(n)}(x^n)\right), y^n\right)$, we can conclude that

$$\lim_{n\to\infty} \frac{1}{n} H\left(\mathsf{g}_{\mathrm{L}}^{(n)}\left(\mathsf{f}_{\mathrm{L}}^{(n)}(X^n)\right), Y^n \middle| X^n \notin \mathcal{V}^{(n)}\right) = H(X,Y). \tag{4.34}$$

Therefore,

$$\lim_{n\to\infty} \frac{1}{n} H\left(Y^n \middle| \mathsf{g}_{\mathrm{L}}^{(n)}\left(\mathsf{f}_{\mathrm{L}}^{(n)}(X^n)\right), X^n \notin \mathcal{V}^{(n)}\right) = H(Y|X). \tag{4.35}$$

Combining (4.24), (4.25), (4.26), and (4.35),

$$H\left(\mathcal{Y} \middle| \hat{\mathcal{X}}\right) = p\epsilon H(Y) + (1 - p\epsilon) H(Y|X),$$

which can be anywhere in the interval $(H(Y|X), (1-\epsilon)H(Y|X) + \epsilon H(Y)]$.

□

Remark 8. *When one sets the asymptotic probability of $\mathcal{V}^{(n)}$ as $p = 1$, the encoder of $\mathsf{f}^{(n)} = \mathsf{f}_{\mathrm{H}}^{(n)}$, the code that maps all incorrectly decoded realizations to a single codeword. That is, all realizations that are decoded incorrectly, i.e., $x^n \in \mathcal{X}^n \backslash \mathcal{S}_{\mathrm{L}}^{(n)}$ map to the same codeword c^*. In this case, $H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = (1-\epsilon)H(Y|X) + \epsilon H(Y)$. This observation suggests that $H(\mathcal{Y}|\hat{\mathcal{X}})$ is affected by what the code does to realizations that would be decoded incorrectly, and furthermore, any code in $\mathcal{D}_\epsilon^{(n)}$ that maps all realizations that are decoded incorrectly to a single codeword yields $H(\mathcal{Y}|\hat{\mathcal{X}}) = (1-\epsilon)H(Y|X) + \epsilon H(Y)$.*

Proof of Theorem 8

The following lemma is a widget that I will use in the proof of Theorem 8.

Lemma 14. *Let $A_1, A_2, \ldots$ be a random process on finite alphabet $\mathcal{A}$. Let $\mathcal{S}^{(n)} \subseteq \mathcal{A}^n, n = 1, 2, \ldots$ be a sequence of sets such that $\lim_{n\to\infty} \Pr\left[A^n \in \mathcal{S}^{(n)}\right] = p, p \in (0,1)$. Let $\mathsf{f}^{(n)}, n = 1, 2, \ldots$ be a sequence of functions on $\mathcal{A}^n$. If*

$$\text{p-}\limsup_{n\to\infty} \frac{1}{n} \iota\left(f^{(n)}(A^n)\right) = \text{p-}\liminf_{n\to\infty} \frac{1}{n} \iota\left(f^{(n)}(A^n)\right) = H \tag{4.36}$$

where H is a positive constant, then

$$\text{p-}\limsup_{n\to\infty} \frac{1}{n} \iota\left(f^{(n)}(A^n) \middle| A^n \in \mathcal{S}^{(n)}\right) \geqslant H. \tag{4.37}$$

Proof. See Section 4.7. □

Proof of Theorem 8. To prove Theorem 8, first notice that it is impossible for $\mathcal{H}(\mathcal{Y}|\hat{\mathcal{X}})$ to be strictly larger than $H(Y)$, because i.i.d. sources satisfy the strong converse property. Also, it is impossible for $\mathcal{H}(\mathcal{Y}|\hat{\mathcal{X}})$ to be strictly smaller than $H(Y|X)$ since $\hat{X}^n - X^n - Y^n$ forms a Markov chain.

Next, I set aside $H = H(Y)$ first, and show that $\mathcal{H}(\mathcal{Y}|\hat{\mathcal{X}}) = H$ is possible for $H \in [H(Y|X), H(Y))$.

I start by showing that there exists a sequence of codes $\{(\mathsf{f}^{(n)}, \mathsf{g}^{(n)})\}_{n=1}^{\infty}$ that satisfy the following properties for any $\delta, \varepsilon > 0$ for large enough blocklength n. For convenience, I assume codewords are converted to their binary expansions. I denote $R' \triangleq H(X,Y) - H$, and $\mathsf{f}_{R'}^{(n)}(x^n) = \lfloor \mathsf{f}^{(n)}(x^n) \rfloor_{nR'}$.

- The code itself satisfies $(\mathsf{f}^{(n)}, \mathsf{g}^{(n)}) \in \mathcal{D}_{\epsilon}^{(n)}$.
- There exists an encoder $\mathsf{f}_Y^{(n)} : \mathcal{Y}^n \mapsto [2^{n(H+\delta)}]$ and a decoder $\mathsf{g}_Y^{(n)} : [2^{nR'}] \times [2^{n(H+\delta)}] \mapsto \mathcal{Y}^n$ such that

$$\mathbb{P}\left[\mathsf{g}_Y^{(n)}\left(\mathsf{f}_{R'}^{(n)}(X^n), \mathsf{f}_Y^{(n)}(Y^n)\right) \neq Y^n\right] \leqslant \varepsilon. \tag{4.38}$$

- For any pair of encoder $\mathsf{f}_Y^{(n)} : \mathcal{Y}^n \mapsto [2^{n(H-\delta)}]$ and decoder $\mathsf{g}_Y^{(n)} : [2^{nR'}] \times [2^{n(H-\delta)}] \mapsto \mathcal{Y}^n$,

$$\mathbb{P}\left[\mathsf{g}_Y^{(n)}\left(\mathsf{f}_{R'}^{(n)}(X^n), \mathsf{f}_Y^{(n)}(Y^n)\right) \neq Y^n\right] \geqslant 1 - \varepsilon. \tag{4.39}$$

That is, the code itself is efficient and reliable in terms of rate and error probability. If we truncate it at rate R', the truncated codeword can be combined with an encoded description of Y^n at any rate above H to produce a reconstruction of Y^n with error probability arbitrarily close to 0; but when the truncated codeword is combined with any encoded description of Y^n at rate below H, the reconstruction error is arbitrarily close to 1. This can be done by consulting the random binning code generation. There exists a $C = O(1)$ such that, if the encoder $\mathsf{f}^{(n)}$ is generated using the classic random binning technique ($\mathsf{F}^{(n)}$ denotes the random variable) with rate R_n being the largest rate satisfying $2^{nR_n} \in \mathbb{Z}_+$ and

$$R_n \leqslant H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} + \frac{C}{n},$$

then the probability of event

$$\left\{\mathbb{P}_{X^n}\left[\mathsf{G}^{(n)}\left(\mathsf{F}^{(n)}(X^n)\right) \neq X^n\right] \leqslant \epsilon\right\} \tag{4.40}$$

is bounded from below by a positive constant that can be made close to 1 (how close is determined by the choice of C). This is a result of [9, Lemma 25]. The proof of [9, Theorem 24] directly implies that the claim above is true; a similar argument also appears in the proof of Lemma 17.

Using random binning to generate $\mathsf{F}_Y^{(n)}$ as well, the second requirement can also be met with probability close to 1. Note that the rate choices make the pair $(R', H+\delta)$ fall in the asymptotic achievable rate region of the Slepian-Wolf code. Multiple results on the behavior of random binning code in that region exists, for example, the proof of [52, Theorem 15.4.1] shows that the average

$$\mathbb{P}_{X^nY^n}\left[\mathsf{G}_J^{(n)}\left(\mathsf{F}_{R'}^{(n)}(X^n), \mathsf{F}_Y^{(n)}(Y^n)\right) \neq (X^n, Y^n)\right] \tag{4.41}$$

can be made arbitrarily close to 0, where $\mathsf{G}_J^{(n)}$ is a joint decoder that reconstructs both sources. Let $\mathsf{G}_Y^{(n)}$ take only the reconstruction of Y^n from the output of $\mathsf{G}_J^{(n)}$, then, the average

$$\mathbb{P}_{X^nY^n}\left[\mathsf{G}_Y^{(n)}\left(\mathsf{F}_{R'}^{(n)}(X^n), \mathsf{F}_Y^{(n)}(Y^n)\right) \neq Y^n\right] \tag{4.42}$$

can also be made arbitrarily close to 0. By Markov's inequality,

$$\begin{aligned}\Pr\left[\mathbb{P}_{X^nY^n}\left[\mathsf{G}_Y^{(n)}\left(\mathsf{F}_{R'}^{(n)}(X^n), \mathsf{F}_Y^{(n)}(Y^n)\right) \leqslant Y^n\right] \geqslant a\right] \\ \geqslant 1 - \frac{\mathbb{E}\left[\mathbb{P}_{X^nY^n}\left[\mathsf{G}_Y^{(n)}\left(\mathsf{F}_{R'}^{(n)}(X^n), \mathsf{F}_Y^{(n)}(Y^n)\right) \neq Y^n\right]\right]}{a},\end{aligned}$$

therefore, the random binning code satisfies the requirement in (4.38) with a probability that one can also bound from below by a constant that can be made close to 1.

On the other hand, $(R', H-\delta)$ falls outside the asymptotic Slepian-Wolf rate region. Due to the strong converse property of $(X^n, Y^n) \sim P_{XY}^n$, no matter how $\mathsf{f}_Y^{(n)}$ is designed, for any decoder that reconstructs (X^n, Y^n) from $(\mathsf{f}_{R'}^{(n)}(X^n), \mathsf{f}_Y^{(n)}(Y^n))$, the probability of error will be arbitrarily close to 1. This includes the decoder that tries to reconstruct Y^n with optimal success probability. But, in order to show that any $\mathsf{g}_Y^{(n)}$ satisfies (4.39), the error probability bound for Y^n only, I still need to show that the case where Y^n is reconstructed correctly but X^n is not does not contribute much

to the joint error probability. I call the reconstruction of X^n from such a code $\hat{X}^n_{R'}$ to distinguish from X^n, the reconstruction by $(\mathsf{f}^{(n)}, \mathsf{g}^{(n)})$, and refer to the reconstruction of Y^n as $\hat{Y}^n$. That is,

$$\mathbb{P}\left[Y^n \neq \hat{Y}^n\right] = \mathbb{P}\left[(X^n, Y^n) \neq \left(\hat{X}^n_{R'}, \hat{Y}^n\right)\right] - \mathbb{P}\left[X^n \neq \hat{X}^n_{R'}, Y^n = \hat{Y}^n\right]. \tag{4.43}$$

I need to show that the random binning design of $(\mathsf{f}^{(n)}, \mathsf{g}^{(n)})$ makes the second term on the right-hand side arbitrarily small in order to show that the left-hand side is arbitrarily close to 1 like the first term on the right-hand side. Consider the maximum likelihood decoder with side information, breaking ties lexicographically,

$$\mathsf{g}^{(n)}_{1,\mathrm{SI}}(c_1, y^n) = \arg \min_{x^n : \mathsf{f}^{(n)}_{R'}(x^n) = c_1} \iota\left(x^n | y^n\right). \tag{4.44}$$

If $\hat{X}^n_{R'} = \mathsf{G}^{(n)}_{1,\mathrm{SI}}(\mathsf{F}^{(n)}_{R'}(X^n), \hat{Y}^n)$ (randomness of $\mathsf{G}^{(n)}_{1,\mathrm{SI}}$ comes from random binning generation of the encoder),

$$\begin{aligned}
\mathbb{P}\left[X^n \neq \hat{X}^n_{R'}, Y^n = \hat{Y}^n\right] &\leqslant \mathbb{P}\Big[\exists \hat{x}^n \prec X^n, \mathsf{F}^{(n)}_{R'}(X^n) = \mathsf{F}^{(n)}_{R'}(\bar{x}^n), \\
&\qquad\qquad \iota\left(\hat{x}^n | \hat{Y}^n\right) \leqslant \iota\left(X^n | \hat{Y}^n\right), \hat{Y}^n = Y^n\Big] \\
&= \mathbb{P}\Big[\exists \hat{x}^n \prec X^n, \mathsf{F}^{(n)}_{R'}(X^n) = \mathsf{F}^{(n)}_{R'}(\bar{x}^n), \\
&\qquad\qquad \iota\left(\hat{x}^n | Y^n\right) \leqslant \iota\left(X^n | Y^n\right), \hat{Y}^n = Y^n\Big] \\
&\leqslant \mathbb{P}\Big[\exists \hat{x}^n \prec X^n, \mathsf{F}^{(n)}_{R'}(X^n) = \mathsf{f}^{(n)}_{R'}(\bar{x}^n), \\
&\qquad\qquad \iota\left(\hat{x}^n | Y^n\right) \leqslant \iota\left(X^n | Y^n\right)\Big]
\end{aligned}$$

which approaches 0 exponentially quickly as n grows, identical to bounding the error probability of random binning code for X^n with perfect side information Y^n at a rate that is asymptotically above $H(X|Y)$. Therefore, if $\mathbb{P}\left[Y^n \neq \hat{Y}^n\right]$ is not arbitrarily close to 1 for all possible designs of decoders, $\mathbb{P}\left[(X^n, Y^n) \neq (\hat{X}^n_{R'}, \hat{Y}^n)\right]$ cannot be either. Then, applying Markov's inequality on

$$1 - \mathbb{P}_{X^n Y^n}\left[\mathsf{G}^{(n)}_Y\left(\mathsf{F}^{(n)}_{R'}(X^n), \mathsf{F}^{(n)}_Y(Y^n)\right) \neq Y^n\right], \tag{4.45}$$

the random binning code satisfies the requirement in (4.39) with a probability that one can also bound from below by a constant that can be made close to 1.

Therefore, random binning makes the probability of satisfying all three requirements bounded away from zero. So there exists a deterministic sequence of codes $(\mathsf{f}^{(n)}, \mathsf{g}^{(n)})$ that satisfy all three requirements.

I proceed to construct a sequence of codes $(\mathsf{f}_*^{(n)}, \mathsf{g}^{(n)})$ (the decoder is kept the same) from this code that satisfies $\mathcal{H}(\mathcal{Y}|\hat{\mathcal{X}}) = H$.

Let $\mathcal{S}_*^{(n)} \triangleq \left\{x^n \in \mathcal{X}^n : \mathsf{g}^{(n)}\left(\mathsf{f}^{(n)}\left(x^n\right)\right) \neq x^n\right\}$ be the set of realizations that $(\mathsf{f}^{(n)}, \mathsf{g}^{(n)})$, the code I just constructed, decodes incorrectly. For all $x^n \notin \mathcal{S}_*^{(n)}$, reuse the same codeword $\mathsf{f}_*^{(n)}(x^n) = \mathsf{f}^{(n)}(x^n)$. For all $x^n \in \mathcal{S}_*^{(n)}$, truncate $\mathsf{f}^{(n)}(x^n)$ into $\mathsf{f}_{R'}^{(n)}(x^n)$ and pad with zeros (or any fixed symbol in the codeword symbol alphabet) to reach the same length as other codewords. Note that this $(\mathsf{f}_*^{(n)}, \mathsf{g}^{(n)})$ has the exact same rate and error probability as $(\mathsf{f}^{(n)}, \mathsf{g}^{(n)})$, so it is also in $\mathcal{D}_\epsilon^{(n)}$.

Furthermore, it also satisfies the second requirement because it does not differ from $(\mathsf{f}^{(n)}, \mathsf{g}^{(n)})$ in the first nR' bits of any codeword. Therefore, as n grows, one can also reconstruct Y^n with arbitrarily low positive error probability from $\mathsf{f}_*^{(n)}(X^n)$ and $\mathsf{f}_Y^{(n)}$. Applying Lemma 11 gives $\mathcal{H}(\mathcal{Y}|\hat{\mathcal{X}}) \leqslant H$.

To show that $\mathcal{H}(\mathcal{Y}|\hat{\mathcal{X}}) \geqslant H$, notice that, due to the second and third requirements that $(\mathsf{f}^{(n)}, \mathsf{g}^{(n)})$ satisfies, Y^n can be reconstructed with side information $\mathsf{f}_{R'}^{(n)}(X^n)$ with vanishing probability if compressed at a rate above H, and the reconstruction error approaches 1 if the rate is below H. This implies

$$\begin{aligned}
\text{p-}\limsup_{n\to\infty} \frac{1}{n}\imath\left(Y^n \middle| \mathsf{f}_{R'}^{(n)}(X^n)\right) &= \text{p-}\liminf_{n\to\infty} \frac{1}{n}\imath\left(Y^n \middle| \mathsf{f}_{R'}^{(n)}(X^n)\right) && (4.46)\\
&= \lim_{n\to\infty} \frac{1}{n} H\left(Y^n \middle| \mathsf{f}_{R'}^{(n)}(X^n)\right) && (4.47)\\
&= H && (4.48)
\end{aligned}$$

by the strong converse theorem (side-information version of [31, Theorem 2.3.2], proof omitted due to exact similarity).

Similar to obtaining (4.31) and (4.32),

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} H\left(\mathsf{f}_{R'}^{(n)}(X^n)\right) &= H(X,Y) - \lim_{n\to\infty} \frac{1}{n} H\left(X^n, Y^n \middle| \mathsf{f}_{R'}^{(n)}(X^n)\right) && (4.49)\\
&= H(X,Y) - H && (4.50)\\
&= R'. && (4.51)
\end{aligned}$$

Since the alphabet size of $\mathsf{f}_{R'}^{(n)}(X^n)$ is $2^{nR'}$, the spectral sup-entropy rate cannot exceed R', which means that

$$\begin{aligned}
\text{p-}\limsup_{n\to\infty} \frac{1}{n}\imath\left(\mathsf{f}_{R'}^{(n)}(X^n)\right) &= \text{p-}\liminf_{n\to\infty} \frac{1}{n}\imath\left(\mathsf{f}_{R'}^{(n)}(X^n)\right) && (4.52)\\
&= \lim_{n\to\infty} \frac{1}{n} H\left(\mathsf{f}_{R'}^{(n)}(X^n)\right) && (4.53)\\
&= R', && (4.54)
\end{aligned}$$

further implying

$$\text{p-}\limsup_{n\longrightarrow\infty}\frac{1}{n}\imath\left(Y^n,\mathsf{f}_{R'}^{(n)}(X^n)\right)=\text{p-}\liminf_{n\longrightarrow\infty}\frac{1}{n}\imath\left(Y^n,\mathsf{f}_{R'}^{(n)}(X^n)\right) \tag{4.55}$$

$$=\lim_{n\longrightarrow\infty}\frac{1}{n}H\left(Y^n,\mathsf{f}_{R'}^{(n)}(X^n)\right) \tag{4.56}$$

$$=H(X,Y). \tag{4.57}$$

Applying Lemma 14 to (4.57),

$$\text{p-}\limsup_{n\longrightarrow\infty}\frac{1}{n}\imath\left(Y^n,\mathsf{f}_{R'}^{(n)}(X^n)\middle|\,X^n\in\mathcal{S}_*^{(n)}\right)\geqslant H(X,Y). \tag{4.58}$$

Applying Lemma 14 to (4.54) and reusing the alphabet size constraint of $2^{nR'}$,

$$\text{p-}\limsup_{n\longrightarrow\infty}\frac{1}{n}\imath\left(\mathsf{f}_{R'}^{(n)}(X^n)\middle|\,X^n\in\mathcal{S}_*^{(n)}\right)=R'. \tag{4.59}$$

Therefore,

$$\text{p-}\limsup_{n\longrightarrow\infty}\frac{1}{n}\imath\left(Y^n\left|\mathsf{f}_{R'}^{(n)}(X^n)\right.,X^n\in\mathcal{S}_*^{(n)}\right)\geqslant H. \tag{4.60}$$

Since $\mathbb{P}\left[X^n\in\mathcal{S}_*^{(n)}\right]\rightarrow\epsilon>0$, due to the construction of $\mathsf{f}^{(n)}$,

$$\imath\left(Y^n\left|\mathsf{f}^{(n)}(X^n)\right.\right)=\imath\left(Y^n\left|\mathsf{f}_{R'}^{(n)}(X^n)\right.,X^n\in\mathcal{S}_*^{(n)}\right) \tag{4.61}$$

with a non-vanishing probability. Therefore,

$$\mathcal{H}\left(\mathcal{Y}|\hat{\mathcal{X}}\right)=\text{p-}\limsup_{n\longrightarrow\infty}\frac{1}{n}\imath\left(Y^n\left|\mathsf{f}^{(n)}(X^n)\right.\right)\geqslant H. \tag{4.62}$$

The final task is to show that there exists a sequence of code such that $\mathcal{H}(\mathcal{Y}|\hat{\mathcal{X}})=H(Y)$. I use the same design that gives $H(\mathcal{Y}|\hat{\mathcal{X}})=\epsilon H(Y)+(1-\epsilon)H(Y|X)$ in the proof of Theorem 7: all realizations in $\mathcal{S}_*^{(n)}$ map to a single codeword. Then, with non-vanishing probability,

$$\imath(Y^n|\hat{X}^n)=\imath(Y^n|\mathsf{f}^{(n)}(X^n)) \tag{4.63}$$

$$=\imath\left(Y^n|\mathcal{E}_n\right), \tag{4.64}$$

where $\mathcal{E}_n$ is the event that X^n maps to that single codeword. As in the proof of Theorem 7, the construction of the code ensures that $\mathbb{P}[\mathcal{E}_n] \to \epsilon$. Furthermore,

$$\begin{aligned}
\imath(Y^n|\mathcal{E}_n) &= \log \frac{1}{P_{Y^n|\mathcal{E}_n}(Y^n)} && (4.65)\\
&= \log \frac{1}{P_{Y^n|\mathbf{1}\{\mathcal{E}_n\}}(Y^n|1)} && (4.66)\\
&= \log \frac{P_{\mathbf{1}\{\mathcal{E}_n\}}(1)}{P_{\mathbf{1}\{\mathcal{E}_n\}|Y^n}(1|Y^n)P_Y^n(Y^n)} && (4.67)\\
&= \imath(Y^n) + \log \mathbb{P}[\mathcal{E}_n] + \log \frac{1}{P_{\mathbf{1}\{\mathcal{E}_n\}|Y^n}(1|Y^n)} && (4.68)\\
&\geqslant \imath(Y^n) + \log \mathbb{P}[\mathcal{E}_n]. && (4.69)
\end{aligned}$$

Since the second term converges to a constant, and $\frac{1}{n}\imath(Y^n)$ converges to $H(Y)$ with probability 1,

$$\mathcal{H}\left(\mathcal{Y}|\hat{\mathcal{X}}\right) \triangleq \text{p-}\limsup_{n\to\infty} \frac{1}{n}\imath(Y^n|\hat{X}^n) \geqslant \text{p-}\liminf_{n\to\infty} \frac{1}{n}\imath(Y^n|\mathcal{E}_n) \geqslant H(Y). \tag{4.70}$$

□

Proof of Theorem 9

The proof of Theorem 9 [2] relies on some definitions and auxiliary results stated below.

The quantization strategy in the following is seen in proofs for universal source coding, see, for example, [53].

Definition 15 (Quantized Distribution Set)**.** *Given a finite alphabet $\mathcal{A}$ and $n \in \mathbb{Z}_+$, the quantized distribution set $\mathcal{S}_Q^{(n)}(\mathcal{A})$ consists of all p.m.f.s $Q : \mathcal{A} \mapsto [0,1]$ such that $Q(a)$ is an integer multiple of $\frac{1}{n}$ for all $a \in \mathcal{A}$.*

Under this definition, the number of distinct quantized distributions satisfies

$$\left|\mathcal{S}_Q^{(n)}(\mathcal{A})\right| \leqslant (n+1)^{|\mathcal{A}|}.$$

Definition 16 (Distribution Quantization)**.** *Given a p.m.f. $P : \mathcal{A} \mapsto [0,1]$ where $\mathcal{A}$ is a finite alphabet, the n-quantization of P for $n \in \mathbb{Z}_+$, $Q^{(n)}(P) = (Q^{(n)}(P,x) : x \in$*

[2] There are easier ways to prove Theorem 9. The proof of Theorem 9 in this thesis is deliberately constructed such that the code for Y^n does not need to consult X^n in any other way than observing $\hat{X}^n$.

$\mathcal{A}$), *and*

$$Q^{(n)}(P,x) = \begin{cases} \frac{1}{n}\lceil nP(x)\rceil, & x \in \mathcal{A}\backslash\{x^*\} \\ 1 - \sum_{x\in\mathcal{A}\backslash\{x^*\}} Q^{(n)}(P,x), & x = x^* \end{cases} \tag{4.71}$$

where $x^* \triangleq \arg\max_{x\in\mathcal{A}} P(x)$ *with ties broken in ascending order on the ordered alphabet* $\mathcal{A}$. *Note that* $n \geqslant |\mathcal{A}|(|\mathcal{A}|-1)$ *implies*

$$Q^{(n)}(P) \in \mathcal{S}_Q^{(n)}(\mathcal{A}), \tag{4.72}$$

where $n \geqslant |\mathcal{A}|(|\mathcal{A}|-1)$ *is needed to ensure that* $Q^{(n)}(P,x)$ *is a p.m.f. on the same support as* P_{XY}.

Lemma 15 (Modified from [53, Lemma 3])**.** *Let* $P_{XY} : \mathcal{X}\times\mathcal{Y} \mapsto [0,1]$ *be a p.m.f. on finite alphabet* $\mathcal{X}\times\mathcal{Y}$. *For all* $\mathcal{S}^{(n)} \subseteq \mathcal{X}^n \times \mathcal{Y}^n$,

$$\mathbb{P}\left[\mathcal{S}^{(n)}\right] = \mathbb{Q}\left[\mathcal{S}^{(n)}\right] \cdot O(1), \tag{4.73}$$

where $\mathbb{P}$ *captures the randomness of* $(X^n,Y^n) \sim P_{XY}^n$ *and* $\mathbb{Q}$ *captures the randomness of* $(X^n,Y^n) \sim Q_{XY}^n$,

$$Q_{XY}(x,y) \triangleq Q^{(n)}(P_{Y|X=x},y)P_X(x) \quad \forall (x,y) \in \mathcal{X}\times\mathcal{Y}.$$

That is, quantizing the conditional distribution does not rescale the probability of any event by more than $O(1)$.

Proof. See Section 4.7. □

Lemma 16. *Consider stationary, memoryless sources* $X_1, X_2, \ldots$ *drawn i.i.d. according to p.m.f.* P_X *on finite alphabet* $\mathcal{X}$ *that satisfies* $V(X) > 0$. *Let* $\left(\mathsf{F}_1^{(n)}, \mathsf{G}_1^{(n)}\right)$ *be a random binning code ensemble with*

$$R_1 = H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}\left(\epsilon - \frac{C'}{\sqrt{n}}\right) - \frac{\log n}{2n} \tag{4.74}$$

Where $C' = B(X) + C(X)$ *is a positive constant (*$B(X)$ *and* $C(X)$ *follow the definitions in [9, Eqn (35-36)]). For a conditional distribution* $P_{Y|X} : \mathcal{X}\times\mathcal{Y} \mapsto [0,1]$, *let* $(\mathsf{F}_2^{(n)}, \mathsf{G}_2^{(n)})$ *be the random binning code ensemble with side information at rate*

$$R_2 = H(Y|X) + \delta_2, \tag{4.75}$$

with δ_2 *being a positive constant. Then, the random code ensemble satisfies*

$$\mathbb{E}_{\mathsf{F}_1^{(n)},\mathsf{G}_1^{(n)}}\left[\mathbb{P}_{X^n}\left[\mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right) \neq X^n\right]\right] \leqslant \epsilon \tag{4.76}$$

and

$$\mathbb{E}_{\mathsf{F}_1^{(n)},\mathsf{G}_1^{(n)},\mathsf{F}_2^{(n)},\mathsf{G}_2^{(n)}}\left[\mathbb{P}_{X^nY^n}\left[\mathsf{G}_2^{(n)}\left(\mathsf{F}_2^{(n)}\left(Y^n\right),\mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right)\right)\neq Y^n\right]\right]\leqslant 2^{-nE_2} \tag{4.77}$$

where $E_2 > 0$ is a function of δ_2.

Proof. The bound (4.76) can be obtained by the exact same strategy as [9] by applying the RCU bound as $C' = B(X) + C(X)$.

The exponential decay of the error probability at the second stage of the onion peeling relies on the coding strategy in Figure 4.2. Due to the choice of R_1 and R_2, there exists a constant δ_{12} such that

$$R_1 + R_2 - H(X,Y) \geqslant \delta_{12} \tag{4.78}$$

for sufficiently large n.

The event $\left\{\mathsf{G}_2^{(n)}\left(\mathsf{F}_2^{(n)}(Y^n),\mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}(X^n)\right)\right)\neq Y^n\right\}$ can be captured by the union of the following error events.

$$\mathcal{E}_2 \triangleq \{\exists \bar{y}^n \in \mathcal{Y}^n\backslash\{Y^n\} \text{ s.t. } \imath(\bar{y}^n|X^n)\leqslant \imath(Y^n|X^n), \mathsf{F}_2^{(n)}(\bar{y}^n) = \mathsf{F}_2^{(n)}(Y^n)\} \tag{4.79}$$

$$\begin{aligned}\mathcal{E}_{12} \triangleq \{&\exists \bar{x}^n \in \mathcal{X}^n\backslash\{X^n\}, \bar{y}^n \in \mathcal{Y}^n\backslash\{Y^n\}\\ &\text{s.t. } \imath(\bar{x}^n,\bar{y}^n)\leqslant \imath(X^n,Y^n), \mathsf{F}_1^{(n)}(\bar{x}^n) = \mathsf{F}_1^{(n)}(X^n), \mathsf{F}_2^{(n)}(\bar{y}^n) = \mathsf{F}_2^{(n)}(Y^n)\}.\end{aligned} \tag{4.80}$$

The probabilities of both events decay exponentially with n since $R_2 - H(Y|X) = \delta_2 > 0$ and (4.78). Specifically, $\mathbb{P}[\mathcal{E}_2]$ exponentially decays, with the error exponent characterized in [54]. Similar strategy works for $\mathcal{E}_{12}$. Therefore, there exists positive constant E_2 such that (4.77) holds.

□

Lemma 17 (Excess Error Probability)**.** *Consider stationary, memoryless sources $X_1, X_2, \ldots$ drawn i.i.d. according to p.m.f. P_X on finite alphabet $\mathcal{X}$ that satisfies $V(X) > 0$. Then, there exists constants $C'' > 0, \delta \in (0,1)$ such that the following is true. Let $\left(\mathsf{F}_1^{(n)},\mathsf{G}_1^{(n)}\right)$ be a random binning code ensemble with*

$$R_1 = H(X) + \sqrt{\frac{V(X)}{n}}Q^{-1}\left(\epsilon - \frac{C'}{\sqrt{n}} - \frac{C''}{\sqrt{n}}\right) - \frac{\log n}{2n}. \tag{4.81}$$

Let $\mathcal{Y}$ be a finite alphabet. For the k-th joint p.m.f. $P_{XY}^{(k)} : \mathcal{X}\times\mathcal{Y}\mapsto[0,1]$ such that

$$P_{XY}^{(k)} \in \mathcal{S}_{\text{cond}}^{(n)} = \left\{P'_{XY} : P'_X = P_X, P'_{Y|X=x} \in \mathcal{S}_Q^{(n)}(\mathcal{Y})\forall x\in\mathcal{X}\right\}.$$

Let

$$H^{(k)}(Y|X) \triangleq \sum_{\substack{x\in\mathcal{X}\\ y\in\mathcal{Y}}} P_{XY}^{(k)}(x,y)\log\frac{P_X(x)}{P_{XY}^{(k)}(x,y)} \tag{4.82}$$

be the conditional entropy with respect to $P_{XY}^{(k)}$. Let $(\mathsf{F}_2^{(n,k)}, \mathsf{G}_2^{(n,k)})$ be a random binning code ensemble with side information at rate $H^{(k)}(Y|X) + \delta_2^{(k)}$ with $\delta_2^{(k)}$s being positive constants. For sufficiently large n, the random code ensembles satisfy

$$\mathbb{P}\left[\mathcal{P}_1 \cup \left\{\bigcup_{k\in\left[\left|\mathcal{S}_{\text{cond}}^{(n)}\right|\right]} \mathcal{P}_2^{(k)}\right\}\right] \leqslant \delta, \tag{4.83}$$

where the $\mathbb{P}$ captures the randomness of $\left(\mathsf{F}_1^{(\mathsf{n})}, \mathsf{G}_1^{(n)}\right), \left(\mathsf{F}_2^{(n,k)}, \mathsf{G}_2^{(n,k)}\right), k \in \left[\left|\mathcal{S}_{\text{cond}}^{(n)}\right|\right]$,

$$\mathcal{P}_1 \triangleq \left\{\mathbb{P}_{X^n}\left[\mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right) \neq X^n\right] > \epsilon\right\}, \tag{4.84}$$

and

$$\mathcal{P}_2^{(k)} \triangleq \left\{\mathbb{P}_{X^nY^n}\left[\mathsf{G}_2^{(n,k)}\left(\mathsf{F}_2^{(n,k)}\left(Y^n\right), \mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right)\right) \neq Y^n\right] > 2^{-n\gamma^{(k)}}\right\} \tag{4.85}$$

for $k \in \left[\left|\mathcal{S}_{\text{cond}}^{(n)}\right|\right]$.

Proof of Lemma 17. For each $P_{XY}^{(k)} \in \mathcal{S}_Q^{(n)}$, the expected error probability of the second stage of the code can be bounded by

$$\mathbb{E}_{\mathsf{F}_1^{(n)},\mathsf{G}_1^{(n)},\mathsf{F}_2^{(n,k)},\mathsf{G}_2^{(n,k)}}\left[\mathbb{P}_{X^nY^n}\left[\mathsf{G}_2^{(n,k)}\left(\mathsf{F}_2^{(n,k)}\left(Y^n\right), \mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right)\right) \neq Y^n\right]\right] \leqslant 2^{-nE_2^{(k)}}. \tag{4.86}$$

By Markov's inequality, for $\gamma^{(k)} \in \left(0, E_2^{(k)}\right)$,

$$\mathbb{P}\left[\mathcal{P}_2^{(k)}\right] \tag{4.87}$$

$$= \mathbb{P}_{\mathsf{F}_1^{(n)},\mathsf{G}_1^{(n)},\mathsf{F}_2^{(n,k)},\mathsf{G}_2^{(n,k)}}\left[\mathbb{P}_{X^nY^n}\left[\mathsf{G}_2^{(n,k)}\left(\mathsf{F}_2^{(n,k)}\left(Y^n\right), \mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right)\right) \neq Y^n\right] > 2^{-n\gamma^{(k)}}\right] \tag{4.88}$$

$$\leqslant 2^{-n\left(E_2^{(k)}-\gamma^{(k)}\right)}. \tag{4.89}$$

That is, the probability of $\mathcal{P}_2^{(k)}$ decays exponentially with n.

To bound the probability of $\mathcal{P}_1$, we use [9, Lemma 25], a direct result from Markov's inequality. For a point-to-point fixed-length source code, by converse result of the finite-blocklength source coding theorem [28], the smallest possible error probability

of a code with rate specified in (4.81) is close to ϵ. Specifically, let $C^{(n)}$ be all codes $\left(\mathsf{f}_1^{(n)}, \mathsf{g}_1^{(n)}\right)$ such that

$$R_1 = H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}\left(\epsilon - \frac{C'}{\sqrt{n}} - \frac{C''}{\sqrt{n}}\right) - \frac{\log n}{2n}. \tag{4.90}$$

Then, the converse of almost lossless point-to-point source codes [28][9] bounds the minimum decoding error as

$$\min_{\left(\mathsf{f}_1^{(n)}, \mathsf{g}_1^{(n)}\right) \in C^{(n)}} \mathbb{P}\left[\mathsf{g}_1^{(n)}\left(\mathsf{f}_1^{(n)}\left(X^n\right)\right) \neq X^n\right] \geqslant \epsilon - \frac{C''}{\sqrt{n}} - \frac{\Delta}{\sqrt{n}} \tag{4.91}$$

for some constant $\Delta > 0$.

Lemma 16 shows that the expected error probability satisfies

$$\mathbb{E}_{\mathsf{F}_1^{(n)}, \mathsf{G}_1^{(n)}}\left[\mathbb{P}_{X^n}\left[\mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right) \neq X^n\right]\right] \leqslant \epsilon - \frac{C''}{\sqrt{n}}. \tag{4.92}$$

Applying [9, Lemma 25],

$$\mathbb{P}[\mathcal{P}_1] \leqslant \frac{\epsilon - \frac{C''}{\sqrt{n}} - \epsilon + \frac{C''}{\sqrt{n}} + \frac{\Delta}{\sqrt{n}}}{\epsilon - \epsilon + \frac{C''}{\sqrt{n}} + \frac{\Delta}{\sqrt{n}}} \tag{4.93}$$

$$= \frac{\Delta}{C'' + \Delta}. \tag{4.94}$$

Choosing $C'' = 2\Delta/\delta - \Delta$ gives

$$\mathbb{P}[\mathcal{P}_1] \leqslant \frac{\delta}{2}. \tag{4.95}$$

Since the cardinality of $\mathcal{S}_{\text{cond}}^{(n)}$ is bounded from above by $(n+1)^{|\mathcal{X}||\mathcal{Y}|}$, for sufficiently large n,

$$\mathbb{P}\left[\bigcup_{k \in \left[\left|\mathcal{S}_{\text{cond}}^{(n)}\right|\right]} \mathcal{P}_2^{(k)}\right] \leqslant \sum_{k \in \left[\left|\mathcal{S}_{\text{cond}}^{(n)}\right|\right]} \mathbb{P}\left[\mathcal{P}_2^{(k)}\right] \leqslant \frac{\delta}{2}. \tag{4.96}$$

Combining (4.95) and (4.96) and applying the union bound again gives the desired result.

□

Proof of Theorem 9. By the differentiability of $Q^{-1}(\cdot)$, the rate in (4.81) satisfies

$$R_1 = H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}(\epsilon) - \frac{\log n}{2n} + O\left(\frac{1}{n}\right). \tag{4.97}$$

Let $\mathcal{S}_{\mathcal{E}} \in \mathcal{X}^n \times \mathcal{Y}^n$ be the set of realizations (x^n, y^n) such that the reconstruction $\hat{y}^n \neq y^n$ under a given code design. By Lemma 15,

$$\mathbb{P}\left[\mathcal{S}_{\mathcal{E}}\right] = \mathbb{Q}\left[\mathcal{S}_{\mathcal{E}}\right] \cdot O(1), \tag{4.98}$$

where $\mathbb{P}$ and $\mathbb{Q}$ are probabilities with respect to distributions P_{XY} and Q_{XY} (as defined in Lemma 15), respectively. This means that, if a deterministic sequence of code achieves exponentially decaying error probability for source distribution Q_{XY}, then the same error exponent can be achieved by the exact same code for source distribution P_{XY}. That is, I can first design a code for each quantized $P_{XY}^{(k)}$ with rate $R_2 = H^{Q}(Y|X) + \delta_2^{(k)}$. Then, for any $P_{XY} = P_{Y|X}P_X$ that is not necessarily in $\mathcal{S}_{\text{cond}}^{(n)}$, I can use the code designed for its quantized version Q_{XY} which is in $\mathcal{S}_{\text{cond}}^{(n)}$. By Lemma 17 and (4.98), the resulting code satisfies the following two criteria with probability at least $1 - \delta$:

1. The first-stage code is efficient and reliable. That is, $\left(\mathsf{F}_1^{(n)}, \mathsf{G}_1^{(n)}\right) \in \mathcal{D}_{\epsilon}^{(n)}$.
2. The 2nd-stage reconstruction error is bounded by $2^{-n\gamma^{(k)}} \cdot O(1)$.

Therefore, there exists a deterministic code that satisfies these two criteria. The conditional entropy based on the quantized distribution $H^{Q}(Y|X)$ is close to $H(Y|X)$ because of the continuity of entropies. Specifically, by the definition of $Q^{(n)}(\cdot)$,

$$\left\|Q^{(n)}\left(P_{XY}\right) - P_{XY}\right\|_2 \leqslant \left\|Q^{(n)}\left(P_{XY}\right) - P_{XY}\right\|_1 = O\left(\frac{1}{n}\right), \tag{4.99}$$

which implies that $\left|H^{Q}(X,Y) - H(X,Y)\right| = O(1/n)$. Also, $\left|H^{Q}(X) - H(X)\right|$. Therefore, the 2nd-stage code used for any $P_{Y|X}$ satisfies

$$R_2^{(k)} = H(Y|X) + \delta_2^{(k)} + O\left(\frac{1}{n}\right). \tag{4.100}$$

Since $\delta_2^{(k)}$ can be made arbitrarily small, according to Lemma 11,

$$\mathcal{H}\left(\mathcal{Y}|\hat{\mathcal{X}}\right) \leqslant H(Y|X). \tag{4.101}$$

Due to the fact that $Y^n \to X^n \to \hat{X}^n$ forms a Markov chain, (4.101) holds with equality. In conclusion, there exists a sequence of code in $\mathcal{D}_{\epsilon}^{(n)}$ such that $\mathcal{H}\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = H(Y|X)$ for universal $P_{Y|X}$. □

Proof of Theorem 10

LDPC codes [55] are a family of linear block error correcting codes that are widely used in channel coding scenario [56]. Its structure enables the use of a suboptimal decoder that operates more efficiently than the computationally heavy [57] maximum likelihood decoder. The LDPC code is originally designed for channel coding scenario, but it can be transplanted to source coding by using the parity check matrix as the encoder [58]. This helps with proving Theorem 10.

I prove Theorem 10 by showing that random LDPC first-stage code generation yields $H\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = \mathcal{H}\left(\mathcal{Y}|\hat{\mathcal{X}}\right) = H(Y|X)$ with certain probability. The non-asymptotic analysis in auxiliary results Theorem 11 and part of Lemma 18 is a dual of the channel code result from [59]. The following definition is identical to the random LDPC matrix generation method in [59] for code alphabet $GF(2)$.

Definition 17 (LDPC Matrix by Uniform Permutation)**.** *An (n, m, ρ, λ) LDPC matrix by uniform permutation is a random $m \times n$ binary matrix generated by the following process. Let $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_{n\lambda})$ be a uniform random permutation of $(1, 2, \ldots, n\lambda)$. Then, construct Φ as follows. For $j \in [m]$ and $k \in [n]$, the entry of Φ at j-th row, k-th column is decided by*

$$h_{j,k} = |\{(j-1)\rho + 1, \ldots, j\rho\} \cap \{\pi_{(k-1)\lambda+1}, \ldots, \pi_{k\lambda}\}| \ mod\ 2. \tag{4.102}$$

This is equivalent to connecting sockets from variable nodes and check nodes of the associated Tanner graph by a uniform random permutation. Further, define

$$\alpha \triangleq \max_{x \in \mathcal{T}^n \setminus \{\mathbf{0}\}} \frac{2^m \bar{S}^n\left(\mathcal{T}^n(x)\right)}{B\left(n, \mathcal{T}^n(x)\right)} \tag{4.103}$$

where $\mathcal{T}^n(x)$ is the set of length-n vectors with the same type [52, Ch.11] as x, $B(n, \mathcal{T})$ is the number of length-n vectors with type $\mathcal{T}$, and $\bar{S}^n(\mathcal{T})$ is the expected number of vectors of type $\mathcal{T}$ that are in the kernel of Φ.

In the proof of Theorem 10, I choose the density of Φ such that $\log \alpha = O(\log n)$. This is possible if the number of 1's in each row/column of Φ increases as $O(\log n)$ [59]. If the density of the matrix is high (e.g. $O(n)$ 1's per row/column), $\log \alpha$ becomes smaller, reducing the rate penalty in Lemma 18 and therefore in Theorem 10. However, even if such a code is also linear, it requires significantly more complicated hardware and software design.

Definition 18 (Uniform Permutation LDPC Random Code Ensemble)**.** *An (n, m, ρ, λ) uniform LDPC random code $\left(\mathsf{F}^{(n)}, \mathsf{G}^{(n)}\right)$ is the code corresponding to an (n, m, ρ, λ)*

random uniform-permutation LDPC matrix Φ*. That is,*

$$\mathsf{F}^{(n)}(x^n) = \Phi x^n \tag{4.104}$$

and

$$\mathsf{G}^{(n)}(c) = \arg \max_{x^n : \Phi x^n = c} P_X^n(x^n) = \arg \min_{x^n : \Phi x^n = c} \imath(x^n) \tag{4.105}$$

with ties broken lexicographically in ascending order.

The following auxiliary results are useful for the proof of Theorem 10.

The random coding union (RCU) bound, stated below, is a non-asymptotic achievability result on abstract sources.

Theorem 11 (LDPC RCU)**.** *For a random variable* X *on alphabet* $\{0,1\}^N$*, let* $(\mathsf{F}^{(N)}, \mathsf{G}^{(N)})$ *be an* (N, M, ρ, λ) *uniform LDPC random code ensemble. Then,*

$$\begin{aligned} &\mathbb{E}\left[\mathbb{P}\left[\mathsf{G}^{(N)}\left(\mathsf{F}^{(N)}(X)\right) \neq X \middle| \mathsf{F}^{(N)}, \mathsf{G}^{(N)}\right]\right] \\ &\quad \leqslant \mathbb{E}\left[\min\left\{1, \frac{\alpha}{2^M}\mathbb{E}\left[\exp\left(\imath(\bar{X})\right)\mathbf{1}\left\{\imath(\bar{X}) \leqslant \imath(X)\right\}|X\right]\right\}\right] \end{aligned} \tag{4.106}$$

where $P_{X\bar{X}}(x, \bar{x}) = P_X(x)P_X(\bar{x})$ *for all* $x, \bar{x} \in \{0,1\}^N$*.*

Proof of Theorem 11. The expected error probability can be bounded as follows.

$$\begin{aligned} &\mathbb{E}\left[\mathbb{P}\left[\mathsf{G}^{(N)}\left(\mathsf{F}^{(n)}(X)\right) \neq X \middle| \mathsf{F}^{(N)}, \mathsf{G}^{(N)}\right]\right] \\ \leqslant &\mathbb{P}\left[\exists \bar{x} \in \{0,1\}^N \backslash \{X\} \text{ s.t. } \imath(\bar{x}) \leqslant \imath(X), \Phi\bar{x} = \Phi X\right] \\ = &\mathbb{P}\left[\bigcup_{\bar{x} \in \{0,1\}^N \backslash \{X\}} \{\imath(\bar{x}) \leqslant \imath(X), \Phi\bar{x} = \Phi X\}\right] \\ = &\mathbb{E}\left[\mathbb{P}\left[\bigcup_{\bar{x} \in \{0,1\}^N \backslash \{X\}} \{\imath(\bar{x}) \leqslant \imath(X), \Phi\bar{x} = \Phi X\}|X\right]\right] \\ \leqslant &\mathbb{E}\left[\min\left\{1, \sum_{\bar{x} \in \{0,1\}^N \backslash \{X\}} \mathbb{P}\left[\{\imath(\bar{x}) \leqslant \imath(X), \Phi\bar{x} = \Phi X\}|X\right]\right\}\right] \\ \leqslant &\mathbb{E}\left[\min\left\{1, \frac{\alpha}{2^M}\sum_{\bar{x} \in \{0,1\}^N} \mathbf{1}\left\{\imath(\bar{x}) \leqslant \imath(X)\right\}\right\}\right] \\ = &\mathbb{E}\left[\min\left\{1, \frac{\alpha}{2^M}\mathbb{E}\left[\exp\left(\imath(\bar{X})\right)\mathbf{1}\left\{\imath(\bar{X}) \leqslant \imath(X)\right\}|X\right]\right\}\right]. \end{aligned}$$

□

Lemma 18. *Consider stationary, memoryless source $X_1, X_2, \ldots$ drawn i.i.d. according to p.m.f. P_X on alphabet $\mathcal{X} = \{0,1\}$ that satisfies $V(X) > 0$. Let $\left(\mathsf{F}_1^{(n)}, \mathsf{G}_1^{(n)}\right)$ be an $(n, m_1, \rho_1, \lambda_1)$ uniform LDPC random code ensemble (with encoder matrix Φ) with*

$$R_1 = \frac{m_1}{n} = H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}\left(\epsilon - \frac{C'}{\sqrt{n}}\right) - \frac{\log n}{2n} + \frac{\log \alpha}{n} \tag{4.107}$$

where $C' = B(X) + C(X)$. For a conditional distribution $P_{Y|X} : \mathcal{X} \times \mathcal{Y} \mapsto [0,1]$, let $(\mathsf{F}_2^{(n)}, \mathsf{G}_2^{(n)})$ be the random binning code ensemble with side information at rate

$$R_2 = H(Y|X) + \delta_2, \tag{4.108}$$

with δ_2 being a positive constant. Then, the random code ensemble satisfies

$$\mathbb{E}_{\mathsf{F}_1^{(n)}, \mathsf{G}_1^{(n)}}\left[\mathbb{P}_{X^n}\left[\mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right) \neq X^n\right]\right] \leqslant \epsilon \tag{4.109}$$

and

$$\mathbb{E}_{\mathsf{F}_1^{(n)}, \mathsf{G}_1^{(n)}, \mathsf{F}_2^{(n)}, \mathsf{G}_2^{(n)}}\left[\mathbb{P}_{X^n Y^n}\left[\mathsf{G}_2^{(n)}\left(\mathsf{F}_2^{(n)}\left(Y^n\right), \mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right)\right) \neq Y^n\right]\right] \leqslant 2^{-nE_2} \tag{4.110}$$

where $E_2 > 0$ is a function of δ_2.

Proof of Lemma 18. By Theorem 11,

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{P}\left[\mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right) \neq X^n \middle| \mathsf{F}_1^{(n)}, \mathsf{G}_1^{(n)}\right]\right] \\
&\quad \leqslant \mathbb{E}\left[\min\left\{1, \frac{\alpha}{2^m}\mathbb{E}\left[\exp\left(\imath\left(\bar{X}^n\right)\right)\mathbf{1}\left\{\imath\left(\bar{X}^n\right) \leqslant \imath\left(X^n\right)\right\}|X^n\right]\right\}\right]
\end{aligned}$$

where $P_{X\bar{X}}^n = P_X^n P_X^n$. Applying [21, Lemma 47] to the term

$$\mathbb{E}\left[\exp\left(\imath\left(\bar{X}^n\right)\right)\mathbf{1}\left\{\imath\left(\bar{X}^n\right) \leqslant \imath\left(X^n\right)\right\}|X^n\right]$$

gives

$$\mathbb{E}\left[\exp\left(\imath\left(\bar{X}^n\right)\right)\mathbf{1}\left\{\imath\left(\bar{X}^n\right) \leqslant \imath\left(X^n\right)\right\}|X^n\right] \leqslant \frac{C(X)}{\sqrt{n}}\exp\left(\imath(X^n)\right)$$

where $C(X)$ is a positive constant determined by P_X. Hence

$$\begin{aligned}
&\mathbb{E}\left[\mathbb{P}\left[\mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right) \neq X^n \middle| \mathsf{F}_1^{(n)}, \mathsf{G}_1^{(n)}\right]\right] \\
\leqslant &\mathbb{E}\left[\min\left\{1, \frac{\alpha C(X)}{2^m\sqrt{n}}\exp\left(\imath(X^n)\right)\right\}\right] \\
= &\mathbb{P}\left[\imath(X^n) > \log\frac{2^m\sqrt{n}}{\alpha C(X)}\right] + \frac{\alpha C(X)}{2^m\sqrt{n}}\mathbb{E}\left[\exp\left(\imath(X^n)\right)\mathbf{1}\left\{\imath(X^n) \leqslant \log\frac{2^m\sqrt{n}}{\alpha C(X)}\right\}\right] \\
\leqslant &\mathbb{P}\left[\imath(X^n) > m + \frac{1}{2}\log n - \log C(X) - \log\alpha\right] + \frac{C(X)}{\sqrt{n}}.
\end{aligned}$$

Choosing

$$m = nH(X) + \sqrt{nV(X)}Q^{-1}\left(\epsilon - \frac{B(X)+C(X)}{\sqrt{n}}\right) - \frac{1}{2}\log n + \log C(X) + \log \alpha$$

where $B(X)$ is the Berry-Esseen constant determined by P_X gives

$$\mathbb{E}_{\mathsf{F}_1^{(n)},\mathsf{G}_1^{(n)}}\left[\mathbb{P}_{X^n}\left[\mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right) \neq X^n\right]\right] \leqslant \epsilon$$

with $C' = B(X) + C(X)$.

Identical to the proof of Lemma 16, the exponential decay of the error probability at the second stage of the onion peeling relies on the coding strategy in Figure 4.2. Due to the choice of R_1 and R_2, there exists a constant δ_{12} such that

$$R_1 + R_2 - H(X,Y) \geqslant \delta_{12} \tag{4.111}$$

for sufficiently large n.

The event $\left\{\mathsf{G}_2^{(n)}\left(\mathsf{F}_2^{(n)}(Y^n), \mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}(X^n)\right)\right) \neq Y^n\right\}$ can be captured by the union of the following error events.

$$\begin{aligned}
\mathcal{E}_2 &\triangleq \left\{\exists \bar{y}^n \in \mathcal{Y}^n \backslash \{Y^n\} \text{ s.t. } \imath(\bar{y}^n|X^n) \leqslant \imath(Y^n|X^n), \mathsf{F}_2^{(n)}(\bar{y}^n) = \mathsf{F}_2^{(n)}(Y^n)\right\} \\
\mathcal{E}_{12} &\triangleq \left\{\exists \bar{x}^n \in \mathcal{X}^n \backslash \{X^n\}, \bar{y}^n \in \mathcal{Y}^n \backslash \{Y^n\}\right. \\
&\qquad \left.\text{s.t. } \imath(\bar{x}^n, \bar{y}^n) \leqslant \imath(X^n, Y^n), \Phi\bar{x}^n = \Phi X^n, \mathsf{F}_2^{(n)}(\bar{y}^n) = \mathsf{F}_2^{(n)}(Y^n)\right\}.
\end{aligned}$$

The probabilities of both events decay exponentially with n since $R_2 - H(Y|X) = \delta_2 > 0$ and (4.111). Specifically, bounding $\mathcal{E}_2$ is identical to bounding the error probability of a random binning code with side information. The error exponent for such random binning code is characterized in [54]. For $\mathcal{E}_{12}$, whereas a better error exponent can be obtained, I provide a simplified argument just to show that

the probability of the event is exponentially decaying.

$$
\begin{aligned}
&\mathbb{P}[\mathcal{E}_{12}] \\
&= \mathbb{P}\left[\bigcup_{\bar{x}^n \neq X^n, \bar{y}^n \neq Y^n} \{\iota(\bar{x}^n, \bar{y}^n) \leqslant \iota(X^n, Y^n), \Phi\bar{x}^n = \Phi X^n, \mathsf{F}_2^{(n)}(\bar{y}^n) = \mathsf{F}_2^{(n)}(Y^n)\}\right] \\
&\leqslant \mathbb{E}\left[\min\left\{1, \sum_{\bar{x}^n \neq X^n, \bar{y}^n \neq Y^n} \mathbb{P}\left[\{\iota(\bar{x}^n, \bar{y}^n) \leqslant \iota(X^n, Y^n), \Phi\bar{x}^n = \Phi X^n, \right.\right.\right. \\
&\qquad\qquad \left.\left.\left. \mathsf{F}_2^{(n)}(\bar{y}^n) = \mathsf{F}_2^{(n)}(Y^n)\} | X^n, Y^n\right]\right\}\right] \\
&\leqslant \mathbb{E}\left[\min\left\{1, \frac{\alpha}{2^{M_1+M_2}} \mathbb{E}\left[\exp(\iota(\bar{X}^n, \bar{Y}^n)) \mathbf{1}\left\{\iota(\bar{X}^n, \bar{Y}^n \leqslant \iota(X^n, Y^n))\right\} | X^n, Y^n\right]\right\}\right] \\
&\leqslant \max(1, \alpha) \mathbb{E}\left[\min\left\{1, \frac{1}{2^{M_1+M_2}} \mathbb{E}\left[\exp(\iota(\bar{X}^n, \bar{Y}^n)) \right.\right.\right. \\
&\qquad\qquad \left.\left.\left. \cdot \mathbf{1}\left\{\iota(\bar{X}^n, \bar{Y}^n \leqslant \iota(X^n, Y^n))\right\} | X^n, Y^n\right]\right\}\right]
\end{aligned}
$$

which is $\max(1, \alpha)$ multiples of a bound for a point-to-point random binning code ensemble for (X^n, Y^n). It can be shown to decay exponentially when $R_1 + R_2 > H(X, Y)$. Therefore, when $\log(\alpha) = O(\log n)$, which is the result of (ρ, λ) choices, there exists positive constant E_2 such that (4.110) holds. □

Proof of Theorem 10. Let

$$
\mathcal{P}_1 \triangleq \left\{\mathbb{P}\left[\mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}(X^n)\right) \neq X^n \middle| \mathsf{F}_1^{(n)}, \mathsf{G}_1^{(n)}\right] > \epsilon\right\} \tag{4.112}
$$

and

$$
\mathcal{P}_2 \triangleq \left\{\mathbb{P}\left[\mathsf{G}_2^{(n)}\left(\mathsf{F}_2^{(n)}(Y^n), \mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}(X^n)\right)\right) \neq Y^n \middle| \mathsf{F}_1^{(n)}, \mathsf{G}_1^{(n)}, \mathsf{F}_2^{(n)}, \mathsf{G}_2^{(n)}\right] > 2^{-n\gamma}\right\}. \tag{4.113}
$$

Here, picking $\gamma < E_2$, the probability of $\mathcal{P}_2$ can easily be bounded using Markov's inequality and Lemma 18.

$$
\mathbb{P}[\mathcal{P}_2] \leqslant \mathbb{P}\left[\mathsf{G}_2^{(n)}\left(\mathsf{F}_2^{(n)}(Y^n), \mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}(X^n)\right)\right) \neq Y^n\right] \cdot 2^{n\gamma} = 2^{-n(E_2-\gamma)} \tag{4.114}
$$

which exponentially decays as $n \to \infty$.

To bound the probability of $\mathcal{P}_1$, I use [9, Lemma 25], a result from Markov's inequality. For a point-to-point fixed-length source code, by converse result of the finite-blocklength source coding theorem, the smallest possible error probability of

a code with rate specified in (4.81) is close to ϵ. Specifically, let $C^{(n)}$ be all codes $\left(\mathsf{f}_1^{(n)}, \mathsf{g}_1^{(n)}\right)$ such that

$$R_1 = H(X) + \sqrt{\frac{V(X)}{n}} Q^{-1}\left(\epsilon - \frac{C'}{\sqrt{n}} - \frac{\Delta_\delta \log \alpha}{\sqrt{n}}\right) - \frac{\log n}{2n} + \frac{\log \alpha}{n} \tag{4.115}$$

for some constant $\Delta_\delta > 0$, and let $C_{\text{Linear}}^{(n)} \subseteq C^{(n)}$ be all linear codes with the same rate. Then, the converse of almost lossless point-to-point code [28][9] implies

$$\begin{aligned}\min_{\left(\mathsf{f}_1^{(n)}, \mathsf{g}_1^{(n)}\right) \in C_{\text{Linear}}^{(n)}} \mathbb{P}\left[\mathsf{g}_1^{(n)}\left(\mathsf{f}_1^{(n)}\left(X^n\right)\right) \neq X^n\right] &\geqslant \min_{\left(\mathsf{f}_1^{(n)}, \mathsf{g}_1^{(n)}\right) \in C^{(n)}} \mathbb{P}\left[\mathsf{g}_1^{(n)}\left(\mathsf{f}_1^{(n)}\left(X^n\right)\right) \neq X^n\right] \\ &\geqslant \epsilon - \frac{(\Delta_\delta + \Delta)\log \alpha}{\sqrt{n}}\end{aligned}$$

for some $\Delta > 0$.

Lemma 16 shows that the expected error probability satisfies

$$\mathbb{E}\left[\mathbb{P}\left[\mathsf{G}_1^{(n)}\left(\mathsf{F}_1^{(n)}\left(X^n\right)\right) \neq X^n \middle| \mathsf{F}_1^{(n)}, \mathsf{G}_1^{(n)}\right]\right] \leqslant \epsilon - \frac{\Delta_\delta \log \alpha}{\sqrt{n}}. \tag{4.116}$$

Applying [9, Lemma 25],

$$\mathbb{P}[\mathcal{P}_1] \leqslant \frac{\epsilon - \frac{\Delta_\delta \log \alpha}{\sqrt{n}} - \epsilon + \frac{(\Delta_\delta + \Delta)\log \alpha}{\sqrt{n}}}{\epsilon - \epsilon + \frac{(\Delta_\delta + \Delta)\log \alpha}{\sqrt{n}}} \tag{4.117}$$

$$= \frac{\Delta}{\Delta_\delta + \Delta}. \tag{4.118}$$

Choice Δ_δ can be made to bound (4.118) from above by an arbitrarily small constant. Here, for simplicity, I choose Δ_δ as a function of $\delta \in (0, 1)$ that makes $\mathbb{P}[\mathcal{P}_1] \leqslant \delta/2$. For n large enough, event $\mathcal{P}_2$ in (4.113) satisfies $\mathbb{P}[\mathcal{P}_2] \leqslant \delta/2$ as well.

Therefore, there exists a sequence of linear codes $(\mathsf{F}_1^{(n)}, \mathsf{G}_1^{(n)})$ satisfying the rate constraint in (4.10) such that one can design $(\mathsf{F}_2^{(n)}, \mathsf{G}_2^{(n)})$ with side information $\hat{X}^n$ to have (exponentially) vanishing error for any rate greater than $H(Y|X)$. Applying Lemma 11 completes the proof. □

4.7 Proof of Auxiliary Results

This section contains proofs of auxiliary results that are primarily technical and tangential to the main themes of this chapter.

Proof of Lemma 12

Proof.

$$\begin{aligned}
\frac{1}{n}H(Y^n|\mathcal{E}) &= \sum_{y^n\in\mathcal{Y}^n} P_{Y^n|\mathcal{E}^{(n)}}(y^n)\log\frac{1}{P_{Y^n|\mathcal{E}^{(n)}}(y^n)} \\
&= \frac{1}{n}\sum_{y^n\in\mathcal{Y}^n}\frac{\mathbb{P}[\mathcal{E}^{(n)}|Y^n=y^n]P_Y^n(y^n)}{\mathbb{P}[\mathcal{E}^{(n)}]}\log\frac{\mathbb{P}[\mathcal{E}^{(n)}]}{\mathbb{P}[\mathcal{E}^{(n)}|Y^n=y^n]P_Y^n(y^n)} \\
&= \frac{1}{n}\log\mathbb{P}[\mathcal{E}^{(n)}] + \frac{1}{n\mathbb{P}[\mathcal{E}^{(n)}]}\sum_{y^n\in\mathcal{Y}^n}\mathbb{P}[\mathcal{E}^{(n)}|Y^n=y^n]P_Y^n(y^n)\log\frac{1}{P_Y^n(y^n)} \\
&\quad + \frac{1}{n\mathbb{P}[\mathcal{E}^{(n)}]}\sum_{y^n\in\mathcal{Y}^n}P_Y^n(y^n)\mathbb{P}[\mathcal{E}^{(n)}|Y^n=y^n]\log\frac{1}{\mathbb{P}[\mathcal{E}^{(n)}|Y^n=y^n]}.
\end{aligned}$$

The first term satisfies $\lim_{n\to\infty}\log\mathbb{P}[\mathcal{E}^{(n)}]/n = 0$. Observing that $a\log\frac{1}{a} \in [0, \frac{1}{e\ln 2}]$ for $a\in[0,1]$, the third term satisfies

$$\lim_{n\to\infty}\frac{1}{n\mathbb{P}[\mathcal{E}^{(n)}]}\sum_{y^n\in\mathcal{Y}^n}P_Y^n(y^n)\mathbb{P}[\mathcal{E}^{(n)}|Y^n=y^n]\log\frac{1}{\mathbb{P}[\mathcal{E}^{(n)}|Y^n=y^n]} = 0$$

as well. The second term can be written as

$$\begin{aligned}
&\frac{1}{n\mathbb{P}[\mathcal{E}^{(n)}]}\sum_{y^n\in\mathcal{Y}^n}\mathbb{P}[\mathcal{E}^{(n)}|Y^n=y^n]P_Y^n(y^n)\log\frac{1}{P_Y^n(y^n)} \\
&\qquad = \frac{1}{\mathbb{P}[\mathcal{E}^{(n)}]}\mathbb{E}\left[\mathbb{P}[\mathcal{E}^{(n)}|Y^n]\frac{\imath(Y^n)}{n}\right] \\
&\qquad = \frac{1}{\mathbb{P}[\mathcal{E}^{(n)}]}\mathbb{E}\left[\mathbb{P}[\mathcal{E}^{(n)}|Y^n]\frac{\imath(Y^n)}{n} - \mathbb{P}[\mathcal{E}^{(n)}]H(Y)\right] + H(Y) \\
&\qquad = \frac{1}{\mathbb{P}[\mathcal{E}^{(n)}]}\mathbb{E}\left[\mathbb{P}[\mathcal{E}^{(n)}|Y^n]\left(\frac{\imath(Y^n)}{n} - H(Y)\right)\right] + H(Y)
\end{aligned}$$

where

$$\begin{aligned}
\left|\mathbb{E}\left[\mathbb{P}[\mathcal{E}^{(n)}|Y^n]\left(\frac{\imath(Y^n)}{n} - H(Y)\right)\right]\right| &\leqslant \mathbb{E}\left[|\mathbb{P}[\mathcal{E}^{(n)}|Y^n]|\left|\frac{\imath(Y^n)}{n} - H(Y)\right|\right] \\
&\leqslant \mathbb{E}\left[\left|\frac{\imath(Y^n)}{n} - H(Y)\right|\right] \\
&\to 0.
\end{aligned}$$

Therefore,

$$\lim_{n\to\infty}\frac{1}{n}H(Y^n|\mathcal{E}^{(n)}) = 0 + 0 + H(Y) + 0 = H(Y).$$

□

Proof of Lemma 13

Proof. By Lemma 12, $\lim_{n\longrightarrow\infty}\frac{1}{n}H(A^n|A^n\in\mathcal{S}^{(n)})=\lim_{n\longrightarrow\infty}\frac{1}{n}H(A^n|A^n\notin\mathcal{S}^{(n)})=H(A)$. Therefore,

$$H\left(f^{(n)}(A^n)\middle|A^n\in\mathcal{S}^{(n)}\right)\leqslant H\left(A^n|A^n\in\mathcal{S}^{(n)}\right)$$

implies

$$\lim_{n\longrightarrow\infty}\frac{1}{n}H\left(f^{(n)}(A^n)\middle|A^n\in\mathcal{S}^{(n)}\right)\leqslant H(A). \tag{4.119}$$

Similarly,

$$\lim_{n\longrightarrow\infty}\frac{1}{n}H\left(f^{(n)}(A^n)\middle|A^n\notin\mathcal{S}^{(n)}\right)\leqslant H(A). \tag{4.120}$$

Notice that the chain rule of entropy gives

$$\begin{aligned}&H\left(f^{(n)}(A^n)\right)+H\left(\mathbf{1}\left\{A^n\in\mathcal{S}^{(n)}\right\}\middle|f^{(n)}(A^n)\right)\\&\qquad=H\left(\mathbf{1}\left\{A^n\in\mathcal{S}^{(n)}\right\}\right)+H\left(f^{(n)}(A^n)\middle|\mathbf{1}\left\{A^n\in\mathcal{S}^{(n)}\right\}\right),\end{aligned}$$

which, combined with the fact that

$$0\leqslant H\left(\mathbf{1}\left\{A^n\in\mathcal{S}^{(n)}\right\}\right)\leqslant 1,$$

implies

$$H\left(f^{(n)}(A^n)\right)-1\leqslant H\left(f^{(n)}(A^n)\middle|\mathbf{1}\left\{A^n\in\mathcal{S}^{(n)}\right\}\right)\leqslant H\left(f^{(n)}(A^n)\right). \tag{4.121}$$

Therefore,

$$\lim_{n\longrightarrow\infty}\frac{1}{n}H\left(f^{(n)}(A^n)\middle|\mathbf{1}\left\{A^n\in\mathcal{S}^{(n)}\right\}\right)=H(A), \tag{4.122}$$

which implies that both (4.119) and (4.120) hold with equality. □

Proof of Lemma 14

Proof. For any $a^n \in \mathcal{S}^{(n)}$,

$$\begin{aligned}
&\imath\left(f^{(n)}(a^n)\,\middle|A^n \in \mathcal{S}^{(n)}\right)\\
&= \log \frac{1}{\Pr\left[f^{(n)}(A^n) = f^{(n)}(a^n)|A^n \in \mathcal{S}^{(n)}\right]}\\
&= \log \frac{\Pr\left[A^n \in \mathcal{S}^{(n)}\right]}{\Pr\left[f^{(n)}(A^n) = f^{(n)}(a^n), A^n \in \mathcal{S}^{(n)}\right]}\\
&= \log \frac{\Pr\left[A^n \in \mathcal{S}^{(n)}\right]}{\Pr\left[f^{(n)}(A^n) = f^{(n)}(a^n)\right]\Pr\left[A^n \in \mathcal{S}^{(n)}|f^{(n)}(A^n) = f^{(n)}(a^n)\right]}\\
&= -\log \frac{1}{\Pr\left[A^n \in \mathcal{S}^{(n)}\right]} + \imath\left(f^{(n)}(a^n)\right) + \log \frac{1}{\Pr\left[A^n \in \mathcal{S}^{(n)}|f^{(n)}(A^n) = f^{(n)}(a^n)\right]}\\
&\geqslant -\log \frac{1}{\Pr\left[A^n \in \mathcal{S}^{(n)}\right]} + \imath\left(f^{(n)}(a^n)\right).
\end{aligned}$$

Therefore, since $\lim_{n\to\infty} \Pr\left[A^n \in \mathcal{S}^{(n)}\right] = p$, for any $\delta > 0$ and sufficiently large n,

$$\frac{1}{n}\imath\left(f^{(n)}(a^n)\,\middle|A^n \in \mathcal{S}^{(n)}\right) \geqslant \frac{1}{n}\imath\left(f^{(n)}(a^n)\right) - \frac{1}{n}\log\frac{1}{p-\delta}.$$

That is, with at least a probability approaching p,

$$\frac{1}{n}\imath\left(f^{(n)}(A^n)\,\middle|A^n \in \mathcal{S}^{(n)}\right) \geqslant \frac{1}{n}\imath\left(f^{(n)}(A^n)\right) - \frac{1}{n}\log\frac{1}{p-\delta}.$$

Combining this with (4.36) gives (4.37). □

Proof of Lemma 15

Proof. Define subsets of the support of P_{XY}, $\mathcal{A}$ and $\mathcal{B}$, as

$$\begin{aligned}
\mathcal{A} &\triangleq \left\{(x,y) : y \neq \arg\max_{y'\in\mathcal{Y}} P_{Y|X}(y'|x)\right\}\\
\mathcal{B} &\triangleq \left\{(x,y) : y = \arg\max_{y'\in\mathcal{Y}} P_{Y|X}(y'|x)\right\}.
\end{aligned}$$

Then, for large enough n, for all $(x,y) \in \mathcal{A}$,

$$0 \leqslant Q(x,y) - P(x,y) \leqslant \frac{P_X(x)}{n},$$

and for all $(x,y) \in \mathcal{B}$,

$$0 \leqslant P_{XY}(x,y) < Q_{XY}(x,y) + \frac{|\mathcal{Y}|-1}{n}P_X(x).$$

For specific (x^n, y^n), let $N(x, y)$ denote the number of occurrence of (x, y) in (x^n, y^n). Then,

$$
\begin{aligned}
&P_{XY}^n(x^n, y^n) \\
&= \prod_{(x,y)\in\mathcal{B}} P_{XY}(x, y)^{N(x,y)} \prod_{(x,y)\in\mathcal{A}} P_{XY}(x, y)^{N(x,y)} \\
&\leqslant \prod_{(x,y)\in\mathcal{B}} P_{XY}(x, y)^{N(x,y)} \prod_{(x,y)\in\mathcal{A}} Q_{XY}(x, y)^{N(x,y)} \\
&\leqslant \prod_{(x,y)\in\mathcal{B}} \left(Q_{XY}(x, y) + \frac{(|\mathcal{Y}| - 1)P_X(x)}{n}\right)^{N(x,y)} \prod_{(x,y)\in\mathcal{A}} Q_{XY}(x, y)^{N(x,y)} \\
&= \prod_{(x,y)\in\mathcal{B}} \left(1 + \frac{(|\mathcal{Y}| - 1)P_X(x)}{nQ_{XY}(x, y)}\right)^{N(x,y)} Q_{XY}(x, y)^{N(x,y)} \prod_{(x,y)\in\mathcal{A}} Q_{XY}(x, y)^{N(x,y)} \\
&= Q_{XY}^n(x, y) \prod_{(x,y)\in\mathcal{B}} \left(1 + \frac{(|\mathcal{Y}| - 1)P_X(x)}{nQ_{XY}(x, y)}\right)^{N(x,y)} \\
&\leqslant Q_{XY}^n(x, y) \left(1 + \frac{(|\mathcal{Y}| - 1)\max_x P_X(x)}{n \min_{(x,y)\in\mathcal{B}} Q_{XY}(x, y)}\right)^n \\
&\leqslant C_0 Q_{XY}^n(x, y)
\end{aligned}
$$

for some positive constant C_0. This is because $(1 + a/n)^n$ converges to a constant upper bound as $n \to \infty$. □

Chapter 5

DISTRIBUTED HYPOTHESIS TESTING AND SOURCE CODING

The materials in this chapter are published in part as [60].

5.1 Introduction

The problem of hypothesis testing with communication constraints, proposed in [32], or distributed hypothesis testing (e.g., [33]), considers the scenario where a hypothesis test uses the compressed description of one source and uncoded version of a potentially dependent source as inputs. An example hypothesis of interest in this and prior works concerns whether the sources are distributed according to certain joint distribution. Studies of this problem include [32][33][34][35][36][61][62][63][64]. The strategy for a blocklength-n hypothesis test on the compressed description of one source, X^n, and the uncompressed observation of another source, Y^n, typically involves generating the compressed description by finding a random variable U that captures some shared information between the two sources, and compressing U^n instead of X^n. While it is possible to set $U = X$, [33] and [35] suggest that this does not always yield the best performance. More precisely, this choice is suboptimal when distinguishing between distributions is the only goal. Note, however, that setting $U = X$ is useful if the receiver wishes both to apply the hypothesis test and to reconstruct X^n when there is, indeed, dependence between the two sources. That scenario, here called distributed hypothesis testing and source coding, is the focus of this chapter.

As an example, consider a vector source sequence $\{(X_i, Y_i)\}_{i=1}^{\infty}$ drawn i.i.d. according to some p.m.f. P_{XY}. The receiver sees a perfect copy of Y^n and an encoded description of $\{X_i\}_{i=1}^n$, but does not know whether $k = 0$, in which case $X_{(k)}^n$ and Y^n are dependent sources observed in the same time frame, or $k > 0$, in which case $X_{(k)}^n$ and Y^n are observed in different time frames and therefore independent. A successful hypothesis test would enable the receiver to decode with low error probability using a rate close to the conditional entropy of X given side information Y to describe X^n if the encoded description indeed corresponds to a dependent source block, and avoid decoding when the absence of dependent side information makes the probability of successful decoding low.

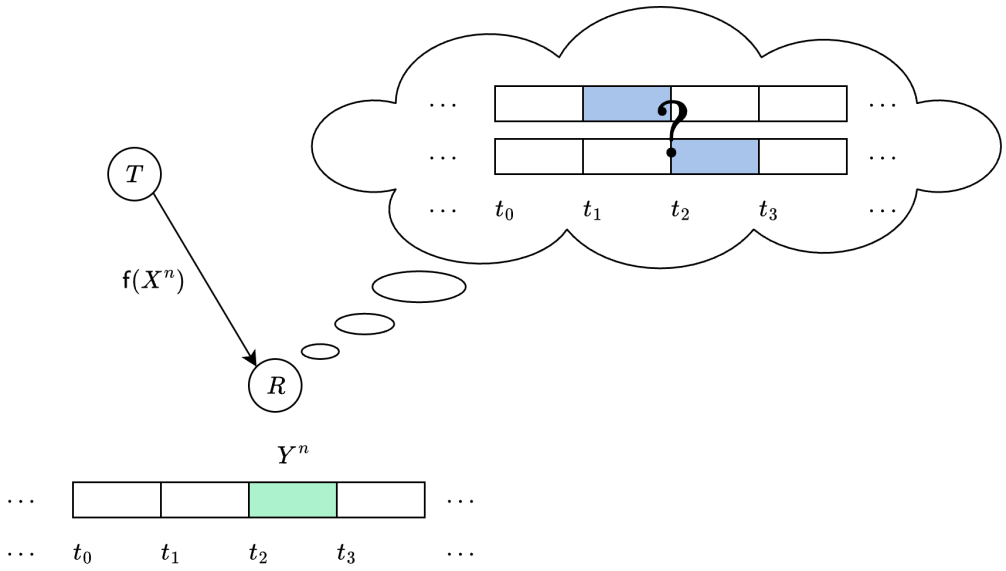


Figure 5.1: Illustration of the scenario where the receiver (R) does not know whether the encoded description received from the transmitter (T) is for a dependent source block or an independent one due to the unknown temporal proximity.

I here study distributed hypothesis testing with dependent source reconstruction assuming use of a random binning source code for X^n. Random binning is popular in many network information theory problems, including problems on hypothesis testing with communication constraints [33][35].

The output statistics of a random binning code have been studied in both point-to-point and muti-source scenarios. In [65], Yassaee *et al.* show that the expected total variation distance between the distribution on the symbols of compressed data and the uniform distribution converges to 0 as the blocklength grows provided the compression rate is smaller than the entropy. In [66], Mojahedian *et al.* strengthen this result by showing the convergence speed to be exponential. In [67], Kavian *et al.* study the output statistics of a random binning code in terms of Tsallis-divergence. While these studies focus on the case when the rate does not converge to source entropy, [68] studies the statistics of a compressed description when the rate approaches the source entropy as $O(1/\sqrt{n})$, though not for a random binning encoder.

This chapter shows that if X^n is compressed using a random binning code of sufficient rate, then, the encoded description, combined with Y^n, can be used to perform a hypothesis test against independence such that the type-II error of this test decays to zero subexponentially with the blocklength. In this case, sufficient rate means that

source X^n is compressed at a rate R_n that converges sufficiently slowly ($O(1/\sqrt{n})$) to the conditional entropy of X given side information Y. I explore the implications of this result on the statistics of source descriptions of random binning codes. I further show that a similar strategy can be used to build hypothesis tests that can distinguish any joint distribution between X and Y from other joint distributions with the same marginals and higher conditional entropies (less dependence). Finally, I show that, within the given framework, with high probability, no two distributions with identical marginals and conditional entropies exceeding the rate are distinguishable.

5.2 Distributed Source Code and Hypothesis Test

In the next two sections, I study the performance of a distributed hypothesis test that uses an encoded description of source X^n and an uncoded vector Y^n to try to distinguish whether (X^n, Y^n) arises from P_{XY}^n or $P_X^n P_Y^n$. Here $\mathcal{X} \times \mathcal{Y}$ is a finite alphabet, P_{XY} is a p.m.f. with support $\mathcal{S} \subseteq \mathcal{X} \times \mathcal{Y}$ for which $P_{XY} \neq P_X P_Y$ and $V(X|Y) > 0$.

Definition 19 (Distributed Source Code and Hypothesis Test)**.** *An (n, M_n) distributed source code and hypothesis test is a triplet $(\mathsf{f}_n, \mathsf{g}_n, \mathsf{t}_n)$, where*

$$\begin{aligned}
\mathsf{f}_n &: \mathcal{X}^n \mapsto [M_n] \\
\mathsf{g}_n &: [M_n] \times \mathcal{Y}^n \mapsto \hat{\mathcal{X}}^n \\
\mathsf{t}_n &: [M_n] \times \mathcal{Y}^n \mapsto \{0, 1\}
\end{aligned}$$

and $\mathcal{X}^n \subseteq \hat{\mathcal{X}}^n$. The probability of decoding incorrectly when $(X^n, Y^n) \sim P_{XY}^n$ is $P_{\text{dec}}^{(n)} \triangleq \mathbb{P}\left[\mathsf{g}_n(\mathsf{f}_n(X^n), Y^n) \neq X^n\right]$. The type-I error probability of the hypothesis test is $P_{\text{I}}^{(n)} \triangleq \mathbb{P}\left[\mathsf{t}_n(\mathsf{f}_n(X^n), Y^n) = 1\right]$. The type-II error probability of the hypothesis test is $P_{\text{II}}^{(n)} \triangleq \mathbb{Q}[\mathsf{t}_n(\mathsf{f}_n(X^n), Y^n) = 0]$. Here, $\mathbb{P}$ and $\mathbb{Q}$ capture the randomness of $(X^n, Y^n) \sim P_{XY}^n$ and $(X^n, Y^n) \sim P_X^n P_Y^n$, respectively.

An $(\mathsf{f}_n, \mathsf{g}_n, \mathsf{t}_n)$ with rate R_n refers to an $(n, \lfloor 2^{nR_n} \rfloor)$ distributed source code and hypothesis test.

I am interested in analyzing the performance of the classic random binning code. In both almost lossless source and channel coding problems, two very popular decoders frequently used to derive achievability bounds are the optimal maximum likelihood decoder and the suboptimal threshold decoder. The difference between the rate needed to achieve a target error probability using the optimal decoder and the rate needed to achieve the same error probability using the threshold decoder

is $O(\log n/n)$ in the achievability bounds under various scenarios [21][26][9]. In the next two sections, I analyze both decoder types. Each has some interesting implications. The analysis of the maximum likelihood decoder supports an argument for the existence of a deterministic code that achieves target error probabilities for both decoding and hypothesis testing. The analysis of the threshold decoder shows that one can use threshold tests on the same quantity, potentially with different thresholds, to perform both hypothesis testing and source code decoding.

5.3 Maximum Likelihood Decoder as Source Decoder

Definition 20, below, defines a sub-class of distributed source code and hypothesis tests in which the encoder f_n for source X^n is a random binning code while the decoder g_n is a maximum likelihood decoder and hypothesis test t_n is a threshold test. Given the random design of f_n and dependence of g_n and t_n on f_n, notation $(\mathsf{F}_n, \mathsf{G}_n, \mathsf{T}_n)$ describes an instance of the resulting code.

Definition 20 (Random Binning Distributed Source Code and Hypothesis Test with Maximum Likelihood Decoder)**.** *A random binning encoder* $\mathsf{F}_n : \mathcal{X}^n \mapsto [M_n]$ *is a source encoder in which each* $x^n \in \mathcal{X}^n$ *is independently assigned a codeword* $\mathsf{F}_n(x^n) \in [M_n]$ *uniformly at random. A maximum likelihood decoder* G_n *and hypothesis test* T_n *for encoder* F_n *are defined as*

$$\mathsf{G}_n(c, y^n) = \arg \max_{x^n \in \mathcal{X}^n, \mathsf{F}_n(x^n)=c} \imath(x^n|y^n) \tag{5.1}$$

$$\mathsf{T}_n(c, y^n) = \begin{cases} 0, & \exists x^n \in \mathcal{X}^n, \mathsf{F}_n(x^n) = c, \imath(x^n|y^n) \leqslant \log \gamma_{\mathsf{T}} \\ 1, & \textit{otherwise}. \end{cases} \tag{5.2}$$

All ties are broken lexicographically in ascending order.

Unless indicated otherwise, I use $\mathbb{P}$ to capture the randomness of both the source vectors when $(X^n, Y^n) \sim P_{XY}^n$ and the choice of a source encoder under the random binning code construction. I use $\mathbb{Q}$ to capture the randomness of both the source vectors when $(X^n, Y^n) \sim P_X^n P_Y^n$ and the random binning code construction.

Theorem 12 (Random Binning with Maximum Likelihood Decoder)**.** *For* $0 < \epsilon_S < \epsilon_H < 1$*, there exist positive constants* C, κ *such that the random binning ensemble* $(\mathsf{F}_n, \mathsf{G}_n, \mathsf{T}_n)$ *with rate*

$$R_n = H(X|Y) + \sqrt{\frac{V(X|Y)}{n}} Q^{-1}(\epsilon_S) - \frac{\log n}{2n} + \frac{C}{n} \tag{5.3}$$

satisfies

$$\mathbb{P}\left[\mathsf{G}_n\left(\mathsf{F}_n(X^n), Y^n\right) = X^n\right] \geqslant 1 - \epsilon_S \tag{5.4}$$

$$\mathbb{P}\left[\mathsf{T}_n(\mathsf{F}_n(X^n), Y^n) = 0\right] \geqslant 1 - \epsilon_H \tag{5.5}$$

$$\mathbb{Q}\left[\mathsf{T}_n\left(\mathsf{F}_n(X^n), Y^n\right) = 0\right] \leqslant \exp\{-\kappa\sqrt{n}\} \tag{5.6}$$

for sufficiently large n.

Proof. The performance of $(\mathsf{F}_n, \mathsf{G}_n)$ is known from prior work [15][9][42]. The third-order achievability bound for a source code with side information [15, Theorem 7, 8] can also be established by a random binning argument like [9, Theorem 5] using the maximum likelihood decoder. The rest of the proof analyzes the behavior of the hypothesis test T_n.

Define random acceptance set $\mathcal{A}^{(n)} \subseteq \mathcal{X}^n \times \mathcal{Y}^n$ as

$$\mathcal{A}^{(n)} \triangleq \left\{(x^n, y^n) : \mathsf{T}_n\left(\mathsf{F}_n\left(x^n\right), y^n\right) = 0\right\}.$$

A type-I error occurs if the true p.m.f. is $\mathbb{P}$ but $(X^n, Y^n) \notin \mathcal{A}^{(n)}$. The probability of this event can be bounded as

$$\begin{aligned}
\mathbb{P}[\mathcal{A}_{\mathsf{c}}^{(n)}] &= \mathbb{P}\left[\mathsf{T}_n(\mathsf{F}_n(X^n), Y^n) = 1\right] \\
&= \mathbb{P}\left[\nexists x^n \in \mathcal{X}^n : \mathsf{F}_n(x^n) = \mathsf{F}_n(X^n), \imath(x^n|Y^n) \leqslant \log\gamma_{\mathsf{T}}\right] \\
&\leqslant \mathbb{P}\left[\imath(X^n|Y^n) > \log\gamma_{\mathsf{T}}\right].
\end{aligned}$$

Setting

$$\log\gamma_{\mathsf{T}} = nH(X|Y) + \tau_{\mathsf{T}}\sqrt{nV(X|Y)} \tag{5.7}$$

and applying the Berry-Esseen inequality give $\mathbb{P}[\mathcal{A}_{\mathsf{c}}^{(n)}] \leqslant \epsilon_H$, where

$$\tau_{\mathsf{T}} = Q^{-1}\left(\epsilon_H - \frac{B}{\sqrt{n}}\right)$$

and $B = \frac{C_0 T}{V^{3/2}}$ is the Berry-Esseen constant.

A type-II error occurs if the true p.m.f. is $\mathbb{Q}$ and $(X^n, Y^n) \in \mathcal{A}^{(n)}$. I bound this error probability as

$$\begin{aligned}
\mathbb{Q}[\mathcal{A}^{(n)}] &= \mathbb{Q}\left[\exists x^n \in \mathcal{X}^n : \mathsf{F}_n(x^n) = \mathsf{F}_n(X^n), \imath(x^n|Y^n) \leqslant \log\gamma_{\mathsf{T}}\right] \\
&\leqslant \mathbb{Q}\left[\exists x^n \in \mathcal{X}^n\backslash\{X^n\} : \mathsf{F}_n(x^n) = \mathsf{F}_n(X^n), \imath(x^n|Y^n) \leqslant \log\gamma_{\mathsf{T}}\right] \\
&\quad + \mathbb{Q}[\imath(X^n|Y^n) \leqslant \log\gamma_{\mathsf{T}}].
\end{aligned} \tag{5.8}$$

To bound the second term in (5.8), first note that the expected value of $\imath(X^n|Y^n)$ under $\mathbb{Q}$ is

$$\begin{aligned}
\mathbb{E}_{\mathbb{Q}}\left[\imath(X|Y)\right] &= \mathbb{E}_{P_X P_Y}\left[\imath(X|Y)\right] \\
&= \sum_{x,y} P_X(x)P_Y(y)\log\frac{1}{P_{X|Y}(x|y)} \\
&= \sum_{x,y} P_X(x)P_Y(y)\log\frac{P_X(x)P_Y(y)}{P_{X|Y}(x|y)P_Y(y)P_X(x)} \\
&= D(P_X P_Y||P_{XY}) + H(X) \\
&> H(X|Y).
\end{aligned}$$

Combining this bound with the choice of γ_{T} in (5.7), Cramér's theorem implies that $\mathbb{Q}[\imath(X^n|Y^n) \leqslant \log\gamma_{\mathsf{T}}]$ decays exponentially with n.

To bound the first term in (5.8), note that

$$\begin{aligned}
&\mathbb{Q}\left[\exists x^n \in \mathcal{X}^n\backslash\{X^n\} : \mathsf{F}_n(x^n) = \mathsf{F}_n(X^n), \imath(x^n|Y^n) \leqslant \log\gamma_{\mathsf{T}}\right] \\
&= \mathbb{Q}\left[\bigcup_{x^n\in\mathcal{X}^n\backslash\{X^n\}} \{\mathsf{F}_n(x^n) = \mathsf{F}_n(X^n), \imath(x^n|Y^n) \leqslant \log\gamma_{\mathsf{T}}\}\right] \\
&\leqslant \frac{1}{M_n}\sum_{x^n\in\mathcal{X}^n, y^n\in\mathcal{Y}^n} P_Y^n(Y^n)\mathbf{1}\left\{\imath(x^n|y^n) \leqslant \log\gamma_{\mathsf{T}}\right\} \qquad (5.9) \\
&= \frac{1}{M_n}\mathbb{E}_{P_{XY}}\left[\frac{1}{P_{X|Y}^n(X^n|Y^n)}\mathbf{1}\{\imath(X^n|Y^n) \leqslant \log\gamma_{\mathsf{T}}\}\right] \\
&\leqslant \frac{1}{M_n}2\left(\frac{\log 2}{\sqrt{2\pi V(X|Y)}} + 2B\right)\frac{1}{\sqrt{n}}\gamma_{\mathsf{T}}. \qquad (5.10)
\end{aligned}$$

Here, (5.9) follows from the union bound and the independent choice of the codewords for different x^n, and (5.10) follows from [21, Lemma 47] which bounds the expectation $\mathbb{E}[\exp\{-Z_n\}\mathbf{1}\{Z_n \geqslant A\}]$ where Z_n is the sum of independent random variables. Since $\epsilon_S < \epsilon_H$ by assumption, for any

$$K < \sqrt{V(X|Y)}\left(Q^{-1}(\epsilon_S) - Q^{-1}(\epsilon_H)\right) \qquad (5.11)$$

and sufficiently large n,

$$\begin{aligned}
\mathbb{Q}[\mathcal{A}_n] &\leqslant \frac{1}{M_n}2\left(\frac{\log 2}{\sqrt{2\pi V(X|Y)}} + 2B\right)\frac{1}{\sqrt{n}}\gamma_{\mathsf{T}} + \exp\{-O(n)\} \\
&\leqslant O(1)\cdot 2^{-\sqrt{nV(X|Y)}\left(Q^{-1}(\epsilon_S)-Q^{-1}(\epsilon_H)\right)} + \exp\{-O(n)\} \qquad (5.12) \\
&\leqslant \exp\{-K\sqrt{n}\},
\end{aligned}$$

where (5.12) follows from the rate R_n and the differentiability of $Q^{-1}(\cdot)$. Therefore, choosing $\kappa = K$ for any such K that satisfies (5.11) completes the proof. □

Corollary 1 (Deterministic Code)**.** *For $0 < \epsilon_S < \epsilon_H < 1$, there exist positive constants C_d, κ_d such that, for rate*

$$R_n \geqslant R_n^* = H(X|Y) + \sqrt{\frac{V(X|Y)}{n}} Q^{-1}(\epsilon_S) - \frac{\log n}{2n} + \frac{C_d}{n}, \tag{5.13}$$

there exists a deterministic distributed source code and hypothesis test $(\mathsf{f}_n, \mathsf{g}_n, \mathsf{t}_n)$ such that

$$\mathbb{P}[\mathsf{g}_n(\mathsf{f}_n(X^n), Y^n) = X^n] \geqslant 1 - \epsilon_S \tag{5.14}$$

$$\mathbb{P}[\mathsf{t}_n(\mathsf{f}_n(X^n), Y^n) = 0] \geqslant 1 - \epsilon_H \tag{5.15}$$

$$\mathbb{Q}[\mathsf{t}_n(\mathsf{f}_n(X^n), Y^n) = 0] \leqslant \exp\{-\kappa_d \sqrt{n}\} \tag{5.16}$$

for sufficiently large n.

Proof. This proof uses [9, Lemma 25], a direct result of Markov's inequality, and the converse for the side information code from [15, Theorem 1], which is tight up to the 3rd order. Choose a constant $C' > 0$; the exact choice is discussed later in the proof. Let the rate R_n be set as

$$R_n = H(X|Y) + \sqrt{\frac{V(X|Y)}{n}} Q^{-1}\left(\epsilon_S - \frac{C'}{\sqrt{n}}\right) - \frac{\log n}{2n} + \frac{C}{n}. \tag{5.17}$$

Then

$$R_n = H(X|Y) + \sqrt{\frac{V(X|Y)}{n}} Q^{-1}\left(\epsilon_S\right) - \frac{\log n}{2n} + \frac{C + \Delta(C')}{n} \tag{5.18}$$

where $\Delta(C') = O(1)$ due to the differentiability of $Q^{-1}(\cdot)$. Let $C_d = C + \Delta(C')$. The side information converse bound ensures that the rate choice in (5.13) implies

$$\mathbb{P}[\mathsf{g}_n(\mathsf{f}_n(X^n), Y^n) \neq Y^n] \geqslant \epsilon_S - \frac{C' + C''}{\sqrt{n}} = \epsilon^* \tag{5.19}$$

for some positive constant C'', no matter how $(\mathsf{f}_n, \mathsf{g}_n)$ are designed. Let notation $\mathbb{P}_{X^nY^n}$ capture only the randomness of X^n and Y^n. Then,

$$\Pr\left(\mathbb{P}_{X^nY^n}[\mathsf{G}_n(\mathsf{F}_n(X^n), Y^n) \neq X^n] \geqslant \epsilon_S\right) \tag{5.20}$$

$$\leqslant \frac{\mathbb{P}[\mathsf{G}_n(\mathsf{F}_n(X^n), Y^n) \neq X^n] - \epsilon^*}{\epsilon_S - \epsilon^*} \tag{5.21}$$

$$\leqslant \frac{\epsilon_S - \frac{C'}{\sqrt{n}} - \epsilon_S + \frac{C'+C''}{\sqrt{n}}}{\epsilon_S - \epsilon_S + \frac{C'+C''}{\sqrt{n}}} \tag{5.22}$$

$$= \frac{C''}{C' + C''}, \tag{5.23}$$

where (5.21) follows from [9, Lemma 25] and (5.22) from the side information converse [15, Theorem 1] and Theorem 12. Therefore, choosing C' so that

$$C' > \frac{\epsilon_H + \epsilon_S}{\epsilon_H - \epsilon_S} C'', \tag{5.24}$$

gives

$$\Pr\left(\mathbb{P}_{X^n Y^n}[\mathsf{G}_n(\mathsf{F}_n(X^n), Y^n) \neq X^n] \geqslant \epsilon_S\right) < \frac{\epsilon_H - \epsilon_S}{2\epsilon_H}. \tag{5.25}$$

Since $\epsilon_H > \epsilon_S$, $(\epsilon_H + \epsilon_S)/2 > \epsilon_S$. Designing T_n to meet error probability constraint $(\epsilon_H + \epsilon_S)/2$ gives

$$\mathbb{P}[\mathsf{T}_n(\mathsf{F}_n(X^n), Y^n) = 0] \geqslant 1 - \frac{\epsilon_H + \epsilon_S}{2} \tag{5.26}$$

so that

$$\Pr(\mathbb{P}_{X^n Y^n}[\mathsf{T}_n(\mathsf{F}_n(X^n), Y^n) = 1] \geqslant \epsilon_H) \leqslant \frac{\epsilon_H + \epsilon_S}{2\epsilon_H}. \tag{5.27}$$

Lastly, the expected value of the type-II error probability decays subexponentially. Therefore, the probability that the type-II error probability decays slower than subexponentially with $\kappa_d < \kappa$ where κ is chosen for type-I error probability $(\epsilon_H + \epsilon_S)/2$, is $o(1)$. Combining this with (5.25) and (5.27) and using the union bound, the probability that the code generated by random binning does not satisfy all of (5.14), (5.15) and (5.16) is strictly smaller than 1, completing the proof. □

Remark 9. *By Theorem 12, there exists a source code with side information at the decoder that achieves the optimal rate up to the 3rd order and for which the encoded description (i.e., the codeword) and the side information can also be used to perform a hypothesis test against independence with type-II error that decays to 0 subexponentially in n and type-I error that meets a constant error probability bound. The resulting performance is better than the performance achievable if one creates a code at the rate described in* (5.13) *solely for the purpose of decoding and then adds a prefix or suffix for hypothesis testing purposes since meeting the hypothesis testing performance goals requires more than* $O(1/n)$.

5.4 Threshold Decoder as Hypothesis Test and Source Decoder

Definition 21, below, defines another sub-class of distributed source code and hypothesis tests in which the encoder f_n for source X^n is a random binning code while the decoder g_n and hypothesis test t_n are threshold tests.

Definition 21 (Random Binning Distributed Source Code and Hypothesis Test with Threshold Decoder)**.** *A random binning encoder* $\mathsf{F}_n : \mathcal{X}^n \mapsto [M_n]$ *is a source encoder in which each* $x^n \in \mathcal{X}^n$ *is independently assigned a codeword* $\mathsf{F}_n(x^n) \in [M_n]$ *uniformly at random. A threshold decoder* G_n *and hypothesis test* T_n *for encoder* F_n *are defined as*

$$\begin{aligned}&\mathsf{G}_n\left(c, y^n\right)\\ &= \begin{cases} x^n, & \exists x^n \in \mathcal{X}^n, \mathsf{F}_n(x^n) = c, \imath(x^n|y^n) \leqslant \log \gamma_{\mathsf{G}} \\ \mathsf{e}, & \textit{otherwise} \end{cases}\end{aligned} \tag{5.28}$$

$$\begin{aligned}&\mathsf{T}_n(c, y^n)\\ &= \begin{cases} 0, & \exists x^n \in \mathcal{X}^n, \mathsf{F}_n(x^n) = c, \imath(x^n|y^n) \leqslant \log \gamma_{\mathsf{T}} \\ 1, & \textit{otherwise}. \end{cases}\end{aligned} \tag{5.29}$$

All ties are broken lexicographically in ascending order.

Although G_n and T_n both employ thresholds on $\imath(x^n|y^n)$, they may use different thresholds for decoding and hypothesis testing.

Theorem 13 (Random Binning with Threshold Decoder)**.** *For* $0 < \epsilon_S < \epsilon_H < 1$*, there exist positive constants* C, κ *such that the random binning ensemble* $(\mathsf{F}_n, \mathsf{G}_n, \mathsf{T}_n)$ *with rate*

$$R_n = H(X|Y) + \sqrt{\frac{V(X|Y)}{n}} Q^{-1}(\epsilon_S) + \frac{C}{n} \tag{5.30}$$

satisfies

$$\mathbb{P}\left[\mathsf{G}_n\left(\mathsf{F}_n(X^n), Y^n\right) = X^n\right] \geqslant 1 - \epsilon_S \tag{5.31}$$

$$\mathbb{P}\left[\mathsf{T}_n(\mathsf{F}_n(X^n), Y^n) = 0\right] \geqslant 1 - \epsilon_H \tag{5.32}$$

$$\mathbb{Q}\left[\mathsf{T}_n\left(\mathsf{F}_n(X^n), Y^n\right) = 0\right] \leqslant \exp\{-\kappa\sqrt{n}\} \tag{5.33}$$

for sufficiently large n.

Proof. Replacing the maximum likelihood decoder with the threshold decoder (5.28) and using

$$\begin{aligned}\log \gamma_{\mathsf{G}} &= nH(X|Y) + \tau_{\mathsf{G}}\sqrt{nV(X|Y)}\\ \tau_{\mathsf{G}} &= Q^{-1}\left(\epsilon_S - \frac{C'}{\sqrt{n}}\right)\end{aligned}$$

where C' is a constant, give (5.30) and (5.31). The argument is identical to the first half of the achievability proof of Theorem 5 in this thesis. The rest of the proof is identical to that of Theorem 12. □

5.5 Implications on Statistics of Random Binning

Theorem 14 highlights an implication of Theorem 13 (or Theorem 12) regarding the statistical dependence of Y^n and a source coded description of X^n. It uses the following definition of the Ω notation which states that one function is asymptotically bounded below by the other.

Definition 22 (Big Ω notation)**.** *Function $f(n) = \Omega(g(n))$ if and only if*

$$\liminf_{n\longrightarrow\infty} f(n)/g(n) > 0.$$

Theorem 14 (Output Statistics of Random Binning)**.** *Given a sequence $\{\mathsf{F}_n\}_{n=1}^{\infty}$ of random binning encoders $\mathsf{F}_n : \mathcal{X}^n \mapsto [\lfloor 2^{nR_n} \rfloor]$, if*

$$R_n = H(X|Y) + O\left(\frac{1}{\sqrt{n}}\right),$$

then

$$\mathbb{E}_{\mathsf{F}_n}\left[D\left(P_{\mathsf{F}_n(X^n)Y^n}||P_{\mathsf{F}_n(X^n)}P_Y^n\right)\right] = \Omega(\sqrt{n}).$$

There are two ways to prove Theorem 14. One is to invoke Theorem 12. The alternative uses a strategy inspired by [68]. The bound obtained by the second method is tighter than that obtained by the first method, but on the same order. I present both proofs in the next two subsections and compare the bounds they yield.

Proof of Theorem 14 using Theorem 12

Proof. From Theorem 12, the codeword from the random binning encoder can be used as input to a hypothesis test that distinguishes $P_{\mathsf{F}_n(X^n)Y^n}$ from $P_{\mathsf{F}_n(X^n)}P_Y^n$. If P and Q are two p.m.f.s on the same support $\mathcal{S}$ and there exists a hypothesis test with acceptance set $\mathcal{A} \subseteq \mathcal{S}$ such that the type-I error probability is $P(\mathcal{A}_\mathsf{c}) = P_\mathrm{I}$ and the type-II error probability is $Q(\mathcal{A}) = P_\mathrm{II}$, then

$$D(P||Q) \geqslant (1 - P_\mathrm{I}) \log \frac{(1 - P_\mathrm{I})}{P_\mathrm{II}} + P_\mathrm{I} \log \frac{P_\mathrm{I}}{1 - P_\mathrm{II}}$$

by the data processing inequality. The Hessian of

$$f(\alpha, \beta) = \alpha \log \frac{\alpha}{\beta} + (1 - \alpha) \log \frac{1 - \alpha}{1 - \beta} \tag{5.34}$$

is

$$\frac{1}{\ln 2}\begin{bmatrix} \frac{1}{\alpha(1-\alpha)} & -\frac{1}{\beta(1-\beta)} \\ -\frac{1}{\beta(1-\beta)} & \frac{\beta^2-2\alpha\beta+\alpha}{\beta^2(1-\beta)^2} \end{bmatrix}, \tag{5.35}$$

from which it can be verified that $f(\cdot,\cdot)$ is jointly convex on $[0,1]\times[0,1]$. Since $R_n = H(X|Y) + O(1/\sqrt{n})$, it satisfies (5.3) in Theorem 12 for some ϵ_S. For any $\epsilon_H < \epsilon_S$,

$$
\begin{aligned}
&\mathbb{E}\left[D\left(P_{\mathsf{F}_n(X^n)Y^n}||P_{\mathsf{F}_n(X^n)}P_Y^n\right)\right] \\
&\geqslant \mathbb{E}\left[\left(1-P_{\mathrm{I}}^{(n)}\right)\log\frac{\left(1-P_{\mathrm{I}}^{(n)}\right)}{P_{\mathrm{II}}^{(n)}} + P_{\mathrm{I}}^{(n)}\log\frac{P_{\mathrm{I}}^{(n)}}{1-P_{\mathrm{II}}^{(n)}}\right] \\
&\geqslant \mathbb{E}\left[1-P_{\mathrm{I}}^{(n)}\right]\log\frac{\mathbb{E}\left[1-P_{\mathrm{I}}^{(n)}\right]}{\mathbb{E}\left[P_{\mathrm{II}}^{(n)}\right]} + \mathbb{E}\left[P_{\mathrm{I}}^{(n)}\right]\log\frac{\mathbb{E}\left[P_{\mathrm{I}}^{(n)}\right]}{1-\mathbb{E}\left[P_{\mathrm{II}}^{(n)}\right]} \qquad (5.36)\\
&= -H\left(\mathbb{E}\left[1-P_{\mathrm{I}}^{(n)}\right]\right) + \mathbb{E}\left[1-P_{\mathrm{I}}^{(n)}\right]\log\frac{1}{\mathbb{E}\left[P_{\mathrm{II}}^{(n)}\right]} + \mathbb{E}\left[P_{\mathrm{I}}^{(n)}\right]\log\frac{1}{1-\mathbb{E}\left[P_{\mathrm{II}}^{(n)}\right]} \\
&\geqslant \mathbb{E}\left[1-P_{\mathrm{I}}^{(n)}\right]\log\frac{1}{\mathbb{E}\left[P_{\mathrm{II}}^{(n)}\right]} - 1 \\
&\geqslant (1-\epsilon_H)\log\frac{1}{2^{-K(\epsilon_H)\sqrt{n}}} - 1 \\
&= (1-\epsilon_H)K(\epsilon_H)\sqrt{n} - 1 \qquad (5.37)\\
&= \Omega(\sqrt{n}).
\end{aligned}
$$

Here (5.36) follows from Jensen's inequality by the convexity of $f(\cdot,\cdot)$, and $K(\epsilon_H)$ is determined by the choice of ϵ_H (see (5.11)). Optimizing over choices of $\epsilon_H < \epsilon_S$ for the maximization of

$$(1-\epsilon_H)\sqrt{V(X|Y)}(Q^{-1}(\epsilon_S) - Q^{-1}(\epsilon_H))$$

gives a bound on the expected KL divergence. □

Alternative proof of Theorem 14

The alternative proof of Theorem 14 is inspired by [68]. It uses Lemma 19 which is a version of [68, Lemma 5] that includes side information. I state Lemma 19 here; the proof appears in Section 5.8.

Lemma 19 (Side Information Intrinsic Randomness)**.** *Let*

$$H_{\mathrm{SI}}(M, P_{XY}) \triangleq \sum_{(x,y):\imath(x|y)\leqslant\log M_n} P_{XY}(x,y)\log\frac{1}{P_{X|Y}(x|y)}.$$

Then

$$\begin{aligned}
&H(\mathsf{f}_n(X^n)|Y^n)\\
&\leqslant H_{\mathrm{SI}}(M_n, P_{XY}^n)\\
&\quad + \mathbb{P}\left[\imath(X^n|Y^n) \geqslant \log M_n\right]\left(\log M_n + \log\frac{1}{\mathbb{P}\left[\imath(X^n|Y^n) \geqslant \log M_n\right]}\right). \qquad (5.38)
\end{aligned}$$

Proof of Theorem 14. For any $\mathsf{f}_n : \mathcal{X}^n \mapsto [M_n]$,

$$\begin{aligned}
&D(P_{\mathsf{f}_n(X^n)Y^n}||P_{\mathsf{f}_n(X^n)}P_Y^n)\\
&= D(P_{\mathsf{f}_n(X^n)Y^n}||U_{M_n}P_Y^n)\\
&\quad - \sum_{c\in[M_n], y^n\in\mathcal{Y}^n} P_{\mathsf{f}_n(X^n)Y^n}(c, y^n)\log\frac{P_{\mathsf{f}_n(X^n)}(c)P_Y^n(y^n)}{U_{M_n}(c)P_Y^n(y^n)}\\
&= D(P_{\mathsf{f}_n(X^n)Y^n}||U_{M_n}P_Y^n) - D(P_{\mathsf{f}_n(X^n)}||U_{M_n}). \qquad (5.39)
\end{aligned}$$

For F_n, since $R_n = H(X|Y) + O(\frac{1}{\sqrt{n}})$ and $H(X|Y) < H(X)$, [67, Theorem 1] shows

$$\lim_{n\longrightarrow\infty}\mathbb{E}_{\mathsf{F}_n}D\left(P_{\mathsf{F}_n(X^n)}||U_{M_n}\right) = 0$$

by showing that the expected Tsallis divergence between $P_{\mathsf{F}_n(X^n)}$ and U_{M_n} converges to zero.

I use Lemma 19 to bound the first term in (5.39) from below.

Define cumulative distribution function Q_n on $\mathbb{R}$ as

$$Q_n(x) \triangleq \mathbb{P}\left[\frac{1}{n}\imath(X^n|Y^n) < H(X|Y) + \frac{x}{\sqrt{n}}\right] \qquad (5.40)$$

for distribution P_{XY}^n. Then,

$$H_{\mathrm{SI}}(M, P_{X|Y}^n) = \int_{-\infty}^{b_n}(\sqrt{n}x + nH(X|Y))Q_n(dx) \qquad (5.41)$$

holds, where $b_n \triangleq \frac{1}{\sqrt{n}}(\log M_n - nH(X|Y))$. Applying Lemma 19, by reasoning similar to that in [68, Theorem 8],

$$\begin{aligned}
&H(\mathsf{f}_n(X^n)|Y^n)\\
&\leqslant \sqrt{n}\int_{-\infty}^{b_n} xQ_n(dx) + nH(X|Y)\\
&\quad + \mathbb{P}\left[\imath(X^n|Y^n) \geqslant \log M_n\right]\left(\sqrt{n}b_n - \log\frac{1}{\mathbb{P}\left[\imath(X^n|Y^n) \geqslant \log M_n\right]}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\left(H\left(\mathsf{f}_n(X^n)|Y^n\right) - nH(X|Y)\right) \\
&\leqslant \int_{-\infty}^{b_n} xQ_n(dx) + \mathbb{P}\left[\imath(X^n|Y^n) \geqslant \log M_n\right] \\
&\qquad \times \left(b_n - \frac{1}{\sqrt{n}}\log\frac{1}{\mathbb{P}\left[\imath(X^n|Y^n) \geqslant \log M_n\right]}\right).
\end{aligned}$$

Taking the limit,

$$\lim_{n\longrightarrow\infty}\frac{1}{\sqrt{n}}\left(H\left(\mathsf{f}_n(X^n)|Y^n\right) - nH(X|Y)\right) \leqslant \int_{-\infty}^{b} xF(dx) + b(1-F(b))$$

where $b \triangleq \lim_{n\longrightarrow\infty} b_n$ and $F \triangleq \lim_{n\longrightarrow\infty} Q_n$. Then

$$\begin{aligned}
&\lim_{n\longrightarrow\infty}\frac{1}{\sqrt{n}}D(P_{\mathsf{f}_n(X^n)}P_Y^n||U_{M_n}P_Y^n) \\
&= \lim_{n\longrightarrow\infty}\frac{1}{\sqrt{n}}\left(\log M - nH(X|Y) - \left(H\left(\mathsf{f}_n(X^n)|Y^n\right) - nH(X|Y)\right)\right) \\
&\geqslant b - \int_{-\infty}^{b} xF(dx) - b + bF(b) \\
&= \int_{-\infty}^{b}(b-x)F(dx). \qquad (5.42)
\end{aligned}$$

For the distribution P_{XY}^n, setting $b = \sqrt{V(X|Y)}Q^{-1}(\epsilon)$, the right hand side of (5.42) equals

$$\begin{aligned}
&\sqrt{V(X|Y)}\int_{-\infty}^{Q^{-1}(\epsilon)}\frac{Q^{-1}(\epsilon)-x}{\sqrt{2\pi}}e^{-x^2/2}dx \\
&= \sqrt{V(X|Y)}\left(Q^{-1}(\epsilon)(1-\epsilon) + \frac{e^{-(Q^{-1}(\epsilon))^2/2}}{\sqrt{2\pi}}\right),
\end{aligned}$$

which is a function of the second-order rate term b and is independent of n. □

Remark 10. *The two proof techniques both show that the expected KL divergence is* $\Omega(\sqrt{n})$. *The method invoking Theorem 12 gives*

$$\mathbb{E}_{\mathsf{F}_n}\left[D\left(P_{\mathsf{F}_n(X^n)Y^n}||P_{\mathsf{F}_n(X^n)}P_Y^n\right)\right] \geqslant \exp\{K\sqrt{n}\} \qquad (5.43)$$

for any

$$K < K_{\mathrm{HT}}^{\star}(\epsilon_S) \triangleq \sup_{0<\epsilon_H<\epsilon_S}(1-\epsilon_H)\sqrt{V(X|Y)}\left(Q^{-1}(\epsilon_S) - Q^{-1}(\epsilon_H)\right)$$

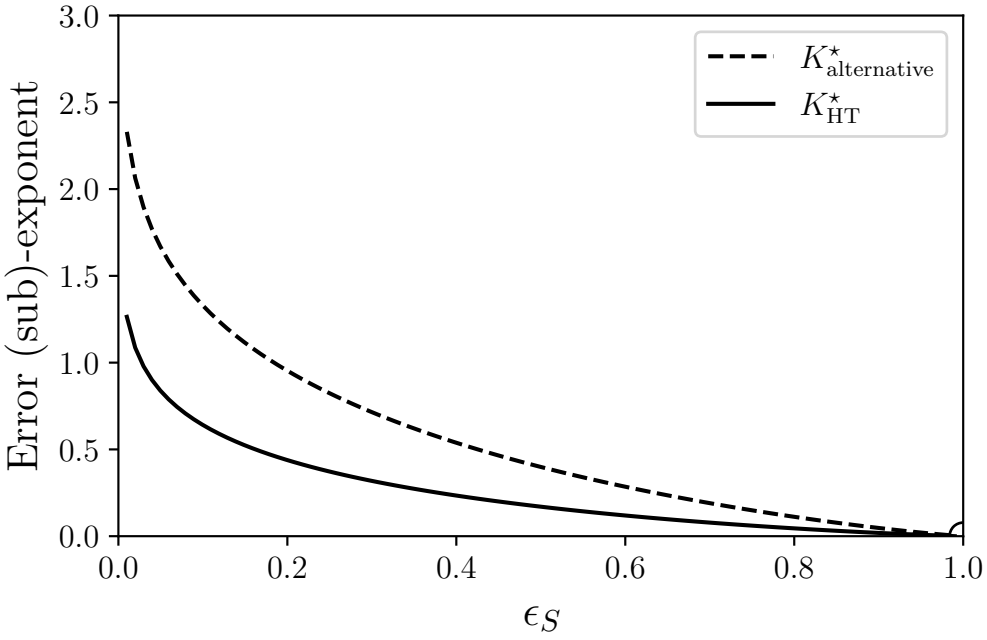


Figure 5.2: Different multipliers of $\sqrt{n}$ in the exponent of expected KL divergence bound from two different proofs of Theorem 14.

while the alternative proof gives a tighter constant, yielding (5.43) *for any*

$$K < K^\star_{\text{alternative}}(\epsilon_S) \triangleq \sqrt{V(X|Y)}\left(Q^{-1}(\epsilon_S)(1-\epsilon_S) + \frac{\exp\{-Q^{-1}(\epsilon_S)^2/2\}}{\sqrt{2\pi}}\right).$$

Here,

$$\epsilon_S = Q\left(\lim_{n\longrightarrow\infty}(R_n - H(X|Y))\sqrt{\frac{n}{V(X|Y)}}\right).$$

The advantage of the first approach, which derives Theorem 14 as an extension of Theorem 12 or Theorem 13, is that it lends insight into the behavior of a 1-shot hypothesis test that does not arise from only the studying of the KL-divergence directly (non-asymptotic behavior of hypothesis test relies on more information than the KL divergence, see, for example, [69]). Namely, it shows that under certain rate constraint, the random binning code yields codeword that can distinguish $P_{\mathsf{F}_n(X^n)Y^n}$ from $P_{\mathsf{F}_n(X^n)}P_Y^n$, which implies a lower bound on the KL-divergence.

Figure 5.2 plots the multiplier in the $\Omega(\sqrt{n})$ terms $K^\star_{\text{HT}}(\epsilon_S)$ and $K^\star_{\text{alternative}}(\epsilon_S)$, showing that the alternative method yields a tighter constant. Figure 5.3 shows how changing ϵ_H optimizes the multiplier of $\sqrt{n}$ when ϵ_S is fixed, with the supremum equaling $K^\star_{\text{HT}}(\epsilon_S)$.

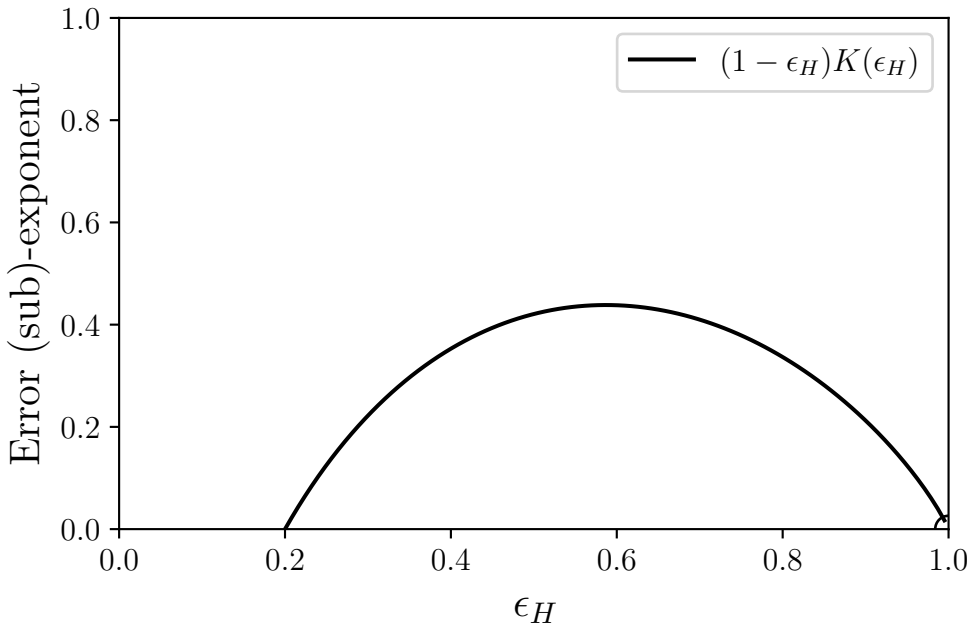


Figure 5.3: Optimization of the multiplier of $\sqrt{n}$ in the proof of Theorem 14 by the existence of a hypothesis test by changing ϵ_H when $\epsilon_S = 0.2$.

5.6 Distributed Hypothesis Testing with Identical Marginals

Just as the random-binning encoded description of X^n can be used to test the independence or dependence of (X^n, Y^n), it can also be used to distinguish between other pairs of distributions with identical marginals. That is, if $\mathbb{P}$ is defined as in Section 5.4 and $\mathbb{Q}$ captures the randomness of random binning and a new p.m.f. Q_{XY} on the same support $\mathcal{S}$ for which $Q_X = P_X, Q_Y = P_Y$, and $H_Q(X|Y) > H_P(X|Y)$, then Theorem 12 and Theorem 13 also hold. This extension is possible since

$$\mathbb{E}_{Q_{XY}}[\imath(X|Y)] = H_Q(X|Y) + D(Q_{XY}||P_{XY}) > H_P(X|Y)$$

which means that the proof of Theorem 13 continues to hold under the new definition of $\mathbb{Q}$. Theorem 14 becomes

Theorem 15 (Output Statistics of Random Binning)**.** *Given a sequence* $\{\mathsf{F}_n\}_{n=1}^{\infty}$ *of random binning encoders* $\mathsf{F}_n : \mathcal{X}^n \mapsto [\lfloor 2^{nR_n} \rfloor]$*, if*

$$R_n = H(X|Y) + O\left(\frac{1}{\sqrt{n}}\right),$$

then

$$\mathbb{E}_{\mathsf{F}}\left[D(P_{\mathsf{F}_n(X^n)Y^n}||Q_{\mathsf{F}_n(X^n)Y^n})\right] = \Omega(\sqrt{n}).$$

The proof method using the existence of hypothesis test stays identical. However, the alternative route needs to be updated. I present the proof of Theorem 15 below.

Proof of Theorem 15. Due to the symmetry among bins in the random binning code design, the expression of interest can be written as

$$\begin{aligned}\mathbb{E}_{\mathsf{F}} D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) &= \mathbb{E}_{\mathsf{F},Y^n} M_n P_{\mathsf{F}_n(X^n)Y^n}(1,Y^n) \log \frac{P_{\mathsf{F}_n(X^n)Y^n}(1,Y^n)}{Q_{\mathsf{F}_n(X^n)Y^n}(1,Y^n)} \\ &= \mathbb{E}_{\mathsf{F}} D\left(P_{\mathsf{F}_n(X^n)Y^n} || U_{M_n} P_Y^n\right) \\ &\quad - \mathbb{E}_{\mathsf{F},Y^n} M_n P_{\mathsf{F}_n(X^n)Y^n}(1,Y^n) \log \frac{Q_{\mathsf{F}_n(X^n)Y^n}(1,Y^n)}{U_{M_n}(1) P_Y^n(Y^n)}\end{aligned} \tag{5.44}$$

where the first term is shown to be $\Omega(\sqrt{n})$ by identical means as Theorem 14.

Let $\mathcal{T}$ be the set of all joint types of sequences on $\mathcal{X}^n \times \mathcal{Y}^n$. Let $\mathcal{T}(y^n)$ be the set of all joint types $T \in \mathcal{T}$ such that there exists an x^n, $(x^n, y^n) \in T$. Notice that

$$\begin{aligned}P_{\mathsf{F}_n(X^n)|Y_n=y^n}(1) &= \sum_{x^n \in \mathcal{X}^n} P_{X|Y}^n(x^n|y^n) \mathbf{1}\{\mathsf{F}_n(X^n) = 1\} \\ &= \sum_{T \in \mathcal{T}(y^n)} P_{X|Y}^n(T) \sum_{x^n : (x^n, y^n) \in T} \mathbf{1}\{\mathsf{F}_n(X^n) = 1\} \\ &= \sum_{T \in \mathcal{T}(y^n)} P_{X|Y}^n(T) B_T(y^n)\end{aligned}$$

where $B_T(y^n) \sim Binomial(\sum_{x^n} \mathbf{1}\{(x^n, y^n) \in T\}, 1/M_n)$. Then, the second term in

(5.44) equals

$$
\begin{aligned}
&\sum_{y^n \in \mathcal{Y}^n} P_Y^n(y^n) \mathbb{E}_{\mathsf{F}} M_n P_{\mathsf{F}_n(X^n)|Y^n=y^n}(1) \log M_n Q_{\mathsf{F}_n(X^n)|Y^n=y^n}(1) \\
&= \sum_{y^n \in \mathcal{Y}^n} P_Y^n(y^n) \mathbb{E}\Bigg[\sum_{T \in \mathcal{T}(y^n)} M_n P_{X|Y}^n(T) B_T(y^n) \\
&\qquad \cdot \log\left(M_n Q_{X|Y}^n(T) B_T(y^n) + \sum_{T' \in \mathcal{T}(y^n) \setminus \{T\}} M_n Q_{X|Y}^n(T') B_{T'}(y^n) \right) \Bigg] \\
&\leqslant \sum_{y^n \in \mathcal{Y}^n} P_Y^n(y^n) \sum_{T \in \mathcal{T}(y^n)} \mathbb{E}\Bigg[M_n P_{X|Y}^n(T) B_T(y^n) \\
&\qquad \cdot \log\left(\max\left(1, \sum_{T' \in \mathcal{T}(y^n) \setminus \{T\}} M_n Q_{X|Y}^n(T') B_{T'}(y^n) \right) \right) \Bigg] \\
&\quad + \sum_{y^n \in \mathcal{Y}^n} P_Y^n(y^n) \sum_{T \in \mathcal{T}(y^n)} \mathbb{E}\Bigg[M_n P_{X|Y}^n(T) B_T(y^n) \\
&\qquad \cdot \log\left(M_n Q_{X|Y}^n(T) B_T(y^n) + \max\left(1, \sum_{T' \in \mathcal{T}(y^n) \setminus \{T\}} M_n Q_{X|Y}^n(T') B_{T'}(y^n) \right) \right) \Bigg] \\
&\quad - \sum_{y^n \in \mathcal{Y}^n} P_Y^n(y^n) \sum_{T \in \mathcal{T}(y^n)} \mathbb{E}\Bigg[M_n P_{X|Y}^n(T) B_T(y^n) \\
&\qquad \cdot \log\left(\max\left(1, \sum_{T' \in \mathcal{T}(y^n) \setminus \{T\}} M_n Q_{X|Y}^n(T') B_{T'}(y^n) \right) \right) \Bigg] . \qquad (5.45)
\end{aligned}
$$

Due to the random binning strategy, the binning of sequences of type T is independent

of the binning of sequences of other types $T' \in \mathcal{T}\backslash\{T\}$. Therefore,

$$
\begin{aligned}
&\sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) \sum_{T\in\mathcal{T}(y^n)} \mathbb{E}\Bigg[M_n P_{X|Y}^n(T) B_T(y^n) \\
&\qquad\qquad \times \log\left(\max\left(1, \sum_{T'\in\mathcal{T}(y^n)\backslash\{T\}} M_n Q_{X|Y}^n(T') B_{T'}(y^n)\right)\right)\Bigg] \\
=&\sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) \sum_{T\in\mathcal{T}(y^n)} \mathbb{E}\left[M_n P_{X|Y}^n(T) B_T(y^n)\right] \\
&\qquad\qquad \times \mathbb{E}\left[\log\left(\max\left(1, \sum_{T'\in\mathcal{T}(y^n)\backslash\{T\}} M_n Q_{X|Y}^n(T') B_{T'}(y^n)\right)\right)\right] \qquad (5.46)\\
\leqslant&\sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) \sum_{T\in\mathcal{T}(y^n)} \mathbb{E}\left[M_n P_{X|Y}^n(T) B_T(y^n)\right] \\
&\qquad\qquad \times \mathbb{E}\left[\log\left(\max\left(1, \sum_{T'\in\mathcal{T}(y^n)} M_n Q_{X|Y}^n(T') B_{T'}(y^n)\right)\right)\right] \qquad (5.47)\\
=&\sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) \mathbb{E}\left[\log\left(\max\left(1, M_n Q_{\mathsf{F}_n(X^n)|Y^n=y^n}(1)\right)\right)\right] \\
&\qquad\qquad \times \sum_{T\in\mathcal{T}(y^n)} \mathbb{E}\left[M_n P_{X|Y}^n(T) B_T(y^n)\right] \qquad (5.48)\\
=&\sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) \mathbb{E}\left[\log\left(\max\left(1, M_n Q_{\mathsf{F}_n(X^n)|Y^n=y^n}(1)\right)\right)\right] \mathbb{E}\left[M_n P_{\mathsf{F}_n(X^n)|Y^n=y^n}(1)\right] \\
&\qquad\qquad (5.49)\\
=&\sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) \mathbb{E}\left[\log\left(\max\left(1, M_n Q_{\mathsf{F}_n(X^n)|Y^n=y^n}(1)\right)\right)\right] \qquad (5.50)\\
\leqslant&\sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) \log\left(\mathbb{E}\left[\max\left(1, M_n Q_{\mathsf{F}_n(X^n)|Y^n=y^n}(1)\right)\right]\right) \qquad (5.51)\\
\leqslant&\sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) \log\left(\mathbb{E}\left[1 + M_n Q_{\mathsf{F}_n(X^n)|Y^n=y^n}(1)\right]\right) \qquad (5.52)\\
=&\sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) \log\left(1 + \mathbb{E}\left[M_n Q_{\mathsf{F}_n(X^n)|Y^n=y^n}(1)\right]\right) \qquad (5.53)\\
\leqslant&\frac{1}{\ln 2} \sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) \mathbb{E}\left[M_n Q_{\mathsf{F}_n(X^n)|Y^n=y^n}(1)\right] \qquad (5.54)\\
=&\frac{1}{\ln 2}. \qquad (5.55)
\end{aligned}
$$

Here, (5.46) results from the independence among binning of different types, (5.47) holds because the set is relaxed, (5.48) and (5.49) rewrite the weighted sum of

binomials into the weight of the bin, (5.50) and (5.55) follow from the fact that the expected weight of a bin is $1/M_n$, (5.51) follows from the concavity of the logarithm and Jensen's inequality, (5.54) holds because the slope of the logarithm is bounded by $1/\ln 2$ from above when the variable is greater than 1.

To show that the difference between the last two terms in (5.45) is also small, I split $\mathcal{T}$ into two subsets. There exists constants $\gamma, \delta > 0$ such that

$$H_P(X|Y) - H_W(X|Y) - D(W_{X|Y}||P_{X|Y}|W_Y) - D(W_{XY}||Q_{XY}|W_Y) < -\delta \quad (5.56)$$

for all distributions W on the same support as P, Q such that $H_W(X|Y) > H_P(X|Y) - \gamma$. This is true because $P_{X|Y} \neq Q_{X|Y}$. Let $\mathcal{T}_1$ be the subset of $\mathcal{T}$ such that $H_T(X|Y) \geqslant H_P(X|Y) - \gamma$, where $H_T(X|Y)$ is the conditional entropy of the empirical distribution of type T. Let $\mathcal{T}_2 = \mathcal{T} \backslash \mathcal{T}_1$.

Since the derivative of the logarithm is bounded from above by $\frac{1}{\ln 2}$ when the variable is greater than or equal to 1,

$$\begin{aligned}
&\log\left(M_n Q^n_{X|Y}(T) B_T(y^n) + \max\left(1, \sum_{T' \in \mathcal{T}(y^n) \backslash \{T\}} M_n Q^n_{X|Y}(T') B_{T'}(y^n)\right)\right) \\
&\quad - \log\left(\max\left(1, \sum_{T' \in \mathcal{T}(y^n) \backslash \{T\}} M_n Q^n_{X|Y}(T') B_{T'}(y^n)\right)\right) && (5.57) \\
\leqslant &\frac{1}{\ln 2} M_n Q^n_{X|Y}(T) B_T(y^n). && (5.58)
\end{aligned}$$

Therefore, for a type $T \in \mathcal{T}(y^n) \cap \mathcal{T}_1$,

$$\begin{aligned}
&\mathbb{E}\Bigg[M_n P^n_{X|Y}(T) B_T(y^n) && (5.59) \\
&\quad \times \log\left(M_n Q^n_{X|Y}(T) B_T(y^n) + \max\left(1, \sum_{T' \in \mathcal{T}(y^n) \backslash \{T\}} M_n Q^n_{X|Y}(T') B_{T'}(y^n)\right)\right)\Bigg] \\
&\quad - \mathbb{E}\left[M_n P^n_{X|Y}(T) B_T(y^n) \log\left(\max\left(1, \sum_{T' \in \mathcal{T}(y^m) \backslash \{T\}} M_n Q^n_{X|Y}(T') B_{T'}(y^n)\right)\right)\right] && (5.60) \\
\leqslant &\frac{1}{\ln 2} \mathbb{E}\left[M_n P^n_{X|Y}(T) B_T(y^n) M_n Q^n_{X|Y}(T) B_T(y^n)\right] && (5.61) \\
= &\frac{1}{\ln 2} M_n^2 P^n_{X|Y}(T) Q^n_{X|Y}(T) \mathbb{E}\left[B_T(y^n)^2\right]. && (5.62)
\end{aligned}$$

The behavior of $P^n_{X|Y}(T)$ is

$$\log P^n_{X|Y}(T) \tag{5.63}$$
$$= \log \prod_{i=1}^{n} P_{X|Y}(x_{Ti}|y_{Ti}) \tag{5.64}$$
$$= \sum_{i=1}^{n} \log P_{X|Y}(x_{Ti}|y_{Ti}) \tag{5.65}$$
$$= \sum_{x,y} N_T(x,y) \log_{X|Y}(x|y) \tag{5.66}$$
$$= n \sum_{i=1}^{n} W^T_{XY}(x,y) \log P_{X|Y}(x|y) \tag{5.67}$$
$$= -n \left(\sum_{i=1}^{n} W^T_{XY}(x,y) \log \frac{W^T_{X|Y}(x|y)}{P_{X|Y}(x|y)} + \sum_{i=1}^{n} W^T_{XY}(x,y) \log \frac{1}{W^T_{X|Y}(x|y)} \right) \tag{5.68}$$
$$= -n \left(D(W^T_{X|Y}||P_{X|Y}|W^T_Y) + H_T(X|Y) \right) \tag{5.69}$$

where W^T_{XY} is the empirical distribution of type T and $N_T(x,y)$ is the number of occurrences of (x,y) in a sequence of type T. Identically,

$$\log Q^n_{X|Y}(T) = -n \left(D(W^T_{X|Y}||Q_{X|Y}|W^T_Y) + H_T(X|Y) \right). \tag{5.70}$$

The second moment of $B_T(y^n)$ is a function of the number of trials, which can be characterized to the first order in the exponent as

$$\sum_{x^n \in \mathcal{X}^n} \mathbf{1}\{(x^n, y^n) \in T\} = \frac{|T|}{|T_Y|} \doteq \frac{2^{nH_T(X,Y)}}{2^{nH_T(Y)}} = 2^{nH_T(X|Y)} \tag{5.71}$$

where T_Y is the type of $y^n \in \mathcal{Y}^n$. Hence,

$$\begin{aligned}
&M_n^2 P^n_{X|Y}(T) Q^n_{X|Y}(T) \mathbb{E}\left[B_T(y^n)^2\right] \\
&\doteq 2^{2nR} 2^{-n\left(D(W^T_{X|Y}||P_{X|Y}|W^T_Y)+H_T(X|Y)\right)} 2^{-n\left(D(W^T_{X|Y}||Q_{X|Y}|W^T_Y)+H_T(X|Y)\right)} \\
&\qquad\qquad \times \left(\frac{2^{2nH_T(X|Y)} + 2^{n(R+H_T(X|Y))}}{2^{2nR}} \right) \\
&\doteq 2^{-n(D(W^T_{X|Y}||P_{X|Y}|W^T_Y)+D(W^T_{X|Y}||Q_{X|Y}|W^T_Y)+2H_T(X|Y))} \left(2^{2nH_T(X|Y)} + 2^{n(H_T(X|Y)+R)} \right) \\
&\doteq 2^{-n(D(W^T_{X|Y}||P_{X|Y}|W^T_Y)+D(W^T_{X|Y}||Q_{X|Y}|W^T_Y)+H_T(X|Y)-\max(H_P(X|Y),H_T(X|Y)))} \\
&\leqslant 2^{-n\delta} \text{ in the exponent}
\end{aligned}$$

due to the definition of $\mathcal{T}_1$ and the setting that $R = H_P(X|Y) + O(1/\sqrt{n})$. The summation of $|\mathcal{T}_1 \cap \mathcal{T}(y^n)|$ such exponentially decaying terms converges to zero, because the number of joint types grows only polynomially with n.

For $T \in \mathcal{T}_2 \cap \mathcal{T}(y^n)$, the number of trials for $B_T(y^n)$, $2^{nH_T(X|Y)}$, is strictly smaller than M_n to the first order in the exponent, making $B_T(y^n) = 0$ significantly more probable than other outcomes. Specifically,

$$1 - \Pr(B_T(y^n) = 0) = 1 - \left(\frac{M_n - 1}{M_n}\right)^{|T|/|T_Y|} \doteq 2^{-n(H_P(X|Y) - H_T(X|Y))}. \tag{5.72}$$

Furthermore,

$$\Pr(B_T(y^n) = k+1) = \Pr(B_T(y^n) = k) \cdot \frac{1}{M_n - 1} \cdot \frac{|T|/|T_Y| - k - 1}{k+1} \tag{5.73}$$

$$< \Pr(B_T(y^n) = k) \frac{|T|/|T_Y|}{M_n} \tag{5.74}$$

$$\leqslant \Pr(B_T(y^n) = k) 2^{-n(R - H_T(X|Y))}. \tag{5.75}$$

Then,

$$
\begin{aligned}
&\mathbb{E}\Bigg[M_n P^n_{X|Y}(T) B_T(y^n) \\
&\qquad \times \log\left(M_n Q^n_{X|Y}(T) B_T(y^n) + \max\left(1, \sum_{T' \in \mathcal{T}(y^n)\setminus\{T\}} M_n Q^n_{X|Y}(T') B_{T'}(y^n)\right)\right)\Bigg] \\
&\quad - \mathbb{E}\left[M_n P^n_{X|Y}(T) B_T(y^n) \log\left(\max\left(1, \sum_{T' \in \mathcal{T}(y^n)\setminus\{T\}} M_n Q^n_{X|Y}(T') B_{T'}(y^n)\right)\right)\right] \\
&= \mathbb{E} \sum_{b_T \in [0:|T|/|T_Y|]} \Pr(B_T(y^n) = b_T) M_n P^n_{X|Y}(T) b_T \\
&\qquad \times \log\left(M_n Q^n_{X|Y}(T) b_T + \max\left(1, \sum_{T' \in \mathcal{T}(y^n)\setminus\{T\}} M_n Q^n_{X|Y}(T') B_{T'}(y^n)\right)\right) \\
&\quad - \mathbb{E} \sum_{b_T \in [0:|T|/|T_Y|]} \Pr(B_T(y^n) = b_T) M_n P^n_{X|Y}(T) b_T \\
&\qquad \times \log\left(\max\left(1, \sum_{T' \in \mathcal{T}(y^n)\setminus\{T\}} M_n Q^n_{X|Y}(T') B_{T'}(y^n)\right)\right) \\
&= \mathbb{E} \sum_{b_T \in [1:|T|/|T_Y|]} \Pr(B_T(y^n) = b_T) M_n P^n_{X|Y}(T) b_T \\
&\qquad \times \log\left(M_n Q^n_{X|Y}(T) b_T + \max\left(1, \sum_{T' \in \mathcal{T}(y^n)\setminus\{T\}} M_n Q^n_{X|Y}(T') B_{T'}(y^n)\right)\right) \\
&\quad - \sum_{b_T \in [1:|T|/|T_Y|]} \Pr(B_T(y^n) = b_T) M_n P^n_{X|Y}(T) b_T \\
&\qquad \times \log\left(\max\left(1, \sum_{T' \in \mathcal{T}(y^n)\setminus\{T\}} M_n Q^n_{X|Y}(T') B_{T'}(y^n)\right)\right). \qquad (5.76)
\end{aligned}
$$

Here, for each T,

$$\begin{aligned}
&\Pr(B_T(y^n)=1)M_nP^n_{X|Y}(T)\\
&\qquad\times\log\left(M_nQ^n_{X|Y}(T)+\max\left(1,\sum_{T'\in\mathcal{T}(y^n)\setminus\{T\}}M_nQ^n_{X|Y}(T')B_{T'}(y^n)\right)\right)\\
&\quad-\Pr(B_T(y^n)=1)M_nP^n_{X|Y}(T)\log\left(\max\left(1,\sum_{T'\in\mathcal{T}(y^n)\setminus\{T\}}M_nQ^n_{X|Y}(T')B_{T'}(y^n)\right)\right)\\
&\leqslant\Pr(B_T(y^n)=1)M_nP^n_{X|Y}(T)\max\left(\log M_nQ_{\mathsf{F}_n(X^n)|Y^n=y^n}(1)+1,1\right) &(5.77)\\
&\leqslant 2^{-n(R-H_T(X|Y))}2^{nR}2^{-n(D(W^T_{X|Y}||P_{X|Y}|W^T_Y)+H_T(X|Y))}(nR+1), &(5.78)
\end{aligned}$$

where (5.77) follows from the fact that $\log(a+b)-\log b\leqslant\max(\log a+1,1)$ when $b\geqslant 1$. The summation over $b_T\in[1:|T|/|T_Y|]$ is bounded from above by

$$\begin{aligned}
&\sum_{b_T=1}^{|T|/|T_Y|}\left(2^{-n(R-H_T(X|Y))}\right)^{b_T}2^{nR}2^{-n(D(W^T_{XY}||P_{XY})+H_T(X|Y))}b_T(nR+1) &(5.79)\\
&\leqslant 2^{-nD(W^T_{X|Y}||P_{X|Y}|W^T_Y)}(nR+1)\sum_{b_T=1}^{\infty}\left(2^{-n(R-H_T(X|Y))}\right)^{b_T-1}b_T &(5.80)\\
&= 2^{-nD(W^T_{X|Y}||P_{X|Y}|W^T_Y)}(nR+1)\frac{1}{\left(1-2^{-n(R-H_T(X|Y))}\right)^2} &(5.81)\\
&\leqslant 2^{-nD(W^T_{X|Y}||P_{X|Y}|W^T_Y)}(2nR+2) &(5.82)\\
&\doteq 2^{-nD(W^T_{X|Y}||P_{X|Y}|W^T_Y)}. &(5.83)
\end{aligned}$$

When y^n is strongly ϵ-typical [52] with respect to P_Y for a small enough constant $\epsilon>0$, $H_T(X|Y)<H_P(X|Y)-\gamma$ implies that (5.83) decays exponentially in n. Combining this with the observation that the number of types in $\mathcal{T}_2\cap\mathcal{T}(y^n)$ grows only polynomially with n, the sum across $\mathcal{T}_2\cap\mathcal{T}(y^n)$ also converges to 0, so does the weighted average across y^n. If y^n is strongly atypical with respect to P_Y, then (5.83) can grow linearly with n, however, the probability of Y^n being atypical with respect to P_Y decays exponentially, beating the growth. Therefore, the difference between the last two terms in (5.45) converges to 0. Combining this with the fact that the first term is bounded from above by a constant,

$$\begin{aligned}
\mathbb{E}_{\mathsf{F}}D\left(P_{\mathsf{F}_n(X^n)Y^n}||Q_{\mathsf{F}_n(X^n)Y^n}\right)&=\mathbb{E}_{\mathsf{F}}D\left(P_{\mathsf{F}_n(X^n)Y^n}||U_{M_n}P^n_Y\right)-O(1) &(5.84)\\
&=\Omega(\sqrt{n})-O(1) &(5.85)\\
&=\Omega(\sqrt{n}), &(5.86)
\end{aligned}$$

completing the proof. □

Remark 11. *The above proof shows that $P_{\mathsf{F}_n(X^n)Y^n}$ and $Q_{\mathsf{F}_n(X^n)Y^n}$ are expected to be distinguishable when $R_n = H_P(X|Y) + O(1/\sqrt{n})$. Intuitively, this is due to the fact that the typical sets under P and Q become disjoint as n increases, and the binning operations of the two sets are done independently. The effect of P and Q being skewed in the same (or different) direction is negligible in the long run.*

In-distinguishability due to over-compression

Intuitively, the better we do compressing samples from a random process, the more the encoder output should look like i.i.d. uniform random bits independent even of the decoder side information; anything else could be compressed further. We therefore expect recognizable traces of a source sequence's identity to persist only when redundancy remains, which in a code for source $\{X_i\}_{i=1}^{\infty}$ with decoder side information $\{Y_i\}_{i=1}^{\infty}$ should require $R > H(X|Y)$. Lemma 20 proves part of this intuition to be correct, showing that when $R < H(X|Y)$, the underlying distribution of $\{X_i\}_{i=1}^{\infty}$ satisfying $H(X|Y) > R$ is indistinguishable from any other p.m.f. with the same marginal under the same constraint.

Definition 23. *Fix distributions P_X on finite alphabet $\mathcal{X}$ and P_Y on finite alphabet $\mathcal{Y}$. Let $\mathcal{P}$ be all distributions P_{XY} on finite support $\mathcal{S} \subseteq \mathcal{X} \times \mathcal{Y}$ with marginal distribution P_Y. Define $\mathcal{P}(R) \subseteq \mathcal{P}$ as*

$$\mathcal{P}(R) \triangleq \{P_{XY} \in \mathcal{P} : H_P(X|Y) \geqslant R\}. \tag{5.87}$$

Lemma 20. *Define a sequence $\{\mathsf{F}_n\}_{n=1}^{\infty}$ of source encoders, where for each n, $\mathsf{F}_n : \mathcal{X}^n \mapsto [\lfloor 2^{nR_n} \rfloor]$ is a random binning encoder. Then, for any $\{R_n\}$ with $\lim_{n\to\infty} R_n = R$ and $\delta, b > 0$,*

$$\mathbb{P}\left[\sup_{P_{XY}, Q_{XY} \in \mathcal{P}(R+\delta)} D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) \geqslant b\right] \to 0.$$

Remark 12. *Distributions P and Q have identical support, which ensures that $D(P_{\mathsf{f}_n(X^n)Y^n} || Q_{\mathsf{f}_n(X^n)Y^n})$ is well-defined for any encoder f_n.*

The proof of Lemma 20 uses a quantization method that partitions $\mathcal{P}(R + \delta)$ into a number of subsets that grows polynomially in n. Like many arguments for universality (for example, [8]), this method relies on the fact that the exponential decay of the KL-divergence between compressed quantized distributions always dominates the polynomial growth of the number of quantized distributions in the long run. The proof is in Section 5.8.

5.7 Summary

In a source code with dependent side information at the decoder, the random binning code can serve well both for compression and for a distributed hypothesis test seeking to distinguish between joint distributions of the distributed inputs. When the compression rate converges to its asymptotic optimum at a rate of $O(1/\sqrt{n})$, the type-II error of the distributed hypothesis test decays subexponentially at $\exp\{-O(\sqrt{n})\}$. This property implies a lower bound on the KL-divergence that describes the distribution of the codeword from random binning. When over-compressed, however, the codeword from random binning is expected to look indistinguishable for different distributions.

5.8 Proofs of Auxiliary Results

Proof of Lemma 19

Proof. I begin by bounding $H(\mathsf{f}_n(X^n)|Y^n)$ as

$$\begin{aligned}
&H(\mathsf{f}_n(X^n)|Y^n)\\
&= \sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) H(\mathsf{f}_n(X^n)|Y^n = y^n)\\
&\leqslant \sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) \Bigg(H(M_n, P_{X^n|Y^n=y^n})\\
&\qquad + P_{X^n|Y^n=y^n}\left[\imath(X^n|y^n) \geqslant \log M_n\right] \qquad (5.88)\\
&\qquad \times \left(\log M_n + \log \frac{1}{P_{X^n|Y^n=y^n}\left[\imath(X^n|y^n) \geqslant \log M_n\right]}\right)\Bigg).
\end{aligned}$$

The inequality follows from [68, Lemma 5] by treating the side-information code under the condition where $Y^n = y^n$ as a point-to-point code for a source drawn from probability distribution $P_{X^n|Y^n=y^n}$. Next, note that

$$\begin{aligned}
&\sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) H(M_n, P_{X^n|Y^n=y^n})\\
&= \sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) \sum_{x^n:\imath(x^n|y^n)\leqslant\log M_n} P_{X^n|Y^n=y^n}(x^n) \log \frac{1}{P_{X^n|Y^n=y^n}(x^n)}\\
&= \sum_{(x^n,y^n):\imath(x^n|y^n)\leqslant\log M_n} P_{XY}^n(x^n,y^n) \log \frac{1}{P_{X|Y}^n(x^n|y^n)}\\
&= H_{\mathrm{SI}}(M_n, P_{XY}^n). \qquad (5.89)
\end{aligned}$$

For $y^n \in \mathcal{Y}^n$, let

$$a = a(y^n) = P_{X^n|Y^n=y^n}\left[\imath(X^n|y^n) \geqslant \log M_n\right].$$

When $Y^n \in \mathcal{Y}^n$ is random, this function is also random, i.e.,

$$A = A(Y^n) = P_{X^n|Y^n}\left[\imath(X^n|Y^n) \geqslant \log M_n\right]. \tag{5.90}$$

Then, by Jensen's inequality, $\mathbb{E}[f(A)] \leqslant f(\mathbb{E}[A])$ since $f(a) = a\log\frac{1}{a}$ is strictly concave in a. Therefore,

$$\begin{aligned}
&\sum_{y^n\in\mathcal{Y}^n} P_Y^n(y^n) P_{X^n|Y^n=y^n}\left[\imath(X^n|y^n) \geqslant \log M_n\right] \\
&\qquad \times \log \frac{1}{P_{X^n|Y^n=y^n}\left[\imath(X^n|y^n) \geqslant \log M_n\right]} \\
&\leqslant \mathbb{P}\left[\imath(X^n|Y^n) \geqslant \log M_n\right] \log \frac{1}{\mathbb{P}\left[\imath(X^n|Y^n) \geqslant \log M_n\right]}.
\end{aligned} \tag{5.91}$$

Combining (5.88), (5.89) and (5.91) completes the proof. □

Proof of Lemma 20

The proof appears below after some related definitions and results.

Definition 24 (ζ-Quantization)**.** *Given a p.m.f. P on finite alphabet $\mathcal{X}$ and a constant $\zeta > 0$, the ζ-quantization of P is defined as*

$$\hat{P}_\zeta(x) \triangleq \begin{cases} \zeta\left\lceil P(x)/\zeta\right\rceil, & x \in \mathcal{X}\backslash\{x^*\} \\ 1 - \sum_{x'\in\mathcal{X}\backslash\{x^*\}} \hat{P}_\zeta(x'), & \textit{otherwise,} \end{cases}$$

where $\lceil\cdot\rceil$ takes the ceiling and $x^* \in \mathcal{X}$ is the realization with largest probability under P. If there is a tie, x^* is chosen to be the largest symbol lexicographically.

For a sufficiently small ζ, $\hat{P}_\zeta(x) > 0$ and therefore $\hat{P}_\zeta(\cdot)$ is a p.m.f..

The following lemmas are useful for the proof of Lemma 20. To begin with, Lemma 21 examines the impact of slight differences in distributions on KL divergence.

Lemma 21. *Let P, Q, P', Q' be p.m.f.s on the same support $\mathcal{X}$. If*

$$\frac{P(x)}{P'(x)}, \frac{Q(x)}{Q'(x)} \in [1-\delta, 1+\delta] \tag{5.92}$$

for all $x \in \mathcal{X}$ and some $0 < \delta < 1$, then

$$|D(P||Q) - D(P'||Q')| \leqslant \delta\left(D(P'||Q') + \frac{2}{e\ln 2}\right) + \log\frac{1+\delta}{1-\delta}. \tag{5.93}$$

Proof.

$$D(P||Q) = \sum_{x\in\mathcal{X}} P(x)\log\frac{P(x)}{Q(x)} \tag{5.94}$$

$$= \sum_{x\in\mathcal{X}} P(x)\log\frac{P'(x)}{Q'(x)} - \sum_{x\in\mathcal{X}} P(x)\log\frac{P'(x)/P(x)}{Q'(x)/Q(x)}. \tag{5.95}$$

Since

$$\frac{1-\delta}{1+\delta} \leqslant \frac{P'(x)/P(x)}{Q'(x)/Q(x)} \leqslant \frac{1+\delta}{1-\delta}, \tag{5.96}$$

the second term in (5.95) satisfies

$$\left|\sum_{x\in\mathcal{X}} P(x)\log\frac{P'(x)/P(x)}{Q'(x)/Q(x)}\right| \leqslant \left|\sum_{x\in\mathcal{X}} P(x)\log\frac{1+\delta}{1-\delta}\right| = \log\frac{1+\delta}{1-\delta}. \tag{5.97}$$

For the first term in (5.95),

$$\sum_{x\in\mathcal{X}} P(x)\log\frac{P'(x)}{Q'(x)} = D(P'||Q') - \sum_{x\in\mathcal{X}} (P'(x)-P(x))\log\frac{P'(x)}{Q'(x)}. \tag{5.98}$$

Here

$$\left|\sum_{x\in\mathcal{X}} (P'(x)-P(x))\log\frac{P'(x)}{Q'(x)}\right| \tag{5.99}$$

$$\leqslant \sum_{x\in\mathcal{X}} |P(x)-P'(x)| \left|\log\frac{P'(x)}{Q'(x)}\right| \tag{5.100}$$

$$\leqslant \delta \sum_{x\in\mathcal{X}} P'(x) \left|\log\frac{P'(x)}{Q'(x)}\right| \tag{5.101}$$

$$= \delta\left(\sum_{x:P'(x)\geqslant Q'(x)} P'(x)\log\frac{P'(x)}{Q'(x)} - \sum_{x:P'(x)<Q'(x)} P'(x)\log\frac{P'(x)}{Q'(x)}\right) \tag{5.102}$$

$$= \delta\left(D(P'||Q') - 2\sum_{x:P'(x)<Q'(x)} Q'(x)\frac{P'(x)}{Q'(x)}\log\frac{P'(x)}{Q'(x)}\right) \tag{5.103}$$

$$\leqslant \delta\left(D(P'||Q') + 2\sum_{x:P'(x)<Q'(x)} Q'(x)\frac{1}{e\ln 2}\right) \tag{5.104}$$

$$\leqslant \delta\left(D(P'||Q') + \frac{2}{e\ln 2}\right), \tag{5.105}$$

where (5.104) follows from the fact that $-a \log a \leqslant \frac{1}{e \ln 2}$ for $a \in [0, 1]$. Combining (5.95), (5.97), (5.98) and (5.105) gives

$$|D(P||Q) - D(P'||Q')| \leqslant \left| D(P'||Q') - \sum_{x \in \mathcal{X}} P(x) \log \frac{P'(x)}{Q'(x)} \right| + \log \frac{1+\delta}{1-\delta} \quad (5.106)$$

$$= \left| \sum_{x \in \mathcal{X}} (P'(x) - P(x)) \log \frac{P'(x)}{Q'(x)} \right| + \log \frac{1+\delta}{1-\delta} \quad (5.107)$$

$$\leqslant \delta \left(D(P'||Q') + \frac{2}{e \ln 2} \right) + \log \frac{1+\delta}{1-\delta}. \quad (5.108)$$

□

The following lemma is a widget to prove the lemma that succeeds it.

Lemma 22 (Convergence)**.** *If* $\mathbb{E}|X_n| \leqslant \exp\{-kn\}$ *for some* $0 < k < \infty$ *and* $|X_n| \leqslant \exp\{k'n\}$ *for some* $0 < k' < \infty$*, then there exists a constant* $p > 1$ *such that* $X_n \xrightarrow{L_p} 0$.

Proof. Write

$$\mathbb{E}|X_n|^p = \mathbb{E}\left[|X_n|^p \mathbf{1}\{|X_n| \leqslant \exp\{-kn/2\}\}\right] + \mathbb{E}\left[|X_n|^p \mathbf{1}\{|X_n| > \exp\{-kn/2\}\}\right].$$

The first term

$$\mathbb{E}\left[|X_n|^p \mathbf{1}\{|X_n| \leqslant \exp\{-kn/2\}\}\right] \leqslant \left(e^{-kn/2}\right)^p = e^{-kpn/2} \to 0.$$

The second term

$$\begin{aligned}
\mathbb{E}\left[|X_n|^p \mathbf{1}\{|X_n| > \exp\{-kn/2\}\}\right] &= \mathbb{E}\left[|X_n|^{(p-1)} |X_n| \mathbf{1}\{|X_n| > \exp\{-kn/2\}\}\right] \\
&= |x_{n,\max}|^{(p-1)} \mathbb{E}[|X_n| \mathbf{1}\{|X_n| > \exp\{-kn/2\}\}] \\
&\leqslant \left(e^{k'n}\right)^{(p-1)} e^{-kn} \\
&= e^{-(k-k'(p-1))n}. \qquad (5.109)
\end{aligned}$$

Since (k, k') are constants, one can always find p sufficiently close to 1 so that (5.109) decays to 0. Combining the two terms completes the proof. □

The following lemma studies the behavior of the expected KL divergence between a pair of overcompressed distributions using the random binning code.

Lemma 23 (Indistinguishable Agents)**.** *Let $P_{XY}, Q_{XY} \in \mathcal{P}(R+\alpha)$ for some $\alpha > 0$. For positive integer n, let $\mathsf{F}_n : \mathcal{X}^n \mapsto [\lfloor 2^{nR} \rfloor]$ be the random binning encoder at rate R. Then, for sufficiently large n,*

$$\mathbb{E}_{\mathsf{F}_n} D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) \leqslant \exp\{-\xi(\alpha)n\}, \tag{5.110}$$

where $\xi(\cdot)$ is a positive and deterministic function of α.

Proof. Denote $M_n = \lfloor 2^{nR} \rfloor$. Let P_{XY} be a p.m.f. on support $\mathcal{S} \subseteq \mathcal{X} \times \mathcal{Y}$ that satisfies $H_P(X|Y) - R \geqslant \alpha > 0$. The total variation distance between $P_{\mathsf{F}_n(X^n)Y^n}$ and $U_{M_n} P_Y^n$ converges to zero [65, Theorem 1], and the convergence speed is exponential [66, Theorem 5]. From the proof of [66, Theorem 5], the exponent is a function of α. That is, for sufficiently large n,

$$\left\| P_{\mathsf{F}_n(X^n)Y^n} - U_{M_n} P_Y^n \right\|_{\mathrm{TV}} \leqslant \exp\left\{-\kappa(\alpha)n\right\}. \tag{5.111}$$

The same is true for Q.

I split the expectation of interest into two terms.

$$\begin{aligned}
&\mathbb{E}_{\mathsf{F}_n} D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) \\
=&\mathbb{E}_{\mathsf{F}_n} \sum_{c \in [M_n]} \sum_{y^n \in \mathcal{Y}^n} P_{\mathsf{F}_n(X^n)Y^n}(c, y^n) \log \frac{P_{\mathsf{F}_n(X^n)Y^n}(c, y^n)}{Q_{\mathsf{F}_n(X^n)Y^n}(c, y^n)} \\
=&M_n \mathbb{E}_{\mathsf{F}_n} \sum_{y^n \in \mathcal{Y}^n} P_{\mathsf{F}_n(X^n)Y^n}(1, y^n) \log \frac{P_{\mathsf{F}_n(X^n)Y^n}(1, y^n)}{Q_{\mathsf{F}_n(X^n)Y^n}(1, y^n)} \\
=&M_n \mathbb{E}_{\mathsf{F}_n} \sum_{y^n \in \mathcal{Y}^n} P_Y^n(y^n) \left(P_{\mathsf{F}_n(X^n)|Y^n}(1|y^n) - U_{M_n}(1)\right) \log \frac{P_{\mathsf{F}_n(X^n)|Y^n}(1|y^n)}{Q_{\mathsf{F}_n(X^n)|Y^n}(1|y^n)} \\
&\quad + \mathbb{E}_{\mathsf{F}_n} \sum_{y^n \in \mathcal{Y}^n} P_Y^n(y^n) \log \frac{P_{\mathsf{F}_n(X^n)|Y^n}(1|y^n)}{Q_{\mathsf{F}_n(X^n)|Y^n}(1|y^n)}.
\end{aligned} \tag{5.112}$$

The second term can be written as

$$\exp\{-\kappa n\} \mathbb{E}_{\mathsf{F}_n, Y^n} \exp\{\kappa n\} \log \frac{P_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n)}{Q_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n)}, \tag{5.113}$$

where the constant κ is set later in this argument. The two random variables in the logarithm, $P_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n)$ and $Q_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n)$, must be either both zero or both nonzero. By the convention $0 \log \frac{0}{0} = 0$, (5.113) equals

$$\exp\{-\kappa n\} \mathbb{E}_{\mathsf{F}_n, Y^n} \exp\{\kappa n\} \log \frac{P^*_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n)}{Q^*_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n)}, \tag{5.114}$$

where

$$P^*_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n) = \begin{cases} P_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n), & \text{if } P_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n) \neq 0 \\ \frac{1}{M_n} & \text{otherwise,} \end{cases} \tag{5.115}$$

the same holds for Q. This construction ensures that

$$\begin{aligned} \mathbb{E}_{\mathsf{F}_n}\left\|P^*_{\mathsf{F}_n(X^n)Y^n} - U_{M_n}P^n_Y\right\|_{\mathrm{TV}} &\leqslant \mathbb{E}_{\mathsf{F}_n}\left\|P_{\mathsf{F}_n(X^n)Y^n} - U_{M_n}P^n_Y\right\|_{\mathrm{TV}} \\ &\leqslant \exp\{-\kappa(\alpha)n\}, \end{aligned} \tag{5.116}$$

the same holds for Q, where the second inequality comes from [66, Theorem 5].

Then,

$$\begin{aligned} &\mathbb{E}_{\mathsf{F}_n,Y^n}\left[\left|\frac{P^*_{\mathsf{F}_n(X^n)Y^n}(1|Y^n)}{U_{M_n}(1)} - 1\right|\right] \\ =&\mathbb{E}_{\mathsf{F}_n}\sum_{y^n\in\mathcal{Y}^n} P^n_Y(y^n)M_n\left|P^*_{\mathsf{F}_n(X^n)|Y^n}(1|y^n) - U_{M_n}(1)\right| \\ =&\mathbb{E}_{\mathsf{F}_n}\sum_{y^n\in\mathcal{Y}^n} P^n_Y(y^n)\sum_{c\in[M_n]}\left|P^*_{\mathsf{F}_n(X^n)|Y^n}(c|y^n) - U_{M_n}(c)\right| \qquad (5.117) \\ =&2\mathbb{E}_{\mathsf{F}_n}\left\|P^*_{\mathsf{F}_n(X^n)Y^n} - U_{M_n}P^n_Y\right\|_{\mathrm{TV}} \\ \leqslant&2\exp\left\{-\kappa(\alpha)n\right\}, \end{aligned}$$

where (5.117) follows from the symmetry among codewords under the random binning code construction. Therefore, the expectation of both $\frac{P^*_{\mathsf{F}_n(X^n)Y^n}(1|Y^n)}{U_{M_n}(1)}$ and $\frac{Q^*_{\mathsf{F}_n(X^n)Y^n}(1|Y^n)}{U_{M_n}(1)}$ converge to 1 exponentially quickly. By Markov's inequality, they converge in probability with exponentially light tails. Combining this with the observation that $|\log P^*_{\mathsf{F}_n(X^n)Y^n}(1|Y^n)|$ and $|\log Q^*_{\mathsf{F}_n(X^n)Y^n}(1|Y^n)|$ can both be at most linearly growing with n, one can conclude that the expectation of $\log\frac{P^*_{\mathsf{F}_n(X^n)Y^n}(1|Y^n)}{Q^*_{\mathsf{F}_n(X^n)Y^n}(1|Y^n)}$ converges to 0 exponentially quickly with any exponent below $\kappa(\alpha)$. Choosing $\kappa = \xi'(\alpha) \triangleq \kappa(\alpha)/2$ makes

$$\mathbb{E}_{\mathsf{F}_n,Y^n}\exp\{\kappa n\}\log\frac{P_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n)}{Q_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n)} \to 0.$$

That is, the second term in (5.112) converges to 0 exponentially quickly.

The first term in (5.112) can be written as $\mathbb{E}_{\mathsf{F},Y^n} V_n W_n$ where

$$V_n \triangleq M_n P^*_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n) - 1$$
$$W_n \triangleq \log \frac{P^*_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n)}{Q^*_{\mathsf{F}_n(X^n)|Y^n}(1|Y^n)}.$$

Notice that

$$\begin{aligned}\mathbb{E}_{\mathsf{F}_n,Y^n}|V_n| &= 2\mathbb{E}_{\mathsf{F}_n}\left\|P^*_{\mathsf{F}_n(X^n)Y^n} - U_{M_n}P^n_Y\right\|_{\mathrm{TV}} \\ &\leqslant 2\exp\{-\kappa(\alpha)n\} \\ &= 2\exp\{-\xi'(\alpha)n\}\cdot\exp\{-\xi'(\alpha)n\}.\end{aligned}$$

Therefore, letting $V'_n \triangleq \exp\{\xi'(\alpha)n\}V_n/2$, $\mathbb{E}_{\mathsf{F}_n,Y^n}|V'_n| \leqslant \exp\{-\xi'(\alpha)n\}$. Also, $|V'_n| \leqslant M_n \exp\{\xi'(\alpha)n\}/2$, which is exponentially large. Applying Lemma 22, there exists a $p > 0$ such that $V'_n \xrightarrow{p} 0$. As for W_n, due to the logarithm, $|W_n|$ is at most linearly growing with n, yet it converges exponentially quickly. Therefore, $W_n \xrightarrow{q} 0$ for all $q < \infty$. Choosing q so that $\frac{1}{p} + \frac{1}{q} = 1$, Hölder's inequality [70] gives L_1 convergence of $V'_n W_n$. Therefore, for sufficiently large n,

$$\mathbb{E}_{\mathsf{F}_n,Y^n} V_n W_n = 2\exp\{-\xi'(\alpha)\mathbb{E}_{\mathsf{F}_n,Y^n} V'_n W_n \leqslant 2\exp\{-\xi'(\alpha)n\}.$$

Therefore, both terms in (5.112) decay exponentially quickly with the same exponent $\xi'(\alpha)$, picking any $\xi(\alpha) < \xi'(\alpha)$ uniformly completes the proof. □

Remark 13. *In [67], Kavian et al. briefly mentioned that the enhancement of the output statistics of random binning (OSRB) result in [65] provided by [66] can be used to argue that the random binning statistics, with sufficiently low rate, is close to uniformity when measured by KL-divergence. Part of the proof of Lemma 23 can be used to formalize that claim.*

With these auxiliary results, I proceed to prove Lemma 20.

Proof of Lemma 20. The set $\mathcal{P}(R+\delta)$ can be uncountable. I first define a sequence of sets, each of which contains a countable number of elements. Let

$$\mathcal{P}^\star_\zeta \triangleq \left\{\hat{P}_{XY} : \exists P_{XY} \in \mathcal{P}(R+\delta) \text{ with } \hat{P}_{X|Y=y,\zeta} = \hat{P}_{X|Y=y} \text{ for all } y \in \mathcal{Y}, \text{ and } \hat{P}_Y = P_Y\right\}$$

be the set of distributions that are quantized from $\mathcal{P}(R+\delta)$ with ζ-quantized conditionals, that is, they have identical marginal P_Y with the distributions in $\mathcal{P}(R+$

δ), but the conditionals $P_{X|Y=y}, \forall y \in \mathcal{Y}$ are ζ-quantized. Use quantizer $\zeta_n = \frac{1}{n^2}$. The quantization ensures that the cardinality of $\mathcal{P}^*_{\zeta_n}$ is countable and grows with n polynomially. Specifically,

$$\left|\mathcal{P}^\star_{\zeta_n}\right| \leqslant (n^2+1)^{(|\mathcal{X}|-1)|\mathcal{Y}|} = O\left(n^{2|\mathcal{X}||\mathcal{Y}|}\right).$$

This is because $P_{X|Y=y}$ for each $y \in \mathcal{Y}$ is quantized differently, and for each specific y, $\hat{P}_{X|Y=y,\zeta_n}(x)$ for each x except the one with largest $P_{X|Y=y}(x)$ can take no more than n^2+1 possible values.

Let $H(X|Y)$ be the conditional entropy with respect to $P_{XY} \in \mathcal{P}(R+\delta)$. Let $\hat{H}(X|Y)$ be the conditional entropy with respect to $\hat{P}_{n,XY}$, where

$$\hat{P}_{n,XY}(x,y) = \hat{P}_{X|Y=y,\zeta_n}(x|y)P_Y(y) \tag{5.118}$$

for all $(x,y) \in \mathcal{S}$. Then, since the support is finite, $\hat{H}(X|Y) = H(X|Y) \pm O\left(\frac{1}{n^2}\right)$, which, given (5.87), guarantees

$$\lim_{n\longrightarrow\infty} \min_{P_{XY}\in\mathcal{P}^\star_{\zeta_n}} H(X|Y) \geqslant R+\alpha \tag{5.119}$$

for any $\alpha < \delta$.

For $P_{XY}, Q_{XY} \in \mathcal{P}^\star_{\zeta_n}$ with same marginal P_Y, $\min_{P_{XY}\in\mathcal{P}^\star_{\zeta_n}} H(X|Y) - R \geqslant \alpha > 0$ implies that

$$\mathbb{E}_{\mathsf{F}_n} D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) \leqslant \exp\{-\xi(\alpha)n\} \tag{5.120}$$

by Lemma 23. Therefore, given $a_0 = \exp\{-a^*n\}$, $0 < a^* < \xi(\alpha)$, and $0 < \xi^\star < \xi(\alpha) - a^*$,

$$\begin{aligned}
&\mathbb{P}\left[\bigcap_{P_{XY},Q_{XY}\in\mathcal{P}^\star_{\zeta_n}} \left\{D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) \leqslant a_0\right\}\right] \\
&= 1 - \mathbb{P}\left[\bigcup_{P_{XY},Q_{XY}\in\mathcal{P}^\star_{\zeta_n}} \left\{D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) > a_0\right\}\right] \\
&\geqslant 1 - \sum_{P_{XY},Q_{XY}\in\mathcal{P}^\star_{\zeta_n}} \mathbb{P}\left[D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) > a_0\right] \\
&\geqslant 1 - \sum_{P_{XY},Q_{XY}\in\mathcal{P}^\star_{\zeta_n}} \frac{\mathbb{E}_{\mathsf{F}_n} D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right)}{a_0} \\
&\geqslant 1 - \sum_{P_{XY},Q_{XY}\in\mathcal{P}^\star_{\zeta_n}} \frac{\exp\{-\xi(\alpha)n\}}{a_0} \\
&\geqslant 1 - \exp\{-\xi^\star n\}
\end{aligned}$$

for sufficiently large n. That is, with probability approaching 1 exponentially quickly, compressed versions of distributions in the quantized set $\mathcal{P}^{\star}_{\zeta_n}$ have exponentially small KL divergence between one-another.

Next, I show how this behavior of the countable, quantized set $\mathcal{P}^{\star}_{\zeta_n}$ determines the behavior of the uncountable set $\mathcal{P}(R+\delta)$.

By the definition of the ζ-quantization method, for small enough ζ, for any p.m.f. $P(x)$ on support $\mathcal{S}$,

$$\max_{x\in\mathcal{S}} \frac{P(x)}{\hat{P}_{\zeta}(x)} \leqslant \max_{x\in\mathcal{S}} \frac{P(x)}{P(x)-(|\mathcal{X}|-1)\zeta} \leqslant \frac{p_{\min}}{p_{\min}-(|\mathcal{X}|-1)\zeta} \tag{5.121}$$

and

$$\min_{x\in\mathcal{S}} \frac{P(x)}{\hat{P}_{\zeta}(x)} \geqslant \min_{x\in\mathcal{S}} \frac{P(x)}{P(x)+\zeta} = \frac{p_{\min}}{p_{\min}+\zeta}.$$

where $p_{\min} \triangleq \min_{x\in\mathcal{S}} P(x) > 0$ by definition of support set $\mathcal{S}$. Due to the construction of quantized distribution $\hat{P}_{n,XY}$,

$$\frac{1}{1+k/n^2} \leqslant \frac{P(x,y)}{\hat{P}_{X|Y=y,\zeta_n}(x)P_Y(y)} = \frac{P(x,y)}{\hat{P}_{n,XY}(x,y)} \leqslant \frac{1}{1-k/n^2} \tag{5.122}$$

where

$$k = \frac{(|\mathcal{X}|-1)}{\min_{(x,y)\in\mathcal{S}} P_{X|Y}(x|y)} > 0. \tag{5.123}$$

Then, notice that

$$\left(\frac{1}{1+O(1/n^2)}\right)^n = e^{n\ln(1+O(1/n^2))} \tag{5.124}$$

$$= e^{n(0+O(1/n^2)+O(1/n^3))} \tag{5.125}$$

$$= e^{O(1/n)} \tag{5.126}$$

$$= 1+O(1/n) \tag{5.127}$$

by Taylor expansion and the differentiability of logarithm and exponential.

Furthermore, for any function $\mathsf{f}: \mathcal{X} \mapsto \mathcal{Y}$, if P, P' satisfies $P_X(x)/P'_X(x) \in [1-\delta, 1+\delta]$ for all $x \in \mathcal{S}$, then $P_{\mathsf{f}(\mathsf{X})}(c)/P'_{\mathsf{f}(X)}(c) \in [1-\delta, 1+\delta]$ for all $c \in \{f(x) : x \in \mathcal{S}\}$. Therefore, there exists a constant $C > 0$ such that

$$1-\frac{C}{n} \leqslant \frac{P_{\mathsf{F}_n(X^n)Y^n}(c, y^n)}{\hat{P}_{n,\mathsf{F}_n(X^n)Y^n}(c, y^n)} \leqslant 1+\frac{C}{n} \tag{5.128}$$

for all P_{XY}, where $\hat{P}_{n,\mathsf{F}_n(X^n)Y^n}(c, y^n)$ is the joint distribution of $\mathsf{F}_n(X^n)$ and Y^n when $(X^n, Y^n) \sim \hat{P}^n_{n,XY}$. Applying Lemma 21 gives

$$\left| D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) - D\left(\hat{P}_{n,\mathsf{F}_n(X^n)Y^n} || \hat{Q}_{n,\mathsf{F}_n(X^n)Y^n}\right)\right| \tag{5.129}$$

$$\leqslant \frac{C}{n}\left(D\left(\hat{P}_{n,\mathsf{F}_n(X^n)Y^n} || \hat{Q}_{n,\mathsf{F}_n(X^n)Y^n}\right) + \frac{2}{e \ln 2}\right) + \log\frac{1 + C/n}{1 - C/n} \tag{5.130}$$

$$= D\left(\hat{P}_{n,\mathsf{F}_n(X^n)Y^n} || \hat{Q}_{n,\mathsf{F}_n(X^n)Y^n}\right) O\left(\frac{1}{n}\right) + O\left(\frac{1}{n}\right), \tag{5.131}$$

which implies that for any constant $a > 0$,

$$\begin{aligned}
&\sup_{P_{XY}, Q_{XY} \in \mathcal{P}(R+\delta)} D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) \\
&\qquad \leqslant \max\left(2 \max_{P_{XY}, Q_{XY} \in \mathcal{P}^\star_{\zeta_n}} D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right), \right. \\
&\qquad\qquad \left. \max_{P_{XY}, Q_{XY} \in \mathcal{P}^\star_{\zeta_n}} D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) + a\right)
\end{aligned}$$

when n is sufficiently large. Therefore,

$$\begin{aligned}
&\mathbb{P}\left[\sup_{P_{XY}, Q_{XY} \in \mathcal{P}(R+\delta)} D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) \geqslant b\right] \\
&\leqslant \mathbb{P}\left[\max_{P_{XY}, Q_{XY} \in \mathcal{P}^\star_{\zeta_n}} D\left(P_{\mathsf{F}_n(X^n)Y^n} || Q_{\mathsf{F}_n(X^n)Y^n}\right) \geqslant \frac{b}{2}\right] \\
&\to 0
\end{aligned}$$

for $b = 2a$, completing the proof. □

Chapter 6

SUMMARY AND TOPICS FOR FUTURE STUDIES

In this thesis, I study several source coding problems that arise from the uncertainty in communication networks.

In Chapter 2, I investigate the problem of asynchronous random access source coding. This problem generalizes the synchronous random access source coding setting, which itself extends the classical multiple access source coding framework. I propose an ARASC that allows a transmitter to encode a block of source observations and begin transmitting the encoded description to the receiver at any time, provided that the transmitter has accumulated enough observations and the channel for codeword transmission is idle. I analyze the performance of the proposed code. Although the performance bound derived in Chapter 2 is not tight in general, it offers a simple characterization of the penalty incurred by removing the synchronization constraint.

In Chapter 3, I propose a 2-stage variable-rate source code that uses unreliable side information to assist the decoding of a source. I analyze the performance of the code under two different models: one representing the worst-case scenario in which the joint distribution of the unreliable side information and the source is known, and the other in which the joint distribution is unknown. By allowing two stages of decoding, my proposed code allows the user to prevent the uncertainty from the unreliable side information from propagating to the reconstruction of the source of interest.

Chapter 4 examines the scenario in Chapter 3 from the perspective of an upstream point-to-point code creating uncertainties for downstream users due to imperfect source reconstructions. In Chapter 4, I study the same scenario from the viewpoint of the almost lossless upstream code. I show that the decoder output of near-optimal almost lossless point-to-point source codes can reveal vastly different amounts of information about a dependent source. Moreover, the code designer can smoothly adjust the amount of such information within a certain range. This implies that the designer of an almost lossless source code has the authority to control the amount of information exposed through the decoder output about dependent sources, thereby influencing the potential performance of downstream users in both beneficial and

detrimental ways as desired.

In Chapter 5, I study the problem of distributed hypothesis testing and source coding. I show that the compressed description of a random binning code serves as an input to a distributed hypothesis test. I analyze both the case of testing against independence and that of testing against an alternative distribution. I demonstrate that when the source is over-compressed, the codewords of a random binning code from different source distributions become similar in distribution. However, when the source is at the boundary between over-compression and under-compression, I show that the codewords of a random binning code from different source distributions are distinguishable with a subexponentially decaying type-II error probability. The result provides insight into the structure of a random binning code.

An interesting direction for future work is to investigate the potential of creating a system that dynamically learns about and adapts to the uncertainty arising from the environment, including synchronization issues, unreliable side information, unknown distributions, and so on. For example, in Chapter 3, the performance of USSC under the unknown model suffers from the decoder's ignorance of the joint distribution of the source and the unreliable side information. But if the decoder is able to refine its understanding of the distribution over time - assuming that the distribution is fixed, slowly varying, or following a predictable pattern - then the USSC could adjust its behavior accordingly and eventually outperform its performance under the worst-case known-distribution model. Another direction of future work involves studying some of the problems in this thesis in the lossy world, especially with almost lossless source code replaced by lossy code in the finite-blocklength regime.

BIBLIOGRAPHY

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948. DOI: 10.1002/j.1538-7305.1948.tb01338.x.

[2] D. Slepian and J. Wolf, "Noiseless coding of correlated information sources," *IEEE Transactions on Information Theory*, vol. 19, no. 4, pp. 471–480, 1973. DOI: 10.1109/TIT.1973.1055037.

[3] F. Willems, "Totally asynchronous slepian-wolf data compression," *IEEE Transactions on Information Theory*, vol. 34, no. 1, pp. 35–44, 1988. DOI: 10.1109/18.2599.

[4] J. Hui and P. Humblet, "The capacity region of the totally asynchronous multiple-access channel," *IEEE Transactions on Information Theory*, vol. 31, no. 2, pp. 207–216, 1985. DOI: 10.1109/TIT.1985.1057012.

[5] B. Rimoldi and R. Urbanke, "A rate-splitting approach to the gaussian multiple-access channel," *IEEE Transactions on Information Theory*, vol. 42, no. 2, pp. 364–375, 1996. DOI: 10.1109/18.485709.

[6] Z. Sun, C. Tian, J. Chen, and K. M. Wong, "Ldpc code design for asynchronous slepian-wolf coding," *IEEE Transactions on Communications*, vol. 58, no. 2, pp. 511–520, 2010. DOI: 10.1109/TCOMM.2010.02.090034.

[7] T. Matsuta and T. Uyematsu, "Coding theorems for asynchronous slepian–wolf coding systems," *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 4774–4795, 2020. DOI: 10.1109/TIT.2020.2974736.

[8] N. Merhav, "Universal decoding for asynchronous slepian-wolf encoding," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2652–2662, 2021. DOI: 10.1109/TIT.2020.3047911.

[9] S. Chen, M. Effros, and V. Kostina, "Lossless source coding in the point-to-point, multiple access, and random access scenarios," *IEEE Transactions on Information Theory*, vol. 66, no. 11, pp. 6688–6722, 2020. DOI: 10.1109/TIT.2020.3005155.

[10] R. C. Yavas, V. Kostina, and M. Effros, "Random access channel coding in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 67, no. 4, pp. 2115–2140, 2021. DOI: 10.1109/TIT.2020.3047630.

[11] K. Tatwawadi, S. S. Bidokhti, and T. Weissman, "On universal compression with constant random access," in *2018 IEEE International Symposium on Information Theory (ISIT)*, 2018, pp. 891–895. DOI: 10.1109/ISIT.2018.8437931.

[12] R. Vestergaard, Q. Zhang, and D. E. Lucani, "Enabling random access in universal compressors," in *IEEE INFOCOM 2021 - IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 2021, pp. 1–6. DOI: 10.1109/INFOCOMWKSHPS51825.2021.9484460.

[13] S. Kamparaju, S. Mastan, and S. Vatedka, "Low-complexity compression with random access," in *2022 IEEE International Conference on Signal Processing and Communications (SPCOM)*, 2022, pp. 1–5. DOI: 10.1109/SPCOM55316.2022.9840790.

[14] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, "Nonasymptotic and second-order achievability bounds for coding with side-information," *IEEE Transactions on Information Theory*, vol. 61, no. 4, pp. 1574–1605, 2015. DOI: 10.1109/TIT.2015.2400994.

[15] L. Gavalakis and I. Kontoyiannis, "Fundamental limits of lossless data compression with side information," *IEEE Transactions on Information Theory*, vol. 67, no. 5, pp. 2680–2692, 2021. DOI: 10.1109/TIT.2021.3062614.

[16] A. Wyner and J. Ziv, "The rate-distortion function for source coding with side information at the decoder," *IEEE Transactions on Information Theory*, vol. 22, no. 1, pp. 1–10, 1976. DOI: 10.1109/TIT.1976.1055508.

[17] A. Kaspi, "Rate-distortion function when side-information may be present at the decoder," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 2031–2034, 1994. DOI: 10.1109/18.340475.

[18] B. G. Kelly and A. B. Wagner, "Reliability in source coding with side information," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5086–5111, 2012. DOI: 10.1109/TIT.2012.2201346.

[19] World Meteorological Organization, *Guide to Meteorological Instruments and Methods of Observation (WMO-No. 8)*. Geneva, Switzerland: World Meteorological Organization, 2018.

[20] V. Strassen, "Asymptotische Abschätzungen in Shannons Informationstheorie," in *Transactions of the Third Prague Conference on Information Theory, Statistical Decision Functions, Random Processes*, Prague: Publishing House of the Czechoslovak Academy of Sciences, 1962, pp. 689–723.

[21] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, 2010. DOI: 10.1109/TIT.2010.2043769.

[22] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Feedback in the non-asymptotic regime," *IEEE Transactions on Information Theory*, vol. 57, no. 8, pp. 4903–4925, 2011. DOI: 10.1109/TIT.2011.2158476.

[23] V. Kostina and S. Verdu, "Fixed-length lossy compression in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 58, no. 6, pp. 3309–3338, 2012. DOI: 10.1109/TIT.2012.2186786.

[24] Y.-W. Huang and P. Moulin, "Finite blocklength coding for multiple access channels," in *2012 IEEE International Symposium on Information Theory Proceedings*, 2012, pp. 831–835. DOI: 10.1109/ISIT.2012.6284677.

[25] V. Kostina and S. Verdú, "Lossy joint source-channel coding in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 59, no. 5, pp. 2545–2575, 2013. DOI: 10.1109/TIT.2013.2238657.

[26] V. Y. F. Tan and O. Kosut, "On the dispersions of three network information theory problems," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 881–903, 2014. DOI: 10.1109/TIT.2013.2291231.

[27] R. Nomura and T. S. Han, "Second-order slepian-wolf coding theorems for non-mixed and mixed sources," *IEEE Transactions on Information Theory*, vol. 60, no. 9, pp. 5553–5572, 2014. DOI: 10.1109/TIT.2014.2339231.

[28] I. Kontoyiannis and S. Verdú, "Optimal lossless data compression: Non-asymptotics and asymptotics," *IEEE Transactions on Information Theory*, vol. 60, no. 2, pp. 777–795, 2014. DOI: 10.1109/TIT.2013.2291007.

[29] V. Kostina, Y. Polyanskiy, and S. Verdú, "Variable-length compression allowing errors," *IEEE Transactions on Information Theory*, vol. 61, no. 8, pp. 4316–4330, 2015. DOI: 10.1109/TIT.2015.2438831.

[30] L. Zhou and M. Motani, "Kaspi problem revisited: Non-asymptotic converse bound and second-order asymptotics," in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, 2017, pp. 1–6. DOI: 10.1109/GLOCOM.2017.8254168.

[31] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer-Verlag Berlin Heidelberg, 2003, ISBN: 3540435816.

[32] R. Ahlswede and I. Csiszar, "Hypothesis testing with communication constraints," *IEEE Transactions on Information Theory*, vol. 32, no. 4, pp. 533–542, 1986. DOI: 10.1109/TIT.1986.1057194.

[33] M. S. Rahman and A. B. Wagner, "On the optimality of binning for distributed hypothesis testing," *IEEE Transactions on Information Theory*, vol. 58, no. 10, pp. 6282–6303, 2012. DOI: 10.1109/TIT.2012.2206793.

[34] E. Haim and Y. Kochman, "Binary distributed hypothesis testing via körner-marton coding," in *2016 IEEE Information Theory Workshop (ITW)*, 2016, pp. 146–150. DOI: 10.1109/ITW.2016.7606813.

[35] S. Watanabe, "On sub-optimality of random binning for distributed hypothesis testing," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 2708–2713. DOI: 10.1109/ISIT50566.2022.9834275.

[36] Y. Kochman, "An improved upper bound for distributed hypothesis testing," in *2024 IEEE International Symposium on Information Theory (ISIT)*, 2024, pp. 2909–2914. DOI: 10.1109/ISIT57864.2024.10619234.

[37] S. Sun and M. Effros, "Asynchronous random access data compression," in *2024 Data Compression Conference (DCC)*, 2024, pp. 482–491. DOI: 10.1109/DCC58796.2024.00056.

[38] L. Kleinrock, *Queueing Systems*. New York: Wiley, 1976.

[39] M. Mecking, "Multiple-access with stripping receivers," in *in Proc. 4th European Mobile Communication Conference (EPMCC*, 2001.

[40] Y.-W. P. Hong, Y.-R. Tsai, Y.-Y. Liao, C.-H. Lin, and K.-J. Yang, "On the throughput, delay, and energy efficiency of distributed source coding in random access sensor networks," *IEEE Transactions on Wireless Communications*, vol. 9, no. 6, pp. 1965–1975, 2010. DOI: 10.1109/TWC.2010.5475341.

[41] M. Zhu, D. G. M. Mitchell, M. Lentmaier, D. J. Costello, and B. Bai, "Error propagation mitigation in sliding window decoding of braided convolutional codes," *IEEE Transactions on Communications*, vol. 68, no. 11, pp. 6683–6698, 2020. DOI: 10.1109/TCOMM.2020.3015945.

[42] S. Sun and M. Effros, "Source coding with unreliable side information in the finite blocklength regime," in *2022 IEEE International Symposium on Information Theory (ISIT)*, 2022, pp. 222–227. DOI: 10.1109/ISIT50566.2022.9834490.

[43] R. Gallager, "A perspective on multiaccess channels," *IEEE Transactions on Information Theory*, vol. 31, no. 2, pp. 124–142, 1985. DOI: 10.1109/TIT.1985.1057022.

[44] R. Shu-Kwan Cheng, "Stripping cdma-an asymptotically optimal coding scheme for l-out-of-k white gaussian channels," in *Proceedings of GLOBECOM'96. 1996 IEEE Global Telecommunications Conference*, 1996, pp. 142–146. DOI: 10.1109/GLOCOM.1996.586797.

[45] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge, 2013.

[46] O. Gazi and A. A. Andi, "The effect of error propagation on the performance of polar codes utilizing successive cancellation decoding algorithm," *Advanced Electromagnetics*, vol. 8, no. 2, pp. 114–120, Jun. 2019. DOI: 10.7716/aem.v8i2.998.

[47] S. Draper, "Universal incremental slepian-wolf coding," in *Proc. 42th Annual Allerton Conference on Communications, Control, and Computing*, Sep. 2004.

[48] S. Sun and M. Effros, "Almost lossless onion peeling data compression," in *2024 Data Compression Conference (DCC)*, 2024, pp. 472–481. DOI: 10.1109/DCC58796.2024.00055.

[49] S. Sun and M. Effros, "Code design in almost lossless onion peeling data compression," in *2025 Data Compression Conference (DCC)*, 2025, pp. 293–302. DOI: 10.1109/DCC62719.2025.00037.

[50] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011. DOI: 10.1017/CBO9780511921889.

[51] A. K. Al Jabri and S. Al-Issa, "Zero-error codes for correlated information sources," in *Crytography and Coding*, M. Darnell, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 1997, pp. 17–22, ISBN: 978-3-540-69668-1.

[52] T. M. Cover and J. A. Thomas, *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA: Wiley-Interscience, 2006, ISBN: 0471241954.

[53] T. Matsuta, T. Uyematsu, and R. Matsumoto, "Universal slepian-wolf source codes using low-density parity-check matrices," in *2010 IEEE International Symposium on Information Theory*, 2010, pp. 186–190. DOI: 10.1109/ISIT.2010.5513253.

[54] R. G. Gallager, "Source coding with side information and universal coding," *Mass. Instit. Tech., Cambridge, MA, USA, Tech. Rep. LIDS-P-937*, 1976.

[55] R. Gallager, "Low-density parity-check codes," *IRE Transactions on Information Theory*, vol. 8, no. 1, pp. 21–28, 1962. DOI: 10.1109/TIT.1962.1057683.

[56] J. H. Bae, A. Abotabl, H.-P. Lin, K.-B. Song, and J. Lee, "An overview of channel coding for 5g nr cellular communications," *APSIPA Transactions on Signal and Information Processing*, vol. 8, e17, 2019. DOI: 10.1017/ATSIP.2019.10.

[57] E. Berlekamp, R. McEliece, and H. van Tilborg, "On the inherent intractability of certain coding problems (corresp.)," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 384–386, 1978. DOI: 10.1109/TIT.1978.1055873.

[58] G. Caire, S. Shamai, and S. Verdú, "Noiseless data compression with low-density parity-check codes," in *Advances in Network Information Theory*, ser. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, G. K. P. Gupta and A. J. van Wijngaarden, Eds., vol. 66, 2004, pp. 263–284.

[59] Y. Liu and M. Effros, "Finite-blocklength and error-exponent analyses for ldpc codes in point-to-point and multiple access communication," *arXiv.org*, May 2020. [Online]. Available: https://arxiv.org/abs/2005.06428.

[60] S. Sun and M. Effros, "Distributed hypothesis testing and the output statistics of random binning," in *Proceedings of the 61st Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, Available at IDEALS: https://hdl.handle.net/2142/130280, Champaign, IL, USA, 2025.

[61] Y. Kochman and L. Wang, "Improved random-binning exponent for distributed hypothesis testing," *IEEE Transactions on Information Theory*, vol. 71, no. 11, pp. 8217–8222, 2025. DOI: 10.1109/TIT.2025.3603269.

[62] I. S. Adamou, E. Dupraz, and T. Matsumoto, *An information-spectrum approach to distributed hypothesis testing for general sources*, 2023. arXiv: 2305.06887 [cs.IT]. [Online]. Available: https://arxiv.org/abs/2305.06887.

[63] E. Dupraz, I. S. Adamou, R. Asvadi, and T. Matsumoto, *Practical short-length coding schemes for binary distributed hypothesis testing*, 2024. arXiv: 2405.07697 [cs.IT]. [Online]. Available: https://arxiv.org/abs/2405.07697.

[64] G. Katz, P. Piantanida, R. Couillet, and M. Debbah, "On the necessity of binning for the distributed hypothesis testing problem," in *2015 IEEE International Symposium on Information Theory (ISIT)*, 2015, pp. 2797–2801. DOI: 10.1109/ISIT.2015.7282966.

[65] M. H. Yassaee, M. R. Aref, and A. Gohari, "Achievability proof via output statistics of random binning," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6760–6786, 2014. DOI: 10.1109/TIT.2014.2351812.

[66] M. M. Mojahedian, A. Gohari, and M. R. Aref, "On the equivalency of reliability and security metrics for wireline networks," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 2713–2717. DOI: 10.1109/ISIT.2017.8007022.

[67] M. Kavian, M. Mahdi Mojahedian, M. Hossein Yassaee, M. Mirmohseni, and M. Reza Aref, "Output statistics of random binning: Tsallis divergence and its applications," *IEEE Transactions on Information Theory*, vol. 72, no. 3, pp. 1521–1542, 2026. DOI: 10.1109/TIT.2026.3657514.

[68] M. Hayashi, "Second-order asymptotics in fixed-length source coding and intrinsic randomness," *IEEE Transactions on Information Theory*, vol. 54, no. 10, pp. 4619–4637, 2008. DOI: 10.1109/TIT.2008.928985.

[69] M. V. Burnashev, "On stein's lemma in hypotheses testing in general non-asymptotic case," en, *Stat. Inference Stoch. Process.*, vol. 26, no. 1, pp. 89–97, Apr. 2023.

[70] P. Billingsley, *Probability and Measure*, 4th ed. New York: John Wiley & Sons, 1995.