Imperial College of Science, Technology and Medicine
Department of Electrical and Electronic Engineering

On the Rate-Distortion-Perception Trade-off and Remote Coordination: The Role of Randomness

Yassine Hamdi

Submitted in part fulfilment of the requirements for the degree of
Doctor of Philosophy in Electrical and Electronic Engineering of the Imperial College London,
July 2026

In the name of Allah[1], the Most Merciful. I dedicate this thesis to my family. I am grateful to my supervisor Deniz Gündüz for his great kindness, motivation, and skills.

[1]in Arabic, Alaha in Aramaic, Eloah in Hebrew, Deus in Latin, God in English, Gudą in proto-Germanic, Hūtá in Sanskrit, Tiānzhǔ in Mandarin.

Copyright

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non Commercial 4.0 International Licence (CC BY-NC).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Declaration of originality

I, Yassine Hamdi, declare that this thesis titled, “On the Rate-Distortion-Perception Trade-off and Remote Coordination: The Role of Randomness", and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at Imperial College of Science, Technology and Medicine.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.

Signed: Yassine Hamdi

Date: July 21, 2026

Abstract

We collect, process, store and communicate ever increasing amounts of data. Lossy compression discards less relevant data under computational, latency, bandwidth or storage constraints. This thesis addresses fundamental problems in lossy compression, focusing on two areas: compression with realism constraints, and distributed coordination.

Classical rate-distortion (RD) theory characterizes the trade-off between compression level (rate) and difference from the original data (distortion), but fails to capture perceptual quality, or realism. Recent work has proposed formulations of a three-way rate-distortion-perception (RDP) trade-off; several of these require randomized algorithms, unlike classical RD. This randomness may be private to the encoder or the decoder, or common to both. First, we characterize the role of private randomness across RDP formulations, showing that encoder private randomness does not improve performance, even when conveyed to the decoder. Second, we analyze the trade-off under the strongest realism constraints in the presence of side information, characterizing its dual role as both a second observation of the source and a source of common randomness. Third, we identify important limitations of existing RDP formulations and propose an alternative framework, based on algorithmic information theory and involving critics of realism. We show that deterministic schemes are optimal when the number of samples a critic observes at a time is practical, and that our formulation is equivalent to previous ones when that number is impractically large.

The second part addresses coordination in distributed systems, where some agents must coordinate without direct communication while others communicate under latency or bandwidth constraints. Coordination requirements admit a mathematical formulation closely resembling the realism constraints above. We study non-interactive remote coordination, in which two agents receive information from two controllers that themselves coordinate without communicating. We develop remote channel synthesis, a technique extending classical channel synthesis, yielding a scheme with improved performance.

Contents

Abstract **iv**

1 Introduction **2**

1.1 Lossy compression with distortion and realism constraints 3

1.1.1 The rate-distortion-perception trade-off 3

1.1.2 The distribution matching formalism for realism 4

1.1.3 The Role of Randomization 6

1.2 Rethinking Realism: From Distributions to Realizations 7

1.3 Coordination problems in Distributed Systems 8

1.4 Thesis Contributions and Organization 10

1.4.1 Overview of Chapter 2 11

1.4.2 Overview of Chapter 3 11

1.4.3 Overview of Chapter 4 12

1.4.4 Overview of Chapter 5 13

1.5 Notation 14

2 Rate-Distortion-Perception Trade-off: The Role of Private Randomness **16**

2.1 Prior work on the role of randomness 16

2.2 Outline 19

2.3 Problem formulation 20

2.4 A soft covering lemma for randomized compressors 25

2.5 The rate-distortion-perception trade-off with near-perfect realism 27

2.5.1 The role of encoder private randomness for finite source alphabets 27

2.5.2 An extension to sources with infinite entropy 29

2.5.3 Per-symbol realism 32

2.6 Achievability proof for Theorem 2.7 33

2.6.1 Modifying a standard code construction 33

2.6.2 Achievability of Theorem 2.7 34

2.7 Extensions to other realism constraints 35

2.7.1 The de-randomization mechanism and its scope 36

2.7.2 Conclusive extensions with limited common randomness 37

2.8 Conclusion 40

Appendices **42**

2.A Proof of Proposition 2.3 42

2.B Converse of Theorem 2.7 44

2.C Rate lower bounds in the converse proof 45

2.D Proof of Theorem 2.11 47

2.D.1 Converse 47

2.D.2 Quantization argument 48

2.D.3 Transition from near-perfect to perfect realism 49

2.D.4 Conclusion 51

2.E Proof of Theorem 2.13 52

2.E.1 Converse - finite source alphabet 52

2.E.2 Converse - source with infinite entropy 52

2.E.3 Achievability - finite source alphabet 52

2.E.4 Achievability - source with infinite entropy 54

2.F Further justifications 55

2.F.1 Proof of Proposition 2.14 55

2.F.2 Local channel synthesis argument 56

2.F.3 Quantizability for Euclidean spaces 57

2.F.4 The uniform integrability constraint is non-degenerate 59

3 Rate-Distortion-Perception Trade-off with Strong Realism Constraints: Role of Side Information and Common Randomness 62

3.1 Introduction 62

3.2 Problem formulation and a first property 65

3.2.1 Definitions 65

3.2.2 Equivalence of the perfect / near-perfect realism problems 68

3.3 Main results 68

3.3.1 Marginal realism when the side information is available at both encoder and decoder 68

3.3.2 Marginal realism when the side information is only available at the decoder 72

3.3.3 When imposing joint realism of the reconstruction and the side information 75

3.3.4 The Gaussian case 77

3.4 Proof of Theorem 3.3 79

3.4.1 Converse 79

3.4.2 Informal outline of the achievability proof 81

3.4.3 Random codebook 81

3.4.4 Distribution $Q^{(1)}$ and soft-covering lemma 82

3.4.5 Constructing a code 84

3.4.6 Distance between $P^{(1)}$ and $Q^{(1)}$ 85

3.4.7 Conclusion 85

3.5 Proof of Theorem 3.9 86

3.5.1 Informal outline 86

3.5.2 Random codebook indexed by a virtual message 86

3.5.3 Distribution $Q^{(1)}$ and soft-covering lemma 87

3.5.4 Construction of a code 89

3.6 Proof of Theorem 3.13 93

3.6.1 Converse . . . 93

3.6.2 Achievability . . . 94

3.7 Proof of Theorem 3.14 . . . 96

3.7.1 Converse . . . 96

3.7.2 Achievability . . . 97

3.8 Proofs of Theorems 3.11 and 3.12 and Corollaries 3.4 and 3.7 . . . 98

3.8.1 Proof of Theorem 3.11 . . . 98

3.8.2 Proof of Theorem 3.12 . . . 102

3.8.3 Proof of Corollary 3.7 . . . 103

3.8.4 Proof of Corollary 3.4 . . . 104

3.9 The Gaussian case . . . 105

3.9.1 The role of common randomness for a Gaussian source in the absence of side information . . . 105

3.9.2 Bi-dimensional Gaussian source and side information . . . 108

3.10 Conclusion . . . 111

Appendices **113**

3.A Measure-theoretic justifications . . . 113

3.B Equivalence between achievability with near-perfect and perfect realism . . . 114

3.B.1 Marginal realism . . . 114

3.B.2 Joint realism . . . 115

3.B.3 Modified distribution . . . 116

3.B.4 Proving perfect joint realism 119

3.B.5 Coupling with Q 121

3.C Further justifications 123

3.C.1 Use of the soft covering lemma when the side information is available only at the decoder 123

3.C.2 Use of the soft covering lemma when the side information is available at both encoder and decoder 125

3.C.3 Computations in the Gaussian case 127

3.C.4 Markov chains in converse proofs 131

4 The Rate-Distortion-Perception Trade-Off with Algorithmic Realism 134

4.1 Introduction 134

4.2 Background on algorithmic information theory 138

4.3 New model for the rate-distortion-perception trade-off 143

4.3.1 The one-shot setting 144

4.3.2 Asymptotic setting 145

4.4 Results 147

4.4.1 Low batch size regime 147

4.4.2 One-shot achievable points 148

4.4.3 Generalizing the distribution matching formalism 150

4.5 Conclusion 151

Appendices **153**

4.A Further background on algorithmic information theory 153

4.A.1 Computability 153

4.A.2 Universal critics and semi-measures 155

4.B Proof of Theorem 4.5 157

4.B.1 Outline 157

4.B.2 Realism performance of a uniformly sampled batch of random reconstructions 157

4.B.3 Further properties of a uniformly sampled batch 160

4.B.4 Proof of Theorem 4.5 162

4.C Achievability of Theorem 4.4 164

4.C.1 Encoder de-randomization 166

4.D Converse of Theorem 4.4 167

4.D.1 Converse proof 167

4.E Proof of Theorem 4.6 169

4.F Standard converse arguments 173

4.G Frequency critics 174

4.G.1 Critic involving the frequency of a specific pattern 174

4.G.2 Critic involving an empirical distribution 176

4.H On the total variation distance 178

4.H.1 Proof of Claim 4.16 178

4.I The birthday paradox . . . 178

4.J Existence of a universal $p^{\otimes *}$-critic . . . 179

4.K Additional semi-computability arguments . . . 182

4.K.1 Proof of Lemma 4.8 . . . 182

4.K.2 Notation for the proof of Lemma 4.9 . . . 182

4.K.3 Assume that f and g are computable . . . 183

4.K.4 Suppose that f is computable and only takes positive values . . . 184

4.K.5 Assume that f and g are lower semi-computable . . . 185

4.K.6 Assume that f and g are semi-computable and non-negative . . . 186

4.K.7 Functions of finite binary strings . . . 188

5 Channel Synthesis Based Non-interactive Remote Coordination 190

5.1 Introduction . . . 190

5.2 Problem formulation . . . 195

5.3 A natural scheme based on channel synthesis . . . 197

5.3.1 The role of common randomness in channel synthesis . . . 197

5.3.2 A simple scheme for non-interactive remote coordination . . . 198

5.3.3 Sub-optimality . . . 199

5.4 Remote Channel Synthesis . . . 199

5.4.1 Outline . . . 199

5.4.2 Problem formulation . . . 201

5.4.3 Main results on the remote channel synthesis problem . . . 202

5.4.4 Scalar schemes v.s. vector schemes 205

5.5 Non-interactive remote coordination scheme based on remote channel synthesis . 206

5.5.1 Scheme description 206

5.5.2 Achievable region 208

5.6 Leveraging Gács-Körner common information between the sources 211

5.6.1 Coordination schemes à la Wyner, improved with channel synthesis . . . 211

5.6.2 A hybrid scheme 213

5.6.3 Main components of the proof of Theorem 5.9 214

5.6.4 Can Gács-Körner common information be leveraged in a scheme based on remote channel synthesis? 215

5.7 Channel synthesis based outer bounds 216

5.8 Conclusion 220

Appendices **222**

5.A Converse direction of Theorem 5.3 222

5.B Achievability direction of Theorem 5.3 227

5.B.1 Random codebook 228

5.B.2 Distribution Q and soft-covering lemma 228

5.B.3 Choosing a codebook 231

5.B.4 Construction of a code 231

5.C Proof of Proposition 5.5 232

5.D Proof of Proposition 5.6 237

5.E Continuation of the proof of Theorem 5.9 . . . 239

5.E.1 On joint typicality in Theorem 5.10 . . . 239

5.E.2 Proof of Proposition 5.11 . . . 240

5.E.3 Proof of Theorem 5.9 . . . 242

6 Conclusion **246**

6.1 Lossy compression with distribution matching constraints . . . 247

6.1.1 The role of randomization . . . 247

6.1.2 Side information and common randomness . . . 248

6.2 Rethinking the realism paradigm . . . 249

6.3 Coordination problems in distributed systems . . . 250

6.4 Unifying themes and future directions . . . 251

6.5 Closing remarks . . . 252

1 Mathematical background **254**

1.1 Probability theory background . . . 254

1.1.1 Conditional probabilities on general alphabets and sequential definitions 254

1.1.2 The notion of Markov chain for general alphabets . . . 257

1.1.3 Density functions . . . 258

1.2 Information quantities on general alphabets . . . 259

1.3 Proofs of lemmas stated in Section 1.1 . . . 263

1.3.1 Proof of Remark 1.2 . . . 263

1.3.2 Proof of Proposition 1.5 263

1.3.3 Proof of Corollary 1.6 265

1.3.4 Proof of Lemma 1.7 266

1.3.5 Proof of Lemma 1.8 267

1.3.6 Proof of Lemma 1.9 267

Bibliography **267**

List of Tables

3.1 Main results of Chapter 3 67

3.2 Corollaries in Chapter 3 67

List of Figures

2.1 The system model, in the case of a fixed-length code with finite rates (Definition 2.3). 21

3.1 The system model. The side information Z^n is always available at the decoder, but not necessarily at the encoder. 63

3.2 Graphical model for the distribution $Q^{(1)}$ appearing in the proof of Theorem 3.3 (E-D-achievable tuples with marginal realism). This is an instance of the graphical model involved in the soft covering lemma with side information [Cuf13, Lemma VII.5]. 82

3.3 Graphical model for the distribution $Q^{(1)}$ appearing in the proof of Theorem 3.9 (D-achievable tuples with marginal realism). This is an instance of the graphical model involved in the soft covering lemma [Cuf13, Lemma VII.4]. 87

3.A.1 Graphical model for $Q^{(1)}$ with quantized variables, where $k_n = \lfloor 2^{n(R+\varepsilon)} \rfloor \times \lfloor 2^{nR'} \rfloor \times \lfloor 2^{nR_c} \rfloor$. The variables inside a dashed oval are independent from variables inside another oval. 113

4.1 The system model for the one-shot setting. 143

4.2 The system model for the asymptotic setting. Index k ranges from 1 to the batch size. The same encoder-decoder pair is used to process each source sample in the batch. 146

5.1 The system model. 191

5.2 Wyner's setup [Wyn75]. 192

5.3 Non-interactive coordination problem, which subsumes those studies by Gács and Körner [GK73] and Berg et. al. [B$^+$21]. 195

5.4 Block coding setup for channel synthesis, a.k.a. the reverse channel coding problem. 197

5.5 Our remote channel synthesis setup. 200

5.6 Lower bound on the gap between the optimal rates in (5.15) — see Appendix 5.C for details. Variables θ and τ denote $q_{Z,Y}(Y \neq Z)$ and $q_{X,Z}(X \neq Z)$, respectively, where $q_{X,Z,Y}$ is a distribution compatible with $(q_{X,Z}, q_{X,Y})$. 205

5.7 First step 207

5.8 Second step 207

5.9 Illustration of the proposed coding scheme. 215

5.B.1 Graphical model for Q. See Equations (5.44) and (5.45). 228

List of Acronyms and Abbreviations

GK Gács-Körner

MSE mean-squared error

RD rate-distortion

RDP rate-distortion-perception

TVD total variation distance

Chapter 1

Introduction

In many use-cases involving image data, one can compress the input in such a way that the resulting image still meets all the requirements in a satisfactory manner. For instance, the resulting modification might be imperceptible to a human end-user, or it might still allow successful detection of the image's main components by an object detection software. This can be generalized to other signals and applications; that is, in order to use a signal for a certain application, it is often not necessary to store or send all the raw data. For many applications, only certain aspects of the input signal are relevant or useful. This suggests not to consider an input signal as an indivisible entity, but rather as a reservoir of information, from which one can ingeniously extract only the relevant parts for the downstream task, a process which is known as *lossy compression*.

Such methods are ubiquitous in the digital age. Lossy compression is typically apropos when a low level of latency (for processing or communication) is imposed (e.g., for media streaming) or in order to optimize storage costs (e.g., in data centers), or to obfuscate certain information for privacy reasons. The ever-growing volumes of data and communication have recently led to an increased interest in lossy compression and to the popularity of certain paradigms known as *semantic communication* or *task-oriented communication* [GKJ+20, G+23]. In the latter, the aim is to discard any part of the data aside from what pertains to meaning (semantics) or information relevant to a specific task. In addition, lossy compression is unavoidable for any

task involving converting analog data into digital data.

There is often no simple way to perfectly dissect a given signal into distinct atoms of information content. Describing and extracting relevant information content are problems that still hold many mysteries, and have inspired much passion in engineers and theorists. It is one of the few areas of engineering with immense practical applications that also has a beautiful mathematical theory; and despite decades of research and significant progress, many open problems remain. In the next sections, we describe the main types of problems that have motivated our work, problems for which there exist many ingenious information extraction techniques. We then conclude Chapter 1 with an outline of the thesis and of our contributions.

1.1 Lossy compression with distortion and realism constraints

1.1.1 The rate-distortion-perception trade-off

The problem of lossy source coding consists in producing a compressed representation of a *source* signal that is close to the original signal according to some measure of *distortion*. Classical rate-distortion (RD) theory, pioneered by Claude Shannon, provides the mathematical foundation for understanding the trade-offs in lossy source coding: how much information can we discard while maintaining acceptable distortion compared to the original signal? This theory has guided the development of compression standards that enable modern digital communication and storage systems. While broadly successful (e.g. [PS11, Say12]), this approach does have certain limitations. One is that the reconstruction might be qualitatively far from the source realization that generated it. When distortion is measured by the mean-squared error (MSE), RD theory predicts that high-frequency components should be discarded, which explains why images compressed via the well-known JPEG algorithm (which produces files with an emblematical .jpeg or .jpg extension) look blurry when compressed into files of small size.

Of course, all distortion measures used in theoretical studies of multimedia compression are

proxies for the measure of real interest, namely how the reconstruction would be perceived by the (usually human) end-user. In some cases, this end-user will prefer a reconstruction that is more distorted according to conventional measures. For instance, MPEG Advanced Audio Coding (ACC) populates high-frequency bands with artificial noise instead of leaving them null [Say12, Sec. 17.4.2], in order to match the power spectrum of the source; this is termed perceptual noise substitution (PNS). These considerations suggest to go beyond the traditional RD formulation, and to study a three-way trade-off between the amount of retained information (or the *compression rate*), the level of distortion, and the *realism* or perceptual quality of the reconstructed version of the signal.

1.1.2 The distribution matching formalism for realism

The general idea of a three-way trade-off between compression rate and distortion and realism has sparked substantial interest since methods based on deep neural networks have been shown to outperform traditional image and video compression codecs [BLS17, MAT+18b, MS20, Kan18, KLL+18, PBK21, ZLO+22, WSK18, RNL+19, DCSMS19, HRTC19, HLX21, LCLK21] under different distortion measures. In [ATM+19], the authors used generative adversarial networks (GANs) to push the limits of image compression in very low rates by synthesizing image content, such as facades of buildings, using a reference image database. This allows the receiver to generate images that resemble the source image semantically, although they may not match the details, providing visually pleasing reconstructions even at very low compression rates. It has been observed — first in the quantization literature, then in neural compression, and finally as a matter of principle — that at such low rates an increase in perceptual quality comes at the cost of increased distortion. We summarize this evidence, as it is precisely what motivates departing from the RD paradigm.

The earliest strand is constructive. Beginning with moment-preserving quantization [DM91] and developed systematically by Li *et al.* [LKK10, LKK11, LKOK12, KZLK13], these works impose on the quantizer an explicit constraint that the distribution of its output match that of the source — that it preserve the source's moments, or, for audio, its spectral density

[LKOK12]. The resulting quantizers are strictly sub-optimal in the RD sense: at a given rate they incur a higher MSE than the RD-optimal quantizer. Yet they are preferred perceptually. The reason is structural: an RD-optimal reconstruction attains its low MSE by shrinking towards a conditional mean, and it is exactly this shrinkage that erases the statistical signature of the source. The deficiency of RD theory is therefore not an imperfection to be tuned away by adjusting the rate — the RD-optimal solution is *systematically* unrealistic.

The same phenomenon reappeared with learned codecs. Tschannen *et al.* [TAL18] trained deep generative compressors under an explicit distribution-preservation requirement, and found that at low rates the MSE-optimal reconstruction is blurred while the distribution-preserving one is sharp and plausible — despite scoring *worse* on MSE. Agustsson *et al.* [ATM+19] pushed this to extreme rates with GANs, synthesizing plausible content rather than reproducing it (see also [SBS18]): the reconstructions convince a human observer while being, by construction, far from the source under any pixel-wise distortion measure. In both cases the codec that a user prefers is one that RD theory would reject.

That this reflects a fundamental obstruction, rather than a limitation of particular codecs, was established by Blau and Michaeli [BM18] in the setting of image restoration. Defining perceptual quality as a divergence between the distribution of an algorithm's outputs and that of natural images, they proved that the perception–distortion function — the best perceptual quality attainable at a given distortion level — is monotonically non-increasing and convex. Consequently, *no* estimator can simultaneously minimize distortion and attain the best perceptual quality, and this holds for *every* distortion measure, so it cannot be circumvented by a more perceptually-motivated choice of d. They corroborated this with a human-opinion study on super-resolution, in which the algorithms with the highest PSNR were judged perceptually the worst. The implication for compression is immediate: optimizing distortion alone — which is the entirety of what RD theory does — provably cannot deliver realism, so realism has to enter the problem formulation as a constraint in its own right.

This is what the rate-distortion-perception (RDP) formalism provides. Motivated by generative modeling, where synthesized images exhibit the statistical properties of the training set,

Matsumoto [Mat18a, Mat19] and Blau and Michaeli [BM19] augmented Shannon's RD function with an explicit constraint on the distribution of the reconstruction, yielding a three-way trade-off between rate, distortion and realism. The classical RD function is recovered as the special case in which the realism constraint is inactive. RDP theory therefore does not overturn RD theory — it strictly contains it — but the operating regime of interest for modern codecs, namely low rate and high realism, is exactly the regime that RD theory leaves unaddressed. If the source signal X is a sample from a distribution p_X, then the requirement of realism of the reconstructed version Y is formulated as $\mathcal{D}(p_X, p_Y) \leq \lambda$, where p_Y is the distribution of Y, λ is a real number, and $\mathcal{D}$ is a measure of discrepancy between distributions, e.g., the total variation distance (TVD) (Definition 1.3) or some other divergence. For very small λ, we refer to this as a *distribution matching constraint*, and as a *perfect realism constraint* if $\lambda = 0$.

1.1.3 The Role of Randomization

A compression scheme consists of a compressor (or encoder), which maps the source signal into a compressed file, and a decompressor (or decoder) which maps that file into a reconstructed version of the signal. In conventional RD theory, no randomization is needed in compression and decompression operations to achieve the optimal RD performance. This simplifies both the analysis and implementation of RD optimal codes. However, for the RDP trade-off with a perfect realism constraint, randomized compression algorithms are necessary in general to achieve the optimal trade-off. Moreover, it was shown in [TA21] and [SLY15a] that when randomized decoders are considered, the availability of common randomness between the encoder and the decoder enables an improvement in performance.

1.2 Rethinking Realism: From Distributions to Realizations

While the distribution matching approach to guaranteeing realism has provided valuable theoretical insights, it has significant limitations that suggest the need for alternative formulations [The24].

Mismatch with human perception: First, while distribution matching constraints are very relevant for assessing the performance of a generative model, they are arguably mismatched to human perception in compression applications. Indeed, under such constraints, it is not meaningful to speak of whether an individual sample is realistic, and realism becomes purely a property of large ensembles of samples. Yet human vision excels at identifying unrealistic aspects of individual images but is blind to many ensemble-level statistical deviations. Humans cannot distinguish natural images from those selected randomly from a large fixed set of realistic images, nor can they detect subtle differences that would be captured by distributional measures.

The common randomness paradox: Second, theoretical results indicate that achieving optimal RDP trade-offs requires large amounts of high-quality common randomness. Yet this does not appear in the state-of-the-art practical schemes. Moreover, real compression systems do not fail to produce realistic images when de-randomized by replacing common randomness with fixed values. This disconnect between theory and practice suggests that the distribution-matching framework may be imposing unnecessarily stringent requirements, which translates into needlessly using additional resources, such as compression rate.

Redefining Realism: These observations motivate a sample-level approach to realism, where we consider critics that assign deficiency scores to individual samples rather than examining distributions. Furthermore, we aim to propose a formulation of realism which:

- Requires that reconstructions be typical in the strongest possible sense,
- Aligns better with human perception by focusing on individual realizations, and

- Eliminates the need for common randomness in practical scenarios.

1.3 Coordination problems in Distributed Systems

Coordination problems appear in various engineering applications, ranging from multi-agent reinforcement learning to federated learning and autonomous systems. For the purpose of introducing the type of problems we focus on to a broad audience, we will give an example from soccer. Through this example, we introduce the notions of statistical coordination, non-interactive coordination, and non-interactive remote coordination. Indeed, this example can be generalized, by replacing players by any kind of agents, such as drones or robots.

The locations of players of both teams at a given moment during the game are represented by a variable S. Consider a situation, corresponding to a value s of S, where Players A and B of one team positioned close to the other team's goal area, and Player A is in possession of the ball. Then, Player A chooses to pass the ball to Player B. Does Player A know exactly what the next move of Player B will be? Perhaps he does not know it with 100% certainty. However, if both players have trained together and have encountered many situations which were similar, in terms of positions of players from their team and of players from an opposing team (i.e., cases where $S \approx s$), then Player A may remember a history, such as:

- In a training session where a similar situation ($S \approx s$) arose, player B received the ball and paused.
- In another similar training situation ($S \approx s$), player B received the ball and directly passed it to another player.
- In an official match where a similar situation ($S \approx s$) arose, player B received the ball and directly passed it to another player.
- In another similar training session ($S \approx s$), player B received the ball and ran with it towards the goal area.

Hence, for similar situations ($S \approx s$), and for the specific choice of action by Player A to pass the ball to Player B, the former knows the corresponding histogram for the reaction of Player B, based on the history. Assume that Player B will behave in the current situation ($S = s$) in the same manner as in similar situations in the history. Then, with sufficient prior observations, Player A can know an estimate of the probability distribution of the choices of actions by Player B, depending on the situation ($S = s$) and the choice of action by Player A. This can be done for other choices of action by Player A and other situations S. Similarly, one can estimate the joint distribution of actions of Player A and Player B. This joint distribution can be agreed upon between the players (e.g., via training). This is a form of *statistical coordination*. For many decision problems[1], it is common for engineers to formulate the goal as optimizing an expected value according to the joint distribution of actions. Hence, optimizing the latter is sufficient to solve the problem (from a theoretical standpoint). The problem of generating samples from a target joint distribution in a distributed manner—distributed coordination—is fundamental to applications ranging from multi-agent reinforcement learning to federated learning and autonomous systems [SB18, TRL21, HG24, H$^+$24a].

In the soccer example, if Players A and B are at distant locations, they may be unable to communicate at a given instant to decide on their next moves, whether verbally or non-verbally. For instance, Player A may be surrounded by players from the opposite team and may be unable to see Player B (and his actions) nor be seen by the latter[2]. This does not prevent coordination. Indeed, a characteristic of great soccer teams is the ability to get out of difficult situations. For instance, it is common for a player (e.g., Player A) on one side of the field to pass the ball to a teammate on the other side of the field (e.g., Player B). Hence, Player B will be anticipating a potential pass from Player A and ingeniously choosing his position or running speed, e.g., based on the locations of the players from the opposite team which are close to him, and based on prior knowledge about Player A acquired during training. Player A also knows, from the history of previous matches, the specific ways in which Player B will be anticipating a potential

[1]Note that this example simply serves to illustrate the engineering problem of statistical coordination. A formal presentation of decision theory is outside the scope of this thesis.

[2]Here, we do not consider the approximate position of Player A as something that is communicated to Player B, but rather as part of the state S, which consists of the locations of all payers. Communication should contain other forms of information about future actions, such as a verbal signal, e.g., "I will pass the ball sometime in the next 10 seconds", or a hand signal telling another player to move in a certain direction.

pass from Player A. For instance, Player B may anticipate the direction towards which Player A may run to get away from surrounding players, based on the positions of the latter and the knowledge of preferences and strategies of Player A acquired during training. The latter can adapt his behavior accordingly. For instance, Player A may attempt to get away from the surrounding players of the opposite team by running in a specific direction which he expects (form his training history) will be favorable to successfully pass the ball to Player B. Top teams demonstrate great aptitude at such statistical coordination without communication, which is called *non-interactive coordination*.

Last but not least, in the specific example of non-interactive coordination given above, where Players A and B cannot observe the locations of all players, receiving timely advice from other teammates or nearby coaches may enhance the coordination abilities of the two players. For instance, a coach may instruct Player A to pass the ball to Player B when observing the positions of certain players from the opposite team which Player A might not have paid attention to. Furthermore, the team captain may anticipate that the coach will give such an instruction to Player A (based on knowledge from training) and might warn Player B that a pass from Player A is imminent. The coach and the team captain may arrive at these decisions without communication. Thus, there can be two layers of non-interactive coordination: between the coach and the captain, and ultimately between Players A and B. This is the type of problem which we have decided to study. More generally, two agents (e.g., drones) may receive timely communication from two controllers having access to more information regarding the situation at hand. This is expounded upon in Chapter 5.

1.4 Thesis Contributions and Organization

Our technical contributions are detailed in Chapters 2, 3, 4 and 5, an overview of which is presented in the remainder of this section. We give concluding remarks in Chapter 6. The mathematical notation used throughout the thesis is described at the end of the present chapter. The Appendix at the end of the thesis presents relevant mathematical background. In addition,

each technical chapter ends with appendices containing details of mathematical proofs.

1.4.1 Overview of Chapter 2

As mentioned in Section 1.1.2, satisfying perfect realism constraints in lossy compression problems requires randomization, contrary to traditional RD problems. This suggests that engineers should depart from the currently dominant RD paradigm and incorporate randomization in codec design. This raises a plethora of questions. Randomization may be performed with a source of common randomness available to both the encoder and the decoder, or with private randomness locally at each terminal. In Chapter 2, we study the role of private randomness. We characterize the five-way trade-off between compression rate, common randomness rate, encoder private randomness rate, decoder private randomness rate, and distortion. We find that encoder private randomness does not impact the optimal asymptotic trade-off when the compression rate is less than the source entropy (i.e., when the compression is lossy). This has implications regarding a common practical implementation of randomization, which consists in sampling a random object — or seed — at the encoder for every new source signal, and sending that object (losslessly) to the decoder. Our results imply that, for an important class of source signals, this form of randomization does not yield superior performance compared to schemes with deterministic encoders.

The work in Chapter 2 has lead to the publication of the following conference paper:
Y. Hamdi, A. B. Wagner, and D. Gündüz. The Rate-Distortion-Perception Trade-off: the Role of Private Randomness. In *IEEE International Symposium on Information Theory*, 2024.

1.4.2 Overview of Chapter 3

As mentioned in Section 1.1.2, the focus on distortion in codecs developed with the RD paradigm is a disadvantage when the compression rate is limited, and the paradigm shift from RD theory to RDP theory has allowed to explore compression rates lower than ever before. Compression schemes based on deep neural networks are very promising for video compression, by maintain-

ing high visual quality at very low compression rates. We provide insight into a problem which bares similarities to video compression with the RDP formalism, specifically with distribution matching constraints. We consider a setup where the compression of a source signal (e.g., a video frame) is aided by another signal (e.g., the compressed file for the previous frame), correlated to the source, and available as side information. When the side information is available at both the encoder and decoder, we characterize its dual role as both a second look on the source and a source of common randomness between the encoder and decoder. We show that this critically depends on the choice of distribution matching constraint, specifically whether this requirement involves the side information. The work in Chapter 3 has lead to the publication of a conference paper and a journal paper listed below:

Y. Hamdi and D. Gündüz. The Rate-Distortion-Perception Trade-off with Side Information. In *IEEE International Symposium on Information Theory*, 2023.

Y. Hamdi, A. B. Wagner and D. Gündüz. Rate-Distortion-Perception Trade-off with Strong Realism Constraints: Role of Side Information and Common Randomness. IEEE Transactions on Information Theory, 2026, Early Access, DOI:10.1109/TIT.2026.3702694.

1.4.3 Overview of Chapter 4

As explained in Section 1.2, the distribution matching formalism for realism has significant drawbacks which call for an alternative framework. As pointed out in [The24], there is extensive algorithmic information theory literature on how to answer the question: does a given sample x come from a distribution p. Based upon this literature, we propose an information-theoretic formalism which incorporates a strong requirement of realism into RD theory. By requiring that reconstructions satisfy all computable critics — a constraint we term *algorithmic realism* — we obtain a framework that overcomes the main drawbacks of the distribution matching formalism.

Indeed, under this framework, reconstructions must satisfy not only simple statistical tests but also subtle limit theorems like the Erdös-Rényi law of runs. A reconstruction with small deficiency under all computable critics must exhibit all the complex patterns and regularities

present in natural source realizations.

The work in Chapter 4 has lead to the publication of a conference paper and a journal paper under review as listed below:

- Y. Hamdi, Aaron B. Wagner and D. Gündüz. The Rate-Distortion-Perception Trade-off with Algorithmic Realism. In *IEEE International Symposium on Information Theory*, 2025.
- Y. Hamdi, Aaron B. Wagner and D. Gündüz. The Rate-Distortion-Perception Trade-off with Algorithmic Realism. Under review for the *IEEE Transactions on Information Theory*. Preprint: arXiv:2511.15255.

1.4.4 Overview of Chapter 5

In Section 1.3, we introduced the notion of statistical coordination and non-interactive coordination. We provided an example of non-trivial coordination between two agents without communication, called non-interactive coordination. Moreover, we showed how said coordination can be improved thanks to information communicated by other parties. In Chapter 5, we study such a problem, which we term *non-interactive remote coordination*.

We first present a scheme based on the technique of channel synthesis, which is central in the broader statistical coordination literature. We develop an extension of this technique, which we term *remote channel synthesis*, which we employ to design a scheme for non-interactive remote coordination. The latter outperforms the initial scheme based on vanilla channel synthesis in certain cases.

The work in Chapter 4 has lead to the publication of a workshop paper and a conference paper listed below. The aforementioned new scheme for non-interactive remote coordination has yet to be published.

- Y. Hamdi, Xueyan Niu, Bo Bai and D. Gündüz. Non-interactive Remote Coordination. In *Workshop on Machine Learning and Compression @NeurIPS 2024*.

- Y. Hamdi and D. Gündüz. Remote Channel Synthesis. In *IEEE Information Theory Workshop*, 2025.

- Y. Hamdi and D. Gündüz. Non-interactive Remote Coordination, to be submitted to the *IEEE Transactions on Information Theory*.

1.5 Notation

Calligraphic letters such as $\mathcal{X}$ denote sets, except in $p_{\mathcal{J}}^{\mathcal{U}}$, which denotes the uniform probability distribution over set $\mathcal{J}$. The cardinality of a finite set $\mathcal{X}$ is denoted $|\mathcal{X}|$. For $\tau \in [0,1]$, we denote by $Ber(\tau)$ the Bernoulli distribution of parameter τ, i.e., the distribution on $\{0,1\}$ for which the probability of 1 is τ. Random variables are denoted using upper case letters such as X and their realizations using lower case letters such as x. For a distribution P, the expression P_X denotes the marginal of variable X while $P_X(x)$ denotes a real number, probability of the event $X = x$. Similarly, $P_{X|Y=y}$ denotes a distribution over $\mathcal{X}$, and $P_{X|Y=y}(x)$ the corresponding real number. The term *channel* is used interchangeably with the term conditional distribution. The operator "$\cdot$" is used to denote the composition of a probability distribution and a conditional distribution, e.g., $P_X \cdot P_{Y|X=x}$, or the composition of two conditional distributions, as rigorously defined in Section 1.1 of Appendix 1. The product notation $\prod$ is also used for sequential composition. The channel resulting from the composition of identical channels is called a *memoryless channel*. Given a positive integer n and a distribution p, we use $p^{\otimes n}$ to denote the distribution of a sequence of n independent random variables, each having distribution p. Any such distribution is said to be memoryless. We write *i.i.d.* as an abbreviation of *independent and identically distributed*. Given a real number τ, we denote by $\lfloor \tau \rfloor$ (resp. $\lceil \tau \rceil$) the largest (resp. smallest) integer less (resp. greater) than or equal to τ. We denote by $[a]$ the set $\{1, ..., \lfloor a \rfloor\}$ and by $\{0,1\}^*$ the set of non-empty finite strings of 0's and 1's. We denote by x^n the finite sequence $(x_1, ..., x_n)$, and use $x_{[n]\setminus k}$ to denote $(x_1, ..., x_{k-1}, x_{k+1}, ..., x_n)$, and $x_{a:b}$ to denote $(x_a, ..., x_b)$. We also use $\mathbf{x}^{(n,b)}$ to denote a batch $\{x_{1:n}^{(k)}\}_{k\in[b]}$ of b strings, each being of length n. We abbreviate $\mathbf{x}^{(1,b)}$ with $\mathbf{x}^{(b)}$. When it is clear from the context, the length of a string x is denoted by $l(x)$.

We denote the set of (strictly) positive reals by $\mathbb{R}_+$, the set of (strictly) positive integers by $\mathbb{N}$, the set of rational numbers by $\mathbb{Q}$. For two sequences $(a_n)_{n\geq 1}$ and $(b_n)_{n\geq 1}$ of positive real numbers, the notation $a_n \underset{n\to\infty}{\simeq} b_n$ means that $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. We use $\equiv$ to denote equality of distributions, e.g., $p_Y \equiv p_X$. We denote by $\|p - q\|_{TV}$ the TVD between distributions p and q (see Definition 1.3 in Appendix 1). The closure of a set is denoted either by an over-lined calligraphic letter, e.g., $\overline{\mathcal{A}}$, or using $cl(\cdot)$, e.g., $cl(\mathcal{A})$. We use base 2 for logarithms, denoted by log, and for objects defined with logarithms. If we use a logarithm in another base b, this will be made explicit via the notation $\log_b$. For random variables X and Y, we use $H(X)$, $H(X|Y)$, and $I(X;Y)$ to denote the entropy, conditional entropy, and mutual information, respectively, defined for general random variables in Section 1.2 of Appendix 1. For any nonempty finite set $\mathcal{X}$, and any distribution p on $\mathcal{X}$, we denote by $p^{\otimes *}$ the function defined on $\cup_{n\in\mathbb{N}}\mathcal{X}^n$, and such that for every $n \in \mathbb{N}$, the restriction of $p^{\otimes *}$ on $\mathcal{X}^n$ is $p^{\otimes n}$. For a finite set $\mathcal{X}$, the empirical distribution of a sequence $x_{1:n}\in\mathcal{X}^n$, denoted $\mathbb{P}_{\mathcal{X}}^{\text{emp}}(x_{1:n})$, is the distribution on $\mathcal{X}$ defined by: for any $y \in \mathcal{X}$,

$$\mathbb{P}_{\mathcal{X}}^{\text{emp}}(x_{1:n})(y) = \frac{1}{n}\big|\{t \in [n] \mid x_t = y\}\big|.$$

Given a distribution P_{X^n} on a n-fold Cartesian product $\mathcal{X}^n$, the *average empirical distribution* of random string X^n, denoted by $\hat{P}_{X^n}$, is the distribution on $\mathcal{X}$ defined by: for any measurable $A \subseteq \mathcal{X}$,

$$\hat{P}_{X^n}(A) = \frac{1}{n}\sum_{t=1}^{n} P_{X_t}(A).$$

Chapter 2

Rate-Distortion-Perception Trade-off: The Role of Private Randomness

A succinct overview of the RDP trade-off has been provided in Section 1.1 of Chapter 1, with a mention of the key role played by randomness. In this chapter, we provide a detailed mathematical formalism, in the *point-to-point* scenario, i.e., with a single source signal, a single encoder, and a single decoder. The latter encompasses both the RDP trade-off with distribution matching constraints and the classical RD trade-off. The main focus of this chapter is on the role of randomization. We provide an outline of the chapter in Section 2.2. In the next section, we provide an overview of prior work.

2.1 Prior work on the role of randomness

In conventional RD theory, the objective is to facilitate the reconstruction of a representation, denoted as $Y^n \triangleq (Y_1, ..., Y_n)$, of a source signal $X^n = (X_1, ..., X_n)$, while optimizing the proximity between the two, measured by a distortion measure $d(X^n, Y^n)$. The *asymptotic regime* $n \to \infty$ has been extensively studied, starting with the work of Claude Shannon, who characterized the optimal asymptotic trade-off between rate and distortion (see Section 2.3 for a formal definition) under certain assumptions on the source signal and distortion (see

Section 2.3). This theory has provided the mathematical foundation for understanding the trade-offs in lossy source coding for more than half a century. Under that framework, it is known that deterministic encoders and decoders are sufficient to achieve the optimal asymptotic RD performance — as well as the optimal performance for *one-shot* ($n = 1$) fixed-length (see Definition 2.3) schemes [Han03]. The systematic study of the impact of perceptual quality constraints and randomization on rate and distortion originated with the works of Li *et al.* [LKK10, LKK11, LKOK12, KZLK13] and Saldi *et al.* [SLY15b, SLY15a]. Therein, perceptual quality, or *realism*, is formalized by requiring the distribution of the reconstruction Y^n to be identical to that of the source, or asymptotically arbitrarily close in TVD. See also the work of Delp *et al.* [DM91]. A more recent line of work lead to the introduction [BM19] — see also [BM18, Mat18a, Mat19] — of a more general distribution-preservation constraint: the problem is then to characterize the optimal rate for which both distortion constraint $d(X^n, Y^n) \leq \Delta$, and realism constraint $\mathcal{D}(P_{X^n}, P_{Y^n}) \leq \lambda$ are met, where $\mathcal{D}$ is a similarity measure, e.g., the TVD or some other divergence. The three-way trade-off between Δ, λ, and the rate, was coined *the RDP* trade-off. We call the above the *strong realism constraint*, and *imperfect strong realism* constraint when $\lambda > 0$ — see also [SCKY24] for a variant. The following weaker version has been recently studied in [CYW+22], where it is called "weak" realism:

$$\forall 1 \leq t \leq n, \quad \mathcal{D}(P_{Y_t}, p_X) \leq \lambda, \tag{2.1}$$

We call this *per-symbol realism*. Chapter 4 is dedicated to the limitations of such definitions for realism, and to alternative formulations.

The different forms of randomness include private randomness at each of the encoder and decoder, and common randomness, available at both terminals. The case where unlimited common randomness is available is well-studied, as described in the remainder of this paragraph. For strong realism, a characterization of the optimal RDP trade-off was proved in [SLY15b] in the case of the *perfect strong realism* constraint $P_{Y^n} \equiv p_X^{\otimes n}$ ($\lambda = 0$), for general distributions p_X. A characterization in the general case ($\lambda \geq 0$) is proved in [CYW+22] for a class of similarity measures. One-shot results are also proved in [KZLK13, TW21, ZQCK21, QZCK22],

[SPC+23]. Note that private sources of randomness are not useful when common randomness is unlimited, since private randomness can be extracted from common randomness. In that case, the optimal trade-off for the strong perfect realism constraint is characterized by the following function [SLY15b], [Wag22]

$$R(\Delta) = \inf_{p_{Y|X}} I(X;Y), \tag{2.2}$$

where the infimum is over all conditional distributions $p_{Y|X}$ such that $\mathbb{E}[d(X,Y)] \leq \Delta$ and $p_Y \equiv p_X$. A formal treatment is provided in subsequent sections of this chapter. As shown in [SLY15a], [Wag22], it turns out that the above RDP function only applies when common randomness is available in a sufficiently large amount — see subsequent sections for details —, even with unconstrained encoder and decoder private randomness. Thus, some care is required when interpreting (2.2) operationally. Moreover, (2.2) also characterizes the optimal trade-off for the *perfect per-symbol realism* constraint $\forall t \in [n], P_{Y_t} \equiv p_X$ (instead of perfect strong realism) in the case of unlimited common randomness [TW21]. However, as proved in [CYW+22], unlimited private randomness is sufficient to achieve the optimal trade-off under that constraint.

The present chapter thoroughly examines the case of limited randomness in the regime $\lambda \to 0$ (see Section 2.2 to follow). While we examine all possible scenarios for the amount of available private and common randomness, there exists prior work on specific scenarios, mentioned in the remainder of this paragraph. The case of limited common randomness has also been studied in [SLY15a] when no decoder private randomness is available. As mentioned in that paper, the corresponding result implies that decoder private randomness and common randomness have a complementary role in the case of perfect strong realism. The case of absent common randomness has also been studied with imperfect strong realism constraints ($\lambda > 0$). See [Mat18a, Mat19] for characterizations with such constraints when neither common randomness nor decoder private randomness are available. See also [CYW+22] for sufficient conditions when decoder private randomness is available but not common randomness. These conditions are also necessary when the reconstruction Y^n is constrained to be i.i.d. [CYW+22]. This concludes

our review of prior literature. The scope of our work is presented in the next section.

2.2 Outline

In the present chapter, we delve deeper into the role of private randomness. We completely characterize the five-way trade-off between compression rate, common randomness rate, encoder private randomness rate, decoder private randomness rate and distortion (see Section 2.3), thereby extending previous results. We consider a source signal that is an i.i.d. sequence $X^n \sim p_X^{\otimes n}$ (see Section 2.3) and the *near-perfect strong realism* and *near-perfect per-symbol realism* constraints:

$$\|P_{Y^n} - p_X^{\otimes n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0 \tag{2.3}$$

$$\max_{1\le t\le n} \|P_{Y_t} - p_X\|_{TV} \underset{n\to\infty}{\longrightarrow} 0, \tag{2.4}$$

where $\|\cdot\|_{TV}$ is the TVD. See Remark 2.12 for an extension of our results to perfect strong (resp. per-symbol) realism constraints when decoder private randomness is unlimited. We first introduce a novel result — which we call a soft covering lemma — regarding the private randomness of stochastic compressors. Then, we show that whether encoder private randomness is available does not impact the optimal asymptotic trade-off between rate and distortion. This holds whatever the resources in terms of common randomness and decoder private randomness are. This implies that in the absence of common randomness, it is not useful, in the limit of large blocklength n, that the encoder include in its message a random seed for a pseudo-random number generator. In other words, the only useful form of shared randomness is common randomness available without extra communication.

The RDP trade-off has strong ties to the channel synthesis problem, a.k.a. reverse channel coding, a.k.a. channel simulation, which we cover in Chapter 5. A channel synthesis scheme was used to prove the first complete characterization [TW21] regarding the RDP trade-off with imperfect realism. In the present chapter, we use proof techniques from the channel synthesis literature: our proofs track those of [Cuf13] (see also [Wag22]), and we use several of the soft

covering lemma variants therein.

The chapter is organized as follows. We give the problem formulation in Section 2.3, and introduce a key lemma in Section 2.4. We present all our results in Section 2.5, and provide a partial proof in Section 2.6. The rest of the proofs is provided in the Appendix.

2.3 Problem formulation

A cornerstone in information theory is the work of Claude Shannon (see e.g. [Sha48]). He proposed a probabilistic formalism for information and its processing, where any input signal is modeled as a random sequence. For example, a choice of text among all possible English texts is modeled as choosing at random a realization of a random sequence of words according to a certain probability distribution — which quantifies how likely a block of consecutive words is, given the previous block. All his results are based on properties of sequences of i.i.d. random variables or more generally stationary ergodic sequences ([CT06, Section 16.8]). An element of a sequence is called a *symbol*, and the set of all possible values of a random symbol is called its *alphabet*. The term *general alphabets* refers to all Polish spaces (see Section 1.1 of Appendix 1), which include all finite sets and all real vector spaces. The terms *set*, *alphabet*, and *(measurable) space* are used interchangeably. Shannon showed the superiority of processing large blocks of consecutive symbols jointly, over processing individual symbols separately. In this section, we present a mathematical formalism which encompasses both Shannon's RD theory, and the RDP trade-off with near-perfect strong realism or per-symbol realism constraints (Eq. 2.3 and 2.4).

Definition 2.1 *Given a space $\mathcal{X}$, a distortion measure is a measurable function $d : \mathcal{X}^2 \to [0, \infty)$ extending to sequences as*

$$d(x^n, y^n) = \tfrac{1}{n} \textstyle\sum_{t=1}^{n} d(x_t, y_t).$$

Definition 2.3 below formalizes the processing of an input signal, involving an encoder a decoder, as shown in Figure 2.1. The compressed file, or *message* M is modeled as an integer that can take at most 2^{nR} different values. This is a formal way of modeling an object that

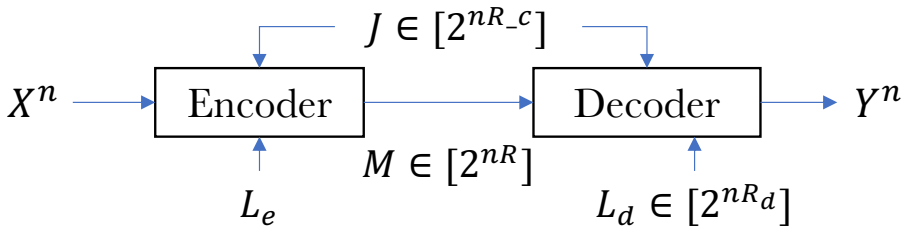


Figure 2.1: The system model, in the case of a fixed-length code with finite rates (Definition 2.3).

admits a binary representation involving at most nR binary digits (*bits*, i.e., 0's and 1's). We call R the *compression rate*, which measures the number of bits used per symbol of the source sequence X^n. Either the encoder private randomness is unconstrained, or it is completely unavailable. We consider common randomness and decoder private randomness variables J and L_d that can take at most 2^{nR_c} and 2^{nR_d} different values, respectively. Parameters R_c and R_d are called the *common randomness rate* and *decoder private randomness rate*, respectively. Definitions 2.2 and 2.3 also contain variants of the aforementioned formalism, which we consider in order to show the robustness of our results: we prove that all our main results hold for all these variants — see subsequent sections. Definitions 2.2 and 2.3 apply to all point-to-point lossy compression problems, including those without distortion nor realism constraints, such as point-to-point problems mentioned in Chapter 5 — except that in the latter the decoder output sequence Y^n is not required to have the same alphabet as the encoder input X^n. Randomness plays a key role in the RDP trade-off but is unnecessary to achieve optimal RD trade-off — see Proposition 2.1 to follow.

Definition 2.2 (Variable-length code) *Given a space $\mathcal{X}$, a positive integer n, and a triplet $(R, R_c, R_d) \in \mathbb{R}_{\geq 0} \times \overline{\mathbb{R}}^2_{\geq 0}$, an (n, R, R_c, R_d) variable-length code is a tuple $(p_J, p_{L_e}, p_{L_d}, F^{(n)}, G^{(n)})$ consisting of distributions p_J, p_{L_e}, p_{L_d} on some Polish spaces $\mathcal{J}, \mathcal{L}_e, \mathcal{L}_d$, a deterministic encoder*

$$F^{(n)} : \mathcal{X}^n \times \mathcal{J} \times \mathcal{L}_e \to \mathcal{M},$$

with some Polish message space $\mathcal{M}$, and a deterministic decoder

$$G^{(n)} : \mathcal{M} \times \mathcal{J} \times \mathcal{L}_d \to \mathcal{X}^n,$$

setting $M := F^{(n)}(X^n, J, L_e)$ *and* $Y^n := G^{(n)}(M, J, L_d)$, *that satisfy*

$$R \geq \tfrac{1}{n}H(M), \quad R_c \geq \tfrac{1}{n}H(J), \quad R_d \geq \tfrac{1}{n}H(L_d).$$

Given a distribution p_X *on* $\mathcal{X}$, *the distribution induced by the code is*

$$P_{X^n,J,L_e,M,L_d,Y^n} := p_X^{\otimes n} \cdot p_J \cdot p_{L_e} \cdot p_{L_d} \cdot \mathbf{1}_{M=F^{(n)}(x^n,j,l_e)} \cdot \mathbf{1}_{Y^n=G^{(n)}(m,j,l_d)}.$$

Note that the product form of the induced distribution above is part of the definition: conditionally on their respective inputs, the encoder's and the decoder's private randomizations are mutually independent, and independent of (X^n, Z^n). *In particular, the kernels* $F^{(n)}$ *and* $G^{(n)}$ *carry no common randomness beyond* J.

Definition 2.3 (Fixed-length code) *Given a triplet* $(R, R_c, R_d) \in \mathbb{R}^3_{\geq 0}$, *an* (n, R, R_c, R_d) *fixed-length code is a variable-length code (Definition 2.2) such that*

$$M \in [2^{nR}] \text{ with probability } 1, \quad p_J \equiv p^{\mathcal{U}}_{[2^{nR_c}]}, \quad p_{L_d} \equiv p^{\mathcal{U}}_{[2^{nR_d}]},$$

where (recall Section 1.5) $[2^{nR}] := \{1, \ldots, \lfloor 2^{nR} \rfloor\}$ *and* $\equiv$ *denotes equality of distributions. Similarly, an* (n, R, R_c, ∞) *fixed-length code is a variable-length code with* $M \in [2^{nR}]$ *with probability* 1 *and* $p_J \equiv p^{\mathcal{U}}_{[2^{nR_c}]}$. *An* (n, R, R_c, R_d) *fixed-length (resp. variable-length) code with non-privately randomized encoding is one for which* L_e *is constant with* P*-probability* 1.

Note that if any of R, R_c, R_d is null, then the corresponding random variable (M or J or L_d) is constant P-almost surely — see e.g. the discussion in [Gra11] above Lemma 7.18. Moreover, no constraint (alphabet, distribution, rate) is imposed on the encoder private randomness of a fixed-length (resp. variable-length) code. As alluded to previously, processing multiple symbols jointly leads to better performance. The larger the *blocklength* n (length of the input sequence) the better. The aim of RD and RDP theory is to study certain perfoamance limits, in the asymptotic setting where the length of the input sequence tends to infinity. This is formalized in the next definition.

Definition 2.4 *Consider a space $\mathcal{X}$, a distribution p_X on $\mathcal{X}$, and a distortion measure d. A tuple $(R, R_c, R_d, \Delta) \in \mathbb{R}_{\geq 0} \times \overline{\mathbb{R}}_{\geq 0}^3$ is said to be achievable with near-perfect realism with fixed-length (resp. variable-length) codes with privately randomized (resp. non-privately randomized) encoding if there exists a sequence of (n, R, R_c, R_d) fixed-length (resp. variable-length) codes with privately randomized (resp. non-privately randomized) encoding with induced distributions $\{P^{(n)}\}_n$ such that*

$$\begin{aligned} \limsup_{n\to\infty} \mathbb{E}_{P^{(n)}}[d(X^n, Y^n)] &\leq \Delta, \\ \|P_{Y^n}^{(n)} - p_X^{\otimes n}\|_{TV} &\underset{n\to\infty}{\longrightarrow} 0. \end{aligned} \tag{2.5}$$

For each of the above notions of achievability, we introduce the corresponding notion of achievability with per-symbol near-perfect realism, defined by replacing (2.5) *by*

$$\max_{t\in[n]} \|P_{Y_t}^{(n)} - p_X\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{2.6}$$

The similar notions of achievability with perfect realism or perfect per-symbol realism are defined by replacing (2.5) *by:*

$$\begin{aligned} &\exists N \in \mathbb{N}, \forall n \geq N,\ P_{Y^n}^{(n)} \equiv p_X^{\otimes n} \quad \textit{or} \\ &\exists N \in \mathbb{N}, \forall n \geq N, \forall t \in [n],\ P_{Y_t}^{(n)} \equiv p_X. \end{aligned} \tag{2.7}$$

This suggests the existence of a *trade-off*: certain combinations of (R, R_c, R_d, Δ) are achievable, and other combinations are not. For instance, reducing the compression rate R may lead to higher distortion — e.g., the more one compresses an image, the less it looks like the original. Removing the realism constraint from the above definition yields similar notions of *achievability with no realism constraint*, which constitute the formalism for the traditional RD trade-off. For the latter, randomness is irrelevant, as stated in the result below, which can be readily deduced from the RD theory literature — e.g., from the proof of [CT06, Theorem 10.2.1] in [CT06, Sections 10.4 & 10.5].

Proposition 2.1 *Assume that $\mathcal{X}$ is finite and that d is bounded. Define the RD region $\mathcal{S}$ as*

$$\left\{ \begin{array}{rcl} (R, \Delta) & : & \exists\, p_{Y|X}, \\ R & \geq & I_p(X;Y) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}. \tag{2.8}$$

Then, the set of tuples (R, R_c, R_d, Δ) achievable with no realism constraint with fixed-length (resp. variable-length) codes with privately randomized (resp. non-privately randomized) encoding has closure

$$\{(R, R_c, R_d, \Delta) \mid (R, \Delta) \in \mathcal{S}\}. \tag{2.9}$$

Remark 2.2 *The region $\mathcal{S}$ is closed. Indeed, $\mathcal{X}$ is finite and reconstructions take values in $\mathcal{X}$, so the set of admissible $p_{Y|X}$ is compact; $I_p(X;Y)$ and $\mathbb{E}_p[d(X,Y)]$ are continuous in $p_{Y|X}$; and the inequalities defining $\mathcal{S}$ are non-strict.*

Such a characterization of a set of achievable tuples is called a *coding theorem*, and specifically a *single-letter* characterization: it does not involve any blocklength n, unlike the corresponding problem. Its different components can be described with the following terminology. A property stating that a set of achievable tuples contains another set is called an *inner bound* or an *achievability* property, and a property stating that a set of achievable tuples is contained in another set is called an *outer bound*.

As argued in [BM19], the use of the TVD to formalize the notion of near-perfect realism in (2.5) is related to the optimal performance of a critic (hypothesis test), as follows. Consider a fixed blocklength n and a distribution P induced by a code. A critic is given either a sample from P_{Y^n}, i.e., an *artificial* sample, or from $p_X^{\otimes n}$, i.e., an *original* sample, such that the distribution of the sample is $\frac{1}{2}P_{Y^n} + \frac{1}{2}p_X^{\otimes n}$. Then, the probability for a critic (hypothesis test) to correctly classify a sample from that mixture, as original or artificial, is at most (e.g., [Nie13])

$$\frac{1}{2} + \frac{1}{2}\|P_{Y^n} - p_X^{\otimes n}\|_{TV}. \tag{2.10}$$

Hence, when the TVD is very small, any critic's performance is essentially no better than that of a coin toss.

In order to study the role of randomness for achievability with near-perfect strong realism or per-symbol realism, we propose a general result regarding randomized lossy compression.

2.4 A soft covering lemma for randomized compressors

Before stating the result, we clarify its scope. It concerns the encoding only: it produces a non-privately randomized encoder that induces nearly the same joint law of (X^n, M, J) and of the empirical distribution of the source and codewords, as measured by the empirical-distribution operator $\hat{P}$ (recall Section 1.5). No reconstruction Y^n and no decoder appear in the statement — the guarantees are on (X^n, M, J) and the codewords alone; any fixed decoder may be composed with the new encoder afterwards, which is how the result is used in Section 2.6.

Proposition 2.3 *Consider a finite input alphabet $\mathcal{X}$, a distribution p_X on $\mathcal{X}$, non-negative reals R, R_c and a sequence $\{F^{(n)}\}_{n\geq 1}$ of encoders corresponding to a sequence of (n, R, R_c, ∞) fixed-length codes. The n-th induced distribution is denoted $P^{(n)}$. If $R < H(X)$, then for any finite alphabet $\mathcal{V}$ and any sequence of deterministic mappings $\mathbf{v}^{(n)} : [2^{nR}] \times [2^{nR_c}] \to \mathcal{V}^n$, there exists a sequence of deterministic maps*

$$
\begin{aligned}
&f^{(n)} : \mathcal{X}^n \times [2^{nR_c}] \to [2^{nR}], \quad \textit{such that} \\
&\left\| \hat{\tilde{P}}^{(n)}_{\mathcal{X}\times\mathcal{V}}[X^n, \mathbf{v}^{(n)}(M,J)] - \hat{P}^{(n)}_{\mathcal{X}\times\mathcal{V}}[X^n, \mathbf{v}^{(n)}(M,J)] \right\|_{TV} \underset{n\to\infty}{\longrightarrow} 0, \\
&\left\| \tilde{P}^{(n)}_{M,J} - P^{(n)}_{M,J} \right\|_{TV} \underset{n\to\infty}{\longrightarrow} 0, \quad \textit{where} \\
&\tilde{P}^{(n)}_{X^n,J,M} := p_X^{\otimes n} \cdot p^{\mathcal{U}}_{[2^{nR_c}]} \cdot \mathbf{1}_{M=f^{(n)}(X^n,J)}
\end{aligned}
\tag{2.11}
$$

(recall the empirical-distribution operator $\hat{P}$ defined in Section 1.5).

In other words, any sequence of compressors, potentially involving private randomness, can be replaced by a sequence of compressors that do not make use of private randomness, with a

minor impact on the induced distribution (for large enough blocklength n). The result follows rather directly from applying the soft covering lemma with a sequence of general sources and channels [Cuf13, Corollary VII.3]. We provide a proof in Appendix 2.A.

Remark 2.4 *In classical rate-distortion theory, encoder randomness is shown to be unnecessary by a simple averaging argument: since the distortion is affine in the code, a random code with average distortion at most Δ admits a deterministic realization with distortion at most Δ. This argument does not carry over to the rate-distortion-perception setting. The realism constraint $P_{Y^n} \approx p_X^{\otimes n}$ constrains the* output distribution*, which for a randomized code is the mixture over the code's randomness; it may hold for that mixture while failing for every individual realization, because fixing the code concentrates the output on finitely many reconstructions and destroys the match to $p_X^{\otimes n}$. One therefore cannot fix the code to control distortion while preserving realism: the code randomness itself acts as the (common) randomness on which realism relies. This is why dispensing with encoder private randomness requires the soft covering argument of Proposition 2.3 rather than an expectation argument, and it indicates that the role of randomness in RDP theory is genuinely different from that in classical RD theory. Whether an analogous phenomenon arises for open problems on the role of randomness in classical RD theory is an interesting direction for future work.*

Remark 2.5 *Consider the setting of Proposition 2.3 and let $\{G^{(n)}\}_{n\geq 1}$ be the sequence of decoders in the initial codes. Assume that there exists a finite alphabet $\mathcal{V}$, a conditional probability $P_{Y|V}$ from $\mathcal{V}$ to $\mathcal{X}$, and a sequence $\{\mathbf{v}^{(n)}\}_{n\geq 1}$ of deterministic maps with $\mathbf{v}^{(n)} : [2^{nR}] \times [2^{nR_c}] \to \mathcal{V}^n$, such that for any (m, j) we have*

$$G^{(n)}_{Y^n|M=m,J=j} \equiv \prod_{t=1}^{n} p_{Y|V=\mathbf{v}_t(m,j)}. \tag{2.12}$$

Then, — e.g. Lemma 1.9 (Appendix 1) —, sequence $\{f^{(n)}\}_{n\geq 1}$ of Proposition 2.3 corresponding

to $\{\mathbf{v}^{(n)}\}_{n\geq 1}$ *satisfies*

$$\begin{gathered}\big\|\hat{\tilde{P}}^{(n)}_{\mathcal{X}^2\times\mathcal{V}}[X^n,Y^n,\mathbf{v}^{(n)}(M,J)]-\hat{P}^{(n)}_{\mathcal{X}^2\times\mathcal{V}}[X^n,Y^n,\mathbf{v}^{(n)}(M,J)]\big\|_{TV}\underset{n\to\infty}{\longrightarrow}0\\ \big\|\tilde{P}^{(n)}_{M,J,Y^n}-P^{(n)}_{M,J,Y^n}\big\|_{TV}\underset{n\to\infty}{\longrightarrow}0,\quad \textit{where}\\ \tilde{P}^{(n)}_{Y^n|M,J,X^n}\equiv P^{(n)}_{Y^n|M,J,X^n}\equiv G^{(n)}_{Y^n|M,J}.\end{gathered}\tag{2.13}$$

Remark 2.6 *The restriction $R < H(X)$ in Proposition 2.3 reflects a genuine limitation of the de-randomization mechanism, which can be seen most simply in the case of no common randomness ($R_c = 0$). The mechanism dispenses with encoder private randomness by drawing, in its place, on the randomness that the source sequence still carries once the message is fixed. When $R < H(X)$, each message is the image of exponentially many source sequences, so this residual randomness is abundant. When $R \geq H(X)$, the number of messages is no smaller than the number of typical source sequences, so the encoder may associate a distinct message to (almost) each such sequence; the source sequence is then essentially determined by the message, and no residual randomness remains to take the place of the encoder private randomness.*

2.5 The rate-distortion-perception trade-off with near-perfect realism

2.5.1 The role of encoder private randomness for finite source alphabets

We have the following characterization, which is an extension of [SLY15a, Theorems 1 & 5]. The achievability is proved in Section 2.6, and the converse in Appendix 2.B.

Theorem 2.7 *Consider a finite source alphabet $\mathcal{X}$, a distribution p_X on $\mathcal{X}$ and a distortion*

measure d. Define the region $\mathcal{S}_D$ of $\overline{\mathbb{R}}^4_{\geq 0}$ as

$$\left\{ \begin{array}{rcl} (R, R_c, R_d, \Delta) & : & \exists\, p_{X,V,Y} \in \mathcal{D}_D, \\ R & \geq & I_p(X;V) \\ R + R_c & \geq & I_p(Y;V) \\ R_d & \geq & H_p(Y|V) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}, \tag{2.14}$$

with $\mathcal{D}_D$ defined as

$$\left\{ \begin{array}{c} p_{X,V,Y} : X \sim p_X,\ p_Y \equiv p_X \\ X - V - Y \\ |\mathcal{V}| \leq |\mathcal{X}|^2 + 1 \end{array} \right\}. \tag{2.15}$$

Denote by $\mathcal{A}_D^{(f)}$ the set of $(R, R_c, R_d, \Delta) \in \mathbb{R}^4_{\geq 0}$ achievable with near-perfect realism with fixed-length codes and by $\mathcal{A}_D^{(u)}$ the set of tuples $(R, R_c, R_d, \Delta) \in \mathbb{R}_{\geq 0} \times \overline{\mathbb{R}}^3_{\geq 0}$ achievable with near-perfect realism with variable-length codes. Denote by $\mathcal{A}_D^{(f,*)}$ the set of tuples (R, R_c, R_d, Δ) in $\mathbb{R}^4_{\geq 0}$ achievable with near-perfect realism with fixed-length codes and non-privately randomized encoding and by $\mathcal{A}_D^{(u,*)}$ the set of tuples $(R, R_c, R_d, \Delta) \in \mathbb{R}_{\geq 0} \times \overline{\mathbb{R}}^3_{\geq 0}$ achievable with near-perfect realism with variable-length codes and non-privately randomized encoding. Then,

1. *The aforementioned sets have identical closures in $\overline{\mathbb{R}}^4_{\geq 0}$:*

$$\overline{\mathcal{A}_D^{(f)}} = \overline{\mathcal{A}_D^{(u)}} = \overline{\mathcal{A}_D^{(f,*)}} = \overline{\mathcal{A}_D^{(u,*)}} = \mathcal{S}_D. \tag{2.16}$$

2. *The same holds if each notion of achievability is replaced by the corresponding achievability with no common randomness ($R_c = 0$) and $\mathcal{S}_D$ is replaced by its intersection with the hyperplane $R_c = 0$.*

3. *The same holds if each notion of achievability is replaced by the corresponding achievability with no decoder private randomness ($R_d = 0$) and $\mathcal{S}_D$ is replaced by its intersection with the hyperplane $R_d = 0$.*

Remark 2.8 *The region $\mathcal{S}_D$ is closed. Indeed, $\mathcal{X}$ is finite and, since mutual information is invariant under relabeling of V, one may restrict V to a fixed set of cardinality $|\mathcal{X}|^2+1$; the set of admissible $p_{X,V,Y}$ is then compact, $I_p(X;V)$ and $I_p(Y;V)$ are continuous in $p_{X,V,Y}$, and the inequalities defining $\mathcal{S}_D$ are non-strict.*

Consequently, for compression with a near-perfect realism constraint, encoder private randomness is not useful whatever the available resources in terms of common randomness and decoder private randomness are. In particular, this holds even if the latter two sources of randomness are not rate-limited or not discrete.

2.5.2 An extension to sources with infinite entropy

Following [Wag22], we use the following assumption in order to handle general alphabets.

Definition 2.5 *[Wag22] Given a source alphabet $\mathcal{X}$, a probability distribution p on $\mathcal{X}$ and a distortion measure d, we say that (d,p) is uniformly integrable if for every $\varepsilon > 0$ there is a $\tau > 0$ such that*

$$\sup_{\mathbb{P}_{X,Y,\xi}} \mathbb{E}[d(X,Y)\cdot\xi] \leq \varepsilon,$$

where the supremum is taken over all distributions $\mathbb{P}_{X,Y,\xi}$ on $\mathcal{X}^2\times\{0,1\}$ satisfying $\mathbb{P}_X \equiv \mathbb{P}_Y \equiv p$ and $\mathbb{P}(\{\xi=1\}) \leq \tau$.

The notion of uniform integrability is reminiscent of a notion in probability theory of the same name that applies to a family of random variables defined over a common probability space. The latter has arisen in RD theory before [Han03, (5.3.2))]. Following [Wag22], an extension of Definition 2.5 — and of Claim 2.9 below — was used in [CYW+22, GPC24, XLCZ25].

Assuming uniform integrability incurs some loss of generality in that there exist (d,p) pairs that are not uniformly integrable yet have a well-defined and finite RD function (see Appendix 2.F.4). Naturally occurring examples are usually uniformly integrable, however. To

wit, (d, p) is uniformly integrable if $\mathcal{X}$ is finite and d does not take infinite values. Moreover, we have:

Claim 2.9 *[Wag22] If $\mathcal{X} \subseteq \mathbb{R}$ and d is the MSE distortion measure, then (d, p) is uniformly integrable if p has a finite fourth moment, e.g., if p is a one-dimensional Gaussian distribution.*

Our approach for turning a privately randomized encoder into a non-privately randomized one requires to work with finite alphabets. To that end, we introduce the following notion, which involves a standard formalism for the notion of arbitrarily fine quantization. A *quantizer* on a measurable space $\mathcal{X}$ is any measurable finite-valued map from $\mathcal{X}$ onto itself.

Definition 2.6 *Consider a source alphabet $\mathcal{X}$, a σ-algebra $\mathcal{B}$ of subsets of $\mathcal{X}$, a probability distribution p on $(\mathcal{X}, \mathcal{B})$ and a distortion measure d. We say that (d, p) is quantizable if the following holds: there exists a sequence $\{\kappa^{(\ell)}\}_{\ell \in \mathbb{N}}$ of quantizers of $\mathcal{X}$ such that the corresponding partitions asymptotically generate $\mathcal{B}$, and for any $\tau, \varepsilon > 0$,*

$$\exists B_\tau \in \mathcal{B}, \quad p(\mathcal{X} \setminus B_\tau) \leq \tau, \tag{2.17}$$

and there exists $L_{\varepsilon,\tau}$ such that for any $\ell \geq L_{\varepsilon,\tau}$,

$$\forall (x, y) \in B_\tau^2, \quad \left| d(x, y) - d\big(\kappa^{(\ell)}(x), \kappa^{(\ell)}(y)\big) \right| \leq \varepsilon, \tag{2.18}$$

$$\forall (x, y) \in \mathcal{X}^2 \setminus B_\tau^2, \quad d\big(\kappa^{(\ell)}(x), \kappa^{(\ell)}(y)\big) \leq d(x, y) + \varepsilon. \tag{2.19}$$

This property is satisfied if $\mathcal{X}$ is finite, and we have the following claim.

Claim 2.10 *If $\mathcal{X}$ is a finite-dimensional real vector space and d denotes the Euclidean distance, then for any distribution p on $\mathcal{X}$ and any $s > 0$, tuple (d^s, p) is quantizable.*

We provide a proof of Claim 2.10 in Appendix 2.F.3. We have the following characterization, which is an extension of [SLY15a, Theorem 1] and [Wag22, Theorem 2]. The proof is provided in Appendix 2.D.

Theorem 2.11 *Consider a Polish source alphabet $\mathcal{X}$, a distribution p_X on $\mathcal{X}$ having infinite entropy, and a distortion measure d such that (d, p_X) is uniformly integrable and quantizable. Define the region $\mathcal{S}_G$ as*

$$\left\{ \begin{array}{rcl} (R, R_c, \Delta) \in \mathbb{R}^3_{\geq 0} & : & \exists\ p_{X,V,Y} \in \mathcal{D}_G \text{ s.t.} \\ R & \geq & I_p(X;V) \\ R + R_c & \geq & I_p(Y;V) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}, \tag{2.20}$$

with $\mathcal{D}_G$ defined as

$$\left\{ \begin{array}{c} p_{X,V,Y} : X \sim p_X,\ p_Y \equiv p_X \\ X - V - Y \end{array} \right\}, \tag{2.21}$$

where the alphabet of V is constrained to be finite. Denote by $\mathcal{A}_G$ the set of triplets $(R, R_c, \Delta) \in \mathbb{R}^3_{\geq 0}$ such that (R, R_c, ∞, Δ) is achievable with near-perfect realism with fixed-length codes, and by $\mathcal{A}_G^{(f,*)}$ the set of triplets $(R, R_c, \Delta) \in \mathbb{R}^3_{\geq 0}$ such that (R, R_c, ∞, Δ) is achievable with near-perfect realism with fixed-length codes and non-privately randomized encoding. Then,

1. *The aforementioned sets have identical closures in $\mathbb{R}^3_{\geq 0}$:*

$$\overline{\mathcal{A}_G^{(f)}} = \overline{\mathcal{A}_G^{(f,*)}} = \overline{S_G}. \tag{2.22}$$

2. *The same holds if each notion of achievability is replaced by the corresponding achievability with no common randomness ($R_c = 0$) and $\mathcal{S}_G$ is replaced by its intersection with the hyperplane $R_c = 0$.*

The assumption of unlimited decoder private randomness is not very restrictive as far as the study of encoder private randomness is concerned. Indeed, for most general alphabets of interest, the TVD between a distribution having finite entropy and a distribution having infinite entropy is equal to 1. Hence, if the source distribution has infinite entropy, then achievability with near-perfect realism requires either R_d or R_c to be infinite (assuming R is finite). Moreover,

the case of unconstrained common randomness is trivial: there is no use for local randomness, since unlimited randomness can be extracted from the common randomness. We conjecture that the assumptions of finite-valued auxiliary variable V and fixed-length codes are not restrictive.

Remark 2.12 *The scheme used in the proof of Theorem 2.11 achieves perfect strong realism, instead of mere near-perfect strong realism. Note that this requires unlimited variable-length decoder private randomness. Hence, the statement of Theorem 2.11 would remain true if the term "near-perfect" were replaced by the term "perfect". A similar equivalence holds for per-symbol realism, regarding Theorem 2.13 to follow, as implied by our proofs.*

2.5.3 Per-symbol realism

The following theorem is an extension of [CYW+22, Theorem 4], which states that under a perfect per-symbol realism constraint, whether common randomness is available does not impact the optimal asymptotic trade-off between rate and distortion. We find that the same can be said of encoder private randomness, if only near-perfect per-symbol realism is required.

Theorem 2.13 *Theorems 2.7 and 2.11 also hold if achievability with near-perfect realism is replaced by achievability with per-symbol near-perfect realism, and each lower bound on $R + R_c$ is removed in $\mathcal{S}_D$ and $\mathcal{S}_G$.*

A proof is provided in Appendix 2.E. Our result also complements [CYW+22, Theorem 2], which states that under the per-symbol realism constraint (2.1), if $\lambda > 0$, then fully determinitic codes are sufficient to achieve the optimal asymptotic trade-off.

2.6 Achievability proof for Theorem 2.7

2.6.1 Modifying a standard code construction

A rather straightforward adaptation of the proof of [Wag22, Theorem 2] yields the following result. See Appendix 2.F.1 for a proof.

Proposition 2.14 *Consider finite alphabets $\mathcal{X}, \mathcal{V}$, a distortion measure d on $\mathcal{X}^2$, a triplet (R, R_c, Δ) in $\mathbb{R}^3_{\geq 0}$, and a distribution $p_{X,Y,V}$ on $\mathcal{X}^2 \times \mathcal{V}$. Assume that $p \in \mathcal{D}_D$ and*

$$R \geq I_p(X;V),\ R + R_c \geq I_p(Y;V),\ \Delta \geq \mathbb{E}_p[d(X,Y)].$$

Then, there exists a sequence $\{\varepsilon_n\}_{n\geq 1}$ of positive reals and a sequence of $(n, R+\varepsilon_n, R_c, \infty)$ fixed-length codes inducing distributions $\{P^{(n)}_{X^n,J,M,Y^n}\}_{n\geq 1}$ such that $\varepsilon_n \underset{n\to\infty}{\to} 0$ and

$$\limsup_{n\to\infty} \mathbb{E}_{P^{(n)}}[d(X^n, Y^n)] \leq \Delta \tag{2.23}$$

$$\|P^{(n)}_{Y^n} - p_X^{\otimes n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0 \tag{2.24}$$

$$\forall (m,j),\ P^{(n)}_{Y^n|M=m,J=j} = \textstyle\prod_{t=1}^n p_{Y|V=v_t^{(n)}(m,j)}, \tag{2.25}$$

$$P^{(n)}\big(\big\|\mathbb{P}^{emp}_{\mathcal{V}}(v^{(n)}(M,J)) - p_V\big\|_{TV} \geq \varepsilon_n\big) \underset{n\to\infty}{\longrightarrow} 0, \tag{2.26}$$

for some sequence of deterministic functions $v^{(n)} : [2^{n(R+\varepsilon)}] \times [2^{nR_c}] \to \mathcal{V}^n$.

As a direct consequence of Proposition 2.3 and Remark 2.5, we can state the following.

Proposition 2.15 *For $R < H(X)$, Proposition 2.14 holds with a sequence of distributions $\{P^{(n)}\}_{n\geq 1}$ such that the n-th is induced by a $(n, R+\varepsilon_n, R_c, \infty)$ fixed-length code with non-privately randomized encoding.*

Interestingly, the proof of Proposition 2.14 (and [Wag22, Theorem 2]) involves the use (twice) of the standard soft covering lemma [Cuf13, Lemma VII.4]. As mentioned in Section 2.4, Propo-

sition 2.3 follows from a more general version of the soft covering lemma [Cuf13, Corollary VII.3]. Hence, Proposition 2.15 follows from multiple uses of soft covering lemmas.

2.6.2 Achievability of Theorem 2.7

By definition, for each of the three statements of Theorem 2.7, we know that $\mathcal{A}_D^{(f)}, \mathcal{A}_D^{(u)}, \mathcal{A}_D^{(u,*)}$ contain $\mathcal{A}_D^{(f,*)}$. In this section, we consider the (more general) setting of the first statement and prove that $\mathcal{S}_D \subseteq \overline{\mathcal{A}_D^{(f,*)}}$. Throughout the section, we explain how our proof implies that the same holds in the respective settings of the two other statements.

Case $R \geq H(X)$. Let $\Delta_{\min}$ denote $\inf_{p \in \mathcal{D}_D} \mathbb{E}_p[d(X,Y)]$. Fix $R' > H(X)$, $\varepsilon > 0$, and $p \in \mathcal{D}_D$ such that $\mathbb{E}_p[d(X,Y)] \leq \varepsilon + \Delta_{\min}$. Since $\mathcal{X}$ is finite and d is bounded, then there is a sequence of deterministic fixed-length codes of rate R' with vanishing error probability $\varepsilon_n := P^{(n)}(\hat{X}^n \neq X^n) \to 0$ [CT06, Section 3.2]. Sampling Y^n from $\prod_{t=1}^n p_{Y|X=\hat{x}_t}$ and ignoring J and L_d yields, for any $R_c, R_d \geq 0$, an (n, R', R_c, R_d) fixed-length code with non-privately randomized (indeed deterministic) encoding, whose induced distribution satisfies $\|P_{Y^n}^{(n)} - p_X^{\otimes n}\|_{TV} \leq \|P_{\hat{X}^n}^{(n)} - p_X^{\otimes n}\|_{TV} \leq \varepsilon_n \to 0$ and $\mathbb{E}_{P^{(n)}}[d(X^n, Y^n)] - \mathbb{E}_p[d(X,Y)] \leq 2\max(d)\,\varepsilon_n \to 0$. Hence $(R', R_c, R_d, \Delta) \in \mathcal{A}_D^{(f,*)}$ for all $R' > H(X)$, all $R_c, R_d \geq 0$, and all $\Delta > \Delta_{\min}$. Taking the closure gives $[H(X), \infty) \times \overline{\mathbb{R}}_{\geq 0}^2 \times [\Delta_{\min}, \infty) \subseteq \overline{\mathcal{A}_D^{(f,*)}}$, which implies

$$\mathcal{S}_D \cap \left([H(X), \infty) \times \overline{\mathbb{R}}_{\geq 0}^3 \right) \subseteq \overline{\mathcal{A}_D^{(f,*)}}. \tag{2.27}$$

These codes use no common randomness and no decoder private randomness, and achieve near-perfect strong realism (hence also near-perfect per-symbol realism), so this covers the settings of all three statements of the theorem and of Theorem 2.13.

It remains to treat $R < H(X)$. Let (R, R_c, R_d, Δ) be a tuple in $\mathcal{S}_D$ and $p_{X,V,Y}$ be a corresponding distribution in $\mathcal{D}_D$. If $R_c = \infty$, we replace it by $H(X)$, if $R_d = \infty$, we replace it with $H(Y|X)$, and if $\Delta = \infty$, we replace it with $\max(d)$, which exists since $\mathcal{X}$ is finite and d is assumed to only take finite values (Definition 2.1). Then, $(R, R_c, R_d, \Delta) \in [0, H(X)) \times \mathbb{R}_{\geq 0}^3$, and we can apply Proposition 2.15. Hereafter, we use the notation of Proposition 2.15. If $R_d = 0$,

then $H_p(Y|V) = 0$, thus from (2.25), for any $n \in \mathbb{N}$, $P^{(n)}$ is the distribution induced by an $(n, R+\varepsilon_n, R_c, 0)$ fixed-length code. Hence, under the setting of the third statement of Theorem 2.7, we have $\mathcal{S}_D \subseteq \overline{\mathcal{A}_D^{(f,*)}}$, which concludes the proof regarding that setting. Moving to the case $R_d > 0$, we can use the same argument as in [Cuf13, Appendix 2] and reach the following claim — see Appendix 2.F.2 for a proof.

Claim 2.16 *From the local channel synthesis lemma [Cuf13, Corollary VII.6],* (2.25) *and* (2.26)*, for any $\gamma > 0$ and any $n \in \mathbb{N}$, there exists a mapping $\tilde{P}_{Y^n|V^n}$ with a fixed-length private randomness of rate $H_p(Y|V) + \gamma$ such that*

$$\left\| P^{(n)}_{X^n,M,J,v^n(M,J)} \cdot \tilde{P}_{Y^n|V^n} - P^{(n)}_{X^n,M,J,v^n(M,J),Y^n} \right\|_{TV} \underset{n\to\infty}{\longrightarrow} 0.$$

Hence, from Lemmas 1.8 and 1.10 (Appendix 1), and since d is bounded, replacing $P^{(n)}_{Y^n|V^n}$ by $\tilde{P}_{Y^n|V^n}$ for every n preserves (2.23) and (2.24), and results in a $(n, R + \varepsilon_n, R_c, R_d + \gamma)$ fixed-length code. This holds for every $\gamma > 0$. Since the common randomness rate used is precisely R_c, both when $R_d = 0$ and when $R_d > 0$, this shows that we have $\mathcal{S}_D \in \overline{A_D^{(f,*)}}$ in both the settings of the first and second statements of Theorem 2.7, which concludes the proof.

2.7 Extensions to other realism constraints

The results of Section 2.5 concern the near-perfect strong realism constraint (2.3) and the near-perfect per-symbol realism constraint (2.4), i.e., the regime $\lambda \to 0$. In this section, we record what our method does and does not yield for related formulations of the realism constraint that have appeared in the literature, including the imperfect strong realism constraint ($\lambda > 0$). We provide definitions and statements only; no proofs are given, as each statement below follows by a direct adaptation of the argument of Section 2.6.

2.7.1 The de-randomization mechanism and its scope

We first isolate the mechanism underlying our results, so as to delineate its scope. Our proof that encoder private randomness is not useful proceeds by *modifying an achievability scheme*: given a sequence of codes with encoder private randomness whose decoders have the codeword-based memoryless form of Remark 2.5, Proposition 2.3 produces a sequence of codes with non-privately randomized encoding whose induced distributions satisfy, by (2.13),

$$\big\|\tilde{P}^{(n)}_{M,J,Y^n} - P^{(n)}_{M,J,Y^n}\big\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{2.28}$$

The modification increases the compression rate, the common randomness rate and the distortion by a vanishing amount, and preserves the joint distribution P_{M,J,Y^n} in the sense of (2.28). In particular, it preserves the marginal P_{Y^n} up to a vanishing TVD, and hence preserves any realism constraint that is a continuous functional of P_{Y^n} with respect to the TVD. The near-perfect strong realism constraint (2.3) is of this form, whether or not the target TVD vanishes: by the triangle inequality, if $\|P^{(n)}_{Y^n} - p_X^{\otimes n}\|_{TV} \leq \lambda + o_n(1)$, then our construction will result in $\|\tilde{P}^{(n)}_{Y^n} - p_X^{\otimes n}\|_{TV} \leq \lambda + o_n(1)$. This observation is what allows the achievability side of our argument to carry over to imperfect strong realism ($\lambda > 0$). The obstruction is on the *converse* side: because our modification acts only on achievability schemes, in the absence of a matching converse for the target constraint (in the case of unlimited encoder private randomness) no conclusion can be drawn on the role of encoder private randomness. This is the case, in particular, for the imperfect strong realism constraint $\limsup_n \mathcal{D}(p_X^{\otimes n}, P_{Y^n}) \leq \lambda$ under a general divergence $\mathcal{D}$ with limited common randomness, whose single-letter characterization remains open [CYW+22, Remark 11].

We stress that (2.28) controls the joint distribution P_{M,J,Y^n}, not the conditional distributions $P_{Y^n|M}$ or $P_{X^n|M}$. Conditioning on an event $\{M = m\}$ of small probability may amplify the TVD in (2.28) arbitrarily. Consequently, our method does not preserve realism constraints defined through such conditional distributions, such as the conditional-distribution perception measure of [SCKY24], whatever the amount of common randomness. Moreover, unlike the marginal-distribution case, here encoder private randomness appears to play an essential role:

the achievability of [SCKY24] uses a stochastic encoder (a posterior reference map) precisely to shape the conditional law $P_{X^n|M}$ on which the constraint depends, and their single-letter characterization is established over stochastic encoders rather than by exhibiting a non-privately randomized one. Whether encoder private randomness can be dispensed with for this formulation is therefore not resolved by existing results, and is not addressed by our method.

Finally, the question addressed by our method — whether encoder private randomness enlarges the achievable region — is only meaningful when common randomness is limited. When unlimited common randomness is available, private randomness of any kind can be extracted from it and is therefore superfluous, so there is no role of encoder private randomness to investigate. This applies, for instance, to the imperfect strong realism trade-off, which is single-letter characterized in that regime for a large class of divergences [CYW+22, Theorem 7], and to the corresponding joint source-channel coding problem over a discrete memoryless channel, for which the separation architecture is then optimal [TCN25]. The conclusive extensions we present next accordingly concern the regime of limited common randomness.

2.7.2 Conclusive extensions with limited common randomness

We now turn to two variants of the imperfect strong realism constraint for which a matching converse is available in the literature under limited common randomness, and for which our method therefore yields a conclusive statement on the role of encoder private randomness. In both variants the reconstruction is constrained more tightly than by the constraint $\limsup_n \mathcal{D}(p_X^{\otimes n}, P_{Y^n}) \leq \lambda$ alone, which is what restores a matching converse.

Reconstruction constrained to be i.i.d.

In the first variant, studied in [SLY15a, XLCZ25], the reconstruction Y^n is required to be i.i.d. — asymptotically, in total variation — with a per-symbol marginal that may depend on the blocklength.

Definition 2.7 *Consider a space $\mathcal{X}$, a distribution p_X on $\mathcal{X}$, a distortion measure d, a divergence $\mathcal{D}$, and a slack $\lambda \geq 0$. A tuple (R, R_c, R_d, Δ) is said to be* achievable with λ-imperfect strong realism and i.i.d. reconstruction *(with a given class of codes and encoding, as in Definition 2.4) if there exist a sequence of (n, R, R_c, R_d) codes of that class, with induced distributions $\{P^{(n)}\}_n$, and a sequence $\{q_n\}$ of distributions on $\mathcal{X}$, such that*

$$\begin{aligned} &\limsup_{n\to\infty} \mathbb{E}_{P^{(n)}}[d(X^n, Y^n)] \leq \Delta, \\ &\left\| P_{Y^n}^{(n)} - q_n^{\otimes n} \right\|_{TV} \underset{n\to\infty}{\longrightarrow} 0, \\ &\limsup_{n\to\infty} \mathcal{D}(p_X, q_n) \leq \lambda. \end{aligned} \tag{2.29}$$

The per-symbol reconstruction marginal q_n is not fixed across blocklengths: as in [XLCZ25], the reconstruction is only required to be i.i.d. for each n, and the realism level is controlled through $\mathcal{D}(p_X, q_n)$. The optimal trade-off admits a single-letter characterization with limited common randomness: for a fixed reconstruction marginal this is output-constrained lossy source coding [SLY15a, Wag22], and the imperfect strong realism trade-off is recovered by taking the infimum over admissible marginals [XLCZ25]. The achievability scheme has the codeword-based memoryless decoder form of Remark 2.5, and the matching converse holds for codes with arbitrary encoder private randomness. Hence our method applies.

Corollary 2.17 *Under the assumptions of [SLY15a, Theorem 1], and for $\lambda \geq 0$, the closure of the set of tuples (R, R_c, R_d, Δ) achievable with λ-imperfect strong realism and i.i.d. reconstruction does not depend on whether encoder private randomness is available. This holds for any common randomness rate R_c and any decoder private randomness rate R_d, the statement being non-trivial when R_c is finite.*

Squared error and squared quadratic Wasserstein distance

In the second variant, the alphabet is $\mathcal{X} \subseteq \mathbb{R}^L$, the distortion measure is the squared error, the divergence is the squared quadratic Wasserstein distance W_2^2, and the strong realism constraint

is imposed on the (not necessarily i.i.d.) reconstruction sequence. A single-letter characterization with limited common randomness was established in [QCYX25].

Definition 2.8 *Consider $\mathcal{X} \subseteq \mathbb{R}^L$, a square-integrable distribution p_X on $\mathcal{X}$, the squared-error distortion measure d, and a slack $\lambda \geq 0$. A tuple (R, R_c, R_d, Δ) is said to be* achievable with λ-imperfect W_2 realism *(with a given class of codes and encoding) if there exists a sequence of (n, R, R_c, R_d) codes of that class, with induced distributions $\{P^{(n)}\}_n$, such that*

$$\begin{aligned} \limsup_{n\to\infty} \tfrac{1}{n}\mathbb{E}_{P^{(n)}}\big[\|X^n - Y^n\|^2\big] &\leq \Delta, \\ \limsup_{n\to\infty} \tfrac{1}{n} W_2^2\big(p_X^{\otimes n}, P_{Y^n}^{(n)}\big) &\leq \lambda. \end{aligned} \tag{2.30}$$

Here a technical point arises that is absent from the TVD case: the modification of Proposition 2.3 controls the reconstruction distribution in TVD, whereas the constraint (2.30) is expressed in W_2^2. Closeness in TVD does not imply closeness in W_2 on unbounded alphabets in general; it does, however, when the normalized second moments $\{\frac{1}{n}\|Y^n\|^2\}_n$ of the reconstruction sequences (and of their de-randomized counterparts) are uniformly integrable, in which case convergence to 0 in TVD implies convergence to 0 in the normalized W_2 (see, e.g., [Vil03, Theorem 7.12]). We stress that this is a stronger requirement than the bound on the mean second moment $\frac{1}{n}\mathbb{E}[\|Y^n\|^2] \leq 2\,\mathbb{E}[\|X\|^2] + 2\Delta + o_n(1)$ that follows from the distortion constraint alone: a bounded mean does not preclude second-moment mass escaping to infinity. A convenient sufficient condition is a uniform per-letter moment bound $\sup_{n,t} \mathbb{E}[\|Y_t\|^{2+\delta}] < \infty$ for some $\delta > 0$, which we secure below (Remark 2.19) under a finite $(2+\delta)$-moment of the source. This is what allows the modification to preserve the constraint (2.30). The achievability scheme of [QCYX25] has the codeword-based memoryless decoder form of Remark 2.5, and the matching converse holds for codes with arbitrary encoder private randomness. Hence our method applies.

Corollary 2.18 *Under the assumptions of [QCYX25, Theorem 1], for $\lambda \geq 0$, and assuming in addition that the source has a finite $(2+\delta)$-moment for some $\delta > 0$ (i.e. $\mathbb{E}[\|X\|^{2+\delta}] < \infty$), the closure of the set of tuples (R, R_c, R_d, Δ) with $R < H(X)$ achievable with λ-imperfect W_2*

realism does not depend on whether encoder private randomness is available. This holds for any common randomness rate R_c and any decoder private randomness rate R_d, the statement being non-trivial when R_c is finite.

Remark 2.19 *The finite $(2+\delta)$-moment hypothesis in Corollary 2.18 secures the uniform integrability required for the passage from TVD to W_2 above, without altering Proposition 2.3. Write $g(v) := \mathbb{E}_{p_{Y|V=v}}[\|Y\|^{2+\delta}]$ for the conditional $(2+\delta)$-moment of the reconstruction given a codeword letter v. Under the codeword-based memoryless decoder $\prod_t p_{Y|V=v_t}$ of Remark 2.5, the reconstruction moment is $\frac{1}{n}\sum_{t=1}^n \mathbb{E}[\|Y_t\|^{2+\delta}] = \mathbb{E}\big[\frac{1}{n}\sum_{t=1}^n g(v_t)\big]$. Since g is unbounded over a continuous codeword alphabet, the de-randomized encoder induces a different distribution over the codewords, and TVD closeness alone does not prevent this distribution from favouring codewords of large g and inflating the moment. This is avoided by a pre-processing of the codebook that leaves the achievable region unchanged. The codewords being drawn from a distribution ρ_V whose induced output marginal is the realism target, the mean $\bar{g} := \mathbb{E}_{\rho_V}[g(V)]$ equals the $(2+\delta)$-moment of that target, which is finite under the source hypothesis. Discard from the codebook every codeword whose empirical mean $\frac{1}{n}\sum_{t=1}^n g(v_t)$ exceeds $2\bar{g}$. By the law of large numbers the discarded fraction of each subcodebook vanishes, so the soft covering, distortion and perception guarantees are preserved up to a vanishing amount. After this pre-processing every surviving codeword satisfies $\frac{1}{n}\sum_{t=1}^n g(v_t) \le 2\bar{g}$, so* for any encoder *— in particular the de-randomized one — the reconstruction satisfies $\frac{1}{n}\sum_{t=1}^n \mathbb{E}[\|Y_t\|^{2+\delta}] \le 2\bar{g}$. By Jensen's inequality this bounds a moment of $\frac{1}{n}\|Y^n\|^2$ of order $1+\delta/2$ uniformly in n, which yields the required uniform integrability. The hypothesis is mild — it holds, for instance, for any Gaussian source — but it is genuinely needed: with only a finite second moment the normalized second moments need not be uniformly integrable, and the passage from TVD to W_2 can fail.*

2.8 Conclusion

We have studied the role of private randomness in the RDP trade-off with near-perfect realism and near-perfect per-symbol realism constraints, in the classical infinite blocklength scenario.

Our work complements previous results on the key role of randomization in these settings. We have characterized the corresponding RDP trade-offs under different situations in terms of the amount of common and private randomness available. Our results show that encoder private randomness is not useful — when the compression rate is lower than the entropy of the source, which is the case in lossy compression. In particular, in that case, if no common randomness is available, then, for large blocklength, it is not useful to send randomness from the encoder. A similar phenomenon was conjectured in [Cuf13] to hold for the channel synthesis problem (see Chapter 5), but this has not yet been proved. The role of encoder private randomness in the finite-blocklength and one-shot compression settings merits further investigation.

In the next chapter, we move beyond the point-to-point setting covered in the present chapter, and consider another signal, correlated with the source, and available as side information. This is linked to the problem of video compression, where a previously compressed video frame can serve as side information. We characterize the latter's role as a source of common randomness, and show that said role is nuanced: it depends on whether the realism constraint involves the side information.

Appendix

2.A Proof of Proposition 2.3

We use the soft covering lemma with a sequence of general sources and channels [Cuf13, Corollary VII.3], which we state for completeness. For any distribution $\Phi_{V,W}$, the information density (Proposition 1.17 in Appendix 1) is denoted

$$i_\Phi(v;w) = \log(\Phi(v,w)/\Phi(v)\Phi(w)).$$

We also use the notation

$$i_\Phi(w) = -\log(\Phi(w)).$$

Lemma 2.20 *[Cuf13, Corollary VII.3] Let $\{\mathcal{U}^{(n)}, \mathcal{V}^{(n)}, \mathcal{W}^{(n)}\}_{n\geq 1}$ be a sequence of finite alphabets and $\{\Phi_{U^{(n)},V^{(n)},W^{(n)}}\}_{n\geq 1}$ be a sequence of distributions, the n-th being on $\mathcal{U}^{(n)}\times\mathcal{V}^{(n)}\times\mathcal{W}^{(n)}$. For every $n\geq 1$, and every $w^{(n)}\in\mathcal{W}^{(n)}$, let $u^{(n)}(w^{(n)})$ be a random variable with distribution $\Phi_{U^{(n)}|W^{(n)}=w^{(n)}}$. Denote the family $\{u^{(n)}(w^{(n)})\}_{w^{(n)}\in\mathcal{W}^{(n)}}$ by $\mathcal{B}^{(n)}$. For every $n\geq 1$, define*

$$\tilde{\Phi}_{W^{(n)},U^{(n)},V^{(n)}} := \Phi_{W^{(n)}} \cdot \mathbf{1}_{U^{(n)}=u^{(n)}(w^{(n)})} \cdot \Phi_{V^{(n)}|W^{(n)},U^{(n)}}.$$

Assume that

$$i_\Phi(W^{(n)}, U^{(n)}; V^{(n)}) - i_\Phi(W^{(n)}) \xrightarrow{\mathcal{P}} -\infty. \tag{2.31}$$

Then, we have

$$\mathbb{E}_{\mathcal{B}^{(n)}}[\|\tilde{\Phi}_{V^{(n)}} - \Phi_{V^{(n)}}\|_{TV}] \underset{n\to\infty}{\longrightarrow} 0. \tag{2.32}$$

For every $n\geq 1$ let $M=\psi^{(n)}(X^n, J, U)$ be a functional representation of $F^{(n)}$. Take $\Phi_{W^{(n)},U^{(n)},V^{(n)}}$ corresponding to $P^{(n)}$ with $W^{(n)}=(X^n, J)$, $U^{(n)}=U$, and $V^{(n)}=(M, J, \mathbb{P}^{\text{emp}}_{X^n,\mathbf{v}^{(n)}(M,J)})$.

Proving (2.31)

Since $V^{(n)}$ is a deterministic function of $(W^{(n)}, U^{(n)})$, it can be easily checked that $1/\Phi_{V^{(n)}}$ is a density for $\Phi_{W^{(n)},U^{(n)},V^{(n)}}$ with respect to $\Phi_{W^{(n)},U^{(n)}} \cdot \Phi_{V^{(n)}}$. Consider $\varepsilon_1, \varepsilon_2 \in (0, H(X)-R)$ such that $\varepsilon_1 + \varepsilon_2 < H(X)-R$.

$$\begin{aligned}
&P^{(n)}\big(i_{P^{(n)}}(M, J, \mathbb{P}^{\text{emp}}_{X^n,\mathbf{v}^{(n)}(M,J)}) \geq n(R + R_c + \varepsilon_1)\big) \\
=\quad &P^{(n)}\big(P^{(n)}(M, J, \mathbb{P}^{\text{emp}}_{X^n,\mathbf{v}^{(n)}(M,J)}) \leq 2^{-n(R+R_c+\varepsilon_1)}\big) \\
\leq\ &\lfloor 2^{nR} \rfloor \cdot \lfloor 2^{nR_c} \rfloor \cdot n^{|\mathcal{X}||\mathcal{V}|} \cdot 2^{-n(R+R_c+\varepsilon_1)} \underset{n\to\infty}{\longrightarrow} 0.
\end{aligned}$$

Since J is uniformly distributed and independent of X^n, and $P^{(n)}_{X^n} \equiv p_X^{\otimes n}$, then for any (x^n, j) we have $i_{P^{(n)}}(x^n, j) = i_{p_X^{\otimes n}}(x^n) + \log(\lfloor 2^{nR_c} \rfloor)$. From the law of large numbers:

$$P^{(n)}(i_{P^{(n)}}(X^n, J) \leq n(H(X) + R_c - \varepsilon_2)) \underset{n\to\infty}{\longrightarrow} 0.$$

When both the above events do not hold, we have

$$i_{P^{(n)}}\big(M, J, \mathbb{P}^{\text{emp}}_{X^n,\mathbf{v}^{(n)}(M,J)}\big) - i_{P^{(n)}}(X^n, J) \leq -n(H(X) - R - \varepsilon_1 - \varepsilon_2).$$

The left-hand side is exactly $i_P(V^{(n)}) - i_P(W^{(n)})$ with $V^{(n)} = \big(M, J, \mathbb{P}^{\text{emp}}_{X^n,\mathbf{v}^{(n)}(M,J)}\big)$ and $W^{(n)} =$

(X^n, J), i.e. the quantity in the hypothesis (2.31) of [Cuf13, Corollary VII.3]. Since $\varepsilon_1 + \varepsilon_2 < H(X) - R$, it tends to $-\infty$ in probability, as required.

Conclusion

For each $n \geq 1$, choose a realization of $\mathcal{B}^{(n)}$ giving a TVD below average, which defines a deterministic

$$f^{(n)} : (x^n, j) \mapsto \psi^{(n)}(x^n, j, u^{(n)}(x^n, j)) \in [2^{nR}]. \tag{2.33}$$

Moreover, for every $n \geq 1$, $\tilde{\Phi}^{(n)}$ defines the same distribution of inputs (X^n, J) as do the distributions $\tilde{P}^{(n)}$ of Proposition 2.3. We conclude using Lemma 1.8 (Appendix 1), and the fact that the empirical distribution is bounded and the average empirical distribution is its expectation, because $\mathcal{X} \times \mathcal{V}$ is finite.

2.B Converse of Theorem 2.7

By definition, for each of the three statements of Theorem 2.7, we know that $\mathcal{A}_D^{(f)}, \mathcal{A}_D^{(f,*)}, \mathcal{A}_D^{(u,*)}$ are included in $\mathcal{A}_D^{(u)}$. Therefore, we only prove that $\mathcal{A}_D^{(u)} \subseteq \mathcal{S}_D$, in the setting of the first statement, and this also implies the same relation in the respective settings of the two other statements. This proof closely tracks that of [Cuf13, Section VI] and the end of [Cuf13, Appendix 2]. Let $(R, R_c, R_d, \Delta) \in \mathbb{R}_{\geq 0} \times \overline{\mathbb{R}}_{\geq 0}^3$ be achievable with near-perfect realism with variable-length codes. Fix $\varepsilon \in (0, 1/4)$. Then, for n large enough, there exists a (n, R, R_c, R_d) code inducing a joint distribution P_{X^n, J, M, Y^n} such that

$$\mathbb{E}_P[d(X^n, Y^n)] \leq \Delta + \varepsilon, \ \|P_{Y^n} - p_X^{\otimes n}\|_{TV} \leq \varepsilon$$
$$H(M) \leq nR, \ H(J) \leq nR_c, \ H(L_d) \leq nR_d.$$

Following the proof of [Wag22, Theorem 2] and [Cuf13, Section VI & Appendix 2], we have the following claim.

Claim 2.21 *By introducing a random index uniformly distributed on* $[n]$*, one can construct a distribution* $P^{(\varepsilon)}_{X,Y,V}$ *on* $\mathcal{X}^2\times\mathcal{V}$ *satisfying* $X-V-Y$*,* $\|P_Y-p_X\|_{TV}\leq\varepsilon$*, and rate bounds*

$$R\geq I(X;V),\ R+R_c\geq I(Y;V)-g(\varepsilon),\ R_d\geq H(Y|V),$$

for some deterministic function g *such that* $g(\varepsilon)\to 0$ *as* $\varepsilon\to 0$*.*

See Appendix 2.C for a proof. We can change V so that $|\mathcal{V}|\leq|\mathcal{X}|^2+1$ while preserving $P^{(\varepsilon)}_{X,Y}$, the Markov chain $X-V-Y$, and quantities $I(X;V), I(Y;V)$ (and thus $H(Y|V)$). This follows from the proof of [Cuf13, Lemma VI.1]. We then conclude with the same argument as in [Cuf13, Section VI], as follows. Consider a vanishing sequence $\{\varepsilon_n\}_{n\geq 1}$ in $(0,1/4)$. Owing to the cardinality bound, all probabilities $P^{(\varepsilon_n)}$ can be considered as points in the compact standard simplex of $\mathbb{R}^{2|\mathcal{X}|+|\mathcal{X}|^2+1}$. Hence, a sub-sequence converges towards some probability $P^*_{X,V,Y}$ in the latter. This distribution satisfies the Markov chain constraint and $P^*_X\equiv p_X\equiv P^*_Y$. Hence, $P^*\in\mathcal{D}_D$. Moreover, since $g(\varepsilon_n)\to 0$ as $n\to\infty$, and from the rate lower bounds in Claim 2.21, we have $(R,R_c,R_d,\Delta)\in\mathcal{S}_D$, which concludes the converse proof.

2.C Rate lower bounds in the converse proof

Here, we provide a proof of Claim 2.21 in Appendix 2.B. Let T be a uniform random variable on $[n]$. Define $V=(M,J,T,Y^{T-1})$. Since $X^n\sim p_X^{\otimes n}$, the distribution of X_T is p_X. From Lemma 1.9 (Appendix 1), we have $\|P_{Y_T}-p_X\|_{TV}\leq\varepsilon$. From the Markov chain $X^n-(M,J)-Y^n$, the distribution P_{X_T,V,Y_T} satisfies the Markov chain X_T-V-Y_T.

Since $P_{X_T,Y_T}\equiv\hat{P}_{\mathcal{X}^2}[X^n,Y^n]$, we have $\mathbb{E}[d(X_T,Y_T)]=\mathbb{E}[d(X^n,Y^n)]\leq\Delta+\varepsilon$. We derive rate lower bounds using the following lemma.

Lemma 2.22 *[Cuf13, Lemma VI.3] For any finite alphabet* $\mathcal{W}$ *and any random sequence* Π_{W^n}

taking values in $\mathcal{W}^n$, if there exists a distribution Γ_W on $\mathcal{W}$ such that

$$\|\Pi_{W^n} - \Gamma_W^{\otimes n}\|_{TV} \leq \varepsilon < \frac{1}{4}, \quad \textit{then,}$$
$$\frac{1}{n}\sum_{t=1}^{n} I_\Pi(W_t; W^{t-1}) \leq 4\varepsilon\Big(\log(|\mathcal{W}|) + \log\Big(\frac{1}{\varepsilon}\Big)\Big),$$

and for any random variable T uniformly distributed on $[n]$ and independent of W^n, we have

$$I_\Pi(W_T; T) \leq 4\varepsilon\Big(\log(|\mathcal{W}|) + \log\Big(\frac{1}{\varepsilon}\Big)\Big).$$

Define

$$g : (0, 1/4) \to (0, \infty),\ \varepsilon \mapsto 4\varepsilon\Big(\log(|\mathcal{X}|) + \log\Big(\frac{1}{\varepsilon}\Big)\Big).$$

We have

$$\begin{aligned}
nR \geq H(M) &\geq I(M; X^n|J) \\
&= I(M, J; X^n) && (2.34)\\
&= \textstyle\sum_{t=1}^n I(M, J; X_t|X_{t+1:n}) \\
&= \textstyle\sum_{t=1}^n I(M, J, X_{t+1:n}; X_t) \\
&\geq \textstyle\sum_{t=1}^n I(M, J; X_t) \\
&= \textstyle\sum_{t=1}^n I(M, J, Y^{t-1}; X_t) && (2.35)\\
&= nI(M, J, Y^{T-1}; X_T|T) && (2.36)\\
&= nI(V; X_T), && (2.37)
\end{aligned}$$

where (2.34) follows from the independence between the common randomness and the sources; and (2.35) follows from Markov chain $X^n - (M, J) - Y^n$; and (2.36) and (2.37) follow from the independence of T and all other variables and from the fact that variables in $\{X_t\}_{t\in[n]}$ are i.i.d..

We also have

$$\begin{aligned} n(R+R_c) &\geq I(M,J;Y^n) \\ &= \sum_{t=1}^{n} I(M,J;Y_t|Y^{t-1}) \\ &= \sum_{t=1}^{n} \big[I(M,J,Y^{t-1};Y_t) - I(Y^{t-1};Y_t)\big] \\ &\geq \sum_{t=1}^{n} I(M,J,Y^{t-1};Y_t) - ng(\varepsilon) \qquad (2.38) \\ &= nI(M,J,Y^{T-1};Y_T|T) - ng(\varepsilon) \\ &= nI(T,M,J,Y^{T-1};Y_T) - nI(T;Y_T) - ng(\varepsilon) \\ &\geq nI(V;Y_T) - 2ng(\varepsilon), \qquad (2.39) \end{aligned}$$

where (2.38) and (2.39) follow from Lemma 2.22. Moreover,

$$\begin{aligned} nR_d \geq H(L_d) \geq I(L_d;Y^n|M,J) &= H(Y^n|M,J) \\ &= \sum_{t=1}^{n} H(Y_t|Y^{t-1},M,J) \\ &= nH(Y_T|V). \end{aligned}$$

2.D Proof of Theorem 2.11

2.D.1 Converse

We know that $\mathcal{A}_G^{(f,*)} \subseteq \mathcal{A}_G^{(f)}$. Moreover, the inclusion $\mathcal{A}_G^{(f)} \subseteq \overline{\mathcal{S}_G}$ is the converse direction of [Wag22, Theorem 2], which is indeed stated for fixed-length codes [Wag22, Definition 1].

2.D.2 Quantization argument

Let (R, R_c, Δ) be a triplet in $\mathcal{S}_G$. Let $p_{X,Y,V}$ be a corresponding distribution from the definition of $\mathcal{S}_G$. Then

$$R \geq I_p(X;V),\ R + R_c \geq I_p(Y;V),\ \Delta \geq \mathbb{E}_p[d(X,Y)]. \tag{2.40}$$

Fix some $\varepsilon > 0$. By assumption, (d, p_X) is quantizable and uniformly integrable. Let τ' be a threshold corresponding to ε as in Definition 2.5. Set $\tau = \tau'/2$. Let $\{\kappa^{(\ell)}\}_{\ell\in\mathbb{N}}$, B_τ and $L_{\varepsilon,\tau}$ be as in Definition 2.6. The former is a sequence of measurable quantizers of $\mathcal{X}$ such that the corresponding partitions asymptotically generate its Borel σ-algebra -i.e. quantization becomes arbitrarily fine as ℓ grows. Therefore, ince the source has infinite entropy, then from [Gra11, Lemma 7.18], there exists L'_ε such that for any $\ell \geq L'_\varepsilon$,

$$H_p(\kappa^{(\ell)}(X)) > R. \tag{2.41}$$

Fix $\ell \geq \max(L'_\varepsilon, L_{\varepsilon,\tau})$. We denote $\kappa^{(\ell)}(X)$ by $[X]$ and $\kappa^{(\ell)}(Y)$ by $[Y]$. Since p satisfies $X - V - Y$ and $p_Y \equiv p_X$, then we have $[X] - V - [Y]$ and $p_{[Y]\equiv p_{[X]}}$. Thus,

$$p_{[X],V,[Y]} \in \mathcal{D}_D, \tag{2.42}$$

where $\mathcal{D}_D$ is the set defined in (2.15), corresponding to source distribution $p_{[X]}$ (instead of p_X). From (2.17) and a union bound, we have

$$p\big((X,Y) \notin B_\tau^2\big) \leq p(X \notin B_\tau) + p(Y \notin B_\tau) \leq 2\tau = \tau', \tag{2.43}$$

then from the uniform integrability we have

$$0 \leq \mathbb{E}_p[d(X,Y)\mathbf{1}_{(X,Y)\notin B_\tau^2}] \leq \varepsilon. \tag{2.44}$$

From (2.18), for every $\ell \geq L_{\varepsilon,\tau}$ and every $(x,y) \in B_\tau^2$, we have

$$\big|d(x,y) - d\big([x]\big),[y]\big)\big| \leq \varepsilon, \tag{2.45}$$

and for every $(x,y) \in \mathcal{X}^2 \setminus B_\tau^2$, we have

$$d\big([x]\big),[y]\big) \leq d(x,y) + \varepsilon. \tag{2.46}$$

From (2.45) and (2.46), we get

$$\mathbb{E}_p[d(X,Y)] + \varepsilon \geq \mathbb{E}_p[d([X],[Y])]. \tag{2.47}$$

Since $\kappa^{(\ell)}$ is deterministic, then from (2.40) and (2.47), we have

$$R \geq I_p([X];V),\ R+R_c \geq I_p([Y];V),\ \Delta+\varepsilon \geq \mathbb{E}[d([X],[Y])]. \tag{2.48}$$

Let $\varepsilon' \in (0,\varepsilon]$ such that

$$R + \varepsilon' < H_p([X]). \tag{2.49}$$

Hence, from (2.42) and since the auxiliary variable V is assumed to be finite-valued, we can apply Proposition 2.15 with distribution $p_{[X],V,[Y]}$ and triplet $(R+\varepsilon', R_c, \Delta+\varepsilon)$. Hereafter, we use the notation of Proposition 2.15.

2.D.3 Transition from near-perfect to perfect realism

Proposition 2.23 *Let n be a positive integer and δ be a positive real. Let $\mathcal{X}$ and $\mathcal{W}$ be two Polish alphabets and p_X be a distribution on $\mathcal{X}$. Let d be a distortion measure such that (d, p_X) is uniformly integrable. Let P_{X^n,W,Y^n} be a distribution on $\mathcal{X}^n \times \mathcal{W} \times \mathcal{X}^n$ satisfying the Markov*

chain property $X^n - W - Y^n$. *Moreover, assume that*

$$\|P_{Y^n} - p_X^{\otimes n}\|_{TV} \leq \delta. \tag{2.50}$$

Then, there exists a conditional distribution $P'_{Y^n|W}$ *such that the distribution* P' *defined by*

$$P'_{X^n,W,Y^n} := P_{X^n,W} \cdot P'_{Y^n|W} \tag{2.51}$$

satisfies

$$\|P'_{X^n,W^n,Y^n} - P_{X^n,W^n,Y^n}\|_{TV} \leq \delta \tag{2.52}$$

$$\text{and } P'_{Y^n} \equiv p_X^{\otimes n}. \tag{2.53}$$

Proof: Appendix 3.B of Chapter 3 is dedicated to the proof of a more general result (Proposition 3.30). ■

For each $n \in \mathbb{N}$, we define

$$\delta_1^{(n)} := \|P_{[Y]^n}^{(n)} - p_X^{\otimes n}\|_{TV},$$

and apply Proposition 2.23 to $P^{(n)}$ with single-letter distribution $p_{[X]}$ and with $W = (M, J)$ and $\delta = \delta_1^{(n)}$. We denote the resulting distribution by $P'^{(n)}$. From the triangle inequality for the TVD, Lemma 1.8 (Appendix 1), and (2.52), we get

$$\left\| \hat{P}'^{(n)}_{[\mathcal{X}]^2}[[X]^n, [Y]^n] - \hat{P}^{(n)}_{[\mathcal{X}]^2}[[X]^n, [Y]^n] \right\|_{TV} \leq \delta_1^{(n)}. \tag{2.54}$$

From the additivity of d, we have

$$\mathbb{E}_{P^{(n)}}[d([X]^n, [Y]^n)] = \mathbb{E}_{\hat{P}^{(n)}_{[\mathcal{X}]^2}[[X]^n,[Y]^n]}[d(X, Y)].$$

Since d does not take infinite values (Definition 2.1) and $[\mathcal{X}]$ is finite, then d is bounded on

$[\mathcal{X}]^2$. Hence, from Lemma 1.10 (Appendix 1), (2.23), and (2.54), we get

$$\mathbb{E}_{\hat{P}'^{(n)}_{[\mathcal{X}]^2}[[X]^n,[Y]^n]}[d(X,Y)] \le \mathbb{E}_{P^{(n)}}[d([X]^n,[Y]^n)] + \delta_2^{(n)}, \tag{2.55}$$

where $\delta_2^{(n)} = 2\max(d)\delta_1^{(n)}$, so that $\delta_2^{(n)} \to 0$. For any $n\in\mathbb{N}$, the following distribution

$$\begin{aligned}\tilde{P}^{(n)}_{X^n,[X]^n,J,M,[Y]^n,Y^n} := p_X^{\otimes n} \cdot \prod_{t=1}^n \mathbf{1}_{[X]_t=[X_t]} \cdot p^{\mathcal{U}}_{[2^{nR_c}]} \cdot P'^{(n)}_{M|[X]^n,J} \\ \cdot \ P'^{(n)}_{[Y]^n|M,J} \cdot \prod_{t=1}^n p_{X|[X]=[Y]_t},\end{aligned}$$

defines a $(n, R+\varepsilon'+\varepsilon_n, R_c, \infty)$ fixed-length code with non-privately randomized encoding satisfying perfect realism, where $[X]^n, [Y]^n$ denote discrete variables, and $[X_t] = \kappa^{(\ell)}(X_t)$. Denote

$$\pi_{X,Y,[X],[Y]} := \hat{\tilde{P}}^{(n)}_{\mathcal{X}^2\times[\mathcal{X}]^2}[X^n, Y^n, [X]^n, [Y]^n]. \tag{2.56}$$

Then, we have

$$\pi_{[X],[Y]} \equiv \hat{P}'^{(n)}_{[\mathcal{X}]^2}[[X]^n, [Y]^n] \quad \text{and} \quad \pi_Y \equiv \pi_X \equiv p_X. \tag{2.57}$$

2.D.4 Conclusion

We have

$$\begin{aligned}&\mathbb{E}_{\tilde{P}^{(n)}}[d(X^n,Y^n)] \\ &= \mathbb{E}_{\hat{\tilde{P}}^{(n)}_{\mathcal{X}^2}[X^n,Y^n]}[d(X,Y)] \qquad (2.58)\\ &= \mathbb{E}_{\pi}[d(X,Y)] \\ &= \mathbb{E}_{\pi}\Big[d(X,Y)\mathbf{1}_{(X,Y)\in B_\tau^2}\Big] \\ &+ \mathbb{E}_{\pi}\Big[d(X,Y)\mathbf{1}_{\ (X,Y)\notin B_\tau^2}\ \Big] \\ &\le \mathbb{E}_{\pi}\Big[d([X],[Y])\mathbf{1}_{\ (X,Y)\in B_\tau^2}\ \Big] + \varepsilon + \varepsilon \qquad (2.59)\\ &\le \mathbb{E}_{\hat{P}'^{(n)}_{[\mathcal{X}]^2}[[X]^n,[Y]^n]}[d([X],[Y])] + 2\varepsilon\end{aligned}$$

$$\leq \Delta + 3\varepsilon, \tag{2.60}$$

where (2.58) follows from the additivity of d; (2.59) follows from (2.45) and (2.44); and (2.60) holds for large enough n from (2.23) and (2.55). For every $n \in \mathbb{N}$, distribution $\tilde{P}^{(n)}$ defines a $(n, R+\varepsilon'+\varepsilon_n, R_c, \infty)$ fixed-length code with non-privately randomized encoding satisfying perfect realism. From the formulation of Proposition 2.14, we have $\delta_{1,2}^{(n)}, \varepsilon_n \to 0$ as $n \to \infty$. Then, from (2.60), and since $\varepsilon' \leq \varepsilon$, tuple $(R+2\varepsilon, R_c, \Delta+3\varepsilon)$ is achievable with perfect realism with fixed-length codes with non-privately randomized encoding. This being true for every $\varepsilon > 0$, we get $(R, R_c, \Delta) \in \overline{\mathcal{A}_G}$, which concludes the proof.

2.E Proof of Theorem 2.13

2.E.1 Converse - finite source alphabet

It is sufficient to use the same proof as that of the converse of Theorem 2.7. Indeed, in the latter, it can be checked that we only ever need information about the joint distribution of the symbols in Y^n to lower bound $R + R_c$.

2.E.2 Converse - source with infinite entropy

It is sufficient to use the same proof as that of the converse of [Wag22, Theorem 2], except that Proposition 2.23 should be applied to (X_T, Y_T) rather than (X^n, Y^n), where T denotes a random index uniformly distributed on $[n]$.

2.E.3 Achievability - finite source alphabet

The case $R \geq H(X)$ follows from the lossless code construction in the proof of Theorem 2.7, since near-perfect strong realism implies near-perfect per-symbol realism. It remains to treat

the case $R < H(X)$. Let $p_{X,V,Y} \in \mathcal{D}_D$ and (R, R_d, Δ) such that

$$R \geq I_p(X;V),\ R_d \geq H_p(Y|V),\ \Delta \geq \mathbb{E}_p[d(X,Y)].$$

Lemma 2.24 *[CYW+22, Lemma 2, Appendix D] Let $\mathcal{X}$ and $\mathcal{V}$ be two Polish alphabets, $p_{X,V}$ be a distribution on $\mathcal{X} \times \mathcal{V}$, and $R > I_p(X;V)$. Then, there exists $K \in [2^{nR}]$ and a codebook*

$$(v^n(m))_{m\in[K]} \subseteq \{v^n \in \mathcal{V}^n | \forall v' \in \mathcal{V}, |\mathbb{P}^{emp}_{v^n}(v') - p_V(v')| \leq \delta p_V(v')\},$$

such that $\|Q_{X^n} - p_X^{\otimes n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0$, where

$$Q_{M,V^n,X^n} := p^{\mathcal{U}}_{[K]} \cdot \mathbf{1}_{V^n = v^n(m)} \cdot \prod_{t=1}^{n} p_{X|V=v_t(m)}. \tag{2.61}$$

Fix $\delta > 0$. As shown in [CYW+22, Appendix D], from Lemma 2.24, there exists a sequence of conditional distributions $(P_{V^n|X^n})_n$ from $\mathcal{X}^n$ to the set of circular shifts of the above codewords, denoted $p_{\check{X}^n|X^n}$ therein, such that

$$\max_{t\in[n]} \|P_{X_t,V_t} - p_{X,V}\|_{TV} \leq \varepsilon_n + \varepsilon_\delta, \text{ with } \varepsilon_n \underset{n\to\infty}{\longrightarrow} 0, \varepsilon_\delta \underset{\delta\to 0}{\longrightarrow} 0, \tag{2.62}$$

$$\text{where } \forall n \in \mathbb{N}, \quad P_{X^n,V^n} := p_X^{\otimes n} \cdot P_{V^n|X^n}. \tag{2.63}$$

We use Lemma 2.24 with a rate of $R + \delta$, hence the set of circular shifts of the corresponding codewords is of size less than $2^{n(R+\delta)}$, for n large enough. We simply send to the decoder the index of the codeword outputted by $P_{V^n|X^n}$. The decoder then applies memoryless channel $p_{Y|V}$. Thus,

$$\max_{t\in[n]} \|P_{X_t,V_t,Y_t} - p_{X,V,Y}\|_{TV} \leq \varepsilon_n + \varepsilon_\delta. \tag{2.64}$$

We then apply Proposition 2.3 and Remark 2.5, which yield a sequence $(P'^{(n)})_n$ of $(n, R{+}\delta, 0, \infty)$ fixed-length codes with non-privately randomized encoding with decoder $\prod p_{Y|V}$, satisfying

$$\max_{t\in[n]} \|P'^{(n)}_{X_t,V_t,Y_t} - p_{X,V,Y}\|_{TV} \leq \varepsilon_n + \varepsilon_\delta + \varepsilon'_n, \tag{2.65}$$

for some vanishing $(\varepsilon'_n)_n$. We know that each codeword (including circular shifts) has an empirical distribution close to p_V in TVD. Hence, using the same argument as in the proof of Theorem 2.7 (Appendix 2.F.2), we can simulate memoryless channel $p_{Y|V}$ with fixed-length private randomness of rate $R_d{+}\delta$, with asymptotically vanishing error in TVD, yielding a sequence $(P''^{(n)})_n$ of $(n, R{+}\delta, 0, R_d{+}\delta)$ fixed-length codes with non-privately randomized encoding, satisfying

$$\max_{t\in[n]} \|P''^{(n)}_{X_t,V_t,Y_t} - p_{X,V,Y}\|_{TV} \leq \varepsilon_n + \varepsilon_\delta + \varepsilon'_n + \varepsilon''_n. \tag{2.66}$$

Since the alphabets are finite and the distortion does not take infinite values, this shows that (R, R_d, Δ) is in the closure of the set of triplets achievable with near-perfect per-symbol realism with no common randomness and no encoder private randomness, as desired.

2.E.4 Achievability - source with infinite entropy

It is sufficient to use the same proof as for Theorem 2.11 (Appendix 2.D) to go from a code on a quantized alphabet to a proper code. Indeed, Proposition 2.23 can be used with $n{=}1$ on each $([X_t], [Y_t])$ to perform the transition from near-perfect per-symbol realism to perfect per-symbol realism. This also yields an average empirical distribution of Y^n equal to p_X. Then, we can use the same distortion bounds, because all TVD bounds are uniform in index t, and imply a bound on the average empirical distribution of $([X]^n, [Y]^n)$.

2.F Further justifications

2.F.1 Proof of Proposition 2.14

We start by proving the following result.

Proposition 2.25 *Consider finite alphabets $\mathcal{X}, \mathcal{V}$, a distortion measure d on $\mathcal{X}^2$, a triplet $(R, R_c, \Delta) \in \mathbb{R}^3_{\geq 0}$, and a distribution $p_{X,Y,V}$ on $\mathcal{X}^2 \times \mathcal{V}$. Assume that $p \in \mathcal{D}_D$ and*

$$R \geq I_p(X;V),\ R + R_c \geq I_p(Y;V),\ \Delta \geq \mathbb{E}_p[d(X,Y)].$$

Then, there exists a sequence $\{\varepsilon_n\}_{n\geq 1}$, a sequence of distributions $\{Q^{(n)}_{X^n,J,M,Y^n}\}_{n\geq 1}$ and a sequence of $(n, R+\varepsilon_n, R_c, \infty)$ fixed-length codes inducing distributions $\{P^{(n)}_{X^n,J,M,Y^n}\}_{n\geq 1}$ such that $\varepsilon_n \underset{n\to\infty}{\to} 0$ and

$$\limsup_{n\to\infty} \mathbb{E}_{Q^{(n)}}[d(X^n, Y^n)] \leq \Delta \tag{2.67}$$

$$\|Q^{(n)}_{Y^n} - p_X^{\otimes n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0 \tag{2.68}$$

$$\|P^{(n)}_{X^\cdot,J,M,Y^n} - Q^{(n)}_{X^\cdot,J,M,Y^n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0 \tag{2.69}$$

$$\forall (m,j),\ P^{(n)}_{Y^n|M=m,J=j} = \textstyle\prod_{t=1}^n p_{Y|V=v_t^{(n)}(m,j)}, \tag{2.70}$$

$$Q^{(n)}\big(\big\|\mathbb{P}^{emp}_{\mathcal{V}}(v^{(n)}(M,J)) - p_V\big\|_{TV} \geq \varepsilon_n\big) \underset{n\to\infty}{\longrightarrow} 0, \tag{2.71}$$

for some sequence of deterministic functions $v^{(n)} : [2^{n(R+\varepsilon)}] \times [2^{nR_c}] \to \mathcal{V}^n$.

Proof: Except for (2.71), the above result follows directly from the random coding proof in [Wag22] — for the interested reader, a detailed standalone proof of this fact in a more general setting is the object of Section 3.5 in Chapter 3. Precisely, $Q^{(n)}$ is Q [Wag22, Eq 76], $P^{(n)}$ is $\tilde{Q}$ [Wag22, Eq 82], (2.67) is [Wag22, Eq. 79], (2.68) is [Wag22, Eq. 77], (2.69) is [Wag22, Eq. 82] and (2.70) is in [Wag22, Eq. 81]. Moreover, one can readily impose (2.71) from the law of large numbers, because $Q^{(n)}$ is constructed from a realization of a random codebook having i.i.d. codewords, each having i.i.d. symbols of distribution p_V. ■

Proposition 2.14 follows from Proposition 2.25. Indeed, from the triangle inequality for the TVD, we have

$$\|P_{Y^n}^{(n)} - p_X^{\otimes n}\|_{TV} \leq \|P_{Y^n}^{(n)} - Q_{Y^n}^{(n)}\|_{TV} + \|Q_{Y^n}^{(n)} - p_X^{\otimes n}\|_{TV} \ .$$

From (2.69) and Lemma 1.8 (Appendix 1), and from (2.68), the above right hand side goes to zero as n goes to infinity, yielding (2.24). From the additivity of d, we have

$$\mathbb{E}_{Q^{(n)}}[d(X^n, Y^n)] = \mathbb{E}_{\hat{Q}_{\mathcal{X}^2}^{(n)}[X^n, Y^n]}[d(X, Y)],$$

and the same for $P^{(n)}$. Moreover, from the triangle inequality for the TVD, Lemma 1.8 (Appendix 1), and (2.69), we have

$$\left\| \hat{P}_{\mathcal{X}^2}^{(n)}[X^n, Y^n] - \hat{Q}_{\mathcal{X}^2}^{(n)}[X^n, Y^n] \right\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{2.72}$$

From Definition 2.1, d does not take infinite values. Since $\mathcal{X}$ is assumed to be finite, then from Lemma 1.10 (Appendix 1), (2.67), and (2.72), we get (2.23). Moreover, (2.26) follows from (2.71) and (2.69).

2.F.2 Local channel synthesis argument

We prove Claim 2.16. Consider a functional representation $Y = \psi(U, V)$ of $p_{Y|V}$. Similarly to the end of [Cuf13, Appendix 2], we apply the local channel synthesis lemma [Cuf13, Corollary VII.6] — which is related to the channel synthesis problem key to Chapter 5 — with codebook distribution P_U and channel $(v, u) \mapsto \psi(v, u)$. Then, for any $\gamma > 0$, there exists a vanishing sequence $(\delta_n)_n$ such that for any $r > \gamma$ and any $n \in \mathbb{N}$, there exists a mapping $\tilde{P}_{Y^n|V^n}$ with a fixed-length private randomness of rate r such that

$$\left\| \tilde{P}_{Y^n|V^n=v^n} - \prod_{t=1}^n p_{Y|V=v_t} \right\|_{TV} \leq \delta_n, \tag{2.73}$$

for any $v^n \in \mathcal{V}^n$ for which $p_{V,Y}^{(v^n)} := P_{v^n}^{\text{emp}} \cdot p_{Y|V}$ satisfies $r - \gamma \geq H_{p^{(v^n)}}(Y|V)$. Fix some $\gamma > 0$. From (2.26) and the continuity of entropy, we have

$$P_{M,J}^{(n)}\big(H_p(Y|V) + \gamma \geq H_{p^{(v^n(M,J))}}(Y|V)\big) \underset{n\to\infty}{\longrightarrow} 1. \tag{2.74}$$

Take $r = R_d + 2\gamma$. For *good* sequences v^n, i.e. those for which the above event holds true, we have upper bound (2.73) since $R_d \geq H_p(Y|V)$. For other sequences v^n, the TVD is upper-bounded by 1. Then, from Lemma 1.7 (Appendix 1), (2.25) and (2.74), we get

$$\big\| P_{v^n(M,J)}^{(n)} \cdot \tilde{P}_{Y^n|V^n} - P_{v^n(M,J),Y^n}^{(n)} \big\|_{TV} \underset{n\to\infty}{\longrightarrow} 0.$$

Since $P^{(n)}$ satisfies Markov chain $(X^n, M, J) - v^n(M,J) - Y^n$, then from Lemma 1.9 (Appendix 1) we get

$$\big\| P_{X^n,M,J,v^n(M,J)}^{(n)} \cdot \tilde{P}_{Y^n|V^n} - P_{X^n,M,J,v^n(M,J),Y^n}^{(n)} \big\|_{TV} \underset{n\to\infty}{\longrightarrow} 0.$$

2.F.3 Quantizability for Euclidean spaces

We provide a proof of Claim 2.10. Let $\mathcal{X}$ be a finite-dimensional real vector space, with Euclidean distance denoted by d, let p be a distribution on $\mathcal{X}$, and let s, ε, τ be positive reals. Since $\mathcal{X}$ is the union of all balls of integer radius centered at the origin, there exists one, denoted B_τ, such that

$$p(\mathcal{X} \setminus B_\tau) \leq \tau. \tag{2.75}$$

Fix an orthonormal basis of $\mathcal{X}$. A *rectangular prism* is a set of points of $\mathcal{X}$ whose coordinates lie in a product of bounded real intervals. Fix $\ell \in \mathbb{N}$. We define a quantizer $\kappa^{(\ell)}$ on $\mathcal{X}$ as follows. Each coordinate axis can be partitioned into half-open intervals of length ℓ^{-s}. Products of such intervals are called *basic rectangular prisms*. The latter form a partition of $\mathcal{X}$. Let B_ℓ denote the closed ball of radius ℓ, centered at the origin. Then, the Euclidean projection onto B_ℓ of

any point x in $\mathcal{X}$ is uniquely defined as the element of B_ℓ having minimal Euclidean distance to x. Map $\kappa^{(\ell)}$ is defined as:

- On each basic rectangular prism $\mathfrak{r}$ which is included in the interior of B_ℓ, define $\kappa^{(\ell)}$ as the constant mapping to the center of $\mathfrak{r}$.
- On each basic rectangular prism having non-empty intersection with the border of B_ℓ, define $\kappa^{(\ell)}$ as the constant mapping to an arbitrarily chosen representative element. Then, the image of each point on the border of B_ℓ is well-defined.
- For any point x in the remainder of $\mathcal{X}$, let x' denote its Euclidean projection onto (the border of) B_ℓ. Define $\kappa^{(\ell)}(x)$ as $\kappa^{(\ell)}(x')$.

Then, $\kappa^{(\ell)}$ is finite-valued and measurable. For any $x \in \mathcal{X}$, we denote $\kappa^{(\ell)}(x)$ by $[x]$. Let $\mathfrak{C}_\ell$ denote the set of basic rectangular prisms included in the interior of B_ℓ. Then, the set $\cup_{\ell\geq 1}\mathfrak{C}_\ell$ asymptotically generates the set of closed rectangular prisms (products of closed intervals). Hence, it asymptotically generates the Borel σ-algebra of $\mathcal{X}$. From the triangle inequality, we have

$$\forall(x,y,\tilde{x},\tilde{y}) \in \mathcal{X}^4, d(x,y) - d(\tilde{x},\tilde{y}) \leq d(x,\tilde{x}) + d(\tilde{y},y). \tag{2.76}$$

Denote by $\tilde{B}_\ell$ the reunion of all basic rectangular prisms having non-empty intersection with B_ℓ. Its diameter is at most $2\ell + 2\ell^{-s}$, hence at most 4ℓ. Simple calculus implies that there exists a constant $\zeta_{s,\varepsilon} > 0$ depending only on s and ε, such that for any $(a,b) \in [0,4\ell]^2$, we have

$$|a^s - b^s| \leq \varepsilon + |a-b|(\zeta_{s,\varepsilon} + s(4\ell)^{s-1}). \tag{2.77}$$

Hence, from (2.76), for any $(x,y) \in \tilde{B}_\ell^2$, we have

$$\begin{aligned} &|d(x,y)^s - d([x],[y])^s| \\ &\leq \varepsilon + \big(d(x,[x]) + d(y,[y])\big)\big(\zeta_{s,\varepsilon} + s(4\ell)^{s-1}\big). \end{aligned}$$

By construction, $\kappa^{(\ell)}$ maps each element x of $\tilde{B}_\ell$ to an element x'' at Euclidean distance at most $\sqrt{\dim(\mathcal{X})}\ell^{-s}$ of x. Hence,

$$\forall (x,y) \in \tilde{B}_\ell^2, \quad |d(x,y)^s - d([x],[y])^s| \leq 2\varepsilon, \tag{2.78}$$

for large enough ℓ. Fix some $(x,y) \in \mathcal{X}^2 \setminus \tilde{B}_\ell^2$. Denote by x' and y' the Euclidean projections of x and y onto B_ℓ. Since the latter is convex and closed, we have $d(x',y') \leq d(x,y)$. We also have $d([x],[x']), d([y],[y']) \leq 2\sqrt{\dim(\mathcal{X})}l^{-s}$. Hence, from (2.76) and (2.77), for large enough ℓ, we have

$$d([x],[y])^s - d([x'],[y'])^s \leq 2\varepsilon.$$

Combining this, $d(x',y') \leq d(x,y)$, and (2.78) for (x',y') gives:

$$\forall (x,y) \in \mathcal{X}^2 \setminus \tilde{B}_\ell^2, \quad d([x],[y])^s \leq d(x,y)^s + 4\varepsilon, \tag{2.79}$$

for large enough ℓ. Moreover, for large enough ℓ, we have $B_\tau \subset \tilde{B}_\ell$. Since (2.75), (2.78) and (2.79) are true for any $\varepsilon, \tau > 0$ for large enough ℓ, this concludes the proof.

2.F.4 The uniform integrability constraint is non-degenerate

We provide a non-degenerate example of a distribution p_X and a distortion measure d such that (d, p_X) is not uniformly integrable. Consider $\mathcal{X} = \mathbb{R} \times \{0,1\}$ — which is Polish —, and let p_X be the product of the Cauchy distribution with zero mean and scale 1 and the Bernoulli distribution with parameter $1/2$. Define the distortion measure

$$d : (x,y) \mapsto |x_1 - y_1| \text{ if } x_2 = y_2 = 0, \text{ and } d(x,y) = 1 \text{ otherwise.} \tag{2.80}$$

First, (d, p_X) is not uniformly integrable. Indeed, consider Y distributed according to p_X and independent from X. Fix $\tau > 0$, and let $\xi \in \{0,1\}$ be independent from (X,Y), having a

Bernoulli distribution of parameter τ. Then, since p_{X_1} is symmetric,

$$\mathbb{E}[d(X,Y)\xi] \geq \mathbb{E}[\xi]\mathbb{E}[|X_1 - Y_1|\mathbf{1}_{X_2=Y_2=0}] \geq \tfrac{\tau}{4}\mathbb{E}[|X_1 - Y_1|] \geq \tfrac{\tau}{4}\mathbb{E}[|X_1 - Y_1|\mathbf{1}_{X_1Y_1\leq 0}] \geq \tfrac{\tau}{8}\mathbb{E}[|X_1|],$$

which is infinite. This being true for any $\tau > 0$, then (d, p_X) is not uniformly integrable.

We conclude by showing that this example is not degenerate, in that there exist jointly distributed X, Y such that $p_Y \equiv p_X$, and $E[d(X,Y)] < \infty$, with arbitrarily low $I(X;Y)$. Consider, the following coupling: fix a number $M \geq 1$. If $|X_1| > M$, then choose $Y_2 = 1 - X_2$ (so the bit is flipped), and Y_1 is drawn from the distribution p_{X_1} conditioned on $\{|X_1| > M\}$. Otherwise, draw Y_1 from the distribution p_{X_1} conditioned on $\{|X_1| \leq M\}$, and draw Y_2 uniformly. Then, $E[d(X,Y)] \leq 2M < \infty$. We move to upper bounding $I(X;Y)$. Since we have the Markov chain $X_2 - X_1 - \mathbf{1}_{\{|X_1|>M\}} - Y_1$, and X_2 is independent of X_1, and Y_2 is independent of (X_1, X_2, Y_1) knowing $\{|X_1| \leq M\}$, then

$$I(X;Y) = I(X_1;Y_1) + I(X_2;Y_1|X_1) + I(X_1,X_2;Y_2|Y_1) \leq H(\mathbf{1}_{\{|X_1|>M\}}) + 0 + Pr(|Y_1| > M)H(Y_2),$$

where the last term comes from writing conditional mutual information in integral form. This is valid when the latter is finite, by [Gra11, (7.32)]. Moreover, the integrand is exactly the mutual infomration with respect to the conditional distribution, because Y_2 is finite, as detailed below:

$$\begin{aligned}
&I(X_1, X_2; Y_2|Y_1) \\
&= \int \ln\Big(\frac{dP_{Y_1,X_1,X_2,Y_2}}{d(P_{Y_1}\cdot P_{X_1,X_2|Y_1}\cdot P_{Y_2|Y_1}}(y_1,x_1,x_2,y_2)\Big) dP_{Y_1,X_1,X_2,Y_2}(y_1,x_1,x_2,y_2) && (2.81)\\
&= \int \ln\Big(\frac{dP_{X_1,X_2,Y_2|Y_1=y_1}}{d(P_{X_1,X_2|Y_1=y_1}\cdot P_{Y_2|Y_1=y_1})}(x_1,x_2,y_2)\Big) dP_{Y_1,X_1,X_2,Y_2}(y_1,x_1,x_2,y_2) && (2.82)\\
&= \int I_{P_{X_1,X_2,Y_2|Y_1=y_1}}(X_1,X_2;Y_2) dP_{Y_1}(y_1), && (2.83)
\end{aligned}$$

where (2.81) follows from [Gra11, (7.32)] because $I(X_1, X_2; Y_2|Y_1) < \infty$; (2.82) holds because we know that the two Radon-Nykodim derivatives are equal P_{Y_1}-almost everywhere whenever the second exists, and because the latter is an information density, which exists when the

corresponding mutual information is finite, by Proposition 1.17; and (2.83) follows again from [Gra11, (7.32)]. Hence, $I(X;Y)$ can be made arbitrarily small by taking M large.

Chapter 3

Rate-Distortion-Perception Trade-off with Strong Realism Constraints: Role of Side Information and Common Randomness

3.1 Introduction

In Section 2.3 of Chapter 2, we presented a mathematical formalism for the point-to-point setting, that encompasses Shannon's traditional RD theory, and the RDP trade-off with perfect or near-perfect realism, in the strong or per-symbol sense. In Proposition 2.1, we stated Shannon's characterization of the optimal RD trade-off, for which randomization is unnecessary. In [WZ76], Wyner and Ziv generalized this result to the case where a side information Z^n, correlated with the source X^n, is available either only at the decoder or at both the encoder and the decoder. The problem is of practical interest because it models settings such as video coding [SPC+23], where a previously compressed frame serves as side information that reduces the rate needed for the current frame; here the side information models such a block of previously compressed data[1]. The focus of this chapter is to revisit the problem through the contemporary paradigm of realism requirements and randomization. The setup is depicted in Fig. 3.1, and is

[1]This is distinct from successive-refinement coding, to which we claim no connection: there, a single source sample is first described coarsely and then refined, whereas here an earlier, fixed reconstruction assists the coding of a *new* source.

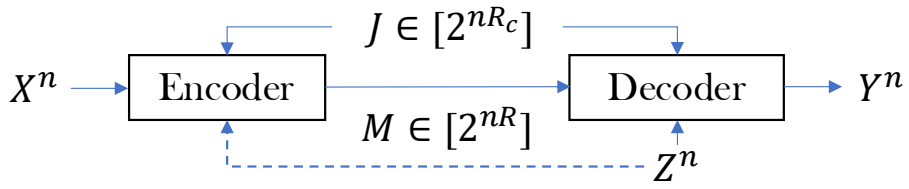


Figure 3.1: The system model. The side information Z^n is always available at the decoder, but not necessarily at the encoder.

quite similar to the one in the previous chapter. We dedicate substantial attention to the role of common randomness, and assume unlimited private randomness (our results from Chapter 2 can be extended to this setup). Indeed, the problem is of theoretical interest due to the dual role that the side information Z^n can play. On the one hand, it provides the receiever with an independent look at the source. On the other hand, it can serve as a source of common randomness (as is apparent in the extreme case in which Z^n is independent of the source).

As proved in Theorem 2.13 of the previous chapter, common randomness is irrelevant when mere per-symbol realism is required. In the present chapter, we consider only strong realism constraints. We prove various coding theorems that elucidate these dual roles and the tension between them. It turns out that the ability of Z^n to simultaneously play both roles depends on the nature of the realism constraint, specifically whether we require $P_{Y^n} \equiv P_{X^n}$ (which we call a *marginal realism constraint*) or $P_{Y^n,Z^n} \equiv P_{X^n,Z^n}$ (which we call a *joint realism constraint*) — the latter being similar to a constraint considered in a one-shot video compression setup with fixed encoder in the case of Euclidean distortion measure [SPC+23]. There is other relevant prior work involving similar settings. A preliminary study of our problem was carried out in [NGBH23] in the case of the marginal realism constraint with side information available at both encoder and decoder, and an inner bound was derived. In [ZQCK21], universal (fixed) encoders and successive refinement are considered under per-symbol realism constraints. Closed form expressions for the latter results are derived in [QZCK22] for binary sources and Hamming distortion measure. One-shot multiple description coding is considered in [KZLK13].

Our main findings are the following. Under a marginal realism constraint, when Z^n is present at both terminals, Z^n can play both of the aforementioned roles in a particularly strong sense: the side information Z^n reduces the rate required for compression, as in conventional RD,

while simultaneously providing a common-randomness rate of $H(Z|Y)$. Conversely, under a joint realism constraint, the side information does not provide any common-randomness at all. Under a marginal constraint with side information available only at the decoder, we provide an inner bound on the achievable region. The challenge in obtaining a conclusive result in this case seems to center around characterizing the extent to which Z^n can also serve as common randomness. As evidence, we provide conclusive results in two special cases for which the ability of Z^n to provide useful common randomness is clear.

We examine the quadratic Gaussian case in detail. When side information Z^n is available and $p_{X,Z}$ is bi-variate Gaussian, we show that for large enough common randomness rate, the RD trade-off is the same whether Z^n is available only at the encoder or at both terminals. This is similar to traditional RD theory — although our result requires sufficient common randomness.

Overall, the main contributions of this chapter can be summarized as follows.

- We characterize the three-way trade-off between compression rate, common randomness rate and distortion in the presence of side information, for strong perfect and near-perfect realism constraints (Definition 3.2 to follow).
- We do so under both the marginal realism and joint realism constraints mentioned above (see Definition 3.2 to follow for details).
- We characterize how the side information plays a dual role as a source of extra common randomness and a second look on the source, when it is available at both terminals, under a marginal realism constraint. We show that this does not hold when the latter is replaced with a joint realism constraint.
- We prove the equivalence between near-perfect marginal (resp. joint) realism and perfect marginal (resp. joint) realism.
- We prove that many of our results hold for general alphabets under mild assumptions.
- When the source and side information are jointly Gaussian, we show that the trade-off is the same whether the side information is available at both terminals, or only at the

decoder, as long as additional common randomness is available at a sufficient rate.

The chapter is organized as follows. We give the problem formulation and state the equivalence between perfect and near-perfect realism in Section 3.2. We present all other results in Section 3.3. We prove our main theorems in Sections 3.4, 3.5, 3.6 and 3.7. Section 3.8 is dedicated to special cases and corollaries, and Section 3.9 to the case of Gaussian distributions. Other proofs are deferred to the appendix.

3.2 Problem formulation and a first property

3.2.1 Definitions

We consider an additive distortion measure (Definition 2.1 from Chapter 2). As shown in Figure 3.1, we consider two cases: side information is either only available at the decoder or available at both the encoder and the decoder. We also study the case where no side information is available at either terminal. Hereafter, we formalize the latter as the special case of constant side information. Throughout this chapter, we do not impose any constraint on each terminal's private randomness, and study the role of common randomness — the role of private randomness in the absence of side information was characterized in Chapter 2. For that reason, in the absence of side information, the definition of an (n, R, R_c) D-code (or E-D-code) stated below reduces to that of an (n, R, R_c, ∞) fixed-length code (Definition 2.3).

Definition 3.1 *Given two measurable spaces with respective source alphabet $\mathcal{X}$ and side information alphabet $\mathcal{Z}$, an (n, R, R_c) D-code (resp. E-D-code) is a privately randomized encoder and decoder couple $(F^{(n)}, G^{(n)})$ consisting of a conditional probability kernel $F^{(n)}_{M|X^n,J}$ (resp. $F^{(n)}_{M|X^n,Z^n,J}$) from $\mathcal{X}^n \times [2^{nR_c}]$ (resp. $\mathcal{X}^n \times \mathcal{Z}^n \times [2^{nR_c}]$) to $[2^{nR}]$, and a conditional probability kernel $G^{(n)}_{Y^n|Z^n,J,M}$ from $\mathcal{Z}^n \times [2^{nR_c}] \times [2^{nR}]$ to $\mathcal{X}^n$. Given a distribution $p_{X,Z}$ on $\mathcal{X} \times \mathcal{Z}$, the following distribution*

$$P_{X^n,Z^n,J,M,Y^n} := p^{\otimes n}_{X,Z} \cdot p^{\mathcal{U}}_{[2^{nR_c}]} \cdot F^{(n)} \cdot G^{(n)}$$

is called the distribution induced by the code. Note that the product form of the induced distribution above is part of the definition: conditionally on their respective inputs, the encoder's and the decoder's private randomizations are mutually independent, and independent of (X^n, Z^n)*. In particular, the kernels* $F^{(n)}$ *and* $G^{(n)}$ *carry no common randomness beyond* J *— mirroring Definition 2.2, in which the encoder and decoder are deterministic maps fed by the explicitly separated variables* J*,* L_e *and* L_d*. An* (n, R) *D-code (resp. E-D-code) with unconstrained common randomness is a triplet* $(p_J, F^{(n)}, G^{(n)})$ *: the common randomness variable* J *is allowed to be defined on any alphabet (instead of* $[2^{nR_c}]$*) — possibly uncountable — and to have any distribution* p_J *(instead of a uniform one).*

The following lemma is a straightforward reformulation.

Lemma 3.1 *Consider a distribution* P_{X^n,Z^n,J,M,Y^n} *on* $\mathcal{X}^n \times \mathcal{Z}^n \times \mathbb{N} \times \mathbb{N} \times \mathcal{X}^n$ *having regular transition kernels* $P_{M|X^n,J}$ *(resp.* $P_{M|X^n,Z^n,J}$*) and* $P_{Y^n|Z^n,J,M}$*. Suppose that there exist non-negative reals* R *and* R_c *such that* $M \in [2^{nR}]$ *and* $J \in [2^{nR_c}]$*,* P*-almost surely. Then,* P *coincides with the distribution induced by* (n, R, R_c) *D-code* $(P_{M|X^n,J}, P_{Y^n|Z^n,J,M})$ *(resp. E-D code* $(P_{M|X^n,Z^n,J}, P_{Y^n|Z^n,J,M})$*) iff* P *satisfies all of the following properties:*

- $P_{X^n,Z^n,J} \equiv p_{X,Z}^{\otimes n} \cdot p_{[2^{nR_c}]}^{\mathcal{U}}$
- $X^n - (Z^n, J, M) - Y^n$
- *(resp.* $Z^n - (X^n, J) - M$ *and* $X^n - (Z^n, J, M) - Y^n$*).*

The formal definition of achievability below is very similar to the one in Chapter 2 (Definition 2.4) with $R_d = \infty$.

Definition 3.2 *Given source and side-information alphabets* $\mathcal{X}$ *and* $\mathcal{Z}$ *and a joint distribution* $p_{X,Z}$ *on* $\mathcal{X} \times \mathcal{Z}$*, a triplet* (R, R_c, Δ) *is said to be D-achievable (resp. E-D-achievable) with near-perfect marginal realism if there exists a sequence* $(F^{(n)}, G^{(n)})_n$ *of* (n, R, R_c) *D-codes (resp. E-*

Table 3.1: Main results of Chapter 3

	E-D-codes	**D-codes**
Marginal realism	Theorem 3.3 $R \geq I_p(X;V\|Z)$ $R+R_c \geq I_p(Y;V\|Z) - H_p(Z\|Y)$ $\Delta \geq \mathbb{E}_p[d(X,Y)]$ $p_Y \equiv p_X$ $X-(Z,V)-Y$	Theorem 3.9 $R \geq I_p(X;V) - I_p(Z;V)$ $R+R_c \geq I_p(Y;V) - I_p(Z;V)$ $\Delta \geq \mathbb{E}_p[d(X,Y)]$ $p_Y \equiv p_X$ $X-(Z,V)-Y$ and $Z-X-V$
Joint realism	Theorem 3.13 $R \geq I_p(X;V\|Z)$ $R+R_c \geq I_p(Y;V\|Z)$ $\Delta \geq \mathbb{E}_p[d(X,Y)]$ $p_{Y,Z} \equiv p_{X,Z}$ $X-(Z,V)-Y$	Theorem 3.14 $R \geq I_p(X;V\|Z)$ $R+R_c \geq I_p(Y;V\|Z)$ $\Delta \geq \mathbb{E}_p[d(X,Y)]$ $p_{Y,Z} \equiv p_{X,Z}$ $X-(Z,V)-Y$ and $Z-X-V$

Table 3.2: Corollaries in Chapter 3

	E-D-codes	**D-codes**
Marginal realism	Corollaries 3.4, 3.5, and 3.7	Theorem 3.11 and 3.12

D-codes) such that

$$\limsup_{n\to\infty} \mathbb{E}_P[d(X^n,Y^n)] \leq \Delta, \tag{3.1}$$

$$\|P_{Y^n} - p_X^{\otimes n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{3.2}$$

When there exists an integer N such that for all $n \geq N$, $P_{Y^n} \equiv p_X^{\otimes n}$, then (R, R_c, Δ) is said to be D-achievable (resp. E-D-achievable) with perfect marginal realism. Each of the above notions has an analogue with joint realism (instead of marginal realism), which is defined by replacing P_{Y^n} by P_{Y^n,Z^n} and $p_X^{\otimes n}$ by $p_{X,Z}^{\otimes n}$. Moreover, each notion of achievability extends to the case of large common randomness rate. A couple (R,Δ) is said to be achievable with finite-rate common randomness if there exists a non-negative real R_c such that (R, R_c, Δ) is achievable. More generally, a couple (R,Δ) is said to be achievable with unconstrained common randomness if the above properties hold for some sequence of (n,R) codes with unconstrained common randomness.

Our main results for the different notions of achievability are listed in Table 3.1. Main corollaries are listed in Table 3.2.

3.2.2 Equivalence of the perfect / near-perfect realism problems

Our first result shows that any sequence of codes that achieves near-perfect realism can be upgraded to one that achieves perfect realism with no asymptotic change to the rates or the distortion. The assumption of unlimited decoder private randomness is key. See Appendix 3.B for the proof, which is very similar to the one for the point-to-point setting in [Wag22] — which was used in Appendix 2.D.3 of Chapter 2, where we pointed to the present chapter for a more general version and a detailed proof. It does not rely on a single-letter characterization of a set of achievable RD tuples.

Theorem 3.2 *Given Polish source and side-information alphabets $\mathcal{X}$ and $\mathcal{Z}$, a joint distribution $p_{X,Z}$ on $\mathcal{X} \times \mathcal{Z}$ and a distortion measure d on $\mathcal{X}$ such that (d, p_X) is uniformly integrable, then a triplet (R, R_c, Δ) is D-achievable (resp. E-D-achievable) with near-perfect marginal realism if and only if it is D-achievable (resp. E-D-achievable) with perfect marginal realism. The same applies for achievability with unconstrained common randomness, and for joint realism.*

Following [Wag22], Theorem 3.2 is used directly in the converse proofs of our main results, stated in Section 3.3, and for achievability proofs, we use a slightly stronger result (Propositions 3.30 and 3.31, Appendix 3.B), in conjunction with the uniform integrability assumption — similarly to Sections 2.D.3 and 2.D.4 from Chapter 2.

3.3 Main results

3.3.1 Marginal realism when the side information is available at both encoder and decoder

In all of Section 3.3, the superscript (m) (resp. (j)) refers to marginal (resp. joint) realism. The proof of the following theorem is provided in Section 3.4.

Theorem 3.3 *Consider a Polish source alphabet $\mathcal{X}$, a finite side-information alphabet $\mathcal{Z}$, a joint distribution $p_{X,Z}$ on $\mathcal{X} \times \mathcal{Z}$ and a distortion measure d such that (d, p_X) is uniformly integrable. Define the region $\mathcal{S}_{E\text{-}D}^{(m)}$ as*

$$\left\{ \begin{array}{rcl} (R, R_c, \Delta) \in \mathbb{R}_{\geq 0}^3 & : & \exists\ p_{X,Z,V,Y} \in \mathcal{D}_{E\text{-}D}^{(m)}\ s.t. \\ R & \geq & I_p(X;V|Z) \\ R + R_c & \geq & I_p(Y;V|Z) - H_p(Z|Y) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}, \tag{3.3}$$

with $\mathcal{D}_{E\text{-}D}^{(m)}$ defined as

$$\left\{ \begin{array}{c} p_{X,Z,V,Y} : (X,Z) \sim p_{X,Z},\ p_Y \equiv p_X \\ X - (Z,V) - Y \end{array} \right\}, \tag{3.4}$$

where the alphabet of V is constrained to be finite. Then, the closure of the set $\mathcal{A}_{E\text{-}D}^{(m)}$ of E-D-achievable triplets (R, R_c, Δ) with perfect or near-perfect marginal realism is equal to the closure of $\mathcal{S}_{E\text{-}D}^{(m)}$.

Corollary 3.4 *Consider the setting of Theorem 3.3. Assume that $p_{X,Z} \equiv p_X \cdot p_Z$. Then, we have*

$$\mathcal{S}_{E\text{-}D}^{(m)} + (0, H(Z), 0) = \mathcal{S}^{(m)} \cap \Big(\mathbb{R}_{\geq 0} \times [H(Z), \infty) \times \mathbb{R}_{\geq 0} \Big), \tag{3.5}$$

where $\mathcal{S}^{(m)}$ is the optimal region in the absence of side information, as stated in Corollary 3.5 to follow, and is defined as

$$\left\{ \begin{array}{rcl} (R, R_c, \Delta) \in \mathbb{R}_{\geq 0}^3 & : & \exists\ p_{X,V,Y} \in \mathcal{D}^{(m)}\ s.t. \\ R & \geq & I_p(X;V) \\ R + R_c & \geq & I_p(Y;V) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}, \tag{3.6}$$

with $\mathcal{D}^{(m)}$ *defined as*

$$\left\{ \begin{array}{c} p_{X,V,Y} : X \sim p_X,\ p_Y \equiv p_X \\ X - V - Y \end{array} \right\}, \tag{3.7}$$

where the alphabet of V *is constrained to be finite.*

The proof is provided in Section 3.8. When the side information is independent from the source, the former is equivalent to a uniform common randomness of rate $H(Z)$. In general (Theorem 3.3), the side information plays a dual role: it provides common randomness independent from the source, as well as a second look on the source. The conditioning with respect to the side information is reminiscent of the result of Wyner and Ziv [WZ76] for lossy source coding. We can also rewrite the second inequality in (3.3) as

$$R + R_c + H_p(Z) \geq I_p(Y; Z, V). \tag{3.8}$$

The left hand side is the effective rate of the codebook used in our scheme (Section 3.4) and the right hand side corresponds to the requirement for the soft covering lemma with side information [Cuf13, Corollary VII.5]. The formal codebook rate is $R + R_c + \log|\mathcal{Z}|$, but atypical strings z^n have a negligible impact on the distortion and TVD. In the absence of side information, we recover [Wag22, Theorem 2] — which also follows from our in-depth analysis in Chapter 2 of the role of randomness in the point-to-point setting. The following result, which is a direct consequence of Theorem 3.3, was also extended in [XLCZ25].

Corollary 3.5 *Consider a Polish source alphabet* $\mathcal{X}$*, a distribution* p_X *on* $\mathcal{X}$ *and a distortion measure* d *such that* (d, p_X) *is uniformly integrable. Then, the closure of the set* $\mathcal{A}^{(m)}$ *of achievable triplets* (R, R_c, Δ) *with perfect or near-perfect marginal realism is equal to the closure of* $\mathcal{S}^{(m)}$ *(Eq. 3.6).*

Remark 3.6 *The rate gains allowed by the availability of a finite-valued side information are at most the following. In the setting of Theorem 3.3, by definition of* $\mathcal{S}^{(m)}_{E\text{-}D}$ *and* $\mathcal{S}^{(m)}$*, for any*

$(R, R_c, \Delta) \in \mathcal{S}^{(m)}_{E\text{-}D}$,

$$(R + I(X;Z), R_c + H(Z|X), \Delta) \in \mathcal{S}^{(m)}.$$

Note that according to our definition (Definition 3.2) of achievability of a triplet (R, R_c, Δ), then when $R_c = 0$ absolutely no common randomness is available. The other extreme, $R_c \to \infty$, is also of interest. We make a distinction between the case of large common randomness rate and that of unconstrained common randomness.

Corollary 3.7 *Consider a Polish source alphabet $\mathcal{X}$, a finite side-information alphabet $\mathcal{Z}$, a joint distribution $p_{X,Z}$ on $\mathcal{X} \times \mathcal{Z}$ and a distortion measure d such that (d, p_X) is uniformly integrable. Then, the closure of the set $\mathcal{A}^{(m)}_{E\text{-}D,\infty}$ of couples (R, Δ) which are E-D-achievable with perfect or near-perfect marginal realism with finite-rate common randomness is the closure of the following region $\mathcal{S}^{(m)}_{E\text{-}D,\infty}$:*

$$\left\{ \begin{array}{rcl} (R,\Delta) \in \mathbb{R}^2_{\geq 0} & : & \exists\ p_{X,Z,V,Y} \in \mathcal{D}^{(m)}_{E\text{-}D}\ s.t. \\ R & \geq & I_p(X;V|Z) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}. \tag{3.9}$$

If we also have that $\mathcal{X}$ is finite, then E-D-achievability with unconstrained common randomness is equivalent to E-D-achievability with finite-rate common randomness, and $\mathcal{S}^{(m)}_{E\text{-}D,\infty}$ rewrites as:

$$\left\{ \begin{array}{rcl} (R,\Delta) \in \mathbb{R}^2_{\geq 0} & : & \exists\ p_{X,Z,Y} \in \mathcal{D}^{(m)}_{\infty}\ s.t. \\ R & \geq & I_p(X;Y|Z) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}, \tag{3.10}$$

with $\mathcal{D}^{(m)}_{\infty}$ defined as

$$\left\{ \quad p_{X,Z,Y} : (X,Z) \sim p_{X,Z},\ p_Y \equiv p_X \quad \right\}. \tag{3.11}$$

The proof is provided in Section 3.8.3. In the absence of side information, region (3.10)

becomes the region corresponding to (2.2) and we recover the result described at the beginning of [SLY15a, Section III.A] regarding the case of finite alphabets. See the latter for further remarks.

Remark 3.8 *If the target distribution has infinite entropy, it is unclear whether it is possible to achieve every point in (the interior of) region* (3.10) *— nor* (2.2) *— with finite-rate common randomness. Whatever common randomness is available, region* (3.10) *is an outer bound of the set of achievable couples — even for variable-length codes — (see the proof of Corollary 3.7). In the absence of side information, then as shown in [SLY15b], region* (2.2) *is achievable with perfect realism using some continuous common randomness variable — and variable-length codes —, for general source alphabets. As can be seen from [Wag22, Proposition 1] (re-stated in Proposition 3.16 to follow, for completeness), for a source with one-dimensional Gaussian distribution, every point in the interior of* (2.2) *is achievable with finite-rate common randomness (see also [SLY15a]). Hence, in that case a discrete common randomness variable with infinite rate is sufficient. Moreover, as stated in Proposition 3.17, when* (X, Z) *is a bi-dimensional Gaussian with non-zero correlation, region* (3.10) *is optimal and there is a finite common randomness rate with which one can achieve any point in its interior.*

3.3.2 Marginal realism when the side information is only available at the decoder

Theorem 3.9 *Consider Polish source and side-information alphabets* $\mathcal{X}$ *and* $\mathcal{Z}$*, a joint distribution* $p_{X,Z}$ *on* $\mathcal{X} \times \mathcal{Z}$ *and a distortion measure* d *such that* (d, p_X) *is uniformly integrable. Define the region* $\mathcal{S}_D^{(m)}$ *as*

$$\left\{ \begin{array}{rcl} (R, R_c, \Delta) \in \mathbb{R}^3_{\geq 0} & : & \exists\ p_{X,Z,V,Y} \in \mathcal{D}_D^{(m)}\ s.t. \\ R & \geq & I_p(X;V) - I_p(Z;V) \\ R + R_c & \geq & I_p(Y;V) - I_p(Z;V) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}, \tag{3.12}$$

with $\mathcal{D}_D^{(m)}$ *defined as*

$$\left\{ \begin{array}{c} p_{X,Z,V,Y} : (X,Z) \sim p_{X,Z},\ p_Y \equiv p_X \\ Z - X - V, \quad X - (Z,V) - Y \\ I_p(Z;V) < \infty \end{array} \right\}, \tag{3.13}$$

where the alphabet of V *is constrained to be Polish. Then, the closure of the set* $\mathcal{A}_D^{(m)}$ *of D-achievable triplets* (R, R_c, Δ) *with perfect or near-perfect marginal realism contains the closure of* $\mathcal{S}_D^{(m)}$.

The proof is provided in Section 3.5.

Remark 3.10 *The first inequality in* (3.12) *is reminiscent of the result of Wyner and Ziv [WZ76] for traditional lossy source coding with side information. We outline the idea of the achievability of Theorem 3.9 and its relation to Wyner–Ziv coding [WZ76]. As in Wyner–Ziv lossy coding, the side information is available only at the decoder, and the description rate is reduced by* binning *the codebook against that side information: the codebook of auxiliary sequences* V^n *is partitioned into bins, only a bin index is described to the decoder, and the decoder uses its side information* Z^n *to identify the intended codeword within the bin. Concretely, in Subsection 3.5.2 we introduce a virtual message* M' *of rate* $R' \approx I_p(Z;V)$*, which plays the role of the bin content that the side information resolves; this is what produces the rate saving* $-I_p(Z;V)$ *in the first inequality of* $\mathcal{S}_D^{(m)}$*, mirroring the corresponding saving in the Wyner–Ziv rate. The point of departure from Wyner–Ziv is the realism constraint: rather than merely reconstructing* X^n *within a distortion, the reconstruction* Y^n *must have (near) the source distribution* $p_X^{\otimes n}$*. This is obtained not by covering but by* soft covering*: the soft covering lemma [Cuf13, Corollary VII.4] is applied so that the codebook, together with the memoryless kernel* $p_{Y|V}$*, synthesizes an output whose distribution is* $p_X^{\otimes n}$ *in total variation, which requires the second rate condition* $R + R_c \geq I_p(Y;V) - I_p(Z;V)$*. Thus the achievability combines Wyner–Ziv binning against the decoder side information with soft-covering for realism. The contrast with the two-terminal case of Theorem 3.3, where the side information is also available at the*

encoder and no such binning is used, is what makes precise the sense in which the role of the side information here parallels that in Wyner–Ziv coding.

The difficulty in obtaining a conclusive characterization appears to be attributable to the difficulty in determining the extent to which the side information can be used to produce common randomness. As evidence of this, we show that the region in (3.12) is optimal in two cases, one for which this question is clear, and one for which it is irrelevant. The first is when the source and side information are conditionally independent with respect to a common component. Previous work has shown that making use of common components can strictly improve distributed compression schemes without realism constraints [WKA11].

Theorem 3.11 *Consider Polish source and side-information alphabets $\mathcal{X}$ and $\mathcal{Z}$, a joint distribution $p_{X,Z}$ on $\mathcal{X} \times \mathcal{Z}$, and a distortion measure d such that (d, p_X) is uniformly integrable. Assume that there exists (deterministic) functions ϕ, ψ such that, under $p_{X,Z}$, the Markov chain $X - \phi(X) - Z$ holds and $\phi(X) = \psi(Z)$ almost surely and $\phi(X)$ has finite support. Then, the achievable region of Theorem 3.9 is optimal, i.e., $\overline{\mathcal{A}}_D^{(m)} = \overline{\mathcal{S}}_D^{(m)}$.*

The proof is provided in Section 3.8.1. The achievable region of Theorem 3.9 is also optimal when common randomness is available at any rate, and either $\mathcal{X}$ or $\mathcal{Z}$ is finite, as follows. The proof is provided in Section 3.8.2.

Theorem 3.12 *Consider Polish source and side-information alphabets $\mathcal{X}$ and $\mathcal{Z}$, a joint distribution $p_{X,Z}$ on $\mathcal{X} \times \mathcal{Z}$ and a distortion measure d such that (d, p_X) is uniformly integrable. If either of $\mathcal{X}$ or $\mathcal{Z}$ is finite, then the closure of the set $\mathcal{A}_{D,\infty}^{(m)}$ of D-achievable couples with finite-rate common randomness with perfect or near-perfect marginal realism is the closure of the following set $\mathcal{S}_{D,\infty}^{(m)}$:*

$$\left\{ \begin{array}{rcl} (R, \Delta) \in \mathbb{R}_{\geq 0}^2 & : & \exists\ p_{X,Z,V,Y} \in \mathcal{D}_D^{(m)}\ s.t. \\ R & \geq & I_p(X;V) - I_p(Z;V) \\ \infty & > & I_p(Y;V) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}. \tag{3.14}$$

3.3.3 When imposing joint realism of the reconstruction and the side information

The perfect joint realism constraint is akin to a perfect marginal realism constraint with conditioning on the realization of the side information. When the side information is available at both terminals, the optimal single-letter region corresponds to the one in the absence of side information with added conditioning on Z, in each mutual information term, in the distributional constraint, and in the Markov chain constraint. The codebook used in the corresponding scheme comprises of one sub-codebook per element of $\mathcal{Z}^n$.

Theorem 3.13 *Consider a Polish source alphabet $\mathcal{X}$, a finite side-information alphabet $\mathcal{Z}$, a joint distribution $p_{X,Z}$ on $\mathcal{X} \times \mathcal{Z}$ and a distortion measure d such that (d, p_X) is uniformly integrable. Define the region $\mathcal{S}^{(j)}_{E\text{-}D}$ as*

$$\left\{ \begin{array}{rcl} (R, R_c, \Delta) \in \mathbb{R}^3_{\geq 0} & : & \exists\ p_{X,Z,V,Y} \in \mathcal{D}^{(j)}_{E\text{-}D}\ s.t. \\ R & \geq & I_p(X;V|Z) \\ R + R_c & \geq & I_p(Y;V|Z) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}, \tag{3.15}$$

with $\mathcal{D}^{(j)}_{E\text{-}D}$ defined as

$$\left\{ \begin{array}{c} p_{X,Z,V,Y} : (X,Z) \sim p_{X,Z},\ p_{Y,Z} \equiv p_{X,Z} \\ X - (Z,V) - Y \end{array} \right\}, \tag{3.16}$$

where the alphabet of V is constrained to be finite. Then, the closure of the set $\mathcal{A}^{(j)}_{E\text{-}D}$ of E-D-achievable triplets (R, R_c, Δ) with perfect or near-perfect joint realism is equal to the closure of $\mathcal{S}^{(j)}_{E\text{-}D}$.

The proof is provided in Section 3.6. We note the absence of entropy term contrary to the case of marginal realism (Theorem 3.3). Hence, in the case of joint realism the side information does not seem to act as a source of common randomness. When the side information is only

available at the decoder, we obtain a full characterization valid in all cases, contrary to the case of marginal realism (Theorems 3.11,3.12).

Theorem 3.14 *Consider Polish source and side information alphabets $\mathcal{X}$ and $\mathcal{Z}$, a joint distribution $p_{X,Z}$ on $\mathcal{X} \times \mathcal{Z}$ and a distortion measure d such that (d, p_X) is uniformly integrable. Define the region $\mathcal{S}_D^{(j)}$ as*

$$\left\{ \begin{array}{rcl} (R, R_c, \Delta) \in \mathbb{R}^3_{\geq 0} & : & \exists\ p_{X,Z,V,Y} \in \mathcal{D}_D^{(j)}\ s.t. \\ R & \geq & I_p(X;V|Z) \\ R + R_c & \geq & I_p(Y;V|Z) \\ \Delta & \geq & \mathbb{E}_p[d(X,Y)] \end{array} \right\}, \tag{3.17}$$

with $\mathcal{D}_D^{(j)}$ defined as

$$\left\{ \begin{array}{c} p_{X,Z,V,Y} : (X,Z) \sim p_{X,Z},\ p_{Y,Z} \equiv p_{X,Z} \\ Z - X - V, \quad X - (Z,V) - Y \\ I_p(Z;V) < \infty \end{array} \right\}, \tag{3.18}$$

where the alphabet of V is constrained to be Polish. Then, the closure of the set $\mathcal{A}_D^{(j)}$ of E-D-achievable triplets (R, R_c, Δ) with perfect or near-perfect joint realism is equal to the closure of $\mathcal{S}_D^{(j)}$.

The proof is provided in Section 3.7.

Remark 3.15 *The results of this chapter concern near-perfect (and perfect) marginal realism, i.e. the regime $\lambda = 0$ of the imperfect strong realism constraint $\limsup_n \mathcal{D}(p_X^{\otimes n}, P_{Y^n}) \leq \lambda$. We discuss here to what extent they carry over to $\lambda > 0$.*

Limited common randomness. *With a finite common randomness rate R_c, our characterizations do not extend to $\lambda > 0$, and this is not a limitation specific to the presence of side information: already in the point-to-point setting (no side information), the imperfect strong realism trade-off under a general divergence lacks a matching converse and its single-letter*

characterization is an open problem [CYW+22, Remark 11]. Since the setting of this chapter recovers that point-to-point problem when the side information is a constant (Corollary 3.5), no matching converse can be expected here either.

Unlimited common randomness. *When common randomness is available at any rate, the trade-off does extend to $\lambda > 0$, for the class of divergences $\mathcal{D}$ that are convex in their second argument, tensorizable and decomposable in the sense of [CYW+22] (which includes the total variation and Kullback–Leibler divergences), when the side information is available at both terminals. The converse is straightforward: the rate bounds in the converse of Theorem 3.3 do not rely on the value of the realism level, and the single-letterization of the imperfect constraint $\frac{1}{n}\mathcal{D}(p_X^{\otimes n}, P_{Y^n}) \leq \lambda$ follows from the convexity, tensorizability and decomposability of $\mathcal{D}$ exactly as in the point-to-point case [CYW+22, Theorem 7], the conditioning on Z being inherited unchanged from our converse. The achievability follows from channel synthesis with side information at both terminals [Cuf13, Corollary VII.5]: it suffices to synthesize, instead of the target $p_{Y|Z,V}^{\otimes n}$ whose output marginal is $p_X^{\otimes n}$, a target whose output marginal is an admissible q_n with $\mathcal{D}(p_X, q_n) \leq \lambda$, which the soft covering lemma with side information achieves under the same rate condition $R + R_c + H_p(Z) \geq I_p(Y; Z, V)$ as in Section 3.4.*

When the side information is available only at the decoder, it is unclear whether the achievability extends to $\lambda > 0$. Our achievability for that setting relies on binning the codebook against the decoder-only side information, and it is not clear that channel synthesis can be carried out against a target output marginal $q_n \neq p_X$ while binning against side information that the encoder does not observe. We leave this case open.

3.3.4 The Gaussian case

The quadratic Gaussian scenario illustrates the difference between the case where no common randomness is available ($R_c = 0$) and the case where it is available at a large or even infinite rate. We start by re-stating a detailed characterization in the absence of side information, proved in [Wag22] — an extension appeared in [XLCZ25].

Proposition 3.16 *[Wag22, Proposition 1] In the absence of side information, if the source*

X is one-dimensional standard Gaussian and $d : (x, y) \mapsto (x - y)^2$, then for $\Delta \in (0, 2)$ and $R_c \geq 0$, then the infimum of rates R such that (R, R_c, Δ) is achievable with near-perfect or perfect marginal realism is

$$R(\Delta, R_c) = \frac{1}{2} \log \frac{1}{1 - \rho^2}, \tag{3.19}$$

where ρ is such that

$$2\rho^2 + 2^{2R_c} - 1 = \sqrt{(2^{2R_c} - 1)^2 + 2^{2R_c}(2 - \Delta)^2}.$$

For completeness[2], we provide a proof in Section 3.9.1. Interestingly, similarly to standard source coding with side information [EGK11, Example 11.2], when the source and side information form a bi-variate Gaussian, then D-achievability is equivalent to E-D-achievability if common randomness is available at a large enough rate.

Proposition 3.17 *Consider a real-valued source and side information with Gaussian joint distribution*

$$p_{X,Z} = \mathcal{N}\left(0, \begin{pmatrix} 1 & \eta \\ \eta & 1 \end{pmatrix}\right),$$

where $\eta \in (-1, 1)$, and the distortion function $d : (x, y) \mapsto (x - y)^2$. Consider Δ in $(0, 2 - 2|\eta|]$, and denote $1 - \Delta/2$ by ρ. Then, the infimum of rates such that (R, Δ) is D- or E-D-achievable with finite-rate common randomness and (near-)perfect marginal realism is:

$$R_D(\Delta) = R_{E\text{-}D}(\Delta) = \frac{1}{2} \log \Big(\frac{1 - \eta^2}{1 - \rho^2}\Big). \tag{3.20}$$

The proof is provided in Section 3.9.2. The numerator in the above logarithm is the same as in standard source coding with side information (see, e.g., [CT06]) and the denominator is the same as in [LKK11], which we recover when the side-information Z is independent from X. The equivalence between D-achievability and E-D-achievability in the setting of Proposition 3.17 breaks down when common randomness is limited. Indeed, consider the case where the

[2][Wag22] is not yet peer-reviewed as of writing this thesis. Our proof tracks the one therein.

source is independent from the side information ($\eta = 0$). On the one hand, the latter is not useful when only available at the decoder, and the corresponding trade-off is given by Proposition 3.16. On the other hand, when the side information is available at both terminals, it acts as a common randomness with infinite rate, which corresponds to the scenario $R_c \to \infty$ in Proposition 3.16.

3.4 Proof of Theorem 3.3

3.4.1 Converse

We prove that $\overline{\mathcal{A}}^{(m)}_{\text{E-D}} \subset \overline{\mathcal{S}}^{(m)}_{\text{E-D}}$ by proving that $\mathcal{A}^{(m)}_{\text{E-D}} \subset \overline{\mathcal{S}}^{(m)}_{\text{E-D}}$. This self-contained proof builds on the approach of [YGA15], where the same Markov chain as in (3.4) is studied — and the components of the present proof are similar to those of the converse proofs in Chapter 2. Let (R, R_c, Δ) be E-D-achievable with near-perfect marginal realism. Then, by Theorem 3.2, it is E-D-achievable with perfect marginal realism. Fix $\varepsilon > 0$. Then, there exists an (n, R, R_c) E-D-code inducing a joint distribution P such that $\mathbb{E}_P[d(X^n, Y^n)] \leq \Delta + \varepsilon$ and $P_{Y^n} \equiv p_X^{\otimes n}$. Let T be a uniform random variable over $[n]$, independent of all other variables. Define $V = (M, J, Z_{[n]\setminus T}, T)$, which has a finite alphabet. Since $(X^n, Z^n) \sim p_{X,Z}^{\otimes n}$, the distribution of (X_T, Z_T) is $p_{X,Z}$. Similarly, we have $P_{Y_T} \equiv p_X$. Distribution P_{X_T,Z_T,V,Y_T} satisfies the Markov chain in $\mathcal{D}^{(m)}_{\text{E-D}}$ (Eq. 3.4) by Lemma 3.39 (Appendix 3.C.4). Thus, $P_{X_T,Y_T,Z_T,V} \in \mathcal{D}^{(m)}_{\text{E-D}}$. It remains to check that the inequalities in the definition of $\mathcal{S}^{(m)}_{\text{E-D}}$ are satisfied. Since $P_{X_T,Y_T} \equiv \hat{P}_{X^n,Y^n}$, we have $\mathbb{E}[d(X_T, Y_T)] = \mathbb{E}[d(X^n, Y^n)] \leq \Delta + \varepsilon$. We first lower bound the rates using quantized variables before taking a limit. Consider any (finite-valued) quantizer on $\mathcal{X}$. The image of a letter x by the latter is denoted $[x]$. We have

$$\begin{aligned} nR \geq H(M) \geq I(M; [X]^n | Z^n, J) &= I(M, J; [X]^n | Z^n) \\ &= \textstyle\sum_{t=1}^n I(M, J; [X]_t | Z^n, [X]_{t+1:n}) \\ &= \textstyle\sum_{t=1}^n I(M, J, [X]_{t+1:n}, Z_{[n]\setminus t}; [X]_t | Z_t) \\ &\geq \textstyle\sum_{t=1}^n I(M, J, Z_{[n]\setminus t}; [X]_t | Z_t) \end{aligned} \tag{3.21}$$

$$= nI(M, J, Z_{[n]\setminus T}; [X]_T | Z_T, T) \tag{3.22}$$

$$= nI(V; [X]_T | Z_T), \tag{3.23}$$

where (3.21) follows from the independence between the common randomness and the sources; and (3.22) and (3.23) follow from the independence of T and all other variables and from the fact that couples in $\{([X]_t, Z_t)\}_{t\in[n]}$ are i.i.d. Moreover, we have

$$\begin{aligned} n[R + R_c + H(p_Z)] \geq I(M, J, Z^n; [Y]^n) &= \sum_{t=1}^{n} I(M, J, Z^n; [Y]_t | [Y]_{1:t-1}) \\ &= \sum_{t=1}^{n} I(M, J, Z^n, [Y]_{1:t-1}; [Y]_t) \qquad (3.24) \\ &\geq \sum_{t=1}^{n} I(M, J, Z^n; [Y]_t) \\ &= nI(M, J, Z_{[n]\setminus T}, Z_T; [Y]_T | T) \\ &= nI(V, Z_T; [Y]_T), \qquad (3.25) \end{aligned}$$

where (3.24) and (3.25) follow from the fact that variables in $\{[Y]_t\}_{t\in[n]}$ are i.i.d. This is true for any quantizer on $\mathcal{X}$. Moreover, by Lemma 1.12 (Appendix 1) we have $I(Z_T; V) \leq H(Z_T) < \infty$. Then, by Proposition 1.16 (Appendix 1) we have

$$R \geq I(X_T; V | Z_T) \text{ and } R + R_c + H(Z_T) \geq I(Y_T; V, Z_T).$$

By Lemma 1.12 we have that $I(Z_T; Y_T) \leq H(Z_T) < \infty$, and that consequently $H(Z_T|Y_T)$ is well-defined and equal to $H(Z_T) - I(Z_T; Y_T)$. Hence, by the chain rule for general alphabets (Proposition 1.14) we have

$$I(Y_T; V, Z_T) - H(Z_T) = I(Y_T; V | Z_T) - H(Z_T | Y_T). \tag{3.26}$$

Therefore, $(R, R_c, \Delta + \varepsilon) \in \mathcal{S}_{E\text{-}D}^{(m)}$. Thus, $(R, R_c, \Delta) \in \overline{\mathcal{S}}_{E\text{-}D}^{(m)}$, which concludes the converse proof.

3.4.2 Informal outline of the achievability proof

We first construct a distribution $Q^{(1)}$ (Subsection 3.4.4), analyzed using the soft covering lemma with side information [Cuf13, Corollary VII.5]. The codebook used consists of one sub-codebook per possible realization z^n in $\mathcal{Z}^n$ (Subsection 3.4.3). Similarly to the approach of [Cuf13, Section V] — for the channel synthesis problem (see Chapter 5) without side information —, we modify $Q^{(1)}$ to obtain a distribution $P^{(1)}$ corresponding to a coding scheme (Subsection 3.4.5). We show that the distance between these two distributions is negligible (Subsection 3.4.6). While most results in [Cuf13] are only proved for finite alphabets, we handle general alphabets, following [Wag22]. In particular, in order to obtain a final distortion bound (Subsection 3.4.7), we use the uniform integrability in conjunction with the equivalence between achievability with near-perfect and perfect realism — similarly to Sections 2.D.3 and 2.D.4 from Chapter 2. The latter argument is needed if the source alphabet is not finite.

3.4.3 Random codebook

Here, we prove that $\overline{\mathcal{S}}^{(m)}_{E\text{-}D} \subseteq \overline{\mathcal{A}}^{(m)}_{E\text{-}D}$ by proving that $\mathcal{S}^{(m)}_{E\text{-}D} \subseteq \overline{\mathcal{A}}^{(m)}_{E\text{-}D}$. Let (R, R_c, Δ) be a triplet in $\mathcal{S}^{(m)}_{E\text{-}D}$. Let $p_{X,Y,Z,V}$ be a corresponding distribution from the definition of $\mathcal{S}^{(m)}_{E\text{-}D}$. Then,

$$R \geq I_p(X;V|Z) \tag{3.27}$$

$$R + R_c \geq I_p(Y;V|Z) - H_p(Z|Y) \tag{3.28}$$

$$\Delta \geq \mathbb{E}_p[d(X,Y)]. \tag{3.29}$$

We start by rewriting these expressions. By Lemma 1.12 (Appendix 1) we have that $I_p(Z;Y) \leq H_p(Z) < \infty$, and that consequently $H_p(Z|Y)$ is well-defined and equal to $H_p(Z) - I_p(Z;Y)$. Hence, by the chain rule for general alphabets (Proposition 1.14) we have

$$I_p(Y;V|Z) - H_p(Z|Y) = I_p(Y;V,Z) - H_p(Z). \tag{3.30}$$

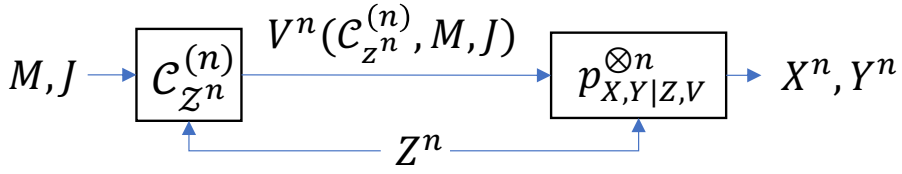


Figure 3.2: Graphical model for the distribution $Q^{(1)}$ appearing in the proof of Theorem 3.3 (E-D-achievable tuples with marginal realism). This is an instance of the graphical model involved in the soft covering lemma with side information [Cuf13, Lemma VII.5].

Moreover, the right hand side of (3.27) can be rewritten. For any quantized version $[X]$ of X,

$$I_p([X]; V|Z) = I_p([X], Z; V, Z) - H_p(Z). \tag{3.31}$$

Since Z is finite-valued, $I(Z;V) < \infty$. Thus, from (3.31) and Proposition 1.16 (Appendix 1),

$$I_p(X; V|Z) = I_p(X, Z; V, Z) - H_p(Z).^3 \tag{3.32}$$

Fix some $\varepsilon > 0$. Throughout, a *codebook* is a finite collection of codewords, i.e. of sequences in $\mathcal{V}^n$, indexed by couples (m, j). We use one codebook per side-information sequence: for every $n \geq 1$, let $(\mathcal{C}_{z^n}^{(n)})_{z^n \in \mathcal{Z}^n}$ be a random family of codebooks, drawn mutually independently across the indices $z^n \in \mathcal{Z}^n$, with the codebook of index z^n having $\lfloor 2^{n(R+\varepsilon)} \rfloor \times \lfloor 2^{nR_c} \rfloor$ i.i.d. codewords sampled from $\prod_{t=1}^n p_{V|Z=z_t}$. The codewords are indexed by couples (m, j). We also denote this random family by $\mathcal{C}_{\mathcal{Z}^n}^{(n)}$, and its distribution by $\mathbb{Q}_{\mathcal{C}_{\mathcal{Z}^n}^{(n)}}$.

3.4.4 Distribution $Q^{(1)}$ and soft-covering lemma

For every positive integer n, we define the distribution $Q^{(1)}$, described in Figure 3.2, as follows:

$$\begin{aligned} Q^{(1)}_{\mathcal{C}_{\mathcal{Z}^n}^{(n)}, M, J, Z^n, V^n, X^n, Y^n} := \; & \mathbb{Q}_{\mathcal{C}_{\mathcal{Z}^n}^{(n)}} \cdot p_{[2^{n(R+\varepsilon)}]}^{\mathcal{U}} \cdot p_{[2^{nR_c}]}^{\mathcal{U}} \cdot p_Z^{\otimes n} \cdot \mathbf{1}_{V^n = v^n(c_{z^n}^{(n)}, m, j)} \\ & \cdot \prod_{t=1}^n p_{X|Z=z_t, V=v_t} \cdot \prod_{t=1}^n p_{Y|Z=z_t, V=v_t}. \end{aligned} \tag{3.33}$$

[3] The identity (3.31) is not a mere application of the chain rule. The alphabet $\mathcal{X}$ may be infinite, and the left-hand side is a *conditional* mutual information; expressing it as a difference of two quantities is legitimate only once these are known to be finite, since otherwise the right-hand side is of the indeterminate form $\infty - \infty$.

Hereafter, when referring to a conditional probability kernel of $Q^{(1)}$, it will be the corresponding one appearing in (3.33) unless stated otherwise. By the definition of $\mathbb{Q}_{\mathcal{C}_{\mathcal{Z}^n}^{(n)}}$, we have $Q^{(1)}_{Z^n,V^n} \equiv p_{Z,V}^{\otimes n}$, and therefore, from (3.33) and the Markov chain in (3.4) we have

$$Q^{(1)}_{X^n,Y^n,Z^n,V^n} \equiv p_{X,Y,Z,V}^{\otimes n}. \tag{3.34}$$

Hence, from (3.29) and the additivity of d, we have

$$\mathbb{E}_{Q^{(1)}}[d(X^n,Y^n)] \leq \Delta. \tag{3.35}$$

Marginals of Y^n and (X^n,Z^n) knowing a codebook

We use the *soft covering lemma* with side information [Cuf13, Corollary VII.5] in its memoryless case. A nearly direct application of these two lemmas gives — see Appendix 3.C.2 for details —, from (3.27), (3.28), (3.30) and (3.32):

Claim 3.18

$$\mathbb{E}_{\mathcal{C}_{\mathcal{Z}^n}^{(n)}}\big[\|Q^{(1)}_{Y^n|\mathcal{C}_{\mathcal{Z}^n}^{(n)}} - p_X^{\otimes n}\|_{TV}\big] \underset{n\to\infty}{\longrightarrow} 0 \tag{3.36}$$

$$\mathbb{E}_{\mathcal{C}_{\mathcal{Z}^n}^{(n)}}\big[\|Q^{(1)}_{J,X^n,Z^n|\mathcal{C}_{\mathcal{Z}^n}^{(n)}} - p^{\mathcal{U}}_{[2^{nR_c}]}p_{X,Z}^{\otimes n}\|_{TV}\big] \underset{n\to\infty}{\longrightarrow} 0. \tag{3.37}$$

This is where the rate condition enters the proof: the soft covering lemma with side information (Appendix 3.C.2) applies to this codebook family precisely under the hypothesis $R + R_c + H_p(Z) \geq I_p(Y;Z,V)$, which is the second inequality defining $\mathcal{S}_{E\text{-}D}^{(m)}$.

Choosing a codebook

From (3.35), Lemma 1.7 (Appendix 1), and the Markov inequality we obtain

$$\mathbb{Q}_{\mathcal{C}_{\mathcal{Z}^n}^{(n)}}\Big(\mathbb{E}_{Q^{(1)}}[d(X^n,Y^n)|\mathcal{C}_{\mathcal{Z}^n}^{(n)}] \leq \Delta + \varepsilon\Big) \geq \varepsilon/(\Delta+\varepsilon). \tag{3.38}$$

In addition, we obtain convergence in probability from (3.36) and (3.37). Combining this with (3.38) gives that for a certain N_0, $\forall n \geq N_0$ there is a codebook $c^{(n)}_{\mathcal{Z}^n,*}$ such that

$$\mathbb{E}_{Q^{(1)}_{X^n,Y^n|\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}}}[d(X^n,Y^n)] \leq \Delta + \varepsilon, \tag{3.39}$$

$$\|Q^{(1)}_{Y^n|\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}} - p_X^{\otimes n}\|_{TV} \leq \varepsilon, \tag{3.40}$$

$$\|Q^{(1)}_{J,X^n,Z^n|\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}} - p^{\mathcal{U}}_{[2^{nR_c}]}p^{\otimes n}_{X,Z}\|_{TV} \leq \varepsilon. \tag{3.41}$$

3.4.5 Constructing a code

In this subsection, we show how the distribution $Q^{(1)}_{|\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}}$ can be modified to lead to a code, with a negligible effect on the two distributions of interest: the marginal of Y^n and the average empirical distribution of (X^n, Y^n). The analysis of the expected distortion is carried out in Section 3.4.7, where we use the uniform integrability in conjunction with the equivalence between near-perfect and perfect realism. By definition (3.33) of $Q^{(1)}_{|\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}}$, we have

$$\begin{aligned} Q^{(1)}_{J,\,Z^n,\,M,\,V^n,\,X^n,\,Y^n|\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}} &\equiv Q^{(1)}_{J,\,Z^n,\,M,\,V^n,\,X^n|\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}} \cdot \prod_{t=1}^{n} p_{Y|Z=z_t,V=v_t} \\ &\equiv Q^{(1)}_{J,\,X^n,\,Z^n|\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}} \cdot Q^{(1)}_{M,\,V^n|J=j,\,X^n=x^n,\,Z^n=z^n,\,\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}} \\ &\quad \cdot \prod_{t=1}^{n} p_{Y|Z=z_t,V=v_t}. \end{aligned} \tag{3.42}$$

We construct a distribution $P^{(1)}$ which differs from $Q^{(1)}_{|\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}}$ in having the correct marginal for (X^n, Z^n, J) as follows:

$$P^{(1)}_{J,X^n,Z^n,M,V^n,Y^n} := p^{\mathcal{U}}_{[2^{nR_c}]} \cdot p^{\otimes n}_{X,\,Z} \cdot Q^{(1)}_{M,\,V^n|J=j,\,X^n=x^n,\,Z^n=z^n,\,\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}} \cdot \prod_{t=1}^{n} p_{Y|Z=z_t,V=v_t}, \tag{3.43}$$

which from (3.42) can be written as

$$P^{(1)}_{J,X^n,Z^n,M,V^n,Y^n} \equiv p^{\mathcal{U}}_{[2^{nR_c}]} \cdot p^{\otimes n}_{X,\,Z} \cdot Q^{(1)}_{M,\,V^n,\,Y^n|J=j,\,X^n=x^n,\,Z^n=z^n,\,\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}}. \tag{3.44}$$

From (3.43), distribution $P^{(1)}$ satisfies

$$(X^n, M, J) - (Z^n, V^n) - Y^n \text{ , and hence } X^n - (Z^n, M, J, V^n) - Y^n. \tag{3.45}$$

Hence, by Lemma 3.1, $P^{(1)}$ defines an $(n, R+\varepsilon, R_c)$ E-D-code.

3.4.6 Distance between $P^{(1)}$ and $Q^{(1)}$

From (3.44) and Lemma 1.9 (Appendix 1), comparing $P^{(1)}$ with $Q^{(1)}$ reduces to comparing marginals, i.e., to (3.41):

$$\left\| P^{(1)}_{M,J,V^n,X^n,Z^n,Y^n} - Q^{(1)}_{M,J,V^n,X^n,Z^n,Y^n|\mathcal{C}^{(n)}_{Z^n}=c^{(n)}_{Z^n,*}} \right\|_{TV} = \left\| P^{(1)}_{J,X^n,Z^n} - Q^{(1)}_{J,X^n,Z^n|\mathcal{C}^{(n)}_{Z^n}=c^{(n)}_{Z^n,*}} \right\|_{TV} \leq \varepsilon.$$

Therefore, by Lemma 1.8 (Appendix 1) with $W = (X^n, Y^n)$ and again $W = Y^n$, we obtain

$$\left\| P^{(1)}_{Y^n} - Q^{(1)}_{Y^n|\mathcal{C}^{(n)}_{Z^n}=c^{(n)}_{Z^n,*}} \right\|_{TV} \leq \left\| P^{(1)}_{X^n,Y^n} - Q^{(1)}_{X^n,Y^n|\mathcal{C}^{(n)}_{Z^n}=c^{(n)}_{Z^n,*}} \right\|_{TV} \leq \varepsilon. \tag{3.46}$$

By the triangle inequality and Lemma 1.8 (Appendix 1) with $W = (X_t, Y_t)$ for $t \in [n]$, we obtain (recall the empirical-distribution operators $\hat{P}$ and $\hat{Q}$ defined Section 1.5)

$$\left\| \hat{P}^{(1)}_{X^n,Y^n} - \hat{Q}^{(1)}_{X^n,Y^n|\mathcal{C}^{(n)}_{Z^n}=c^{(n)}_{Z^n,*}} \right\|_{TV} \leq \left\| P^{(1)}_{X^n,Y^n} - Q^{(1)}_{X^n,Y^n|\mathcal{C}^{(n)}_{Z^n}=c^{(n)}_{Z^n,*}} \right\|_{TV} \leq \varepsilon. \tag{3.47}$$

3.4.7 Conclusion

Since (d, p_X) is uniformly integrable, then from (3.39), (3.40), (3.46) and (3.47), for every $n \geq N_0$, we can use Proposition 3.30 with $Q = Q^{(1)}_{|\mathcal{C}^{(n)}_{Z^n}=c^{(n)}_{Z^n,*}}$, $P = P^{(1)}$, $\varepsilon_1 = \varepsilon_2 = \varepsilon_3 = \varepsilon_4 = \varepsilon$. Then, there is a sequence of $(n, R+\varepsilon, R_c)$ E-D-codes with the same encoder as $P^{(1)}$, and inducing a distribution $P^{(2)}$ such that for large enough n we have

$$\mathbb{E}_{P^{(2)}}[d(X^n, Y^n)] \leq \Delta + \varepsilon + \sup_{\mathbb{P}_{X,Y,\xi}} \mathbb{E}[d(X,Y)\cdot\xi] \text{ and } P^{(2)}_{Y^n} \equiv p_X^{\otimes n}, \tag{3.48}$$

where the supremum is taken over all distributions $\mathbb{P}_{X,Y,\xi}$ on $\mathcal{X}^2 \times \{0,1\}$ satisfying $\mathbb{P}_X \equiv \mathbb{P}_Y \equiv p_X$ and $\mathbb{P}(\{\xi = 1\}) \leq 3\varepsilon$. This proves the E-D-achievability of $(R+\varepsilon, R_c, \Delta+\varepsilon+s(\varepsilon))$, where $s(\varepsilon)$ is the supremum in (3.48) — which does not depend on n. This is true for every $\varepsilon > 0$. By the uniform integrability assumption, we have $s(\varepsilon) \to 0$ if $\varepsilon \to 0$. Therefore, we obtain $(R, R_c, \Delta) \in \overline{\mathcal{A}}^{(m)}_{E\text{-}D}$, as desired.

3.5 Proof of Theorem 3.9

3.5.1 Informal outline

Similarly to the approach of [SCP16, Section IV], we introduce a virtual message M' with rate R' generated and used by the encoder, but not transmitted (Subsection 3.5.2). The decoder will then guess it. The overall proof has a structure that is similar to Section 3.4. We start with a distribution $Q^{(1)}$ (Subsection 3.5.3), analyzed using the basic soft covering lemma [Cuf13, Lemma VII.4] — a different lemma compared to Section 3.4 —, then change it little by little to obtain an intermediate distribution $Q^{(2)}$ (Subsection 3.5.4) and a distribution $P^{(1)}$ corresponding to a coding scheme (Subsection 3.5.4). Following [Wag22], in order to obtain a final distortion bound, we use uniform integrability in conjunction with the equivalence between achievability with near-perfect and perfect realism.

3.5.2 Random codebook indexed by a virtual message

Here, we prove that $\overline{\mathcal{S}}^{(m)}_D \subseteq \overline{\mathcal{A}}^{(m)}_D$ by proving that $\mathcal{S}^{(m)}_D \subseteq \overline{\mathcal{A}}^{(m)}_D$. Let (R, R_c, Δ) be a triplet in $\mathcal{S}^{(m)}_D$. Let $p_{X,Y,Z,V}$ be a corresponding distribution from the definition of $\mathcal{S}^{(m)}_D$. Then, $I_p(Z;V) < \infty$, and

$$R \geq I_p(X;V) - I_p(Z;V) \tag{3.49}$$

$$R + R_c \geq I_p(Y;V) - I_p(Z;V) \tag{3.50}$$

$$\Delta \geq \mathbb{E}_p[d(X,Y)]. \tag{3.51}$$

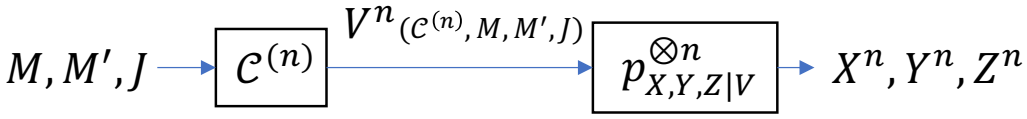


Figure 3.3: Graphical model for the distribution $Q^{(1)}$ appearing in the proof of Theorem 3.9 (D-achievable tuples with marginal realism). This is an instance of the graphical model involved in the soft covering lemma [Cuf13, Lemma VII.4].

Fix some $\varepsilon > 0$. We introduce a rate R' corresponding to a virtual message M', as follows. If $I_p(Z;V) = 0$ then set $R' = 0$. Otherwise, fix R' in $\big(I_p(Z;V) - \varepsilon, I_p(Z;V)\big) \cap \big(0, +\infty\big)$. Then, from (3.49), (3.50), the Markov chain property $Z - X - V$, and the chain rule for general alphabets (Proposition 1.14, Appendix 1) we have:

$$R + R' + \varepsilon > I_p(X;V) = I_p(X,Z;V), \tag{3.52}$$

$$R + R' + R_c + \varepsilon > I_p(Y;V), \tag{3.53}$$

where the equality in (3.52) follows from the Markov chain $Z - X - V$ of (3.13) and the chain rule for general alphabets (Proposition 1.14). For every $n \geq 1$, let $\mathcal{C}^{(n)}$ be a random codebook with $\lfloor 2^{n(R+\varepsilon)} \rfloor \times \lfloor 2^{nR'} \rfloor \times \lfloor 2^{nR_c} \rfloor$ i.i.d. codewords sampled from $p_V^{\otimes n}$. The codewords are indexed by triples (m, m', j). We denote this random codebook distribution by $\mathbb{Q}_{\mathcal{C}^{(n)}}$.

3.5.3 Distribution $Q^{(1)}$ and soft-covering lemma

For every positive integer n, we define the distribution $Q^{(1)}$, described in Figure 3.3, as follows:

$$\begin{aligned} Q^{(1)}_{\mathcal{C}^{(n)},M,M',J,Z^n,V^n,X^n,Y^n} := \; & \mathbb{Q}_{\mathcal{C}^{(n)}} \cdot p^{\mathcal{U}}_{[2^{n(R+\varepsilon)}]} \cdot p^{\mathcal{U}}_{[2^{nR'}]} \cdot p^{\mathcal{U}}_{[2^{nR_c}]} \cdot \mathbf{1}_{V^n = v^n(\mathcal{C}^{(n)},m,m',j)} \\ & \cdot \prod_{t=1}^{n} p_{X|V=v_t} \cdot \prod_{t=1}^{n} p_{Z|X=x_t} \cdot \prod_{t=1}^{n} p_{Y|Z=z_t,V=v_t}. \end{aligned} \tag{3.54}$$

Hereafter, when referring to a conditional probability kernel of $Q^{(1)}$, it will be the corresponding one appearing in (3.54) unless stated otherwise. By the definition of $\mathbb{Q}_{\mathcal{C}^{(n)}}$, we have

$Q^{(1)}_{V^n} \equiv p_V^{\otimes n}$, and therefore, from (3.54) and the Markov chains in (3.13), we have

$$Q^{(1)}_{V^n,X^n,Z^n,Y^n} \equiv p^{\otimes n}_{V,X,Z,Y}. \tag{3.55}$$

Hence, from (3.51) and the additivity of d, we have

$$\mathbb{E}_{Q^{(1)}}[d(X^n, Y^n)] \leq \Delta. \tag{3.56}$$

Marginals of Y^n and (X^n, Z^n) knowing a codebook

We use the *soft covering lemma* for general alphabets [Cuf13, Corollary VII.4] in its memoryless case. A nearly direct application gives — see Appendix 3.C.1 for details —, from (3.52) and (3.53):

Claim 3.19

$$\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|Q^{(1)}_{Y^n|\mathcal{C}^{(n)}} - p_X^{\otimes n}\|_{TV}\right] \underset{n\to\infty}{\longrightarrow} 0 \tag{3.57}$$

$$\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|Q^{(1)}_{J,X^n,Z^n|\mathcal{C}^{(n)}} - p^{\mathcal{U}}_{[2^{nR_c}]}p^{\otimes n}_{X,Z}\|_{TV}\right] \underset{n\to\infty}{\longrightarrow} 0. \tag{3.58}$$

Decoding of M'

The virtual message M' is to be decoded as $\hat{M}'$, conditionally independent from M', X^n, Y^n knowing $M, J, Z^n, \mathcal{C}^{(n)}$: in the case of finite alphabets,

$$Q^{(1)}_{\hat{M}'|\mathcal{C}^{(n)}=c^{(n)},\, M=m,\, J=j,\, Z^n=z^n}$$

is a joint $p_{V,Z}$-typicality decoder for sub-codebook $(v^n(c^{(n)}, m, a, j))_a$ (see, e.g., [CT06, page 201]). By the standard random coding argument for channel coding (see, e.g., [CT06, Section 7.7]) and using the fact that $R' < I_p(Z;V)$ (by construction), we can obtain

$$Q^{(1)}_{\hat{M}',M'}\left(\hat{M}' \neq M'\right) \underset{n\to\infty}{\longrightarrow} 0 \tag{3.59}$$

in the case of finite alphabets.

Claim 3.20 *For general sources, there exists a decoder satisfying* (3.59).

See Appendix 3.A for the details and proof. The probability in (3.59) can be rewritten as an expectation over $\mathcal{C}^{(n)}$. Since the convergence towards zero in expectation implies a convergence in probability for a non-negative variable, we obtain:

$$Q^{(1)}\big(M' \neq \hat{M}' \big|\, \mathcal{C}^{(n)}\big) \xrightarrow[n\to\infty]{\mathcal{P}} 0. \tag{3.60}$$

Choosing a codebook

From (3.56) and the Markov inequality we obtain

$$\mathbb{Q}_{\mathcal{C}^{(n)}}\Big(\mathbb{E}_{Q^{(1)}}[d(X^n,Y^n)|\mathcal{C}^{(n)}] \leq \Delta+\varepsilon\Big) \geq \varepsilon/(\Delta+\varepsilon). \tag{3.61}$$

In addition, similarly to (3.60), we obtain convergence in probability from (3.57) and (3.58). Combining this with (3.61) gives that for a certain N_0, $\forall n \geq N_0$ there is a codebook $c_*^{(n)}$ such that

$$\begin{aligned}
\mathbb{E}_{Q^{(1)}_{X^n,Y^n|\mathcal{C}^{(n)}=c_*^{(n)}}}[d(X^n,Y^n)] &\leq \Delta+\varepsilon,\\
\|Q^{(1)}_{Y^n|\mathcal{C}^{(n)}=c_*^{(n)}} - p_X^{\otimes n}\|_{TV} &\leq \varepsilon,\\
\|Q^{(1)}_{J,X^n,Z^n|\mathcal{C}^{(n)}=c_*^{(n)}} - p^{\mathcal{U}}_{[2^{nR_c}]}p_{X,Z}^{\otimes n}\|_{TV} &\leq \varepsilon, \qquad (3.62)\\
Q^{(1)}_{M',\hat{M}'|\mathcal{C}^{(n)}=c_*^{(n)}}(\hat{M}' \neq M') &\leq \varepsilon. \qquad (3.63)
\end{aligned}$$

3.5.4 Construction of a code

In this subsection, we show how the distribution $Q^{(1)}_{|\mathcal{C}^{(n)}=c_*^{(n)}}$ can be modified to lead to a code, with a negligible effect on the error for decoding M' and on the two distributions of interest: the marginal of Y^n and the average empirical distribution of (X^n,Y^n). The analysis of the

expected distortion is carried out in Section 3.4.7, where we use the uniform integrability in conjunction with the equivalence between near-perfect and perfect realism.

Effect of using the decoded virtual message $\hat{M}'$

For every positive integer n we define the following distribution which differs from $Q^{(1)}_{|\mathcal{C}^{(n)}=c_*^{(n)}}$ in that Y^n is sampled using $\hat{M}'$ instead of M' :

$$Q^{(2)}_{M,M',J,V^n,X^n,Z^n,\hat{M}',Y^n} := p^{\mathcal{U}}_{[2^{n(R+\varepsilon)}]} \cdot p^{\mathcal{U}}_{[2^{nR'}]} \cdot p^{\mathcal{U}}_{[2^{nR_c}]} \cdot \mathbf{1}_{V^n=v^n(c_*^{(n)},m,m',j)} \cdot \prod_{t=1}^{n} p_{X|V=v_t} \cdot \prod_{t=1}^{n} p_{Z|X=x_t} \cdot Q^{(1)}_{\hat{M}'|M=m,\,J=j,\,Z^n=z^n,\mathcal{C}^{(n)}=c_*^{(n)}} \cdot \prod_{t=1}^{n} p_{Y|Z=z_t,V=v_t(c_*^{(n)},m,\hat{m}',j)} \tag{3.64}$$

By definition (3.54) of $Q^{(1)}_{|\mathcal{C}^{(n)}}$ this can be rewritten as

$$Q^{(2)}_{M,M',J,V^n,X^n,Z^n,\hat{M}',Y^n} \equiv Q^{(1)}_{M,\,M',\,J,\,V^n,\,X^n,\,Z^n|\mathcal{C}^{(n)}=c_*^{(n)}} \cdot Q^{(1)}_{\hat{M}'|M=m,\,J=j,\,Z^n=z^n,\mathcal{C}^{(n)}=c_*^{(n)}} \cdot \prod_{t=1}^{n} p_{Y|Z=z_t,V=v_t(c_*^{(n)},m,\hat{m}',j)}, \tag{3.65}$$

which can be further rewritten — e.g. by Corollary 1.3 — as

$$Q^{(2)}_{M,M',J,Z^n,V^n,X^n,Y^n,\hat{M}'} \equiv Q^{(1)}_{M,M',J,Z^n,V^n,X^n,\hat{M}'|\mathcal{C}^{(n)}=c_*^{(n)}} \cdot \prod_{t=1}^{n} p_{Y|Z=z_t,V=v_t(c_*^{(n)},m,\hat{m}',j)}. \tag{3.66}$$

The following lemma appears in [SCP16] with finite alphabets, and also holds for general ones: it is a direct consequence of the standard relation between coupling and the TVD (see, e.g., [Lin92, Chapter I, Theorem 5.2]).

Lemma 3.21 *Let P_{UWL} be a distribution on an alphabet of the form $\mathcal{U}\times\mathcal{U}\times\mathcal{L}$ and let $\eta \in (0,1)$. If $P(U \neq W) \leq \eta$ we have $\|P_{UL} - P_{WL}\|_{TV} \leq \eta$.*

Using Lemma 3.21 on $Q^{(1)}$ with $U = M', W = \hat{M}'$, $L = (M, J, Z^n, X^n)$ we have from (3.63)

$$\left\|Q^{(1)}_{M',M,J,Z^n,X^n|\mathcal{C}^{(n)}=c_*^{(n)}} - Q^{(1)}_{\hat{M}',M,J,Z^n,X^n|\mathcal{C}^{(n)}=c_*^{(n)}}\right\|_{TV} \leq \varepsilon.$$

As a consequence, and by construction of $Q^{(2)}$ and its similarity to that of $Q^{(1)}$, we have by Lemma 1.9 (Appendix 1) with $L = Y^n$:

$$\left\| Q^{(1)}_{M',\,M,\,J,\,Z^n,\,X^n,\,Y^n|\mathcal{C}^{(n)}=c_*^{(n)}} - Q^{(2)}_{\hat{M}',\,M,\,J,\,Z^n,\,X^n,\,Y^n} \right\|_{TV} \leq \varepsilon.$$

Hence, by Lemma 1.8 (Appendix 1) with $W = (X^n, Y^n)$ we have

$$\left\| Q^{(1)}_{X^n,\,Y^n|\mathcal{C}^{(n)}=c_*^{(n)}} - Q^{(2)}_{X^n,\,Y^n} \right\|_{TV} \leq \varepsilon. \tag{3.67}$$

Since $Q^{(2)}_{J,X^n,Z^n} \equiv Q^{(1)}_{J,X^n,Z^n|\mathcal{C}^{(n)}=c_*^{(n)}}$ we also have from (3.62):

$$\|Q^{(2)}_{J,X^n,Z^n} - p^{\mathcal{U}}_{[2^{nR_c}]} p^{\otimes n}_{X,Z}\|_{TV} \leq \varepsilon. \tag{3.68}$$

Finalizing the code construction

From (3.64) distribution $Q^{(2)}$ satisfies

$$Z^n - X^n - (M, M', J, V^n), \text{ and hence } Z^n - (X^n, J) - M. \tag{3.69}$$

From (3.65), when considering variables other than M', V^n we obtain — e.g., Corollary 1.3:

$$Q^{(2)}_{J,\,X^n,\,Z^n,\,M,\,\hat{M}',\,Y^n} \equiv Q^{(2)}_{J,\,X^n,\,Z^n,\,M} \cdot Q^{(2)}_{\hat{M}',\,Y^n|M=m,\,J=j,\,Z^n=z^n}.$$

From Markov chain (3.69), this can be rewritten as

$$Q^{(2)}_{J,\,X^n,\,Z^n,\,M,\,\hat{M}',\,Y^n} \equiv Q^{(2)}_{J,\,X^n,\,Z^n} \cdot Q^{(2)}_{M|J=j,\,X^n=x^n} \cdot Q^{(2)}_{\hat{M}',\,Y^n|M=m,\,J=j,\,Z^n=z^n}. \tag{3.70}$$

We define the distribution $P^{(1)}$ achieving near-perfect marginal realism. It differs from $Q^{(2)}$ in having the correct marginal for (X^n, Z^n, J): fix a transition kernel of $Q^{(2)}$ for M knowing

(X^n, J), and let

$$P^{(1)}_{J,X^n,Z^n,M,\hat{M}',Y^n} := p^{\mathcal{U}}_{[2^{nR_c}]} \cdot p^{\otimes n}_{X,Z} \cdot Q^{(2)}_{M|J=j,\,X^n=x^n} \cdot Q^{(2)}_{\hat{M}',\,Y^n|M=m,\,J=j,\,Z^n=z^n}, \tag{3.71}$$

which from (3.70) can be written as

$$P^{(1)}_{J,X^n,Z^n,M,\hat{M}',Y^n} = p^{\mathcal{U}}_{[2^{nR_c}]} \cdot p^{\otimes n}_{X,Z} \cdot Q^{(2)}_{M,\,\hat{M}',\,Y^n|J=j,\,X^n=x^n,\,Z^n=z^n}. \tag{3.72}$$

From (3.71) distribution $P^{(1)}$ satisfies

$$Z^n - (X^n, J) - M \quad \text{and} \quad X^n - (M, J, Z^n) - (\hat{M}', Y^n). \tag{3.73}$$

The alphabet of M is $[2^{n(R+\varepsilon)}]$ throughout this proof by construction of the random codebook $\mathcal{C}^{(n)}$. Therefore, from (3.72)-(3.73) and Lemma 3.1, distribution $P^{(1)}$ defines an $(n, R+\varepsilon, R_c)$ D-code. From (3.72) and Lemma 1.9 (Appendix 1), comparing $P^{(1)}$ with $Q^{(2)}$ reduces to comparing marginals, i.e., to (3.68):

$$\big\|P^{(1)}_{M,\,M',\,J,\,X^n,\,Z^n,\,\hat{M}',\,Y^n} - Q^{(2)}_{M,\,M',\,J,\,X^n,\,Z^n,\,\hat{M}',\,Y^n}\big\|_{TV} = \big\|P^{(1)}_{J,X^n,Z^n} - Q^{(2)}_{J,X^n,Z^n}\big\|_{TV} \leq \varepsilon. \tag{3.74}$$

Therefore, by Lemma 1.8 (Appendix 1) with $W = Y^n$ and $W = (X^n, Y^n)$, the triangle inequality and (3.67), we obtain

$$\big\|P^{(1)}_{Y^n} - Q^{(1)}_{Y^n|\mathcal{C}^{(n)}=c^{(n)}_*}\big\|_{TV} \leq \big\|P^{(1)}_{X^n,Y^n} - Q^{(1)}_{X^n,Y^n|\mathcal{C}^{(n)}=c^{(n)}_*}\big\|_{TV} \leq 2\varepsilon. \tag{3.75}$$

By the triangle inequality and Lemma 1.8 (Appendix 1) with $W = (X_t, Y_t)$ for $t \in [n]$, we obtain

$$\big\|\hat{P}^{(1)}_{X^n,Y^n} - \hat{Q}^{(1)}_{X^n,Y^n|\mathcal{C}^{(n)}=c^{(n)}_*}\big\|_{TV} \leq \big\|P^{(1)}_{X^n,Y^n} - Q^{(1)}_{X^n,Y^n|\mathcal{C}^{(n)}=c^{(n)}_*}\big\|_{TV} \leq 2\varepsilon. \tag{3.76}$$

The end of the proof is identical to Subsection 3.4.7 and is omitted for the purpose of conciseness.

3.6 Proof of Theorem 3.13

3.6.1 Converse

We prove that $\overline{\mathcal{A}}^{(j)}_{E\text{-}D} \subset \overline{\mathcal{S}}^{(j)}_{E\text{-}D}$ by proving that $\mathcal{A}^{(j)}_{E\text{-}D} \subset \overline{\mathcal{S}}^{(j)}_{E\text{-}D}$. We proceed as in the case of marginal realism (Section 3.4.1). Let (R, R_c, Δ) be E-D-achievable with near-perfect joint realism. Then, by Theorem 3.2, it is E-D-achievable with perfect joint realism. Fix $\varepsilon > 0$. Then, there exists an (n, R, R_c) E-D-code inducing a joint distribution P such that $\mathbb{E}_P[d(X^n, Y^n)] \leq \Delta + \varepsilon$ and $P_{Y^n,Z^n} \equiv p_{X,Z}^{\otimes n}$. Let T be a uniform random variable over $[n]$. Define V as the tuple $(M, J, Z_{[n]\setminus T}, T)$, which has a finite alphabet. Since $(X^n, Z^n) \sim p_{X,Z}^{\otimes n}$, the distribution of (X_T, Z_T) is $p_{X,Z}$. Similarly, we have $P_{Y_T,Z_T} \equiv p_{X,Z}$. Distribution P_{X_T,Z_T,V,Y_T} satisfies the Markov chain in $\mathcal{D}^{(j)}_{E\text{-}D}$ (Eq. 3.16) by Lemma 3.39 (Appendix 3.C.4). Thus, $P_{X_T,Y_T,Z_T,V} \in \mathcal{D}^{(j)}_{E\text{-}D}$. It remains to check that the inequalities in the definition of $\mathcal{S}^{(j)}_{E\text{-}D}$ are satisfied. Since $P_{X_T,Y_T} \equiv \hat{P}_{X^n,Y^n}$, we have $\mathbb{E}[d(X_T, Y_T)] = \mathbb{E}[d(X^n, Y^n)] \leq \Delta + \varepsilon$. We first lower bound the rates using quantized variables before taking the limit. Consider any (finite-valued) quantizer on $\mathcal{X}$. The image of a letter x by the latter is denoted $[x]$. We know that independence is preserved by deterministic mappings. Hence,

$$
\begin{aligned}
nR \geq H(M) \geq I(M; [X]^n | Z^n, J) &= I(M, J; [X]^n | Z^n) && (3.77)\\
&= \textstyle\sum_{t=1}^n I(M, J; [X]_t | Z^n, [X]_{t+1:n})\\
&= \textstyle\sum_{t=1}^n I(M, J, [X]_{t+1:n}, Z_{[n]\setminus t}; [X]_t | Z_t)\\
&\geq \textstyle\sum_{t=1}^n I(M, J, Z_{[n]\setminus t}; [X]_t | Z_t)\\
&= nI(M, J, Z_{[n]\setminus T}; [X]_T | Z_T, T) && (3.78)\\
&= nI(V; [X]_T | Z_T), && (3.79)
\end{aligned}
$$

where (3.77) follows from the independence between the common randomness and the sources; and (3.78) and (3.79) follow from the independence of T and all other variables and from the

fact that couples in $\{([X]_t, Z_t)\}_{t\in[n]}$ are i.i.d. Moreover, we have

$$\begin{aligned} n(R+R_c) \geq H(M,J) \geq I(M,J;[Y]^n|Z^n) &= \sum_{t=1}^{n} I(M,J;[Y]_t|Z^n,[Y]_{1:t-1}) \\ &= \sum_{t=1}^{n} I(M,J,[Y]_{1:t-1};[Y]_t|Z^n) \qquad (3.80) \\ &\geq \sum_{t=1}^{n} I(M,J;[Y]_t|Z^n) \\ &= \sum_{t=1}^{n} I(M,J,Z_{[n]\setminus t};[Y]_t|Z_t) \qquad (3.81) \\ &= nI(M,J,Z_{[n]\setminus T};[Y]_T|Z_T,T) \\ &= nI(V;[Y]_T|Z_T), \qquad (3.82) \end{aligned}$$

where (3.80), (3.81) and (3.82) follow from the fact that couples in $\{([Y]_t, Z_t)\}_{t\in[n]}$ are i.i.d. and that T is independent from (Y^n, Z^n). This is true for any quantizer on $\mathcal{X}$. Moreover, since Z^n is finite-valued, we have $I(Z_T;V) < \infty$. Then, by Proposition 1.16 we have

$$R \geq I(X_T;V|Z_T) \quad \text{and} \quad R+R_c \geq I(Y_T;V|Z_T).$$

Therefore, $(R, R_c, \Delta+\varepsilon) \in \mathcal{S}^{(j)}_{E\text{-}D}$. Hence, $(R, R_c, \Delta) \in \overline{\mathcal{S}}^{(j)}_{E\text{-}D}$, which concludes the converse proof.

3.6.2 Achievability

We prove that $\overline{\mathcal{S}}^{(j)}_{E\text{-}D} \subseteq \overline{\mathcal{A}}^{(j)}_{E\text{-}D}$ by proving that $\mathcal{S}^{(j)}_{E\text{-}D} \subseteq \overline{\mathcal{A}}^{(j)}_{E\text{-}D}$. We proceed exactly as in Section 3.4. Let $(R, R_c, \Delta) \in \mathcal{S}^{(j)}_{E\text{-}D}$. Let $p_{X,Y,Z,V}$ be a corresponding distribution from the definition of $\mathcal{S}^{(j)}_{E\text{-}D}$. Then,

$$R \geq I_p(X;V|Z) \qquad (3.83)$$

$$R+R_c \geq I_p(Y;V|Z) \qquad (3.84)$$

$$\Delta \geq \mathbb{E}_p[d(X,Y)]. \qquad (3.85)$$

For every $n \geq 1$, we define codebook $(\mathcal{C}^{(n)}_{z^n})_{z^n \in \mathcal{Z}^n}$ and distribution $Q^{(1)}$ exactly as in Section 3.4. Similarly to (3.32) therein, Inequalities (3.83) and (3.84) rewrite as

$$R \geq I_p(X, Z; V, Z) - H_p(Z) \quad \text{and} \quad R + R_c \geq I_p(Y, Z; V, Z) - H_p(Z).$$

Therefore, in the same manner as in Section 3.4, the soft covering lemma with side information gives

$$\begin{aligned} \mathbb{E}_{\mathcal{C}^{(n)}_{\mathcal{Z}^n}}\big[\|Q^{(1)}_{Y^n,Z^n|\mathcal{C}^{(n)}_{\mathcal{Z}^n}} - p^{\otimes n}_{X,Z}\|_{TV}\big] &\underset{n\to\infty}{\longrightarrow} 0 \\ \mathbb{E}_{\mathcal{C}^{(n)}_{\mathcal{Z}^n}}\big[\|Q^{(1)}_{J,X^n,Z^n|\mathcal{C}^{(n)}_{\mathcal{Z}^n}} - p^{\mathcal{U}}_{[2^{nR_c}]}p^{\otimes n}_{X,Z}\|_{TV}\big] &\underset{n\to\infty}{\longrightarrow} 0. \end{aligned}$$

Similarly to (3.39), (3.40), (3.41), there exists N_0 such that for $n \geq N_0$ there is a codebook $c^{(n)}_{\mathcal{Z}^n,*}$ such that

$$\mathbb{E}_{Q^{(1)}_{X^n,Y^n|\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}}}[d(X^n, Y^n)] \leq \Delta + \varepsilon, \tag{3.86}$$

$$\|Q^{(1)}_{Y^n,Z^n|\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}} - p^{\otimes n}_{X,Z}\|_{TV} \leq \varepsilon, \tag{3.87}$$

$$\|Q^{(1)}_{J,X^n,Z^n|\mathcal{C}^{(n)}_{\mathcal{Z}^n}=c^{(n)}_{\mathcal{Z}^n,*}} - p^{\mathcal{U}}_{[2^{nR_c}]}p^{\otimes n}_{X,Z}\|_{TV} \leq \varepsilon. \tag{3.88}$$

We construct the E-D-code $P^{(1)}$ exactly as in Section 3.4, and by the same arguments that brought inequality (3.46), we obtain from (3.88):

$$\big\|P^{(1)}_{M,J,V^n,X^n,Z^n,Y^n} - \ Q^{(1)}_{M,J,V^n,X^n,Z^n,Y^n}\big\|_{TV} = \big\|P^{(1)}_{J,X^n,Z^n} - Q^{(1)}_{J,X^n,Z^n}\big\|_{TV} \leq \varepsilon. \tag{3.89}$$

This and (3.87) imply that $P^{(1)}$ achieves near-perfect joint realism. We conclude as in Section 3.4.7, but using Proposition 3.31 instead of Proposition 3.30, with (3.86), (3.87) and (3.89).

3.7 Proof of Theorem 3.14

3.7.1 Converse

The converse proof is identical to that of Theorem 3.13 (E-D-codes with (near)-perfect joint realism), where bounds on rates and distortion do not rely on any Markov chain properties, and only rely on independence properties, which are preserved by quantization. We only add the following arguments, related to the difference between $\mathcal{D}_{E\text{-}D}^{(j)}$ and $\mathcal{D}_{D}^{(j)}$ (Eq. 3.16 and 3.18) regarding the single-letter distributions. First, Markov chain $Z_T - X_T - V$ is satisfied by Corollary 3.38 (Appendix 3.C.4). Second, the quantization argument in bounds on rates is done with three quantizers, on $\mathcal{X}, \mathcal{Z}$, and $\mathcal{Z}^{n-1}$, rather than just on $\mathcal{X}$. Lastly, we have $I(Z_T; V) < \infty$. This is proved in the remainder of this converse argument. Fix $n \in \mathbb{N}$, and quantizers κ_1 and κ_2, on $\mathcal{Z}$ and $\mathcal{Z}^{n-1}$, respectively. Then, we have

$$\begin{aligned} I(M, J, \kappa_2(Z_{[n]\setminus T}), T; \kappa_1(Z_T)) &= I(M, J, \kappa_2(Z_{[n]\setminus T}); \kappa_1(Z_T)|T) \\ &= \tfrac{1}{n}\textstyle\sum_{t=1}^{n} I(M, J, \kappa_2(Z_{[n]\setminus t}); \kappa_1(Z_t)) \\ &= \tfrac{1}{n}\textstyle\sum_{t=1}^{n} I(M; \kappa_1(Z_t)|J, \kappa_2(Z_{[n]\setminus t})) \\ &\leq H(M), \end{aligned}$$

where the first two equalities use the independence of T from all other variables; the third uses the independence of J from $[Z]^n$ and the fact that variables in $\{[Z]_t\}_{t\in[n]}$ are i.i.d.; and the final inequality follows from known properties of finite alphabets. Indeed, independence is preserved by deterministic quantization. This equality is true for any quantizers on $\mathcal{Z}$ and $\mathcal{Z}^{n-1}$. Then, from Proposition 1.16,

$$I(V; Z_T) \leq H(M),$$

which is finite because M is finite-valued (Lemma 1.12).

3.7.2 Achievability

We prove that $\overline{\mathcal{S}}_D^{(j)} \subseteq \overline{\mathcal{A}}_D^{(j)}$ by proving that $\mathcal{S}_D^{(j)} \subseteq \overline{\mathcal{A}}_D^{(j)}$. We proceed exactly as in the marginal realism case (Section 3.5). Let $(R, R_c, \Delta) \in \mathcal{S}_D^{(j)}$ and p be a corresponding distribution from (3.17).

$$R \geq I_p(X;V|Z) \tag{3.90}$$

$$R + R_c \geq I_p(Y;V|Z) \tag{3.91}$$

$$\Delta \geq \mathbb{E}_p[d(X,Y)]. \tag{3.92}$$

By assumption, p satisfies the Markov chain $Z - X - V$ of (3.18) and $I_p(Z;V) < \infty$. Therefore, by the chain rule for general alphabets (Proposition 1.14), inequalities (3.90) and (3.91) rewrite as

$$R \geq I_p(X,Z;V) - I_p(Z;V) \quad \text{and} \quad R + R_c \geq I_p(Y,Z;V) - I_p(Z;V). \tag{3.93}$$

Hence, we can define rate R', codebook $\mathcal{C}^{(n)}$ and distribution $Q^{(1)}$ exactly as in Section 3.5 and obtain, from (3.93) and the soft covering lemma:

$$\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|Q^{(1)}_{J,X^n,Z^n|\mathcal{C}^{(n)}} - p^{\mathcal{U}}_{[2^{nR_c}]}p^{\otimes n}_{X,Z}\|_{TV}\right] \underset{n\to\infty}{\longrightarrow} 0 \tag{3.94}$$

$$\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|Q^{(1)}_{Y^n,Z^n|\mathcal{C}^{(n)}} - p^{\otimes n}_{X,Z}\|_{TV}\right] \underset{n\to\infty}{\longrightarrow} 0. \tag{3.95}$$

In the same manner as in Section 3.5, for a certain N_0, $\forall n \geq N_0$ there is a codebook $c_*^{(n)}$ such that

$$\mathbb{E}_{Q^{(1)}_{X^n,Y^n|\mathcal{C}^{(n)}=c_*^{(n)}}}[d(X^n,Y^n)] \leq \Delta + \varepsilon, \tag{3.96}$$

$$\|Q^{(1)}_{Y^n,Z^n|\mathcal{C}^{(n)}=c_*^{(n)}} - p^{\otimes n}_{X,Z}\|_{TV} \leq \varepsilon. \tag{3.97}$$

$$\|Q^{(1)}_{J,X^n,Z^n|\mathcal{C}^{(n)}=c_*^{(n)}} - p^{\mathcal{U}}_{[2^{nR_c}]}p^{\otimes n}_{X,Z}\|_{TV} \leq \varepsilon. \tag{3.98}$$

$$Q^{(1)}_{M',\hat{M}'|\mathcal{C}^{(n)}=c_*^{(n)}}(\hat{M}' \neq M') \leq \varepsilon. \tag{3.99}$$

We construct distribution $Q^{(2)}$ and D-code $P^{(1)}$ exactly as in Section 3.5, and by the same arguments that brought inequalities (3.67) and (3.74), we obtain

$$\left\| P^{(1)}_{M,M',J,X^n,Z^n,\hat{M}',Y^n} - Q^{(1)}_{M,M',J,X^n,Z^n,\hat{M}',Y^n} \right\|_{TV} \leq 2\varepsilon. \tag{3.100}$$

Hence, $P^{(1)}$ achieves near-perfect joint realism. We conclude as in Section 3.4.7, but using Proposition 3.31 instead of Proposition 3.30, with (3.96), (3.97) and (3.100).

3.8 Proofs of Theorems 3.11 and 3.12 and Corollaries 3.4 and 3.7

3.8.1 Proof of Theorem 3.11

Achievability follows from Theorem 3.9. Moving to the converse, we show that D-achievability with near-perfect realism marginal for joint source and side information distribution $p_{X,Z}$ implies D-achievability with near-perfect marginal realism for joint source and side information distribution $p_{X,\phi(X)}$. This is the object of Subsection 3.8.1. For the latter type of achievability we have the outer bound of Theorem 3.3 (for E-D-codes with marginal realism). We conclude, in Subsection 3.22, by rewriting the latter bound into the form appearing in (3.12). Throughout this section, we manipulate Markov chains in a way that is correct for general alphabets (see Appendix 1), but do not justify this fact explicitly for the sake of conciseness.

Using only the common part is sufficient

Let (R, R_c, Δ) be achievable for $p_{X,Z}$ and P be induced by a corresponding code. Then, P satisfies

$$Z^n - (X^n, J) - M, \text{ hence } ((\psi(Z_t))_{t\in[n]}, Z^n) - (X^n, J) - M \tag{3.101}$$

$$X^n - (M, J, Z^n) - Y^n, \text{ hence } X^n - (M, J, Z^n, (\psi(Z_t))_{t\in[n]}) - Y^n. \tag{3.102}$$

We add to the decoder a variable $\tilde{Z}^n$ defined by

$$P_{\tilde{Z}^n|X^n,Z^n,J,M,Y^n} := \prod_{t=1}^{n} p_{Z|\psi(Z)=\psi(z_t)}. \tag{3.103}$$

We use the initial decoder $P_{Y^n|M,J,Z^n}$, applied to $\tilde{Z}^n$ instead of Z^n, outputting variable $\tilde{Y}^n$ as follows:

$$P_{\tilde{Y}^n|X^n,Z^n,J,M,Y^n,\tilde{Z}^n} := P_{Y^n|M=m,J=j,Z^n=\tilde{z}^n}. \tag{3.104}$$

By construction P satisfies

$$X^n - (M, J, (\psi(Z_t))_{t\in[n]}) - (\tilde{Z}^n, \tilde{Y}^n), \text{ hence } X^n - (M, J, (\psi(Z_t))_{t\in[n]}, \tilde{Z}^n) - \tilde{Y}^n. \tag{3.105}$$

We claim that replacing (Z^n, Y^n) by $(\tilde{Z}^n, \tilde{Y}^n)$ in the initial code does not change the joint distribution:

Claim 3.22

$$P_{X^n,\tilde{Z}^n,J,M,\tilde{Y}^n,(\psi(Z_t))_{t\in[n]}} \equiv P_{X^n,Z^n,J,M,Y^n,(\psi(Z_t))_{t\in[n]}}. \tag{3.106}$$

Proof: From (3.102), (3.104) and (3.105), it is sufficient to prove that

$$P_{X^n,\tilde{Z}^n,J,M,(\psi(Z_t))_{t\in[n]}} \equiv P_{X^n,Z^n,J,M,(\psi(Z_t))_{t\in[n]}}. \tag{3.107}$$

By definition (3.103) of $\tilde{Z}^n$ and the fact that $P_{Z^n,(\psi(Z_t))_{t\in[n]}} \equiv p_{Z,\psi(Z)}^{\otimes n}$, it is sufficient to show that P satisfies $Z^n - (\psi(Z_t))_{t\in[n]} - (X^n, M, J)$. We denote $(\psi(Z_t))_{t\in[n]}$ by S^n. By the chain rule for general alphabets (Proposition 1.14), we have

$$\begin{aligned} I(Z^n; X^n, M, J, S^n) &= I(Z^n; X^n, J, S^n) + I(Z^n; M|X^n, J, S^n) \\ &= I(Z^n; X^n, S^n) + I(Z^n; J|X^n, S^n) + I(Z^n; M|X^n, J, S^n) \\ &= I(Z^n; S^n) + I(Z^n; X^n|S^n) + I(Z^n; J|X^n, S^n) + I(Z^n; M|X^n, J, S^n). \end{aligned}$$

By assumption, S^n is finite-valued, hence the first term is finite (by basic Lemma 1.12). It remains to prove that the last three terms are null. Since $P_{X^n,Z^n} \sim p_{X,Z}^{\otimes}$, then by assumption P satisfies

$$\forall t \in [n],\ Z_t - \psi(Z_t) - X_t,\ \text{thus}\ Z^n - S^n - X^n \tag{3.108}$$

by the chain rule (Proposition 1.14). Since J is independent from (X^n, Z^n) and ψ is deterministic, then J is independent from (X^n, Z^n, S^n). Hence, $I(Z^n; J|X^n, S) = 0$ by the chain rule. From (3.101) and the chain rule we have $I(Z^n; M|X^n, J, S^n) = 0$. ■

From (3.101), (3.105), Claim 3.22 and Lemma 3.1, distribution $P_{X^n,J,M,\tilde{Y}^n,(\psi(Z_t))_{t\in[n]}}$ is the distribution induced by a D-code for joint source and side information distribution $p_{X,\phi(X)}$. Moreover, by Claim 3.22, the message and common randomness rates are the same as for P. Thus, (R, R_c, Δ) is D-achievable with near-perfect marginal realism for joint source and side information distribution $p_{X,\phi(X)}$.

Conclusion

Hence, since $\phi(X)$ is finite-valued, triplet (R, R_c, Δ) satisfies the converse of Theorem 3.3 (for E-D-codes with marginal realism). Therefore, there exists $\varepsilon > 0$ and a distribution $p_{X,\phi(X),V,Y}$ satisfying

$$R + \varepsilon \geq I_p(X; V|\phi(X)) \tag{3.109}$$

$$R + R_c + \varepsilon \geq I_p(Y; V|\phi(X)) - H_p(\phi(X)|Y) \tag{3.110}$$

$$\Delta + \varepsilon \geq \mathbb{E}_p[d(X, Y)]$$

$$X - (\phi(X), V) - Y \tag{3.111}$$

$$p_Y \equiv p_X. \tag{3.112}$$

It remains to show that these properties can be rewritten into the form appearing in definition (3.12) of region $\mathcal{S}_D^{(m)}$ — for joint source and side information distribution $p_{X,Z}$. Define variable

$\tilde{Z}$ by

$$p_{\tilde{Z}|X,V,\phi(X),Y} := p_{Z|\psi(Z)=\phi(x)}. \tag{3.113}$$

Hence,

$$p_{\tilde{Z},X,\phi(X)} \equiv p_{X,\phi(X)} \cdot p_{Z|\psi(Z)=\phi(x)} \equiv p_{Z,X,\phi(X)}, \tag{3.114}$$

where the last equality follows from $X - \phi(X) - Z$. Define $\tilde{V} = (V, \phi(X))$. By definition (3.113) of $\tilde{Z}$, we have $X - (V, Y, \phi(X)) - \tilde{Z}$. Therefore, from (3.111) we have $X - (V, \phi(X)) - (Y, \tilde{Z})$, hence

$$X - (\tilde{V}, \tilde{Z}) - Y. \tag{3.115}$$

By definition (3.113) of $\tilde{Z}$, we have $\tilde{Z} - X - V$. We also trivially have $\tilde{Z} - (X, V) - \phi(X)$, hence

$$\tilde{Z} - X - \tilde{V}. \tag{3.116}$$

Moreover,

$$\begin{aligned} I(\tilde{Z};\tilde{V}) = I(\tilde{Z};V,\phi(X)) = I(\tilde{Z};\phi(X)) + I(\tilde{Z};V|\phi(X)) &= I(\tilde{Z};\psi(\tilde{Z})) \\ &= H(\psi(\tilde{Z})) \\ &= I(X;\phi(X)) \qquad (3.117) \\ &= I(\phi(X);Y) + H(\phi(X)|Y) \qquad (3.118) \end{aligned}$$

where (3.118) follows from the fact that $\phi(X)$ is finite-valued. This implies that $I(\tilde{Z};\tilde{V}) < \infty$. Combining this with (3.112), (3.114), (3.115) and (3.116) gives $p_{X,\tilde{Z},\tilde{V},Y} \in \mathcal{D}_D^{(m)}$ (Eq. 3.13). To conclude, from (3.116),

$$\begin{aligned} I(X;\tilde{V}) = I(X,\tilde{Z};\tilde{V}) = I(X;V,\phi(X)) + I(\tilde{Z};V,\phi(X)|X) &= I(X;V,\phi(X)) \\ &= I(X;\phi(X)) + I(X;V|\phi(X)) \end{aligned}$$

and

$$I(Y;\tilde{V}) = I(Y;\phi(X)) + I(Y;V|\phi(X)),$$

hence from (3.109), (3.110), (3.117), and (3.118), we obtain

$$R + \varepsilon \geq I(X;\tilde{V}) - I(\tilde{Z};\tilde{V}) \;\text{ and }\; R + R_c + \varepsilon \geq I(Y;\tilde{V}) - I(\tilde{Z};\tilde{V}).$$

This concludes the proof.

3.8.2 Proof of Theorem 3.12

Achievability

Consider a point (R, Δ) in $\mathcal{S}^{(m)}_{D,\infty}$. By definition there exists $p_{X,Z,V,Y} \in \mathcal{D}^{(m)}_D$ (Eq. 3.13) such that

$$R \geq I_p(X;V|Z) \text{ and } \infty > I_p(Y;V) \text{ and } \Delta \geq \mathbb{E}_p[d(X,Y)].$$

Since $p_{X,Z,V,Y} \in \mathcal{D}^{(m)}_D$ we also have $I_p(Z;V) < \infty$ and $Z - X - V$, hence

$$I_p(X;V|Z) = I_p(X;V) - I_p(Z;V) \;\text{ and }\; \infty > I_p(Y;V) - I_p(Z;V).$$

Fix some $R_c \geq \max(0, I_p(Y;V) - I_p(Z;V))$. Then, $(R, R_c, \Delta) \in \mathcal{S}^{(m)}_D$. By Theorem 3.9, we have $\mathcal{S}^{(m)}_D \subseteq \overline{\mathcal{A}}^{(m)}_D$. Hence, $(R, \Delta) \in \overline{\mathcal{A}}^{(m)}_{D,\infty}$.

Converse

We are to prove that $\overline{\mathcal{A}}^{(m)}_{D,\infty} \subseteq \overline{\mathcal{S}}^{(m)}_{D,\infty}$. To do so, we can incorporate quantization on the side information in the proof of the converse of Theorem 3.3 (E-D-codes with marginal realism), where bounds on rates and distortion do not rely on any Markov chain properties. We only

add the following arguments, related to the difference between the constraints (3.4) and (3.13) regarding the single-letter distributions. First, the Markov chain $Z_T - X_T - V$ is satisfied by Lemma 3.38 (Appendix 3.C.4). Second, we have $I(Z_T;V) < \infty$ by the exact same argument as in the converse proof of Theorem 3.14 (D-codes with joint realism), which does not rely on joint realism. Third, we have $I(Y_T;V) < \infty$ simply if either of $\mathcal{X}$ or $\mathcal{Z}$ is finite. Indeed, in that case, either Y_T or V is finite-valued. Finally, the quantization argument for the lower bound on the rate is done with three quantizers, on $\mathcal{X}$, $\mathcal{Z}$, and $\mathcal{Z}^{n-1}$, rather than just on $\mathcal{X}$.

3.8.3 Proof of Corollary 3.7

Optimality of region (3.9)

The converse is a direct consequence of the converse direction of Theorem 3.3, and the fact that the alphabet of V is finite. The achievability follows from that of Theorem 3.3 by taking $R_c = H(Z) + H(V)$ — which is finite because the alphabets of Z and V are finite.

Optimality of region (3.10)

For the achievability, it is sufficient to provide a proof in the case of finite-rate common randomness. Achievability follows from that of region (3.9), and by defining V as equal to Y. For the converse, it is sufficient to prove it in the case of unconstrained common randomness. Consider a couple (R, Δ) E-D achievable with near-perfect marginal realism with unconstrained common randomness. In the converse proof of Theorem 3.3, except for the finiteness of $\mathcal{V}$ which is not needed here, the assumption of finite-rate common randomness is not necessary as long as R_c is replaced by ∞. In particular, note that Theorem 3.2 (on the equivalence between perfect and near-perfect realism) is stated with both unconstrained and finite-rate common randomness. The only change that is needed is to replace the common randomness variable J by a quantized version in the derivation — which preserves independence relations —, then take a limit. Hence, for any $\varepsilon > 0$ there exists a point in $\mathcal{D}_{E\text{-}D}^{(m)}$ (Eq. 3.4) such that $(R + \varepsilon, \Delta + \varepsilon)$ satisfies

the first and third inequality in definition (3.3) of $\mathcal{S}^{(m)}_{E\text{-}D}$. The conclusion follows from the data processing inequality, using the Markov chain property appearing in definition (3.4) of $\mathcal{D}^{(m)}_{E\text{-}D}$.

3.8.4 Proof of Corollary 3.4

Consider an element (R, R_c, Δ) of $\mathcal{S}^{(m)}_{E\text{-}D}$. Let $p_{X,Z,V,Y}$ be a corresponding element of $\mathcal{D}^{(m)}_{E\text{-}D}$ (Eq. 3.4). Then, all the constraints appearing in (3.6) and (3.7) are satisfied with auxiliary random variable $\tilde{V} = (V, Z)$, and common randomness rate $R_c + H(Z)$. Thus, $(R, R_c + H(Z), \Delta) \in \mathcal{S}^{(m)}$. Hence,

$$\mathcal{S}^{(m)}_{E\text{-}D} + (0, H(Z), 0) \subseteq \mathcal{S}^{(m)} \cap \mathbb{R}_{\geq 0} \times [H(Z), \infty] \times \mathbb{R}_{\geq 0}.$$

Moving to the reverse direction, consider an element (R, R_c, Δ) of $\mathcal{S}^{(m)} \cap \mathbb{R}_{\geq 0} \times [H(Z), \infty] \times \mathbb{R}_{\geq 0}$, and a corresponding $p_{X,V,Y}$ in $\mathcal{D}^{(m)}$ (Eq. 3.7). Define $p_{Z,X,V,Y} := p_Z \cdot p_{X,V,Y}$. Then, we have $I_p(X;V) = I_p(X;V|Z)$, $I_p(Y;V) = I_p(Y;V|Z) - H_p(Z|Y) + H_p(Z)$, and

$$Z - X - V - Y, \text{ thus } (X, Z) - V - Y, \text{ hence } X - (Z, V) - Y.$$

Therefore, all constraints in (3.3) and (3.4) are satisfied with common randomness rate $\tilde{R}_c = R_c - H(Z)$. Thus, $(R, R_c - H(Z), \Delta) \in \mathcal{S}^{(m)}_{E\text{-}D}$. Hence,

$$\mathcal{S}^{(m)} \cap \mathbb{R}_{\geq 0} \times [H(Z), \infty] \times \mathbb{R}_{\geq 0} \subseteq \mathcal{S}^{(m)}_{E\text{-}D} + (0, H(Z), 0).$$

3.9 The Gaussian case

3.9.1 The role of common randomness for a Gaussian source in the absence of side information

We now prove Proposition 3.16. Without loss of generality, we consider ρ to be non-negative. Simple algebra shows that ρ is the unique solution in $[0, 1)$ to

$$1 - \frac{\Delta}{2} = \rho\sqrt{1 - 2^{-2R_c}(1-\rho^2)}, \tag{3.119}$$

which is a quadratic equation in ρ^2.

Consider ρ satisfying (3.119), and R satisfying (3.19). We choose a distribution $p_{X,V,Y}$ with V being one-dimensional standard Gaussian such that (X, V, Y) are jointly Gaussian, satisfy the Markov chain $X - V - Y$, and

$$\mathbb{E}[XV] = \rho, \tag{3.120}$$

$$\mathbb{E}[VY] = \sqrt{1 - 2^{-2R_c}(1-\rho^2)} =: \tilde{\rho}, \tag{3.121}$$

with Y also being standard Gaussian. Then, $p_{X,V,Y}$ satisfies all the conditions of set $\mathcal{D}^{(m)}$ (Eq. 3.7), except for the finiteness of the alphabet of V. Nevertheless, from Corollary 3.5 and Theorem 3.9 in the case of constant side information, the closure of optimal region $\mathcal{S}^{(m)}$ (Eq. 3.6) would not change if the alphabet of the auxiliary random variable V were not constrained to be finite, but simply Polish.

Claim 3.23 *Through simple computations, one can obtain*

$$\mathbb{E}[(X-Y)^2] = \Delta \quad \textit{and} \quad I(X;V) = \frac{1}{2}\log\frac{1}{1-\rho^2} \quad \textit{and} \quad I(Y;V) = R_c + \frac{1}{2}\log\frac{1}{1-\rho^2}.$$

The proof of this claim is deferred to Appendix 3.C.3. Therefore, $(R, R_c, \Delta) \in \overline{\mathcal{S}}^{(m)}$. Thus, $(R, R_c, \Delta) \in \overline{\mathcal{S}}^{(m)}_{\text{E-D}}$, assuming constant side information. The latter property is the starting

point, in Subsection 3.4.3, of the achievability proof of Theorem 3.3. From that proof, and in particular from its conclusion in Subsection 3.4.7, we have that for any $\varepsilon > 0$, the triplet $(R+\varepsilon, R_c, \Delta+\varepsilon)$ is achievable with perfect marginal realism. By combining codes with different ε, we obtain that for any $\delta > 0$, the triplet $(R + \delta, R_c, \Delta)$ is achievable with perfect marginal realism. Since R satisfies (3.19), the direct part of the proposition is proved.

Moving to the reverse direction, suppose $(R, R_c, \Delta) \in \mathcal{A}^{(m)}$. Since (d, p_X) is uniformly integrable (Claim 2.9 in Chapter 2), then by Corollary 3.5 we have $(R, R_c, \Delta) \in \overline{\mathcal{S}}^{(m)}$. Therefore, for arbitrarily small $\varepsilon > 0$, we have $(R + \varepsilon, R_c + \varepsilon, \Delta + \varepsilon) \in \mathcal{S}^{(m)}$. Fix such a $\varepsilon > 0$ satisfying $\Delta + \varepsilon < 2$, which exists since $\Delta \in (0, 2)$. Then, by definition of $\mathcal{S}^{(m)}$ there exists $p_{X,V,Y} \in \mathcal{D}^{(m)}$ (Eq. 3.7) such that

$$R + \varepsilon \geq I(X;V) \tag{3.122}$$

$$R_c + R + 2\varepsilon \geq I(Y;V) \tag{3.123}$$

$$\Delta + \varepsilon \geq \mathbb{E}[(X-Y)^2]. \tag{3.124}$$

Now define

$$\rho = \sqrt{\mathbb{E}[\mathbb{E}[X|V]^2]} \tag{3.125}$$

$$\tilde{\rho} = \sqrt{\mathbb{E}[\mathbb{E}[Y|V]^2]}. \tag{3.126}$$

Since R and R_c are finite but $I(X;X)$ and $I(Y;Y)$ are infinite, then each of X and Y is not a deterministic function of V. Since $\mathbb{E}[X^2] = \mathbb{E}[Y^2] = 1$, this implies

$$\rho, \tilde{\rho} \in [0, 1). \tag{3.127}$$

Claim 3.24 *Using the entropy-maximizing property of the Gaussian distribution, one can ob-*

tain

$$R + \varepsilon \geq \frac{1}{2} \log \frac{1}{1-\rho^2} \tag{3.128}$$
$$R + R_c + 2\varepsilon \geq \frac{1}{2} \log \frac{1}{1-\tilde{\rho}^2}.$$

The proof of this claim is deferred to Appendix 3.C.3. The second inequality can be written as

$$R + 2\varepsilon \geq \frac{1}{2} \log \frac{2^{-2R_c}}{1-\tilde{\rho}^2}. \tag{3.129}$$

Turning to the distortion constraint, we have — see Appendix 3.C.3:

Claim 3.25 *By Cauchy-Schwarz inequality, one can obtain*

$$\Delta + \varepsilon \geq 2 - 2\rho\tilde{\rho}. \tag{3.130}$$

Thus, from (3.127), (3.128), (3.129) and (3.130) we have

$$R + 2\varepsilon \geq \inf_{\rho,\tilde{\rho}\in[0,1)} \max\left(\frac{1}{2} \log \frac{1}{1-\rho^2}, \frac{1}{2} \log \frac{2^{-2R_c}}{1-\tilde{\rho}^2}\right)$$
$$\text{s.t.} \quad \Delta + \varepsilon \geq 2 - 2\rho\tilde{\rho}. \tag{3.131}$$

Since $s \mapsto -\log(1-s^2)$ is continuous and diverges when $s \to 1$, and the constraint $\Delta + \varepsilon \geq 2 - 2\rho\tilde{\rho}$ does not prevent any of $\rho, \tilde{\rho}$ from increasing with the other being fixed, then the optimization space can be reduced to the set of $(\rho, \tilde{\rho}) \in [0,1)^2$ such that

$$\frac{1}{2} \log \frac{1}{1-\rho^2} = \frac{1}{2} \log \frac{2^{-2R_c}}{1-\tilde{\rho}^2}, \quad \text{i.e.,} \quad \tilde{\rho} = \sqrt{1 - 2^{-2R_c}(1-\rho^2)}.$$

Then, there exists such a $(\rho_\varepsilon, \tilde{\rho}_\varepsilon)$ achieving near-optimality:

$$R + 3\varepsilon \geq \frac{1}{2} \log \frac{1}{1-\rho_\varepsilon^2}. \tag{3.132}$$

Claim 3.26 *By taking $\varepsilon \to 0$ and using a continuity and monotonicity argument, we obtain the desired:*

$$R \geq \frac{1}{2} \log \frac{1}{1-(\rho^*)^2},$$

where ρ^* is the unique solution in $[0,1)$ to

$$1 - \frac{\Delta}{2} = \rho \sqrt{1 - 2^{-2R_c}(1-\rho^2)}. \tag{3.133}$$

The proof is provided in Appendix 3.C.3.

3.9.2 Bi-dimensional Gaussian source and side information

We now prove Proposition 3.17. Fix $\Delta \in (0, 2-2|\eta|]$. Since we know that $R_D(\Delta) \geq R_{E\text{-}D}(\Delta)$, it is sufficient to prove that $R_{E\text{-}D}(\Delta)$ is lower bounded by the right hand side of (3.20) and that $R_D(\Delta)$ is upper bounded by the latter. We start with the lower bound.

Claim 3.27 $R_{E\text{-}D}(\Delta) \geq \frac{1}{2}\log((1-\eta^2)/(1-\rho^2))$.

Proof: Fix $\varepsilon > 0$. Consider a triplet (R, R_c, Δ) E-D-achievable with near-perfect marginal realism. From Theorem 3.2, (R, R_c, Δ) is E-D-achievable with perfect marginal realism. Then, there exists $n \in \mathbb{N}$ and an (n, R, R_c) E-D-code inducing a joint distribution P such that $P_{Y^n} \equiv p_X^{\otimes n}$ and $\mathbb{E}_P[d(X^n, Y^n)] \leq \Delta + \varepsilon$, i.e., $(1/n)\sum_{t=1}^n \mathbb{E}[X_t Y_t] \geq \rho_\varepsilon$, where $\rho_\varepsilon = 1-(\Delta+\varepsilon)/2$. Then, similarly to Claim 3.24 — see its proof in Appendix 3.C.3 —, we can lower bound $I(X_t; Y_t)$ as follows.

$$\begin{aligned} nR \geq I(M; Y^n|Z^n) \geq I(X^n; Y^n|Z^n) &\geq I(X^n; Y^n) - I(X^n; Z^n) \\ &\geq \sum_{t=1}^n I(X_t; Y_t) + \frac{n}{2}\log(1-\eta^2) \\ &\geq \sum_{t=1}^n \frac{1}{2} \log \frac{1-\eta^2}{1-\mathbb{E}[\mathbb{E}[X_t|Y_t]^2]} \end{aligned} \tag{3.134}$$

$$\geq \sum_{t=1}^{n} \frac{1}{2} \log \frac{1-\eta^2}{1-\mathbb{E}[Y_t \mathbb{E}[X_t|Y_t]]^2} \tag{3.135}$$

$$\geq \frac{n}{2} \log \frac{1-\eta^2}{1-(\frac{1}{n}\sum_{t=1}^{n} \mathbb{E}[X_t Y_t])^2} \tag{3.136}$$

$$\geq \frac{n}{2} \log \frac{1-\eta^2}{1-\rho_\varepsilon^2},$$

where (3.134) follows from the chain rule for mutual information and from the fact that $P_{X^n} \equiv p_X^{\otimes n}$; and (3.135) follows from Cauchy-Schwarz inequality; and (3.136) follows from Jensen inequality. This being true for every $\varepsilon > 0$, and any triplet (R, R_c, Δ) E-D-achievable with near-perfect marginal realism, the claim is proved. ■

It remains to prove that $R_D(\Delta)$ is upper bounded by the right hand side of (3.20).

Case 1: $\eta = 0$

We have

$$R(\Delta, R_c) \underset{R_c \to \infty}{\longrightarrow} \tfrac{1}{2} \log(1/(1-\rho^2)), \tag{3.137}$$

with $R(\Delta, R_c)$ as defined in Proposition 3.16 and $\rho = 1 - \Delta/2$. Therefore, by Proposition 3.16, for any $\varepsilon > 0$ there exists a finite rate $R_c^{(\varepsilon)}$ such that $(R^{(\varepsilon)}, R_c^{(\varepsilon)}, \Delta)$ is D-achievable with perfect marginal realism, where

$$R^{(\varepsilon)} = \tfrac{1}{2} \log(1/(1-\rho^2)) + \varepsilon. \tag{3.138}$$

Hence, we have

$$R_D(\Delta) \leq \tfrac{1}{2} \log(1/(1-\rho^2)).$$

Case 2: $\eta \neq 0$

Fix a Δ in $(0, 2-2|\eta|]$. Let

$$(Z, X, V) = \left(\eta\tilde{X} + \sqrt{1-\eta^2}\tilde{Z}, \tilde{X}, b\tilde{X} + \sqrt{1-b^2}\tilde{V}\right) \tag{3.139}$$

$$Y = \rho^{-1}\mathbb{E}[X|Z, V], \tag{3.140}$$

where $(\tilde{Z}, \tilde{X}, \tilde{V})$ is three-dimensional standard Gaussian, with the real b chosen as follows.

Claim 3.28 *For a suitable choice of b in $(0, 1)$, one can find*

$$\mathbb{E}\left[\mathbb{E}[X|Z, V]^2\right] = \rho^2 \tag{3.141}$$

$$p_{X,Z,V,Z} \in \mathcal{D}_D^{(m)} \text{ (Eq. 3.13).} \tag{3.142}$$

Claim 3.29

$$I_p(X; V|Z) = \tfrac{1}{2}\log((1-\eta^2)/(1-\rho^2)), \tag{3.143}$$

$$\mathbb{E}[d(X, Y)] = \Delta \tag{3.144}$$

$$I(Y; V) < \infty. \tag{3.145}$$

The proofs are deferred to Appendices 3.C.3 and 3.35. Moreover, from Claim 2.9 (Chapter 2), (d, p_X) is uniformly integrable. Therefore, by Theorem 3.9 there exists a finite R_c such that

$$(I_p(X; V|Z), R_c, \Delta) \in \overline{\mathcal{A}^{(m)}}_D. \tag{3.146}$$

Then, for any $\varepsilon > 0$, there exists a sequence of $(n, I_p(X; V|Z) + \varepsilon, R_c + \varepsilon)$ D-codes achieving average distortion $\Delta + \varepsilon$ asymptotically, and perfect marginal realism. By considering a vanishing sequence $(\varepsilon_n)_{n\geq 0}$ and combining the latter codes, we obtain that for any $\delta > 0$, the triplet $(I_p(X; V|Z) + \delta, R_c + \delta, \Delta)$ is D-achievable with perfect marginal realism. Hence, the proof is

concluded:

$$R_D(\Delta) \leq \tfrac{1}{2}\log((1-\eta^2)/(1-\rho^2)).$$

3.10 Conclusion

In this chapter, we have studied the role of side information and common randomness in the RDP trade-off with strong perfect and near-perfect realism constraints. Our work complements previous results on the key role of randomization and common randomness regarding both the point-to-point setting — such as the results in Chapter 2 — and similar distributed settings. We have considered the traditional problem of compressing a memoryless source X^n in the presence of side information Z^n, studied by Wyner and Ziv, with the additional requirement of strong perfect or near-perfect strong realism on the reconstruction (a.k.a. marginal realism). We also introduced and studied a constraint of joint realism of the reconstruction and the side information. We have characterized the RD trade-off under perfect or near-perfect realism constraints when common randomness is available between the encoder and the decoder at a limited rate or is absent. Our results show how the similarity or dissimilarity between our problem and that of Wyner and Ziv depends on the common randomness rate. The entropy term present in our single-letter characterization under the marginal realism constraint when Z^n is available at both terminals somewhat quantifies the role of the side information as a source of common randomness. We showed that the side information cannot play said role under a joint perfect or near-perfect realism constraint. These results provide insight into the problem of video compression in the absence of (another) common source of randomness. We have also showed that when X and Z are jointly Gaussian and large enough common randomness rate is available, the RD trade-off is the same whether Z^n is available only at the encoder or at both terminals, but that this breaks down in the absence of common randomness, thereby further highlighting the role of common randomness.

In Chapters 2 and 3, we have provided extensive results showing that randomization is detrimental in the RDP trade-off with distribution matching constraints. In starking constrast with

that theory, there are state-of-the-art image compression algorithms which do not use large amounts of common randomness, and many can reach state-of-the-art performance without using any randomization at all. In the next chapter, we pin-point a source of this discrepancy by showing that the strong perfect or near-perfect realism constraint is equivalent to requiring that reconstructions pass a range of incomputable tests of realism (critics) — as well as computable ones — that involve jointly inspecting an impractically large number of samples. We characterize the optimal trade-off under a novel realism constraint which only considers computable (or nearly computable) critics that jointly inspect a practical number of samples.

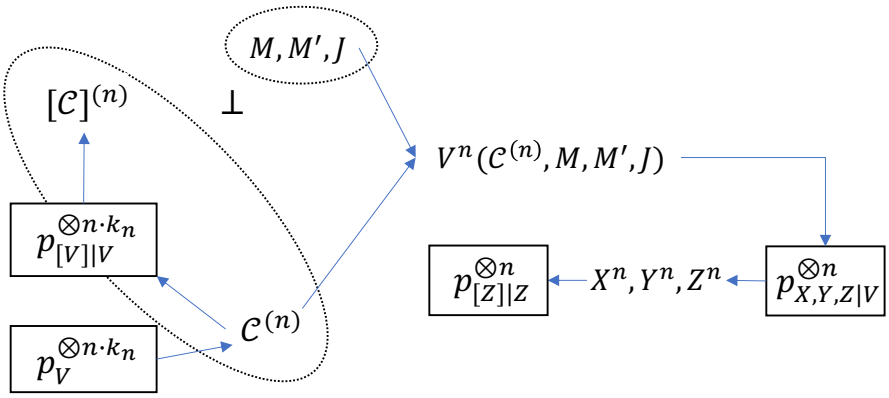


Figure 3.A.1: Graphical model for $Q^{(1)}$ with quantized variables, where $k_n = \lfloor 2^{n(R+\varepsilon)} \rfloor \times \lfloor 2^{nR'} \rfloor \times \lfloor 2^{nR_c} \rfloor$. The variables inside a dashed oval are independent from variables inside another oval.

Appendix

3.A Measure-theoretic justifications

Decoder for bin index M' in the case of infinite alphabets

We prove Claim 3.20. By definition of R', if $I_p(Z;V) = 0$ then $R' = 0$ and the alphabet of M' is a singleton, and therefore M' can be decoded perfectly for any choice of decoder. Suppose the contrary. Then, by definition of R' we have $R' < I_p(Z;V)$. Then, from Proposition 1.16 there exists a (measurably) quantized tuple $([Z],[V])$ such that $I_p([Z];[V]) > R'$. As depicted in Figure 3.A.1, we extend the definition of $Q^{(1)}$ from (3.33) by:

$$p_{V,X,Y,Z,[V],[Z]} = p_{V,X,Y,Z} \cdot p_{[V]|V} \cdot p_{[Z]|Z} \text{ and} \tag{3.147}$$

$$Q^{(1)}_{[Z]^n,[\mathcal{C}]^{(n)},\hat{M}'|\mathcal{C}^{(n)}=c^{(n)},M=m,M'=m',J=j,X^n=x^n,Y^n=y^n,Z^n=z^n}\left([z]^n,[c]^{(n)},\hat{m}'\right)$$
$$=\prod_{k=1}^{n} p_{[Z]|Z=z_k}\left([z]_k\right)\prod_{\tilde{j},\tilde{m},\tilde{m}'}\prod_{k=1}^{n} p_{[V]|V=v_k(c^{(n)},\tilde{m},\tilde{m}',\tilde{j})}\left([v]_k\right)$$
$$P^{D}_{\hat{M}'|[\mathcal{C}]^{(n)}=[c]^{(n)},M=m,J=j,[Z]^n=[z]^n}\left(\hat{m}'\right), \tag{3.148}$$

where P^D represents, for each (m,j), the joint typicality decoder of [CT06, page 201] for the joint distribution $p_{[V],[Z]}$ and codebook $([V]^n([\mathcal{C}]^{(n)},m,a,j))_a$. Note that all transition kernels in the right-hand-side are regular, because each represents a measurable deterministic mapping. It can be checked that this setting is that of the channel random coding proof of [CT06, Section 7.7]. Moreover, for any (m,j) the distribution of the corresponding random sub-codebook of $[\mathcal{C}]^{(n)}$ does not depend on (m,j). Therefore, we have

$$\sup_{(m,j)} Q^{(1)}_{\hat{M}',M'|M=m,J=j}\left(\hat{M}' \neq M'\right) \underset{n\to\infty}{\longrightarrow} 0. \tag{3.149}$$

3.B Equivalence between achievability with near-perfect and perfect realism

3.B.1 Marginal realism

In order to prove the part of Theorem 3.2 pertaining to marginal realism, we prove the following result, which we also use in Sections 3.4 and 3.5.

Proposition 3.30 *Let n be a positive integer and $\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4$ be positive reals. Let $\mathcal{X}$ and $\mathcal{W}$ be two Polish alphabets and p_X be a distribution on $\mathcal{X}$. Let d be a distortion measure on $\mathcal{X}^2$. Let Q_{X^n,Y^n} be a distribution on $\mathcal{X}^n\times\mathcal{X}^n$ satisfying*

$$\mathbb{E}_Q[d(X^n,Y^n)] \leq \Delta+\varepsilon_1 \text{ and } \|Q_{Y^n}-p_X^{\otimes n}\|_{TV}\leq\varepsilon_2,$$

for some real Δ. Let P_{X^n,W,Y^n} be a distribution on $\mathcal{X}^n\times\mathcal{W}\times\mathcal{X}^n$ satisfying $\hat{P}_{X^n}\equiv p_X$ and the

Markov chain property $X^n - W - Y^n$. *Moreover, assume that*

$$\|P_{Y^n} - Q_{Y^n}\|_{TV} \leq \varepsilon_3 \text{ and } \|\hat{P}_{X^n,Y^n} - \hat{Q}_{X^n,Y^n}\|_{TV} \leq \varepsilon_4.$$

Then, there exists a conditional distribution $P'_{Y^n|W}$ *such that the distribution* P' *defined by*

$$P'_{X^n,W,Y^n} := P_{X^n,W} \cdot P'_{Y^n|W}$$

satisfies

$$\mathbb{E}_{P'}[d(X^n, Y^n)] \leq \Delta + \varepsilon_1 + \sup_{\mathbb{P}_{X,Y,\xi}} \mathbb{E}[d(X,Y) \cdot \xi] \quad \text{and} \quad P'_{Y^n} \equiv p_X^{\otimes n},$$

where the supremum is taken over all distributions $\mathbb{P}_{X,Y,\xi}$ *on* $\mathcal{X}^2 \times \{0,1\}$ *satisfying* $\mathbb{P}_X \equiv \mathbb{P}_Y \equiv p_X$ *and* $\mathbb{P}(\{\xi = 1\}) \leq \varepsilon_2 + \varepsilon_3 + \varepsilon_4$.

Proposition 3.30 implies the part of Theorem 3.2 pertaining to marginal realism, by taking $Q \equiv P$ at fixed blocklength, and $\mathcal{W} = [2^{nR}] \times [2^{nR_c}] \times \mathcal{Z}^n$, $W = (M, J, Z^n)$. Proposition 3.30 is proved in the next subsection.

3.B.2 Joint realism

In order to prove the part of Theorem 3.2 pertaining to joint realism, we prove the following result, which we also use in Sections 3.6 and 3.7.

Proposition 3.31 *Let* n *be a positive integer and* $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ *be positive reals. Let* $\mathcal{X}, \mathcal{Z}, \mathcal{W}$ *be Polish alphabets and* $p_{X,Z}$ *be a joint distribution on* $\mathcal{X} \times \mathcal{Z}$. *Let* d *be a distortion measure on* $\mathcal{X}^2$. *Let* Q_{X^n,Y^n,Z^n} *be a distribution on* $\mathcal{X}^n \times \mathcal{X}^n \times \mathcal{Z}^n$ *satisfying*

$$\mathbb{E}_Q[d(X^n, Y^n)] \leq \Delta + \varepsilon_1 \tag{3.150}$$

$$\text{and } \|Q_{Y^n,Z^n} - p_{X,Z}^{\otimes n}\|_{TV} \leq \varepsilon_2, \tag{3.151}$$

for some real Δ. *Let* P_{X^n,Z^n,W,Y^n} *be a distribution on* $\mathcal{X}^n \times \mathcal{Z}^n \times \mathcal{W} \times \mathcal{X}^n$ *satisfying* $\hat{P}_{X^n} \equiv p_X$, $P_{Z^n} \equiv p_Z^{\otimes n}$, *and the Markov chain property* $X^n - (Z^n, W) - Y^n$. *Moreover, assume that*

$$\|P_{Y^n,Z^n} - Q_{Y^n,Z^n}\|_{TV} \leq \varepsilon_3 \tag{3.152}$$

$$\|\hat{P}_{X^n,Y^n} - \hat{Q}_{X^n,Y^n}\|_{TV} \leq \varepsilon_4. \tag{3.153}$$

Then, there exists a conditional distribution $P'_{Y^n|Z^n,W}$ *such that the distribution* P' *defined by*

$$P'_{X^n,Z^n,W,Y^n} := P_{X^n,Z^n,W} \cdot P'_{Y^n|Z^n,W} \tag{3.154}$$

satisfies

$$\mathbb{E}_{P'}[d(X^n, Y^n)] \leq \Delta + \varepsilon_1 + \sup_{\mathbb{P}_{X,Y,\xi}} \mathbb{E}[d(X,Y) \cdot \xi] \quad \textit{and} \quad P'_{Y^n,Z^n} \equiv p_{X,Z}^{\otimes n},$$

where the supremum is taken over all distributions $\mathbb{P}_{X,Y,\xi}$ *on* $\mathcal{X}^2 \times \{0,1\}$ *satisfying* $\mathbb{P}_X \equiv \mathbb{P}_Y \equiv p_X$ *and* $\mathbb{P}(\{\xi = 1\}) \leq \varepsilon_2 + \varepsilon_3 + \varepsilon_4$.

Proposition 3.31 implies the part of Theorem 3.2 pertaining to joint realism, by taking $Q \equiv P$ at fixed blocklength, and $\mathcal{W} = [2^{nR}] \times [2^{nR_c}]$, $W = (M, J)$. Moreover, Proposition 3.30 is a direct consequence of Proposition 3.31 applied with a singleton $\mathcal{Z}$. The remainder of Appendix 3.B is dedicated to the proof of Proposition 3.31.

3.B.3 Modified distribution

Consider the setting of Proposition 3.31. If $P_{Y^n,Z^n} \equiv p_{X,Z}^{\otimes n}$, we just set $P' := P$. Assume this is not the case. From the Markov property $X^n - (W, Z^n) - Y^n$ the conditional distribution of Y^n knowing all other variables reduces to $P_{Y^n|Z^n,W}$. In order to ensure measurability properties regarding conditional densities, we consider a transition kernel $P_{Y^n|Z^n,W}$ constructed as follows. Define the σ-finite measure $\mu_{Y^n,Z^n,W} = P_{Y^n,Z^n,W} + p_{X,Z}^{\otimes n} \cdot P_W$. Then, $\mu_{Y^n,Z^n,W}$ dominates $P_{Y^n,Z^n,W}$, and μ_{Y^n,Z^n} dominates $p_{Y,Z}^{\otimes n}$. Fix densities $dP_{Y^n,Z^n,W}/d\mu_{Y^n,Z^n,W}$ and $dp_{Y,Z}^{\otimes n}/d\mu_{Y^n,Z^n}$ and transition kernels $\mu_{W|Z^n}$ and $\mu_{Y^n|Z^n,W}$. This determines a unique transition kernel $\mu_{Y^n,W|Z^n}$, and

thus a unique $\mu_{Y^n|Z^n}$. From Corollary 1.6 (Appendix 1), we can construct a density function $dP_{Z^n,W}/d\mu_{Z^n,W}$ and a regular transition kernel $P_{Y^n|Z^n,W}$ of $P_{Y^n,Z^n,W}$ such that the following holds: for all $(z^n, w) \in \mathcal{Z}^n \times \mathcal{W}$, the measure $\mu_{Y^n|Z^n=z^n,W=w}$ dominates $P_{Y^n|Z^n=z^n,W=w}$, and there exists a choice of conditional density functions such that the map

$$(y^n, z^n, w) \mapsto \frac{dP_{Y^n|Z^n=z^n,W=w}}{d\mu_{Y^n|Z^n=z^n,W=w}}(y^n) \tag{3.155}$$

is measurable. Similarly, one can construct a transition kernel $P_{W|Z^n}$ such that the following holds: for all $z^n \in \mathcal{Z}^n$, the measure $\mu_{W|Z^n=z^n}$ dominates $P_{W|Z^n=z^n}$, and there exists a choice of conditional density functions such that the map

$$(z^n, w) \mapsto \frac{dP_{W|Z^n=z^n}}{d\mu_{W|Z^n=z^n}}(w) \tag{3.156}$$

is measurable. The transition kernels $P_{Y^n|Z^n,W}$ and $P_{W|Z^n}$ determine a unique transition kernel $P_{Y^n,W|Z^n}$, and thus a unique $P_{Y^n|Z^n}$. Then, from the measurability of the maps in (3.155) and (3.156), and from the generalized Fubini theorem (Theorem 1.1 in Appendix 1), one can readily check that for any $z^n \in \mathcal{Z}^n$, the measure $\mu_{Y^n|Z^n=z^n}$ dominates $P_{Y^n|Z^n=z^n}$, with the following map

$$y^n \mapsto \int_{\mathcal{W}} \frac{dP_{W|Z^n=z^n}}{d\mu_{W|Z^n=z^n}}(w) \frac{dP_{Y^n|Z^n=z^n,W=w}}{d\mu_{Y^n|Z^n=z^n,W=w}}(y^n) d\mu_{W|Y^n=y^n,Z^n=z^n}$$

as density, denoted $dP_{Y^n|Z^n=z^n}/d\mu_{Y^n|Z^n=z^n}$. Moreover, from the measurability of the maps in (3.155) and (3.156), and from Theorem 1.1, the map

$$(y^n, z^n) \mapsto \frac{dP_{Y^n|Z^n=z^n}}{d\mu_{Y^n|Z^n=z^n}}(y^n) \tag{3.157}$$

is measurable. Similarly, one can construct a transition kernel $p_{Y^n|Z^n}$ such that the following holds: for all $z^n \in \mathcal{Z}^n$, the measure $\mu_{Y^n|Z^n=z^n}$ dominates $p_{Y^n|Z^n=z^n}$, and there exists a choice of conditional density functions such that the map

$$(y^n, z^n) \mapsto \frac{dp_{Y^n|Z^n=z^n}}{d\mu_{Y^n|Z^n=z^n}}(y^n) \tag{3.158}$$

is measurable. For any $z^n \in \mathcal{Z}^n$, let

$$\mathcal{X}_{z^n}^+ = \left\{ y^n \in \mathcal{X}^n \text{ s.t. } \frac{dP_{Y^n|Z^n=z^n}}{d\mu_{Y^n|Z^n=z^n}}(y^n) > \frac{dp_{Y^n|Z^n=z^n}}{d\mu_{Y^n|Z^n=z^n}}(y^n) \right\}$$
$$\text{and } \theta_{z^n} : \mathcal{X}_{z^n}^+ \to [0,1),\ y^n \mapsto \frac{dp_{Y^n|Z^n=z^n}/d\mu_{Y^n|Z^n=z^n}(y^n)}{dP_{Y^n|Z^n=z^n}/d\mu_{Y^n|Z^n=z^n}(y^n)}.$$

Then,

$$(y^n, z^n) \mapsto \mathbf{1}_{\mathcal{X}_{z^n}^+}(y^n) \text{ is measurable, and } (y^n, z^n) \mapsto \mathbf{1}_{\mathcal{X}_{z^n}^+}(y^n)\theta_{z^n}(y^n) \text{ is measurable} \quad (3.159)$$

In the remainder of Appendix 3.B, we implicitly use the following lemma, which follows from the measurability of singletons in Polish spaces (which are metrizable, therefore T4, thus T1).

Lemma 3.32 *Let $\mathcal{W}_1$ and $\mathcal{W}_2$ be Polish spaces. Let $\phi : \mathcal{W}_1 \times \mathcal{W}_2 \to \mathbb{R}$ be a measurable map. Then, for any $w_1 \in \mathcal{W}_1$, the map $w_2 \mapsto \phi(w_1, w_2)$ is measurable.*

Let B_1 denote the set of all $z^n \in \mathcal{Z}^n$ such that $P_{Y^n|Z^n=z^n} \neq p_{Y^n|Z^n=z^n}$. From the definition of the TVD (Eq. 1.5), we have, for any $z^n \in \mathcal{Z}^n$,

$$\|P_{Y^n|Z^n=z^n} - p_{Y^n|Z^n=z^n}\|_{TV} = \int_{\mathcal{X}^n \setminus \mathcal{X}_{z^n}^+} \left(\frac{dp_{Y^n|Z^n=z^n}}{d\mu_{Y^n|Z^n=z^n}}(y^n) - \frac{dP_{Y^n|Z^n=z^n}}{d\mu_{Y^n|Z^n=z^n}}(y^n) \right) d\mu_{Y^n|Z^n=z^n}. \quad (3.160)$$

Hence, from the measurability of the maps in (3.157) and (3.158), and from Theorem 1.1, the map $z^n \mapsto \|P_{Y^n|Z^n=z^n} - p_{Y^n|Z^n=z^n}\|_{TV}$ is measurable. Therefore, the set B_1 is measurable. For any w, and any $z^n \notin B_1$, let $P'_{Y^n|Z^n=z^n,W=w}(A) = P_{Y^n|Z^n=z^n,W=w}(A)$ for any Borel set $A \subseteq \mathcal{X}^n$. Otherwise, for any Borel set A we define $P'_{Y^n|Z^n=z^n,W=w}(A)$ as

$$P_{Y^n|Z^n=z^n,W=w}(A \setminus \mathcal{X}_{z^n}^+) + \int_{A \cap \mathcal{X}_{z^n}^+} \theta_{z^n}(y^n) dP_{Y^n|Z^n=z^n,W=w}(y^n) + \phi_{z^n,w}\Gamma_{z^n}(A), \text{ where}$$
$$\Gamma_{z^n}(A) = \frac{\int_{A \setminus \mathcal{X}_{z^n}^+} (dp_{Y^n|Z^n=z^n}/d\mu_{Y^n|Z^n=z^n}(y^n) - dP_{Y^n|Z^n=z^n}/d\mu_{Y^n|Z^n=z^n}(y^n)) d\mu_{Y^n|Z^n=z^n}}{\int_{\mathcal{X}^n \setminus \mathcal{X}_{z^n}^+} (dp_{Y^n|Z^n=z^n}/d\mu_{Y^n|Z^n=z^n}(y^n) - dP_{Y^n|Z^n=z^n}/d\mu_{Y^n|Z^n=z^n}(y^n)) d\mu_{Y^n|Z^n=z^n}}$$
$$\text{and } \phi_{z^n,w} = \int_{\mathcal{X}_{z^n}^+} 1 - \theta_{z^n}(y^n) dP_{Y^n|Z^n=z^n,W=w}(y^n). \quad (3.161)$$

This is well-defined by the measurability of the maps in (3.157) and (3.158), and by (3.159)

(and Lemma 3.32). From (3.160), the denominator in the definition of $\Gamma_{z^n}(A)$ is non-zero for $z^n \in B_1$. We check that this defines a probability measure for every (z^n, w) such that $z \in B_1$. Fix such a pair. By definition of $\mathcal{X}^+_{z^n}$ and θ_{z^n}, we have $P'_{Y^n|Z^n=z^n,W=w}(A) \geq 0$ for any Borel set A. In addition, $P'_{Y^n|Z^n=z^n,W=w}(\mathcal{X}^n)$ equals

$$\begin{aligned}
&P_{Y^n|Z^n=z^n,W=w}(\mathcal{X}^n \backslash \mathcal{X}^+_{z^n}) \; + \int_{\mathcal{X}^+_{z^n}} \theta_{z^n} dP_{Y^n|Z^n=z^n,W=w} + \int_{\mathcal{X}^+_{z^n}} 1 - \theta_{z^n} dP_{Y^n|Z^n=z^n,W=w} \\
&= P_{Y^n|Z^n=z^n,W=w}(\mathcal{X}^n \backslash \mathcal{X}^+_{z^n}) \; + \int_{\mathcal{X}^+_{z^n}} 1 dP_{Y^n|Z^n=z^n,W=w} \\
&= 1.
\end{aligned}$$

Thus, $P'_{Y^n|Z^n=z^n,W=w}$ is a probability measure. Moreover, for any Borel set A, the map $(z^n, w) \mapsto P'_{Y^n|Z^n=z^n,W=w}(A)$ is measurable because

- B_1 is measurable
- $P_{Y^n|Z^n,W}$ is a regular transition kernel
- All other integrals are measurbale functions of (z^n, w) from Theorem 1.1, (3.159), and the measurability of the maps in (3.157) and (3.158).

Hence, $P'_{Y^n|Z^n,W}$ is a regular transition kernel. We define the joint distribution for the new code as

$$P'_{X^n,\,Z^n,\,W,\,Y^n} = P_{X^n,\,Z^n,\,W} \cdot P'_{Y^n|Z^n,\,W}. \tag{3.162}$$

3.B.4 Proving perfect joint realism

We show that P' satisfies $P'_{Y^n,Z^n} \equiv p^{\otimes n}_{X,Z}$. By construction, this is true if $P_{Y^n,Z^n} \equiv p^{\otimes n}_{X,Z}$. Assume that this is not the case. Fix measurable measurable sets $A \subseteq \mathcal{X}^n$ and $B \subseteq \mathcal{Z}^n$. From the

definition (3.161) of $P'_{Y^n|Z^n=z^n,W=w}$, we have, for any $(w,z^n)\in\mathcal{W}\times\mathcal{Z}^n$,

$$\begin{aligned}
&\mathbf{1}_{B\cap B_1}(z^n)P'_{Y^n|Z^n=z^n,W=w}(A)\\
&=\mathbf{1}_{B\cap B_1}(z^n)\cdot\Big[P_{Y^n|Z^n=z^n,W=w}(A\backslash\mathcal{X}^+_{z^n})\ +\int_{A\cap\mathcal{X}^+_{z^n}}\theta_{z^n}(y^n)dP_{Y^n|Z^n=z^n,W=w}(y^n)+\phi_{z^n,w}\Gamma_{z^n}(A)\Big]\\
&=\mathbf{1}_{B\cap B_1}(z^n)\\
&\quad\cdot\int\Big[\mathbf{1}_{A\backslash\mathcal{X}^+_{z^n}}(y^n)+\mathbf{1}_{A\cap\mathcal{X}^+_{z^n}}(y^n)\ \theta_{z^n}(y^n)+\Gamma_{z^n}(A)\mathbf{1}_{\mathcal{X}^+_{z^n}}(1-\theta_{z^n}(y^n))\Big]dP_{Y^n|Z^n=z^n,W=w}.
\end{aligned}$$

From (3.162), similarly to P we have $P'_{Z^n}\equiv p_Z^{\otimes n}$. Then, by the generalized Fubini theorem (Theorem 1.1), integrating both sides according to $P_{Z^n,W}$ gives

$$\begin{aligned}
&P'_{Y^n,Z^n}(A\times(B\cap B_1))=P'_{Y^n,Z^n,W}(A\times(B\cap B_1))\\
&=\int\Big[\mathbf{1}_{(A\backslash\mathcal{X}^+_{z^n})\times(B\cap B_1)}+\mathbf{1}_{(A\cap\mathcal{X}^+_{z^n})\times(B\cap B_1)}\ \theta_{z^n}(y^n)+\Gamma_{z^n}(A)\mathbf{1}_{\mathcal{X}^+_{z^n}\times(B\cap B_1)}(1-\theta_{z^n}(y^n))\Big]dP_{Y^n,Z^n,W}\\
&=\int\Big[\mathbf{1}_{(A\backslash\mathcal{X}^+_{z^n})\times(B\cap B_1)}+\mathbf{1}_{(A\cap\mathcal{X}^+_{z^n})\times(B\cap B_1)}\ \theta_{z^n}(y^n)+\Gamma_{z^n}(A)\mathbf{1}_{\mathcal{X}^+_{z^n}\times(B\cap B_1)}(1-\theta_{z^n}(y^n))\Big]dP_{Y^n,Z^n}\\
&=\int\mathbf{1}_{B\cap B_1}(z^n)\int\Big[\mathbf{1}_{A\backslash\mathcal{X}^+_{z^n}}+\mathbf{1}_{A\cap\mathcal{X}^+_{z^n}}\ \theta_{z^n}(y^n)\Big]dP_{Y^n|Z^n=z^n}dP_{Z^n}(z^n)\\
&\quad+\int\mathbf{1}_{B\cap B_1}(z^n)\Gamma_{z^n}(A)\int\mathbf{1}_{\mathcal{X}^+_{z^n}}(1-\theta_{z^n}(y^n))dP_{Y^n|Z^n=z^n}dP_{Z^n}(z^n)\\
&=\int\mathbf{1}_{B\cap B_1}(z^n)\Big[\int\mathbf{1}_{A\backslash\mathcal{X}^+_{z^n}}dP_{Y^n|Z^n=z^n}+\int\mathbf{1}_{A\cap\mathcal{X}^+_{z^n}}dp_{Y^n|Z^n=z^n}\Big]dP_{Z^n}(z^n)\\
&\quad+\int\mathbf{1}_{B\cap B_1}(z^n)\Gamma_{z^n}(A)\int_{\mathcal{X}^+_{z^n}}\Big[dp_{Y^n|Z^n=z^n}/d\mu_{Y^n|Z^n=z^n}(y^n)\\
&\qquad\qquad-dP_{Y^n|Z^n=z^n}/d\mu_{Y^n|Z^n=z^n}(y^n)\Big]d\mu_{Y^n|Z^n=z^n}dP_{Z^n}(z^n)\\
&=\int\mathbf{1}_{B\cap B_1}(z^n)\Big[\int\mathbf{1}_{A\backslash\mathcal{X}^+_{z^n}}dP_{Y^n|Z^n=z^n}+\int\mathbf{1}_{A\cap\mathcal{X}^+_{z^n}}dp_{Y^n|Z^n=z^n}\\
&\quad+\int\mathbf{1}_{A\backslash\mathcal{X}^+_{z^n}}(dp_{Y^n|Z^n=z^n}-dP_{Y^n|Z^n=z^n})\Big]dP_{Z^n}(z^n)\\
&=p_{X,Z}^{\otimes n}(A\times(B\cap B_1)),
\end{aligned}$$

where throughout we omit null terms corresponding to values of z^n which are not in B_1. Moreover,

$$\begin{aligned}
P'_{Y^n,Z^n}(A\times(B\backslash B_1))&=P'_{Y^n,Z^n,W}(A\times(B\backslash B_1))\\
&=\int\mathbf{1}_{B\backslash B_1}(z^n)\int\mathbf{1}_A(y^n)dP'_{Y^n|Z^n=z^n,W=w}(y^n)dP'_{Z^n,W}(z^n,w)
\end{aligned}$$

$$
\begin{aligned}
&= \int \mathbf{1}_{B\backslash B_1}(z^n) \int \mathbf{1}_A(y^n) dP_{Y^n|Z^n=z^n,W=w}(y^n) dP_{Z^n,W}(z^n, w) \\
&= \int \mathbf{1}_{B\backslash B_1}(z^n) \int \mathbf{1}_A(y^n) dP_{Y^n|Z^n=z^n}(y^n) dP_{Z^n}(z^n) \\
&= \int \mathbf{1}_{B\backslash B_1}(z^n) \int \mathbf{1}_A(y^n) dp_{Y^n|Z^n=z^n}(y^n) dp_Z^{\otimes n}(z^n) \\
&= p_{X,Z}^{\otimes n}(A \times (B\backslash B_1)).
\end{aligned}
$$

Hence, $P'_{Y^n,Z^n}(A\times B) = p_{X,Z}^{\otimes n}(A\times B)$. This being true for any measurable (Borel) sets $A \subseteq \mathcal{X}^n$ and $B \subseteq \mathcal{Z}^n$, then, from the uniqueness property in Theorem 1.1, P'_{Y^n,Z^n} and $p_{X,Z}^{\otimes n}$ are equal on the product σ-algebra of the (Borel) σ-algebras of $\mathcal{X}^n$ and $\mathcal{Z}^n$. As mentioned in Appendix 1, the latter is equal to the Borel σ-algebra of the product Polish space $\mathcal{X}^n \times \mathcal{Z}^n$. Hence, $P'_{Y^n,Z^n} \equiv p_{X,Z}^{\otimes n}$. The last step consists of obtaining the distortion bound by coupling P' with Q.

3.B.5 Coupling with Q

We first compare P' to P. For every z^n, w and every set $A \in \mathcal{X}^n$ we have:

$$
\begin{aligned}
&P_{Y^n|Z^n=z^n,W=w}(A) - P'_{Y^n|Z^n=z^n,W=w}(A) \\
&\leq \int_{A\cap\mathcal{X}_{z^n}^+} 1 - \theta_{z^n}(y^n) dP_{Y^n|Z^n=z^n,W=w}(y^n) - \phi_{z^n,w}\Gamma_{z^n}(A).
\end{aligned}
$$

Both terms in the right hand side being non-negative and smaller than or equal to $\phi_{z^n,w}$, we have

$$
\left\| P'_{Y^n|Z^n=z^n,W=w} - P_{Y^n|Z^n=z^n,W=w} \right\|_{TV} \leq \phi_{z^n,w}.
$$

When integrating this over all variables except Y^n, we obtain by definition (3.162) of P' and Lemma 1.7 (Appendix 1) on the one hand and by definition (3.161) of ϕ and Fubini's theorem on the other hand:

$$
\begin{aligned}
\left\| P'_{X^n,Z^n,W,Y^n} - P_{X^n,Z^n,W,Y^n} \right\|_{TV} &\leq \int \mathbf{1}_{\mathcal{X}_{z^n}^+}[1-\theta^n(y^n)] dP_{Y^n,Z^n}(y^n, z^n) \\
&= \int\int \mathbf{1}_{\mathcal{X}_{z^n}^+}[1-\theta^n(y^n)] dP_{Y^n|Z^n=z^n} dP_{Z^n}(z^n)
\end{aligned}
$$

$$= \int \left\| P_{Y^n|Z^n=z^n} - p_{Y^n|Z^n=z^n} \right\|_{TV} dP_{Z^n}(z^n)$$
$$= \left\| P_{Y^n,Z^n} - p_{X,Z}^{\otimes n} \right\|_{TV}.$$

Therefore, by the triangle inequality and assumptions (3.151) and (3.152) we obtain

$$\left\| P'_{X^n,Z^n,W,Y^n} - P_{X^n,Z^n,W,Y^n} \right\|_{TV} \leq \varepsilon_2 + \varepsilon_3.$$

Therefore, by Lemma 1.8 (Appendix 1) and the triangle inequality we obtain

$$\left\| \hat{P}'_{X^n,Y^n} - \hat{P}_{X^n,Y^n} \right\|_{TV} \leq \varepsilon_2 + \varepsilon_3.$$

By applying again the triangle inequality together with assumption (3.153) we obtain

$$\left\| \hat{P}'_{X^n,Y^n} - \hat{Q}_{X^n,Y^n} \right\|_{TV} \leq \varepsilon_2 + \varepsilon_3 + \varepsilon_4. \tag{3.163}$$

We then use the following standard coupling lemma (see, e.g., [Lin92, Chapter I, Theorem 5.2]): from (3.163) there exists a quadruple $(X, Y, \tilde{X}, \tilde{Y})$ with distribution denoted by ρ having marginals $\rho_{X,Y} \equiv \hat{P}'_{X^n,Y^n}$ and $\rho_{\tilde{X},\tilde{Y}} \equiv \hat{Q}_{X^n,Y^n}$ and such that $\rho((X,Y) \neq (\tilde{X},\tilde{Y})) \leq \varepsilon_2 + \varepsilon_3 + \varepsilon_4$. We then have

$$\begin{aligned}
\mathbb{E}_{P'}[d(X^n,Y^n)] = \mathbb{E}_{\hat{P}'_{X^n,Y^n}}[d(X,Y)] &= \mathbb{E}_\rho[d(X,Y)] \\
&= \mathbb{E}_\rho\Big[d(\tilde{X},\tilde{Y})\mathbf{1}_{(X,Y)=(\tilde{X},\tilde{Y})}\Big] \\
&\quad + \mathbb{E}_\rho\Big[d(X,Y)\mathbf{1}_{(X,Y)\neq(\tilde{X},\tilde{Y})}\Big] \\
&\leq \mathbb{E}_{\hat{Q}_{X^n,Y^n}}[d(X,Y)] + \mathbb{E}_\rho\Big[d(X,Y)\mathbf{1}_{(X,Y)\neq(\tilde{X},\tilde{Y})}\Big] \\
&\leq \mathbb{E}_Q[d(X^n,Y^n)] + \sup_{\mathbb{P}_{X,Y,\xi}} \mathbb{E}\Big[d(X,Y)\cdot\xi\Big],
\end{aligned}$$

where the supremum is over all distributions $\mathbb{P}_{X,Y,\xi}$ on $\mathcal{X}^2 \times \{0,1\}$ satisfying $\mathbb{P}_X \equiv \mathbb{P}_Y \equiv p_X$ and $\mathbb{P}(\{\xi = 1\}) \leq \varepsilon_2 + \varepsilon_3 + \varepsilon_4$, and the last inequality holds because P' satisfies perfect realism $P'_{Y^n} = p_X^{\otimes n}$ and from (3.162) we have $\hat{P}'_{X^n} \equiv \hat{P}_{X^n} \equiv p_X$. We conclude by assumption (3.150).

3.C Further justifications

3.C.1 Use of the soft covering lemma when the side information is available only at the decoder

Statement

We state the soft covering lemma for general alphabets as it appears in [Cuf13], but specialize it to memoryless distributions and to the case of finite mutual information:

Lemma 3.33 *[Cuf13, Corollary VII.4] Let $\mathcal{V}$ be an alphabet and ρ_V a distribution on $\mathcal{V}$. Let R be a non-negative real number and let $(k_n)_{n\geq 1}$ be a sequence of positive integers satisfying $k_n \underset{n\to\infty}{\sim} 2^{nR}$. For every positive integer n let $\mathcal{E}^{(n)}$ be a randomly generated collection of k_n mutually independent sequences in $\mathcal{V}^n$ each drawn according to $\rho_V^{\otimes n}$. Its distribution is denoted $\Gamma_{\mathcal{E}^{(n)}}$. The sequences are indexed by some set $\mathcal{I}$ of size k_n and in a realization $e^{(n)}$ of $\mathcal{E}^{(n)}$ the sequence with index i is denoted by $v^n(e^{(n)}, i)$. A memoryless channel $(\rho_{W|V=v})_{v\in\mathcal{V}}$ induces an output distribution defined as*

$$\Gamma_{\mathcal{E}^{(n)},I,V^n,W^n} := \Gamma_{\mathcal{E}^{(n)}} \cdot p^{\mathcal{U}}_{[k_n]} \cdot \mathbf{1}_{V^n=v^n(e^{(n)},i)} \cdot \prod_{t=1}^{n} \rho^{(w_t)}_{W|V=v_t} . \tag{3.164}$$

Assume that $I_\rho(V;W) < \infty$. If $R > \overline{I}_\rho(V;W)$ then

$$\mathbb{E}_{\mathcal{E}^{(n)}}\big[\|\Gamma_{W^n|\mathcal{E}^{(n)}} - \rho_W^{\otimes n}\|_{TV}\big] \underset{n\to\infty}{\longrightarrow} 0, \tag{3.165}$$

where ρ_W is the marginal of $\rho_{V,W} = \rho_V \cdot \rho_{W|V}$.

Note that the result is stated with the spectral sup-mutual information:

$$\overline{I}(V;W) = \inf\Big\{\tau :\ \rho_{V,W}^{\otimes n}\Big((1/n)\ i_{\rho_{V,W}^{\otimes n}}(V^n,W^n) > \tau\Big) \to 0\Big\},$$

where $i_{p_{V,W}^{\otimes n}}$ is as in Proposition 1.17. Since it is assumed that $I_\rho(V;W) < \infty$, and hence

$I_{\rho^{\otimes n}}(V;W) < \infty$ by the chain rule (Proposition 1.14), then by Proposition 1.17 $i_{p_{Z,V}}$ and $i_{p_{Z,V}^{\otimes n}}$ are well-defined and $i_{p_{Z,V}}$ is $p_{Z,V}$-integrable. Hence, the law of large numbers applies: by the multiplicativity of the Radon-Nikodym derivative for product distributions we have

$$\forall (z^n, v^n), \quad i_{p_{Z,V}^{\otimes n}}(z^n, v^n) = \sum_{t=1}^{n} i_{p_{Z,V}}(z_t, v_t),$$

thus $\overline{I}_\rho(V;W) = I_\rho(V,W)$.

Use of the lemma

We proceed as in [Cuf13] for the proof of our Claim 3.19. We first use Lemma 3.33 with $\rho = p$, $\mathcal{E}^{(n)} = \mathcal{C}^{(n)}$, $k_n = \lfloor 2^{n(R+\varepsilon)} \rfloor \times \lfloor 2^{nR'} \rfloor \times \lfloor 2^{nR_c} \rfloor$, $I = (M, M', J)$, $W^n = Y^n$ and R replaced by $R+\varepsilon+R'+R_c$. The corresponding output distribution is exactly $Q^{(1)}_{\mathcal{C}^{(n)},M,M',J,V^n,Y^n}$, by definition (3.54) of the latter. Indeed, from (3.54) distribution $Q^{(1)}$ satisfies $(\mathcal{C}^{(n)}, M, M', J) - V^n - Y^n$ and from (3.55) we have $Q^{(1)}_{V^n,Y^n} \equiv p_{V,Y}^{\otimes n}$. Note that, from Lemma 1.7 (Appendix 1), the expectation in (3.165) does not depend on the choice of transition kernel $\Gamma_{W^n|\mathcal{E}^{(n)}}$. Thus, from (3.53) and Lemma 3.33 together with the realism property $p_Y \equiv p_X$ from (3.13) we obtain (3.57).

Second, we use Lemma 3.33 as follows. We fix a positive integer j. For every n such that $2^{nR_c} \geq j$, we set $\mathcal{E}^{(n)}$ to be the sub-codebook of $\mathcal{C}^{(n)}$ corresponding to j, denoted $\mathcal{C}_j^{(n)}$, and set $\rho = p$, $k_n = \lfloor 2^{n(R+\varepsilon)} \rfloor \times \lfloor 2^{nR'} \rfloor$. We further set $I = (M, M', j)$, $W = (X, Z)$ and replace R in the lemma by $R+\varepsilon+R'$. The corresponding output distribution is exactly $Q^{(1)}_{\mathcal{C}_j^{(n)},M,M',J,V^n,X^n,Z^n|J=j}$, by definition (3.54) of the latter. Indeed, from (3.54) distribution $Q^{(1)}$ satisfies $(\mathcal{C}^{(n)}, M, M', J) - V^n - (X^n, Z^n)$ and from (3.55) we have $Q^{(1)}_{V^n,X^n,Z^n} \equiv p_{V,X,Z}^{\otimes n}$. Thus, from (3.52) and Lemma 3.33 we have

$$\forall j, \quad \mathbb{E}_{\mathcal{C}_j^{(n)}}\left[\| Q^{(1)}_{X^n,Z^n|\mathcal{C}_j^{(n)},J=j} - p_{X,Z}^{\otimes n} \|_{TV}\right] \underset{n\to\infty}{\longrightarrow} 0. \tag{3.166}$$

Moreover, since $Q^{(1)}_{\mathcal{C}^{(n)},J} \equiv \mathbb{Q}_{\mathcal{C}^{(n)}} p^{\mathcal{U}}_{[2^{nR_c}]}$, then by Lemma 1.7 (Appendix 1) the TVD between

$Q^{(1)}_{\mathcal{C}^{(n)},J,X^n,Z^n}$ and $\mathbb{Q}_{\mathcal{C}^{(n)}} p^{\mathcal{U}}_{[2^{nR_c}]} p^{\otimes n}_{X,Z}$ rewrites as

$$\sum_{j=1}^{\lfloor 2^{nR_c} \rfloor} \frac{1}{\lfloor 2^{nR_c} \rfloor} \mathbb{E}_{\mathcal{C}_j^{(n)}} \left[\| Q^{(1)}_{X^n,Z^n|\mathcal{C}_j^{(n)},J=j} - p^{\otimes n}_{X,Z} \|_{TV} \right].$$

Finally, by construction of $\mathcal{C}^{(n)}$ the distribution of sub-codebook $\mathcal{C}_j^{(n)} = (V^n(\mathcal{C}^{(n)}, m, m', j))_{m,m'}$ is independent of j. Hence, all expectations in the above sum are identical. Using (3.166) we obtain (3.58).

3.C.2 Use of the soft covering lemma when the side information is available at both encoder and decoder

Statement

Lemma 3.34 *[Cuf13, Corollary VII.5] Let $\mathcal{U}, \mathcal{V}$ be alphabets, with $\mathcal{U}$ discrete, and $\rho_{U,V}$ a distribution on $\mathcal{U} \times \mathcal{V}$. Let R be a positive real number and let $(k_n)_{n\geq 1}$ be a sequence of positive integers satisfying $k_n \underset{n\to\infty}{\sim} 2^{nR}$. For every positive integer n and every sequence u^n in $\mathcal{U}^n$, let $\mathcal{E}_{u^n}$ be a randomly generated collection of k_n mutually independent sequences in $\mathcal{V}^n$, each drawn according to $\prod_{t=1}^n \rho_{V|U=u_t}$. The sequences are indexed by some set $\mathcal{I}$ of size k_n and in a realization e_{u^n} of $\mathcal{E}_{u^n}$, the sequence with index i is denoted by $v^n(e_{u^n}, i)$. We also assume that for every n, the collections $\mathcal{E}_{u^n}$ are mutually independent. We denote this random family by $\mathcal{E}_{\mathcal{U}^n}$ and a realization by $e_{\mathcal{U}^n}$. Its distribution is denoted $\Gamma_{\mathcal{E}_{\mathcal{U}^n}}$. A memoryless channel $(\rho_{W|U=u,V=v})_{u\in\mathcal{U},v\in\mathcal{V}}$ induces an output distribution defined as*

$$\Gamma_{\mathcal{E}_{\mathcal{U}^n},I,U^n,V^n,W^n} := \Gamma_{\mathcal{E}_{\mathcal{U}^n}} \cdot p^{\mathcal{U}}_{[k_n]} \cdot \rho_U^{\otimes n} \cdot \mathbf{1}_{V^n = v^n(e_{u^n},i)} \cdot \prod_{t=1}^n \rho_{W|U=u_t,V=v_t} .$$

If $R + H_\rho(U) > I_\rho(U,V;W)$ then

$$\mathbb{E}_{\mathcal{E}_{\mathcal{U}^n}} \left[\| \Gamma_{W^n|\mathcal{E}_{\mathcal{U}^n}} - \rho_W^{\otimes n} \|_{TV} \right] \underset{n\to\infty}{\longrightarrow} 0,$$

where ρ_W is the marginal of $\rho_{U,V,W} = \rho_{U,V} \cdot \rho_{W|U,V}$.

Use of the lemma

We proceed as in the case of marginal realism (Appendix 3.C.1) for the proof of our Claim 3.18. First, we use Lemma 3.34 with $\mathcal{U} = \mathcal{Z}$, $\mathcal{E}^{(n)}_{\mathcal{U}^n} = \mathcal{C}^{(n)}_{\mathcal{Z}^n}$, $k_n = \lfloor 2^{n(R+\varepsilon)} \rfloor \times \lfloor 2^{nR_c} \rfloor$, $I = (M, J)$, $\mathcal{W} = \mathcal{X}$, $\rho_{U,V,W} = p_{Z,V,Y}$, and R replaced by $R + \varepsilon + R_c$. The corresponding output distribution $\Gamma_{\mathcal{E}_{\mathcal{U}^n}, I, U^n, V^n, W^n}$ is exactly $Q^{(1)}_{\mathcal{C}^{(n)}_{\mathcal{Z}^n}, M, J, Z^n, V^n, Y^n}$, by definition (3.33) of the latter. Indeed, from (3.33) distribution $Q^{(1)}$ satisfies $(\mathcal{C}^{(n)}_{\mathcal{Z}^n}, M, J) - V^n - Y^n$ and from (3.34) we have $Q^{(1)}_{V^n,Y^n} \equiv p^{\otimes n}_{V,Y}$. Thus, from (3.28), (3.30) and Lemma 3.34 together with the realism property $p_Y \equiv p_X$ from (3.13) we obtain (3.36).

Second, we use Lemma 3.34 as follows. We fix a positive integer j. For every n such that $2^{nR_c} \geq j$ we set $\mathcal{E}^{(n)}_{\mathcal{U}^n}$ to be the sub-codebook of $\mathcal{C}^{(n)}_{\mathcal{Z}^n}$ corresponding to j, denoted $\mathcal{C}^{(n)}_{\mathcal{Z}^n,j}$, and set $\rho = p$, $k_n = \lfloor 2^{n(R+\varepsilon)} \rfloor$. We further set $I = (M, j)$, $W = (X, Z)$. The corresponding output distribution $\Gamma_{\mathcal{E}_{\mathcal{U}^n}, I, U^n, V^n, W^n}$ is exactly $Q^{(1)}_{\mathcal{C}^{(n)}_{\mathcal{Z}^n,j}, M, J, V^n, X^n, Z^n | J=j}$, thus from (3.27), (3.32) and Lemma 3.34 we have

$$\forall j, \quad \mathbb{E}_{\mathcal{C}^{(n)}_{\mathcal{Z}^n,j}} \left[\| Q^{(1)}_{X^n,Z^n | \mathcal{C}^{(n)}_{\mathcal{Z}^n,j}, J=j} - p^{\otimes n}_{X,Z} \|_{TV} \right] \underset{n\to\infty}{\longrightarrow} 0. \tag{3.167}$$

Moreover, since $Q^{(1)}_{\mathcal{C}^{(n)}_{\mathcal{Z}^n}, J} \equiv \mathbb{Q}_{\mathcal{C}^{(n)}_{\mathcal{Z}^n}} p^{\mathcal{U}}_{[2^{nR_c}]}$, then by Lemma 1.7 (Appendix 1) the TVD between $Q^{(1)}_{\mathcal{C}^{(n)}_{\mathcal{Z}^n}, J, X^n, Z^n}$ and $\mathbb{Q}_{\mathcal{C}^{(n)}_{\mathcal{Z}^n}} p^{\mathcal{U}}_{[2^{nR_c}]} p^{\otimes n}_{X,Z}$ rewrites as

$$\sum_{j=1}^{\lfloor 2^{nR_c} \rfloor} \frac{1}{\lfloor 2^{nR_c} \rfloor} \mathbb{E}_{\mathcal{C}^{(n)}_{\mathcal{Z}^n,j}} \left[\| Q^{(1)}_{X^n,Z^n | \mathcal{C}^{(n)}_{\mathcal{Z}^n,j}, J=j} - p^{\otimes n}_{X,Z} \|_{TV} \right].$$

Finally, by construction of $\mathcal{C}^{(n)}_{\mathcal{Z}^n}$ the distribution of sub-codebook $\mathcal{C}^{(n)}_{\mathcal{Z}^n,j} = (V^n(\mathcal{C}^{(n)}_{\mathcal{Z}^n}, m, j))_m$ is independent of j. Hence, all expectations in the above sum are identical. Using (3.167) we obtain (3.37).

3.C.3 Computations in the Gaussian case

Proof of Claim 3.23

$$\begin{aligned}
\mathbb{E}[(X-Y)^2] &= \mathbb{E}[(X-\rho V+\rho V-\tilde{\rho}V+\tilde{\rho}V-Y)^2] \\
&= \mathbb{E}[(X-\rho V)^2]+\mathbb{E}[(\rho V-\tilde{\rho}V)^2]+\mathbb{E}[(\tilde{\rho}V-Y)^2] \\
&= 1-\rho^2+(\rho-\tilde{\rho})^2+1-\tilde{\rho}^2 \\
&= 2-2\rho\tilde{\rho} \\
&= 2-2\rho\sqrt{1-2^{-2R_c}(1-\rho^2)} \\
&= \Delta
\end{aligned}$$

$$\begin{aligned}
I(X;V) &= \frac{1}{2}\log\frac{1}{1-\rho^2} \\
I(Y;V) &= \frac{1}{2}\log\frac{1}{1-\tilde{\rho}^2} = \frac{1}{2}\log\frac{1}{2^{-2R_c}(1-\rho^2)} = R_c+\frac{1}{2}\log\frac{1}{1-\rho^2}.
\end{aligned}$$

Proof of Claim 3.24

Although the alphabet $\mathcal{V}$ of V is finite — from the definition of $\mathcal{D}^{(m)}$ (Eq. 3.7) —, we provide a general proof for a Polish alphabet $\mathcal{V}$, to which we refer in other parts of the chapter. Sufficient conditions are that $I(X;V)<\infty$ and p_X is a one-dimensional Gaussian distribution. Since $I(X;V)<\infty$, then from Proposition 1.17, $p_X \times p_V$ dominates $p_{X,V}$. Therefore, from Corollary 1.6, there exists a regular transition kernel $p_{X|V}$ such that for every $v \in \mathcal{V}$, p_X dominates $p_{X|V=v}$, and $(x,v) \mapsto dp_{X|V=v}/dp_X(x)$ is measurable. Thus, for any v, the Lebesgue measure μ dominates $p_{X|V=v}$, and $(x,v) \mapsto dp_{X|V=v}/d\mu(x)$ is measurable. Hence, from Theorem 1.1 for the joint distribution of $(V, X-\mathbb{E}[X|V])$, the distribution of $X-\mathbb{E}[X|V]$ has a density with respect to the Lebesgue measure. Therefore, from Proposition 1.17, and since the density of

the Gaussian distribution p_X is non-zero everywhere, we have

$$\begin{aligned} R+\varepsilon \geq I(X;V) &= h(X) - \mathbb{E}_{v\sim p_V}[h(p_{X|V=v})] \\ &= h(X) - \mathbb{E}_{v\sim p_V}[h(p_{X-E[X|V]|V=v})] \\ &\geq \frac{1}{2}\log(2\pi e) - \mathbb{E}\Big[\frac{1}{2}\log\Big(2\pi e \mathbb{E}\big[(X-\mathbb{E}[X|V])^2\big|V\big]\Big)\Big] \\ &\geq \frac{1}{2}\log\frac{1}{\mathbb{E}[(X-\mathbb{E}[X|V])^2]} \qquad (3.168) \\ &= \frac{1}{2}\log\frac{1}{1-\mathbb{E}[\mathbb{E}[X|V]^2]} \\ &= \frac{1}{2}\log\frac{1}{1-\rho^2}, \end{aligned}$$

where (3.168) follows from Jensen's inequality for $-\log$. The second inequality of Claim 3.24 is obtained similarly.

Proof of Claim 3.25

$$\begin{aligned} \Delta+\varepsilon \geq \mathbb{E}[(X-Y)^2] &= \mathbb{E}[(X-\mathbb{E}[X|V])^2] + \mathbb{E}[(\mathbb{E}[X|V]-\mathbb{E}[Y|V])^2] + \mathbb{E}[(\mathbb{E}[Y|V]-Y)^2] \\ &\geq (1-\rho^2) + (1-\tilde{\rho}^2) + \Big(\sqrt{\mathbb{E}[\mathbb{E}[X|V]^2]} - \sqrt{\mathbb{E}[\mathbb{E}[Y|V]^2]}\Big)^2 \\ &= (1-\rho^2) + (1-\tilde{\rho}^2) + (\rho-\tilde{\rho})^2 \\ &= 2 - 2\rho\tilde{\rho}, \end{aligned}$$

where the second inequality follows from Cauchy-Schwarz.

Proof of Claim 3.26

Define the function

$$\phi : [0,1) \to [0,1),\ s \mapsto s\sqrt{1-2^{-2R_c}(1-s^2)}, \qquad (3.169)$$

which is strictly increasing and bijective. Thus, its inverse ϕ^{-1} is continuous. Since $\Delta \in (0,2)$ and we have chosen ε such that $\Delta + \varepsilon \in (0,2)$, we can define $\rho^* = \phi^{-1}(1-\Delta/2)$ and $\rho_\varepsilon^* = \phi^{-1}(1-(\Delta+\varepsilon)/2)$. Since $(\rho_\varepsilon, \tilde{\rho}_\varepsilon)$ satisfies the constraint of (3.131) and the equality (3.132), we have $\phi(\rho_\varepsilon^*) \leq \phi(\rho_\varepsilon)$, which implies $\rho_\varepsilon \geq \rho_\varepsilon^*$. Therefore, from (3.132), we have

$$R + 3\varepsilon \geq \frac{1}{2}\log\frac{1}{1-(\rho_\varepsilon^*)^2}.$$

This being true for arbitrarily small ε, then by continuity of ϕ^{-1}, we obtain the desired inequality:

$$R \geq \frac{1}{2}\log\frac{1}{1-(\rho^*)^2}.$$

Proof of Claim 3.28

Since $\rho \geq |\eta|$, we can choose

$$b = \sqrt{(\rho^2-\eta^2)/(1+\eta^2\rho^2-2\eta^2)}. \tag{3.170}$$

From (3.139), vector (Z,X,V) is Gaussian and we have $(X,Z) \sim p_{X,Z}$, the chain $Z - X - V$ and $I(Z;V) < \infty$ (since $|\eta b| < 1$). Then, Y has a Gaussian distribution with zero mean and we have $X - (Z,V) - Y$. Hence, $p_{X,Z,V,Z} \in \mathcal{D}_D^{(m)}$. To finish the proof of Claim 3.28, it only remains to compute the variance of Y, i.e. to prove (3.141).

Lemma 3.35 *Let $(\vec{x}_1, \vec{x}_2)$ be a Gaussian vector with mean $(\vec{\mu}_1, \vec{\mu}_2)$ and covariance matrix*

$$\begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{12} \end{pmatrix}$$

Then, conditioned on $\vec{x}_2 = \vec{a}$, variable $\vec{x}_1$ is Gaussian with:

$$\overline{\mu}_{|\vec{a}} = \vec{\mu}_1 + \Sigma_{12}\Sigma_{22}^{-1}(\vec{a} - \vec{\mu}_2) \tag{3.171}$$

$$\overline{\Sigma} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}, \tag{3.172}$$

with the covariance matrix $\overline{\Sigma}$ not depending on $\vec{a}$.

For the centered Gaussian vector (Z, X, V), we obtain

$$\mathrm{Var}(X|Z) = 1 - \eta \cdot 1 \cdot \eta = 1 - \eta^2 \text{ and} \tag{3.173}$$

$$\mathbb{E}[X|Z,V] = 0 + \begin{pmatrix} \eta & b \end{pmatrix} \begin{pmatrix} \frac{1}{1-\eta^2 b^2} & \frac{-\eta b}{1-\eta^2 b^2} \\ \frac{-\eta b}{1-\eta^2 b^2} & \frac{1}{1-\eta^2 b^2} \end{pmatrix} \begin{pmatrix} Z \\ V \end{pmatrix}. \tag{3.174}$$

From (3.139), $\mathbb{E}[Z^2] = \mathbb{E}[V^2] = 1$ and $\mathbb{E}[ZV] = \eta b$. Therefore, from (3.174) we have

$$\begin{aligned} \mathbb{E}[\mathbb{E}[X|Z,V]^2] &= \frac{1}{(1-\eta^2 b^2)^2} \mathbb{E}\big[[\eta(1-b^2)Z + b(1-\eta^2)V]^2\big] \\ &= \frac{1}{(1-\eta^2 b^2)^2} \mathbb{E}\big[\eta^2(1-b^2)^2 + b^2(1-\eta^2)^2 + 2\eta^2 b^2(1-\eta^2)(1-b^2)\big] \end{aligned} \tag{3.175}$$

Define $\lambda = 1 - \eta^2$ and $\tilde{\rho} = 1 - \rho^2$. Then, from (3.170) we have

$$(1 + \eta^2\rho^2 - 2\eta^2)b^2 = \lambda - \tilde{\rho}, (1 + \eta^2\rho^2 - 2\eta^2)(1 - b^2) = \tilde{\rho}\lambda,$$

$$(1 + \eta^2\rho^2 - 2\eta^2)(1 - \eta^2 b^2) = \lambda^2, 1 + \eta^2\rho^2 - 2\eta^2 = -\tilde{\rho} + \lambda(1 + \tilde{\rho}).$$

Therefore, by multiplying numerator and denominator by $(1+\eta^2\rho^2-2\eta^2)^2$ in (3.175) we obtain

$$\begin{aligned} &\mathbb{E}[\mathbb{E}[X|Z,V]^2] \\ &= \frac{1}{\lambda^4}\Big[(1-\lambda)\tilde{\rho}^2\lambda^2 + (\lambda - \tilde{\rho})\big[-\tilde{\rho} + \lambda(1+\tilde{\rho})\big]\lambda^2 + 2(1-\lambda)(\lambda-\tilde{\rho})\lambda\tilde{\rho}\lambda\Big] \\ &= \frac{1}{\lambda^2}\Big[(1-\lambda)\tilde{\rho}^2 + \big[\tilde{\rho}^2 - \lambda\tilde{\rho}(2+\tilde{\rho}) + \lambda^2(1+\tilde{\rho})\big] + 2\tilde{\rho}\big[-\tilde{\rho} + (1+\tilde{\rho})\lambda - \lambda^2\big]\Big] \\ &= \frac{1}{\lambda^2}\Big[1\cdot(\tilde{\rho}^2 + \tilde{\rho}^2 - 2\tilde{\rho}^2) + \lambda\cdot\big[-\tilde{\rho}^2 - \tilde{\rho}(2+\tilde{\rho}) + 2\tilde{\rho}(1+\tilde{\rho})\big] + \lambda^2\cdot\big[1 + \tilde{\rho} - 2\tilde{\rho}\big]\Big] \\ &= 1 - \tilde{\rho} = \rho^2. \end{aligned} \tag{3.176}$$

This concludes the proof of (3.141) and Claim 3.28.

Proof of Claim 3.29

We compute $I(X;V|Z)$ and $\mathbb{E}[d(X,Y)]$. By translation invariance of differential entropy we have

$$h(X|Z,V) = h\big(X - \mathbb{E}[X|Z,V]\big|Z,V\big).$$

Since uncorrelated Gaussian vectors are independent, we obtain $h(X|Z,V) = h(X - \mathbb{E}[X|Z,V])$. Since $\mathbb{E}[X^2] = 1$ we obtain

$$\mathrm{Var}\big(X - \mathbb{E}[X|Z,V]\big) = 1 - \mathbb{E}[\mathbb{E}[X|Z,V]^2].$$

The mutual information computation is concluded by using (3.173) and (3.176). Also,

$$\begin{aligned}\mathbb{E}[d(X,Y)] = \mathbb{E}[(X - \rho^{-1}\mathbb{E}[X|Z,V])^2] &= 1 - 2\rho^{-1}\mathbb{E}\big[\mathbb{E}[X|Z,V]^2\big] + \rho^{-2}\mathbb{E}\big[\mathbb{E}[X|Z,V]^2\big] \\ &= 1 - 2\rho + 1 \\ &= \Delta. \end{aligned} \tag{3.177}$$

Finally, we show that $I(Y;V) < \infty$ when $\eta \neq 0$. By construction, (Y,V) is a Gaussian vector. Since both Y and V have non-zero variance (equal to 1), we only need to prove that the covariance matrix of (Y,V) is non-singular, i.e. that $\mathbb{E}[YV] < 1$. By the Cauchy-Schwarz inequality, this is true if Y and V are not linearly dependent. From (3.174), Y is a linear combination of Z and V where the coefficient of Z is non-zero since $|\eta|, b \in (0,1)$. We conclude by noticing that, by construction, Z and V are not linearly dependent.

3.C.4 Markov chains in converse proofs

The following results will be used during mutual information computations.

Proposition 3.36 *Let A, B, C, D be subsets of $[n]$ such that $A \subseteq B$, $C \subseteq [n]\backslash A$ and $D \subseteq$*

$[n]\backslash B$. *If* P_{X^n,Z^n,M,J,Y^n} *is the distribution induced by a D-code, then*

$$Z_A \perp\!\!\!\perp (J, M, X_C, Z_D)|X_B.$$

In particular, one can choose $A = B = \{t\}$, $D = [n] \setminus \{t\}$ and $C = \emptyset$.

Proof: Since P is induced by a D-code we have $J \perp\!\!\!\perp (X^n, Z^n)$ and $Z^n \perp\!\!\!\perp M|J, X^n$. These imply inequalities between Shannon quantities with respect to P as follows:

$$\begin{aligned}
I(Z_A; M, J, X_C, Z_D|X_B) &= I(Z_A; Z_D) + I(Z_A; J|X_B, Z_D) + I(Z_A; M, X_C|J, X_B, Z_D) \\
&\leq 0 + I(Z_A, X_B, Z_D; J) + I(Z_A; M, X_{[n]\backslash A}|J, X_B, Z_D) \\
&\leq 0 + I(Z_A; X_{[n]\backslash A}|J, X_B, Z_D) + I(Z_A; M|J, X_B, X_{[n]\backslash A}, Z_D) \\
&\leq I(Z_A, Z_{D\cap A}, X_{B\cap A}; X_{[n]\backslash A}, X_{B\cap([n]\backslash A)}, Z_{D\cap([n]\backslash A)}|J) \\
&\quad + I(Z_A, Z_D; M|J, X^n) \\
&= 0.
\end{aligned}$$

■

Further similar conditional independence relations can be derived using the following lemma:

Lemma 3.37 *For random variables* K, L, W *we have*

$$W \perp\!\!\!\perp (K, L) \text{ if and only if } W \perp\!\!\!\perp K|L \text{ and } W \perp\!\!\!\perp L.$$

Using standard techniques with a uniform index, Proposition 3.36 implies the following:

Corollary 3.38 *Let If* P_{X^n,Z^n,M,J,Y^n} *is the distribution induced by a D-code and* T *be uniformly distributed on* $[n]$ *and independent of* (X^n, Z^n, M, J, Y^n). *Let* $(B(t))_{1\leq t\leq n}, (C(t))_{1\leq t\leq n}$ *be families of subsets of* $[n]$ *such that for all* $t \in [n]$ *we have* $t \notin B(t)$ *and* $t \notin C(t)$. *Then,*

$$Z_T \perp\!\!\!\perp (M, J, T, X_{B(T)}, Z_{C(T)})|X_T.$$

In particular, one can choose $B(t) = \emptyset$, and $C(t) = [n] \setminus \{t\}$.

As a direct implication of Markov chain $X^n - (M, J, Z^n) - Y^n$ we obtain:

Corollary 3.39 *Let P_{X^n,Z^n,M,J,Y^n} be the distribution induced by a D-code or an E-D-code and T be uniformly distributed on $[n]$ and independent of (X^n, Z^n, M, J, Y^n). For any $t \in [n]$ let $V_t = (M, J, t, Z_{[n]\setminus\{t\}})$. Then,*

$$Y_T \perp\!\!\!\perp X_T | V_T, Z_T.$$

Chapter 4

The Rate-Distortion-Perception Trade-Off with Algorithmic Realism

4.1 Introduction

In the RDP theory literature, the two main kinds of asymptotic realism constraints that have been considered (in conjunction with a traditional additive distortion constraint) are the ones studied in Chapter 2, namely strong realism and per-symbol realism. As mentioned in Chapter 2, for an i.i.d. source, strong realism imposes

$$\limsup_{n\to\infty} \mathcal{D}(P_{Y_{1:n}}^{(n)}, p_X^{\otimes n}) \leq \lambda, \tag{4.1}$$

and per-symbol near-perfect realism imposes

$$\limsup_{n\to\infty} \sup_{1\leq t\leq n} \mathcal{D}(P_{Y_t}^{(n)}, p_X) \leq \lambda, \tag{4.2}$$

where λ is a positive real and $\mathcal{D}$ is a divergence between distributions. Moreover, in this framework, ultimate realism corresponds to the case $\lambda = 0$, where the above constraints become

$$\mathcal{D}(P_{Y_{1:n}}^{(n)}, p_X^{\otimes n}) \underset{n\to\infty}{\longrightarrow} 0, \tag{4.3}$$

and

$$\sup_{1\le t\le n} \mathcal{D}(P_{Y_t}^{(n)}, p_X) \underset{n\to\infty}{\longrightarrow} 0. \tag{4.4}$$

As mentioned in Section 2.1, imposing (4.3) or (4.4) in addition to the usual distortion constraint naturally gives rise — in the simplest scenario of unlimited randomness — to the RDP function [LKK11, SLY15b, SLY15a, TW21]:

$$R^{(1)}(\Delta) := \min_{\substack{p_{Y|X} \text{ s.t.} \\ p_X = p_Y, \\ \mathbb{E}_p[d(X,Y)] \le \Delta}} I_p(X;Y). \tag{4.5}$$

The general case ($\lambda \ge 0$) gives rise to a natural extension of the above RDP function [TW21, CYW+22]. The emerging consensus [TW21, ZQCK21, CYW+22, SPC+23] that realism constraints are captured by strong or per-symbol realism, with the resulting fundamental limit being the RDP function, has two significant limitations, which are the focus of this chapter.

First, the goal of a realism constraint is to ensure that the reconstructed image appears realistic to a human observer. Yet (4.1)-(4.4) are arguably mismatched to human perception, as already noted by Chen *et al.* [CYW+22, Remark 11]. Under each of (4.1)-(4.4), it is not meaningful to speak of whether an individual realization is realistic; realism is a property of the ensemble. While human vision can quickly identify unrealistic aspects of individual images, it is blind to many of the ensemble-level deviations that are central to (4.1)-(4.4). Humans would not be able to distinguish the ensemble of natural images from those selected at random from a very large but fixed set, for instance, nor would they be able to tell, by sight alone, if an ensemble of images is slightly more, or slightly less, compressible under a given algorithm such as JPEG than the ensemble of natural images.

Second, theoretical studies that adopt (4.3) as realism constraint indicate that large amounts of high-quality common randomness between the encoder and the decoder are required in order to achieve the RDP function [SLY15a].[1] Yet no work has experimentally observed the need for a

[1] It is conjectured that the same can be said regarding (4.1), although this was only proved [CYW+22, Remark 11] under the assumption that the reconstruction Y^n is i.i.d..

large amount of common randomness. Indeed, it would be surprising if a practical scheme ceased producing realistic images when it was de-randomized by replacing the common randomness with a fixed realization.

We argue in favor of replacing (4.1)-(4.4) with a realization-based realism constraint. Specifically, we consider a *critic* that assigns a realism deficiency (or atypicality) score to each reconstruction. We then require that the expected deficiency score is asymptotically bounded. This is more aligned with human perception and, as we shall see, it eliminates the need for common randomness for all practical purposes. Realization-based theoretical frameworks have been considered previously in [SLY15a, CYW+22], although the latter were restricted to comparing the empirical distribution of the source and the reconstruction, or merely requiring that the reconstruction lie in the support of the source distribution. Specifically, if $\hat{P}_{y_{1:n}}$ is the empirical distribution of $y_{1:n}$, then we could use $\mathcal{D}(\hat{P}_{y_{1:n}}, p_X)$ as the critic. This approach, as well as adopting (4.2) or (4.4), would fail to impose any constraint on the order of the symbols in the reconstruction sequence $y_{1:n}$. For instance, the reconstruction $101010\cdots$ would be deemed a realistic reconstruction if the source is Bernoulli(1/2), when in fact this realization is atypical in all but the weakest sense.

Inspired by Theis [The24], we impose a stronger constraint in which we require that every computable critic is bounded. Any reconstruction that meets such a constraint must be typical in a strong sense. It must asymptotically have the correct kth-order marginals for any fixed k (see Proposition 4.1 to follow). But it also must be consistent with other, more subtle limit theorems, such as the Erdős-Rényi law of runs and its generalizations: for an n-length i.i.d. sequence of Bernoulli(p) random variables, with n large, the length of the longest run of ones is within a small additive gap of $\log_{1/p} n$ with high probability [GSW86, GO80, Boy72]. The concentration around $\log_{1/p} n$ is sufficiently tight that it can be used to distinguish truly random bits from a human-generated sequence; the runs in the latter are inevitably too short [Sch90, R$\acute{\text{e}}$78]. A reconstruction that has small realism deficiency under every computable critic must have a longest run that is close to $\log_{1/p} n$. It must likewise satisfy every other limit theorem satisfied by the source distribution, including those not yet discovered. [2]

[2] Another realization-based theoretical formalism [QW24] was recently proposed as an alternative to (4.1)-

Although such a realism constraint is very strong, we show that it is essentially free: it is satisfied with high probability by a compression scheme constructed through random selection in the usual fashion. In particular, under this realism constraint, the RDP function in (4.5) can be achieved with deterministic codes. The idea is that each codeword is realistic according to a given critic because it is generated according to the source distribution. To argue that it is realistic according to all computable critics, we use the fact that the number of such critics is countable[3] and thus we can create a "universal" critic whose realism deficiency is a convex combination of all computable deficiencies. If the expected deficiency of this universal critic is asymptotically bounded then it must be bounded for each of the constituent critics.

Any such universal critic $\delta : \cup_{n\in\mathbb{N}} \mathcal{X}^n \to \mathbb{R}$ is itself incomputable and can be shown to be related to the Kolmogorov complexity $x_{1:n} \mapsto K(x_{1:n})$ (see Remark 4.11 to follow). Using the notion of Kolmogorov complexity in the formulation of an engineering problem inevitably raises the objection that $K(x_{1:n})$ is not computable [CT06, Sec. 14.7]. This is not a concern in the present application because it is the critic, not the compression scheme, that is called upon to compute $K(x_{1:n})$. By showing that a code can achieve the RDP function without common randomness while satisfying this most powerful of critics, we show that it can also satisfy any weaker critic, such as a human observer. This is akin to how proofs of information-theoretic security assume adversaries with unlimited computational power.

The realization-based approach to realism espoused in this chapter might appear to be of a fundamentally different nature from the divergence-based approach in (4.1)-(4.4). In fact, it is possible to simultaneously generalize the two by extending the framework to allow the critic to examine batches of reconstructions at a time [The24]. Specifically, when the blocklength is n, we assume that B_n source realizations are generated and passed through the encoder and decoder. The critic then examines the resulting B_n reconstructions. For $B_n = 1$, this

(4.4), and proposes to use a single computable critic — from a specific class of critics. Our formalism is broader since we require that every property of the i.i.d. source process that is computably testable be satisfied — in [QW24], the source is a doubly infinite stationary process. As demonstrated in Sections 4.2 and 4.3, our approach is theoretically principled, and the numerous examples we provide show that it has many benefits. [QW24] proposes quite an interesting critic construction, which has contributed to the design of a state-of-the-art codec [B$^+$25].

[3]This comes from the fact that the set of lower semi-computable functions (see Appendix 4.A for background on computability) is countable, these functions admitting an effective enumeration [LV19, pp. 269-270].

reduces to our original formulation, and no common randomness is needed to achieve the RDP function. On the other hand, if B_n grows very quickly with n, then the batch essentially reveals the n-length distribution of the reconstructions, and the realism constraint is equivalent to the divergence constraint in (4.3). In particular, common randomness is needed to achieve the RDP function. More generally, we show that if B_n grows subexponentially with n, then common randomness is not needed to achieve the RDP function while it is necessary if B_n grows exponentially with n with a sufficiently large rate. Characterizing the extent to which common randomness is needed in the intermediate regime between these two extremes is an interesting theoretical problem raised by this work. In the case of subexponential B_n, these findings shed light on why the need for common randomness has not been observed in practice: to satisfy a realization-based critic such as a human-observer, common randomness is only needed if the critic considers impractically large batches at a time.

In Section 4.2, we provide some background on the formalism for critics in algorithmic information theory. In Section 4.3, we introduce our new formalism for the RDP trade-off. In Section 4.4, we state our main results, namely Theorems 4.4, 4.5, and 4.6. All proofs are deferred to the Appendix.

4.2 Background on algorithmic information theory

The theory of p-critics and universal critics has recently been brought to the attention of the machine vision community via [The24]. We refer to it for a high-level and insightful presentation of the topic and its usefulness in diverse machine learning tasks (generative modeling, outlier detection). Relevant background on computability theory is provided in Appendix 4.A. Throughout the chapter, we assume that the source X follows a distribution p_X on a finite set $\mathcal{X}$. Algorithmic information theoretic notions are commonly defined for binary strings. These can be extended to strings of elements of $\mathcal{X}$ by fixing an injection from $\mathcal{X}$ to $\{0,1\}^s$, for some $s \in \mathbb{N}$. We assume that p_X is a computable function from $\mathcal{X}$ to $(0,1)$.

Definition 4.1 *Consider a finite set $\mathcal{X}$. Let p be a distribution on $\mathcal{X}$ such that $\forall x \in \mathcal{X}, p(x) >$*

0. *A p-critic is a function* $\delta : \mathcal{X} \to \mathbb{R}$, *such that*

$$\sum_{x \in \mathcal{X}} p(x) 2^{\delta(x)} \leq 1. \tag{4.6}$$

A $p^{\otimes *}$ *-critic is a function* $\delta : \cup_{n \in \mathbb{N}} \mathcal{X}^n \to \mathbb{R}$, *such that for every input dimension* $n \in \mathbb{N}$, *we have*

$$\sum_{x \in \mathcal{X}^n} p^{\otimes n}(x) 2^{\delta(x)} \leq 1. \tag{4.7}$$

The notion of $p^{\otimes *}$-critic in Definition 4.1 is used to study an asymptotic regime in Section 4.3.2. For any $n \in \mathbb{N}$, let $\mathbf{1}_{\mathcal{X}^n} : \cup_{n \in \mathbb{N}} \mathcal{X}^n \to \{0, 1\}$ be such that $\mathbf{1}_{\mathcal{X}^n}(x) = 1$ iff $x \in \mathcal{X}^n$. Note that for any probability distribution π on $\mathbb{N}$, the mixture $\tilde{p} := \sum_{n \in \mathbb{N}} \pi(n) \mathbf{1}_{\mathcal{X}^n} p^{\otimes n}$ is a probability measure. By multiplying (4.7) by $\pi(n)$, and summing over n, we obtain

$$\sum_{x \in \cup_{n \in \mathbb{N}} \mathcal{X}^n} \tilde{p}(x) 2^{\delta(x)} \leq 1. \tag{4.8}$$

Thus a $p^{\otimes *}$-critic can be viewed as a critic over $\mathcal{X}^*$.

A critic is essentially a log-likelihood ratio: given a p-critic δ, setting $q : x \mapsto p(x) 2^{\delta(x)}$ gives

$$\delta(x) = \log \left(\frac{q(x)}{p(x)} \right) \tag{4.9}$$

$$\text{and} \quad \sum_{x \in \mathcal{X}} q(x) \leq 1. \tag{4.10}$$

Conversely, given any sequence of sub-probability distributions $\{q_n\}$, with the n-th having the same support as $p_X^{\otimes n}$,

$$x_{1:n} \mapsto \log \frac{q_n(x_{1:n})}{p_X^{\otimes n}(x_{1:n})} \tag{4.11}$$

is a $p^{\otimes *}$-critic. Setting $q_n = q^{\otimes n}$ in (4.11) for some distribution q over $\mathcal{X}$ having the same

support as p yields the $p^{\otimes *}$-critic

$$\delta(x_{1:n}) = \log \frac{q^{\otimes n}(x_{1:n})}{p_X^{\otimes n}(x_{1:n})} = \sum_{i=1}^{n} \log \frac{q(x_i)}{p_X(x_i)}. \tag{4.12}$$

If $X_{1:n}$ is indeed i.i.d p_X, then $\delta(X)$ tends linearly to negative infinity, while if it is actually i.i.d q, then it tends linearly to infinity. Such a test detects "large" deviations in the empirical distribution of the source. With additional effort, we can construct critics that test for smaller-scale deviations in higher-order statistics.

Proposition 4.1 *Consider a finite set $\mathcal{X}$. Let q be a computable distribution on $\mathcal{X}$ such that $\forall x \in \mathcal{X}, q(x) > 0$. Let e_0 be a single symbol in $\mathcal{X}$, considered as a pattern of interest. For any $n \in \mathbb{N}$ and any $x_{1:n} \in \mathcal{X}^n$, let $S(x_{1:n})$ denote the number of occurrences of e_0 in $x_{1:n}$. Define map $\phi : \cup_{n\in\mathbb{N}} \mathcal{X}^n \to \mathbb{N}_{\geq 0}$ by, $\forall n \in \mathbb{N}, \forall x_{1:n} \in \mathcal{X}^n$,*

$$x_{1:n} \mapsto \left\lceil \log \left\lceil |S(x_{1:n}) - q(e_0)n| \;/\; \sqrt{n} \right\rceil \right\rceil \tag{4.13}$$

if $S(x_{1:n}) \neq q(e_0)n$, and $x_{1:n} \mapsto 0$ otherwise. Then, $\phi - 2\log(\phi + 3)$ is a computable $q^{\otimes *}$-critic.

The proof is provided in Appendix 4.G. If $X_{1:n}$ is indeed i.i.d. p_X, then this critic's output will remain bounded in expectation as $n \to \infty$. But if its distribution is such that the empirical fraction of e_0 symbols departs from $q(e_0)$ by, say, $1/n^{1/2-\epsilon}$, then the output will grow without bound.

The critic in (4.12) is in some sense testing whether a realization fails to satisfy the prediction of the law of large numbers. The critic in Proposition 4.1 is testing a form of the central limit theorem. More generally, any theorem showing that source realizations must satisfy a certain property with high probability can potentially be embodied in a critic. We illustrate this with the Erdős-Rényi theory of runs.

Proposition 4.2 *Suppose $\mathcal{X} = \{0, 1\}$ and p_X is Bernoulli(q), where $0 < q \leq 1/2$. Let R_n denote the length of the longest run of ones in $X_{1:n}$. Then there exists a computable, bounded*

sequence $\{a_n\}$*, such that*

$$\delta_{run}(x_{1:n}) = \log\bigl(1 + |R_n - \log_{1/q} n|\bigr) - a_n \tag{4.14}$$

is a computable $p^{\otimes *}$*-critic.*

Proof: Note that R_n is an $\mathbb{N}_{\geq 0}$-valued map of $x_{1:n}$, obtained by finitely many comparisons of symbols, hence is a computable map to $\mathbb{N}_{\geq 0}$ (Definition 4.6), and therefore also a computable map to $\mathbb{R}$ by Lemma 4.8(iii). The map $n \mapsto \log_{1/q} n$ is a computable map to $\mathbb{R}$ by Lemma 4.9(ii). Define the sequence

$$a_n = \log\Bigl(1 + \sqrt{\mathrm{Var}(R_n)} + |\mathbb{E}[R_n] - \log_{1/q} n|\Bigr). \tag{4.15}$$

For each n, the quantities $\mathbb{E}[R_n]$ and $\mathrm{Var}(R_n)$ are finite sums of products of the computable reals $\{p_X^{\otimes n}(x_{1:n})\}_{x_{1:n}\in\mathcal{X}^n}$ and of the integers $\{R_n(x_{1:n})\}_{x_{1:n}\in\mathcal{X}^n}$, hence are computable maps to $\mathbb{R}$ by Lemma 4.9(ii); a_n is then a computable map to $\mathbb{R}$ by the same lemma. We then have

$$\begin{aligned}
\mathbb{E}[2^{\delta_{\mathrm{run}}(X_{1:n})}] &= 2^{-a_n}\bigl(1 + \mathbb{E}[|R_n - \log_{1/p} n|]\bigr) && (4.16)\\
&\leq 2^{-a_n}\Bigl(1 + \sqrt{\mathrm{Var}(R_n)} + |\mathbb{E}[R_n] - \log_{1/p} n|\Bigr) && (4.17)\\
&\leq 1. && (4.18)
\end{aligned}$$

It only remains to show that $\{a_n\}$ is bounded, but this follows from the asymptotic characterization of $\mathbb{E}[R_n]$ and $\mathrm{Var}(R_n)$ due to Gordon, Shilling and Waterman [GSW86, Thm. 2], building on earlier results of Guibas and Odlyzko [GO80] and Boyd [Boy72]. ∎

If $X_{1:n}$ is i.i.d. Bernoulli(r), where $r \neq q$, then R_n will grow as $\log_{1/r} n$, so $\delta_{\mathrm{run}}(X_{1:n}) \to \infty$. Thus the critic in Proposition 4.2 captures atypicality associated with the longest run. As a more subtle example, suppose that that $X_{1:n}$ is generated according to the distribution in which bits are selected independently and uniformly except that if a run of length $\log\log n$ of either zeros or ones occurs, then the next symbol deterministically ends the run.[4] For any k, the k-th

[4]This process is a reasonable model for human-generated sequences of random bits [Sch90, R78́].

order empirical distribution of this process will eventually be close to that of k i.i.d. uniform bits. But this distribution's deviation from true Bernoulli trials will eventually be detected by δ_{run}.

Critics can also be obtained from lossless compressors. Let $f : \mathcal{X} \to \{0,1\}^*$ denote a prefix-free lossless compression algorithm, such as one of the various Lempel-Ziv algorithms [CT06, Sec. 13.4]. Then

$$\delta(x) = \log \frac{1}{p_X(x)} - |f(x)|, \tag{4.19}$$

where $|f(x)|$ is the length of $f(x)$, is a p_X-critic by Kraft's inequality [CT06, Thm. 5.2.1]. This critic measures the extent to which x is atypically difficult to compress under f. Similarly, if $K(x)$ denotes Kolmogorov complexity [CT06, Chapter 14], then

$$\delta(x) = \log \frac{1}{p_X(x)} - K(x), \tag{4.20}$$

would also give a valid p_X-critic [CT06, Lemma 14.3.1]. We will come back to this specific critic in Remarks 4.11 and 4.23 in the appendices — see also [The24, Eq. 8] and [LV19, Theorem 4.3.5].

Evidently the set of computable critics is quite expansive. Yet it is worth noting that it does not include every critic of interest. Fix a sequence of codes $\{(F_n, G_n)\}$ and let q_n denote the distribution of the reconstruction when the blocklength is n. Then define the $p^{\otimes *}$-critic

$$\delta_\star(x_{1:n}) = \log \frac{q_n(x_{1:n})}{p_X^{\otimes n}(x_{1:n})}, \tag{4.21}$$

for $x_{1:n}$ in $\mathcal{X}^n$. This is a critic that directly tests whether $x_{1:n}$ is more likely to be a source string or a reproduction. This critic is not computable in general because it requires implementing a countable family of algorithms, one for each blocklength. It is computable if the sequence of codes is such that q_n can be computed, however, and in general, $\delta_\star$ can also be approximated by a computable critic if we consider very large batches.

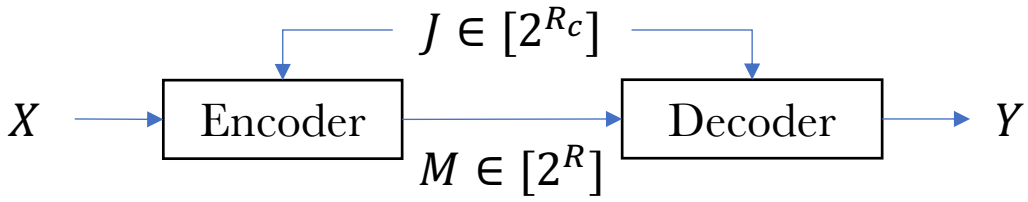


Figure 4.1: The system model for the one-shot setting.

Remark 4.3 *Definition 4.1 is a composite of two prominent definitions of a $p^{\otimes *}$-critic, [LV19, Definitions 4.3.8 & 2.4.1]. The former definition is precisely* (4.8)*, where weights $\{\pi(n)\}_{n\in\mathbb{N}}$ are assumed to be given. The latter definition ([LV19, Definition 2.4.1]) is:*

$$\forall n \in \mathbb{N}, \forall C \in \mathbb{N}_{\geq 0},\ p^{\otimes n}\big(\delta(X) \geq C\big) \leq 2^{-C}, \tag{4.22}$$

where the $p^{\otimes *}$-critic δ is required to take only non-negative integer values. The latter definition is more suitable for a Shannon-theoretic coding formulation, where each blocklength is viewed as a problem unto itself, while the former definition is more succinct and more convenient in certain respects [LV19, p. 220]. Definition 4.1 combines the blocklength-independence of [LV19, Definition 2.4.1] with the mathematical convenience of [LV19, Definition 4.3.8]. As such, replacing Definition 4.1 with [LV19, Definition 2.4.1] would not change the mathematical notion of realism which we adopt in Section 4.3.2, and thus not change our coding theorems.

4.3 New model for the rate-distortion-perception trade-off

The problem formulation in Chapter 2 is very similar to the one presented in this section. In the present chapter, we dedicate particular attention to the common randomness rate — similarly to Chapter 3 — and assume that private randomness is either unlimited or completely unavailable.

4.3.1 The one-shot setting

In this sub-section, we consider the one-shot scenario (blocklength $n = 1$), depicted in Figure 4.1. Given a coding scheme, each item in a batch of source samples is compressed separately, and realism is measured based on the resulting batch of reconstructions. The formal definition of a coding scheme for the general case ($n \geq 1$) is the one in Chapter 2: for ease of notation, we use the term (n, R, R_c) code instead of the term (n, R, R_c, ∞) fixed-length code (Definition 2.3), i.e., no restriction is imposed on (encoder and decoder) private randomness — similarly to Chapter 3. The definition in the case $n = 1$ is a simple special case, but is nonetheless stated as a standalone definition (Definition 4.2), in order to avoid confusion between batch size and blocklength: a given coding scheme is designed for a specific blocklength, and batches of samples are only considered when assessing performance. When the blocklength is 1, it is omitted in notation.

Definition 4.2 *Given non-negative reals R and R_c, an (R, R_c) code is a privately randomized encoder and decoder couple (F, G) consisting of a conditional distribution $F_{M|X,J}$ from $\mathcal{X} \times [2^{R_c}]$ to $[2^R]$, and a conditional distribution $G_{Y|M,J}$ from $[2^R] \times [2^{R_c}]$ to $\mathcal{X}$. The distribution*

$$P := p_X \cdot p^{\mathcal{U}}_{[2^{R_c}]} \cdot F_{M|X,J} \cdot G_{Y|M,J}$$

is called the distribution induced by the code. Moreover, such a code is said to be deterministic if $R_c = 0$ and mappings F, G are deterministic.

We propose a new RDP trade-off, formalized in the following two definitions.

Definition 4.3 *We extend d into an additive distortion measure on batches of elements of $\mathcal{X}$:*

$$\forall B \in \mathbb{N},\ \forall (\mathbf{x}^{(B)}, \mathbf{y}^{(B)}) \in \mathcal{X}^B \times \mathcal{X}^B, \quad d(\mathbf{x}^{(B)}, \mathbf{y}^{(B)}) := \tfrac{1}{B} \sum_{k=1}^{B} d(x^{(k)}, y^{(k)}).$$

Definition 4.4 *Consider a positive integer B, and a $p_X^{\otimes B}$-critic δ. A tuple (R, Δ, C) is said to be δ-achievable with algorithmic realism if there exists some $R_c \in \mathbb{R}_{\geq 0}$ and an (R, R_c) code such that the distribution P induced by the code satisfies*

$$\mathbb{E}_{P^{\otimes B}}\left[d(\mathbf{X}^{(B)}, \mathbf{Y}^{(B)})\right] \leq \Delta \quad \textit{and} \tag{4.23}$$

$$\mathbb{E}_{P^{\otimes B}}[\delta(\mathbf{Y}^{(B)})] \leq C, \tag{4.24}$$

where $\mathbf{X}^{(B)} \in \mathcal{X}^B$ denotes a batch of B i.i.d. source samples, and $\mathbf{Y}^{(B)}$ the batch of corresponding reconstructions produced by the code (with each source sample being compressed separately). If the code is deterministic, then we say that (R, Δ, C) is δ-achievable with a deterministic code.

The main difference with the original three-way RDP trade-off (Section 1.1 in Chapter 1) of [BM19] pertains to the realism constraint. In the latter formulation, the realism constraint is $\mathcal{D}(p_X, P_Y) \leq C$, where $\mathcal{D}$ is some divergence. Intuitively, that constraint corresponds to the special case of infinite batch size in the RDP trade-off proposed in Definition 4.4, since the discrete distributions p_X and P_Y can be approximated arbitrarily well using a large enough number of samples. In that sense, our proposed RDP framework generalizes the original one, through involving elements of practical realism metrics, such as the number B of samples which are inspected, and a scoring function δ which is required to be approximable via an algorithm. Theorem 4.6 to follow constitutes a rigorous statement of this intuition. In the next section, we define an asymptotic notion of achievability. We provide achievable points in the sense of Definition 4.4 in Section 4.4.2.

4.3.2 Asymptotic setting

As in previous chapters, we consider the compression of a source distributed according to $p_X^{\otimes n}$. As mentioned in the previous sub-section, the definition of a code is the one in Chapter 2: we write (n, R, R_c) code instead of (n, R, R_c, ∞) fixed-length code (Definition 2.3). We characterize the RDP trade-off in asymptotic settings where both n and the batch size go to infinity. The extension of d into an additive distortion measure on finite sequences, and batches of finite

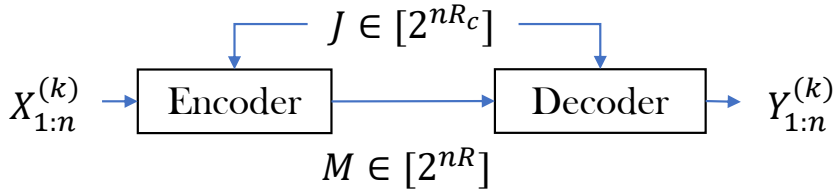


Figure 4.2: The system model for the asymptotic setting. Index k ranges from 1 to the batch size. The same encoder-decoder pair is used to process each source sample in the batch.

sequences, follows from Definition 4.3. The setup is depicted in Figure 4.2. Given a coding scheme, each item in a batch of source samples is compressed separately, and realism is measured based on the resulting batch of reconstructions. We define asymptotic achievability as follows. See Appendix 4.A for background on computability.

Definition 4.5

A quadruplet $(R, R_c, \{B_n\}_{n\geq 1}, \Delta)$ is said to be asymptotically achievable with algorithmic realism if for any $\varepsilon > 0$, there exists a sequence of codes $\{(F^{(n)}, G^{(n)})\}_n$, the n-th being $(n, R+\varepsilon, R_c)$, such that the sequence $\{P^{(n)}\}_n$ of distributions induced by the codes satisfies

$$\limsup_{n\to\infty} \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[d(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)}) \right] \leq \Delta + \varepsilon \tag{4.25}$$

(where $(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)})$ is a batch $\{(X_{1:n}^{(k)}, Y_{1:n}^{(k)})\}_{k\in[B_n]}$ of i.i.d. pairs, each distributed according to $P^{(n)}$), and such that for any lower semi-computable $p_X^{\otimes *}$-critic δ,

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[\delta(\mathbf{Y}^{(n,B_n)}) \right] < \infty. \tag{4.26}$$

We say that $(R, \{B_n\}_{n\geq 1}, \Delta)$ is achievable with a fully deterministic scheme if for each n, the code $(F^{(n)}, G^{(n)})$ is fully deterministic.

When considering the source distribution, then, by Jensen's inequality, the expected score given by a $p_X^{\otimes *}$-critic δ satisfies

$$\mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[\delta(\mathbf{X}^{(n,B_n)}) \right] \leq 0. \tag{4.27}$$

In contrast, in the examples in the previous section, we observed that when $\mathbf{Y}^{(n,B_n)}$ has a

different distribution from the source then we expect $\delta(\mathbf{Y}^{(n,B_n)})$ to grow unboundedly with n — see also standard algorithmic information theory literature, e.g., [LV19, Definition 2.4.3]. This motivates the choice of requiring that the expected realism discrepancy remain bounded with n, as opposed to requiring that it remain below some fixed threshold as in Def. 4.4. It is unclear *a priori* that even this nonspecific constraint can be satisfied at any rate below the entropy of the source since it must hold for every computable critic. We shall see, in fact, that one can satisfy this constraint with deterministic codes that achieve the RDP function. Crucial to this result is that the bound in (4.26) is not imposed uniformly over the critics.

4.4 Results

4.4.1 Low batch size regime

The following theorem states that the RDP function $R^{(1)}(\cdot)$ (defined in (4.5)) which naturaly arises in the distribution matching formalism, also characterizes the optimal trade-off in our asymptotic setting, when the batch size is not impractically large.

Theorem 4.4 *Consider a sequence $\{B_n\}_{n\geq 1}$ of positive integers such that*

$$\log(B_n)/n \underset{n\to\infty}{\longrightarrow} 0. \tag{4.28}$$

For any $\Delta \in \mathbb{R}_+$, let $R(\Delta)$ be the infimum of rates R such that there exists $R_c \in \mathbb{R}_{\geq 0}$ such that $(R, R_c, \{B_n\}_{n\geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism. Moreover, for any $\Delta \in \mathbb{R}_+$, let $R_*(\Delta)$ be the infimum of rates R such that $(R, \{B_n\}_{n\geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism with fully deterministic codes. Then, for any $\Delta \in \mathbb{R}_+$ such that $R^{(1)}(\Delta) < H_p(X)$, we have

$$R(\Delta) = R_*(\Delta) = R^{(1)}(\Delta). \tag{4.29}$$

The proof is provided in Appendices 4.C and 4.D. The strength of this result lies in how

stringent constraint (4.26) is: a single compression scheme satisfies a performance guarantee for every relevant $p_X^{\otimes *}$-critic, and deterministic schemes are sufficient. Moreover, one can find such a scheme for any batch size sequence which is sub-exponential in the dimension n of the source, i.e. for all regimes where the batch size is not impractically large. To prove the achievability direction of Theorem 4.4, we leverage the existence of a *universal* $p_X^{\otimes *}$-critic δ_0 (see Appendix 4.A.2), which is one of the great successes of algorithmic information theory. Indeed, it is sufficient to construct a scheme which achieves (4.26) only for such a δ_0, which is more sensitive than all relevant $p_X^{\otimes *}$-critics. It is a very strong critic, stronger than can be implemented in practice, which is another strength of Theorem 4.4.

4.4.2 One-shot achievable points

For theoretical interest, we provide a family of points which are achievable, in the sense of Definition 4.4, without any statistical assumption on the source distribution p_X. For the sake of gleaning intuition, one can consider the following example.

- $\mathcal{X}$ is a finite set of images, e.g. the set of all images of a given resolution, with a finite range for pixels (finite precision).
- d is the MSE between pixel values.
- B is the number of images inspected by the critic at a time.
- R_1 is the number of bits into which a given image is compressed.

Theorem 4.5 *Consider a finite set $\mathcal{X}$ such that $|\mathcal{X}| \geq 2$, a computable distribution p_X on $\mathcal{X}$ such that $\forall x \in \mathcal{X}, p_X(x)>0$, a positive integer B, some $R > \log(B)/\log(\mathcal{X})$, some $\Delta \in \mathbb{R}_+$, and a $p_X^{\otimes B}$-critic δ. Consider any conditional transition kernel $p_{Y|X}$ from $\mathcal{X}$ to $\mathcal{X}$ satisfying*

$$p_Y \equiv p_X, \quad \mathbb{E}_p[d(X,Y)] \leq \Delta. \tag{4.30}$$

Then, for any $\varepsilon \in (0, \Delta/2)$, *and any* $\gamma > 0$, *the triplet* (R_1, Δ_1, C_1) *is* δ*-achievable, with a* $(R_1, 0)$ *code, where*

$$R_1 := R \log(|\mathcal{X}|) \tag{4.31}$$

$$\Delta_1 := \Delta + \varepsilon + \frac{6\Delta}{\varepsilon} \max(d) \cdot \eta_{R,\gamma} \tag{4.32}$$

$$C_1 := \frac{3\Delta}{\varepsilon} \left[1 + \Big(\frac{B^2}{\lfloor 2^{R_1} \rfloor} + 2B\eta_{R,\gamma} \Big) \cdot \max_x B \log \frac{2}{p_X(x)} \right] \tag{4.33}$$

$$\eta_{R,\gamma} := p(\mathcal{A}_{R,\gamma}) + 2^{-\gamma \log(|\mathcal{X}|)/2} \tag{4.34}$$

$$\mathcal{A}_{R,\gamma} := \Big\{ (x,y) \in \mathcal{X}^2 \mid \log \Big(\frac{p_{X,Y}(x,y)}{p_X(x) p_Y(y)} \Big) - \log(\lfloor 2^{R_1} \rfloor) > -\gamma \log(|\mathcal{X}|) \Big\}, \tag{4.35}$$

with the convention $0/0 := 1$.

The proof is provided in Appendix 4.B. The term $B^2/\lfloor 2^{R_1} \rfloor$ is an upper bound on the probability that two source samples in the batch are compressed into the same message. This is related to the so-called *birthday paradox* (see Appendix 4.I). The term $\max_x B \log(1/p_X(x))$ is an upper bound on the output of δ, which follows from Definition 4.1.

Theorem 4.5 provides insights on the asymptotic regime of Theorem 4.4. Consider the limit of large $|\mathcal{X}|$, with fixed $R, \Delta, \varepsilon, \gamma$, and with $\log(B) = o(\log |\mathcal{X}|)$. We know that

$$\mathbb{E}_p \Big[\log \Big(\frac{p_{X,Y}(x,y)}{p_X(x) p_Y(y)} \Big) \Big] = I_p(X;Y). \tag{4.36}$$

Hence, if this log-likelihood ratio concentrates well, and if $R_1 > I_p(X;Y)$, as in the definition of $R^{(1)}(\cdot)$ in (4.5), then $p(\mathcal{A}_{R,\gamma})$ is small for small enough γ. In such an asymptotic regime, we obtain $\Delta_1 \approx \Delta$, and $C_1 = O(1)$. Therefore, the assumption in Theorem 4.4, that the source is of the form $p_X^{\otimes n}$ for some large n, is only used to ensure fast concentration of the log-likelihood ratio. Hence, Theorem 4.4 can be extended to a larger set of sources. In the next section, we present our last main result, which pertains to an asymptotic regime with large batch size.

4.4.3 Generalizing the distribution matching formalism

In this section, we present a result which connects our proposed formalism for the RDP trade-off to the distribution matching formalism — studied extensively in Chapter 2 —, and concludes our findings regarding the role of common randomness.

The definition of asymptotic achievability with near-perfect or perfect realism is the one in Chapter 2: as in Definitions 4.4 and 4.5, we write (n, R, R_c) code instead of (n, R, R_c, ∞) fixed-length code (Definition 2.3), i.e., no restriction is imposed on (encoder and decoder) private randomness. As stated in the theorem below, in a certain large batch size regime, asymptotic achievability with algorithmic realism (Definition 4.5) is equivalent to asymptotic achievability with near-perfect realism. The proof is provided in Appendix 4.E.

Theorem 4.6 *Consider a computable increasing sequence $\{B_n\}_{n\geq 1}$ of positive integers such that*

$$\frac{B_n}{|\mathcal{X}|^n} \to \infty. \tag{4.37}$$

Then, for any $R_c \in \mathbb{R}_{\geq 0}$, and any $(R, \Delta) \in (\mathbb{R}_+)^2$, tuple $(R, R_c, \{B_n\}_{n\geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism if and only if (R, R_c, Δ) is asymptotically achievable with near-perfect realism, if and only if (R, R_c, Δ) is asymptotically achievable with perfect realism.

Hence, Theorem 4.6, similarly to the finding in [The24], shows that for large batch size, our formalism is equivalent to the distribution matching formalism. Hence, the former is a generalization of the latter.

Remark 4.7 *Our proof of Theorem 4.6 implies that any code satisfying the algorithmic realism constraint* (4.26) *with batch sizes as in* (4.37) *must also satisfy the near-perfect strong realism constraint. We show that the converse holds when unlimited decoder private randomness is available — so that achievability with near-perfect strong realism implies achievability with perfect strong realism (Theorem 3.2 in Chapter 3 with singleton $\mathcal{Z}$), which in turn implies achievability with algorithmic realism.*

4.5 Conclusion

In this chapter, we have proposed an alternative to the realism constraints considered in the previous chapter, which are the most common in the literature, but have several limitations. We require that reconstructions asymptotically satisfy all computable critics of randomness. This includes critics inspecting the frequency of any pattern, but also critics inspecting subtle properties, such as ones pertaining to the Erdős-Rényi law of runs. Moreover, distribution matching constraints are equivalent to requiring that any hypothesis test for the distributions of the source and reconstruction have poor performance (see Section 2.3 in Chapter 2). Hence, this unnecessarily requires that the reconstructions satisfy incomputable critics — as well as computable ones. Theorem 4.4 states that common randomness does not improve the trade-off under our formalism, in all regimes where the batch size is not impractically large with respect to the dimension n of the source. This offers a compelling explanation for the fact that the need for large amounts of common randomness has not been observed in practice. Theorem 4.6 states that when the batch size is extremely large, our formalism is equivalent to the distribution matching formalism [SLY15a, Wag22, CYW+22], hence common randomness is useful in that regime. Thus, Theorems 4.4 and 4.6 indicate that, in order to understand the role of randomization in lossy compression with realism constraints, the focus should be shifted to the size of the batch inspected by the critic. A continuation of our work could be to investigate realism metrics, where particular attention would be given to the choice of the batch size. This could lead to highlighting specific strengths and weaknesses of existing realism metrics. It may also inspire a critical assessment of the relative performance of existing compression schemes, depending on the choice of realism metric. Another continuation could be to more precisely characterize the amount of randomness needed as a function of the batch size. Furthermore, possible extensions of our setup include compression with side information — similarly to Chapter 3 —-, and other distributed settings.

In the next chapter, we consider problems where information is to be ingeniously extracted from one or more sources to satisfy communication constraints (latency, bandwidth) while achieving coordination between distributed agents, which is naturally formalized as a near-

perfect realism constraint on joint distributions. As such, these problems have strong ties to lossy compression with realism constraints. Furthermore, as noted in Section 2.2, our proofs in Chapter 2 — and indeed Chapter 3 — have substantial similarities with proofs from the distributed coordination literature.

Appendix

4.A Further background on algorithmic information theory

4.A.1 Computability

This definition matches [LV19, Definition 1.7.4], except for the definition of a computable real number, which we adapted from [LV19, Exercise 1.7.22], and for the definition of a computable set, which matches that of [LV19, page 32].

Definition 4.6 *Consider a subset $\mathcal{E}$ of $\mathbb{N}_{\geq 0}$. A map f from $\mathcal{E}$ into $\mathbb{N}_{\geq 0}^3$ is said to be computable if there exists a Turing machine which, given as input any $x \in \mathcal{E}$, halts after finitely many steps and outputs the exact value $f(x)$ [LV19, Section 1.7.1]; its behavior on inputs outside $\mathcal{E}$ is immaterial. This notion extends to functions having as domain other common countable sets, such as $\mathbb{N}_{\geq 0}^k$ for $k \in \mathbb{N}$, and $\{0,1\}^*$, or any subset thereof, by identifying elements of these sets with non-negative integers via some reference bijections. Consider a computable map f from a subset $\mathcal{E}$ of $\mathbb{N}_{\geq 0}$ into $\{0,1\} \times \mathbb{N}_{\geq 0} \times \mathbb{N}$. Then, composing f with $(s,a,b) \mapsto (2s-1)a/b$ yields a map from $\mathcal{E}$ to $\mathbb{Q}$, which is said to be a computable map from $\mathcal{E}$ to $\mathbb{Q}$. A map f from a subset $\mathcal{E}$ of $\mathbb{N}_{\geq 0}$ into $\mathbb{R}$ is said to be lower semi-computable if there exists a computable function φ from $\mathcal{E} \times \mathbb{N}$ into $\mathbb{Q}$, such that*

$$\forall x \in \mathcal{E}, \varphi(x,k) \underset{k\to\infty}{\to} f(x), \qquad \text{and} \qquad \forall x \in \mathcal{E}, \forall k \in \mathbb{N}, \quad \varphi(x,k+1) \geq \varphi(x,k).$$

Moreover, f is said to be a computable map from $\mathcal{E}$ to $\mathbb{R}$ if both f and $-f$ are lower semi-computable. A real number λ is said to be computable if the constant function $f : \mathbb{N}_{\geq 0} \to \mathbb{R},\ n \mapsto \lambda$ is a computable function from $\mathbb{N}_{\geq 0}$ to $\mathbb{R}$. A (possibly infinite) subset $\mathcal{X}$ of $\mathbb{N}_{\geq 0}$, is said to be computable if there exists a computable function f from $\mathbb{N}_{\geq 0}$ to $\{0,1\}$, which returns 1 if its input is in $\mathcal{X}$, and 0 otherwise.

Definition 4.6 introduces two notions which should not be conflated. A map into $\mathbb{N}_{\geq 0}$ or $\mathbb{Q}$ is computable when a Turing machine returns its *exact* value in finitely many steps. A map into $\mathbb{R}$ is lower semi-computable, resp. computable, when it is *approximated* by a computable sequence of rationals, from below, resp. from below and from above. For a $\mathbb{Q}$-valued map, the former requirement is strictly stronger than the latter: computability as a $\mathbb{Q}$-valued map implies computability as an $\mathbb{R}$-valued map (Lemma 4.8(iii) below), but the converse fails. The following two lemmas allow to construct (semi-)computable functions. Proofs are deferred to Appendix 4.K.

Lemma 4.8 *Let $\mathcal{E}$ denote a non-empty subset of $\mathbb{N}_{\geq 0}$, and let f and g denote computable maps from $\mathcal{E}$ to $\mathbb{Q}$ (Definition 4.6).*

(i) The maps $f+g$, $f-g$, fg, $|f|$, $\max(f,g)$, $\min(f,g)$, $\lceil f \rceil$ and $\lfloor f \rfloor$ are computable maps from $\mathcal{E}$ to $\mathbb{Q}$. If g does not take the value 0, then so is f/g.

(ii) The maps $x \mapsto \mathbf{1}_{f(x)\geq g(x)}$ and $x \mapsto \mathbf{1}_{f(x)=g(x)}$ are computable maps from $\mathcal{E}$ to $\{0,1\}$. Consequently, if h is a computable map from $\mathcal{E}$ to $\mathbb{Q}$, then so is the map equal to f on $\{x : f(x) \geq g(x)\}$ and to h elsewhere.

(iii) Every computable map from $\mathcal{E}$ to $\mathbb{Q}$ is a computable map from $\mathcal{E}$ to $\mathbb{R}$, and in particular is lower semi-computable.

(iv) Let p and q denote positive integers, and assume f only takes non-negative values. Then $x \mapsto \lceil f(x)^{p/q} \rceil$ and $x \mapsto \lfloor f(x)^{p/q} \rfloor$ are computable maps from $\mathcal{E}$ to $\mathbb{N}_{\geq 0}$.

Lemma 4.9 *Let $\mathcal{E}$ denote a non-empty subset of $\mathbb{N}_{\geq 0}$, and let f and g denote functions from $\mathcal{E}$ to $\mathbb{R}$.*

(i) If f and g are both lower semi-computable, then functions $f+g$, $\lceil f \rceil$, and 2^f are lower semi-computable. If, in addition, f and g only take non-negative values, then fg and $2^f/(3+f)^2$ are lower semi-computable. If, in addition, f only takes positive values, then $\log(f)$ is lower semi-computable.

(ii) If f and g are both computable, then functions $f+g$, fg, and $|f|$ are computable. If, in addition, f only takes positive values, then functions $1/f$, $\log(f)$, and $f^{1/b}$ are computable, for any positive integer b.

(iii) Let $\mathcal{X}$ be a computable finite subset of $\{0,1\}^*$. If f is a lower semi-computable function from $\{0,1\}^*$ into $\mathbb{R}$, then the function $\tilde{f} : \{0,1\}^* \to \mathbb{R}$ which is null outside of $\cup_{n\in\mathbb{N}}\mathcal{X}^n$, and is defined by

$$\forall x \in \cup_{n\in\mathbb{N}}\mathcal{X}^n, \quad \tilde{f}(x) = \sum_{y\in\mathcal{X}^{l(x)}} f(y),$$

is lower semi-computable. Moreover, if p is a lower semi-computable probability measure on $\mathcal{X}$, then $p^{\otimes *}$ is lower semi-computable.

4.A.2 Universal critics and semi-measures

Definition 4.7 *Given a finite set $\mathcal{W}$, a function $f : \mathcal{W} \to [0,1]$ is a semi-measure if*

$$\sum_{w\in\mathcal{W}} f(w) \leq 1.$$

It is said to be a lower semi-computable semi-measure if f is a semi-measure and f is lower semi-computable.

The following theorem, corresponds to Definition 4.3.2, Equation (4.2), and Theorems 4.3.1 and 4.3.3 in [LV19]. It introduces the notion of *universal $p^{\otimes *}$-critic*, used in [The24].

Theorem 4.10 *Consider a finite set $\mathcal{X}$, each element of which is identified with a string in $\{0,1\}^s$, for some $s \in \mathbb{N}$. Let p be a computable distribution on $\mathcal{X}$ such that $\forall x \in \mathcal{X}, p(x) > 0$. Then, there exists a $p^{\otimes *}$-critic δ_0 (which is not necessarily lower semi-computable), such that*

for any lower semi-computable $p^{\otimes *}$-critic δ, there exists a constant c_δ such that

$$\forall x \in \bigcup_{n\in\mathbb{N}} \mathcal{X}^n, \quad \delta_0(x) \geq \delta(x) - c_\delta. \tag{4.38}$$

Any $p^{\otimes *}$-critic satisfying (4.38) *is called a universal $p^{\otimes *}$-critic.*

Since our definitions are slightly different from the classical ones — see Remark 4.3 —, we provide a proof of Theorem 4.10 in Appendix 4.J. See Remark 4.22 for a comparison of $p^{\otimes *}$-critics which are universal according to different formulations. Such a critic δ_0 is one of the best measures of realism deficiency according to p, in the limit of arbitrarily long strings. If a critic δ identifies a certain amount of deficiency in a given string, then δ_0 will identify at least as much deficiency, up to some additive constant. Intuitively, δ_0 is sensitive to all properties of randomness according to p. The existence of such a δ_0 constitutes a remarkable property of the set of all lower semi-computable $p_X^{\otimes *}$-critics (which is infinite).

Remark 4.11 *[LV19, Theorem 4.3.3] The proof of Theorem 4.10 (Appendix 4.J) involves a mixture* $\mathbf{m}$ *called a universal semi-measure. It can be used as a prior distribution, which has been shown to be relevant in machine learning applications involving realism, such as outlier detection and generative modeling [The24]. The universal semi-measure* $\mathbf{m}$ *can be chosen in such a way that*

$$\forall x \in \{0,1\}^*, \quad |-\log(\mathbf{m}(x)) - K(x)| \leq c, \tag{4.39}$$

for some constant c, where K is the Kolmogorov complexity [LV19, Section 3.1]. Property (4.39) *constitutes a strong result, since the Kolmogorov complexity is only defined up to a constant -we omit the corresponding details, for the purpose of clarity. The map $x \mapsto \log(1/p(x)) - K(x)$ is sometimes considered to be an approximation of a universal $p^{\otimes *}$-critic, see, e.g., [The24, Eq. 8], and see Appendix 4.J for a rigorous treatment.*

4.B Proof of Theorem 4.5

4.B.1 Outline

To show the achievability of a tuple (R_1, Δ_1, C_1), we consider a set of random reconstructions, and study its realism properties in Section 4.B.2. Then, we show the existence of a suitable choice of realizations of the latter reconstructions in Section 4.B.3. In Section 4.B.4, we prove Theorem 4.5 by proposing a compression scheme achieving a close-to-uniform sampling from the set of reconstructions. For the remainder of Section 4.B, we fix a finite set $\mathcal{X}$ such that $|\mathcal{X}| \geq 2$, a computable distribution p_X on $\mathcal{X}$ such that $\forall x \in \mathcal{X}, p_X(x)>0$, a positive integer B, and a $p_X^{\otimes B}$-critic δ.

4.B.2 Realism performance of a uniformly sampled batch of random reconstructions

Random candidate reconstructions

Given a positive real R_1, let $\mathcal{C}$ be a family of $\lfloor 2^{R_1} \rfloor$ i.i.d. variables, each sampled from p_X. The m-th variable is denoted $y(\mathcal{C}, m)$. We denote their joint distribution by $Q_{\mathcal{C}}$. Given a realization c of $\mathcal{C}$, we consider a batch $\mathbf{y}^{(B)}$ of B elements of c, sampled uniformly with replacement. Then, we compute the batch's realism score $\delta(\mathbf{y}^{(B)})$. This is formalized in the following lemma, which gives an upper bound of the expected score with respect to $Q_{\mathcal{C}}$.

Lemma 4.12 *Consider a positive real $R_1 \in (\log(B), \infty)$, and the following pmf.*

$$\begin{aligned} &Q_{\mathcal{C},\mathbf{M}^{(B)},\mathbf{Y}^{(B)}}\left(\{y(m')\}_{m' \in [\lfloor 2^{R_1} \rfloor]}, \mathbf{m}^{(B)}, \mathbf{y}^{(B)}\right) \\ &\quad := \Big(\prod_{m'=1}^{\lfloor 2^{R_1} \rfloor} p_X(y(m'))\Big) \cdot \frac{1}{\lfloor 2^{R_1} \rfloor^B} \cdot \prod_{k=1}^{B} \mathbf{1}_{y^{(k)}=y(m^{(k)})}. \end{aligned} \tag{4.40}$$

Then, we have

$$\mathbb{E}_Q[\delta_+(\mathbf{Y}^{(B)})] \leq 1 + \frac{B^2}{\lfloor 2^{R_1} \rfloor} \max_x B \log \frac{2}{p_X(x)}. \tag{4.41}$$

where

$$\delta_+(\mathbf{Y}^{(B)}) = \max(0, \delta(\mathbf{Y}^{(B)})). \tag{4.42}$$

The remainder of Section 4.B.2 is dedicated to the proof of Lemma 4.12. Fix $R_1 > \log(B)$.

Preliminaries

Claim 4.13 *For any distribution p on a finite set, any p-critic δ satisfies*

$$\mathbb{E}_p[2^{\delta_+(X)}] \leq 2 \tag{4.43}$$

$$\mathbb{E}_p[\delta_+(X)] \leq 1. \tag{4.44}$$

$$\max(\delta_+) \leq \max_x \log \frac{2}{p_X(x)}. \tag{4.45}$$

Proof: From (4.6), we have

$$1 \geq \mathbb{E}_p[2^{\delta(X)}].$$

Thus, we have

$$\begin{aligned}
1 &\geq \mathbb{E}_p[2^{\delta(X)}] && (4.46)\\
&\geq \mathbb{E}_p[2^{\delta_+(X)}\mathbf{1}_{\delta(X)\geq 0}] && (4.47)\\
&= \mathbb{E}_p[2^{\delta_+(X)}\mathbf{1}_{\delta(X)\geq 0} + \mathbf{1}_{\delta(X)<0}] - \mathbb{E}_p[\mathbf{1}_{\delta(X)<0}] && (4.48)\\
&= \mathbb{E}_p[2^{\delta_+(X)}] - \mathbb{E}_p[\mathbf{1}_{\delta(X)<0}] && (4.49)\\
&\geq \mathbb{E}_p[2^{\delta_+(X)}] - 1. && (4.50)
\end{aligned}$$

This establishes (4.43), which implies (4.44) by Jensen's inequality. From (4.43), for any x in

$\mathcal{X}$, we have

$$2 \geq \mathbb{E}_p[2^{\delta_+(X)}] \geq p(x)2^{\delta_+(x)}.$$

This proves (4.45). ■

Realism performance

Claim 4.14 *Since* $R_1 > \log(B)$, *a simple bound yields,*

$$(p^{\mathcal{U}}_{[\lfloor 2^{R_1} \rfloor]})^{\otimes B}(M^{(1)}, ..., M^{(B)} \text{ 2 by 2 distinct}) \geq 1 - \frac{B^2}{\lfloor 2^{R_1} \rfloor}.$$

See Appendix 4.I for a proof. From the definition (Section 4.B.2) of Q, for any measurable (Borel) set $\mathcal{E}$ of real numbers,

$$\begin{aligned} Q\Big(\Big\{\delta_0\big(\{y(\mathcal{C}, M^{(k)})\}_{k\in[B]}\big) \in \mathcal{E}\Big\} \\ \Big|\Big\{M^{(1)}, ..., M^{(B)} \text{ 2 by 2 distinct}\Big\}\Big) \\ = p_X^{\otimes B}\Big(\delta_0\big(\mathbf{X}^{(B)}\big) \in \mathcal{E}\Big). \end{aligned} \tag{4.51}$$

Therefore, we have

$$\begin{aligned} &\mathbb{E}_Q[\delta_+(\{y(\mathcal{C}, M^{(k)})\}_{k\in[B]})] \\ &= \sum_{\mathbf{m}^{(B)}} \mathbb{E}_Q[\mathbf{1}_{\mathbf{M}^{(B)}=\mathbf{m}^{(B)}}\delta_+(\{y(\mathcal{C}, m^{(k)})\}_{k\in[B]})] \\ &= \sum_{\mathbf{m}^{(B)}} \mathbb{E}_Q[\mathbf{1}_{\mathbf{M}^{(B)}=\mathbf{m}^{(B)}}]\mathbb{E}_Q[\delta_+(\{y(\mathcal{C}, m^{(k)})\}_{k\in[B]})] \\ &= \sum_{\{m^{(k)}\}_{k\in[B]} \text{ 2 by 2 } \neq} (p^{\mathcal{U}}_{[\lfloor 2^{R_1} \rfloor]})^{\otimes B}(\mathbf{M}^{(B)}=\mathbf{m}^{(B)})\mathbb{E}_{p_X^{\otimes B}}[\delta_+(\mathbf{X}^{(B)})] \\ &+ \sum_{\{m^{(k)}\}_{k\in[B]} \text{ not 2 by 2 } \neq} (p^{\mathcal{U}}_{[\lfloor 2^{R_1} \rfloor]})^{\otimes B}(\mathbf{M}^{(B)}=\mathbf{m}^{(B)})\mathbb{E}_Q[\delta_+(\{y(\mathcal{C}, m^{(k)})\}_{k\in[B]})] \\ &\leq \mathbb{E}_{p_X^{\otimes B}}[\delta_+(\mathbf{X}^{(B)})] + \max(\delta_+)(p^{\mathcal{U}}_{[\lfloor 2^{R_1} \rfloor]})^{\otimes B}(M^{(1)}, ..., M^{(B)}\text{not 2 by 2 } \neq) \\ &\leq 1 + \frac{B^2}{\lfloor 2^{R_1} \rfloor} \max_x B \log \frac{2}{p_X(x)}, \end{aligned} \tag{4.52}$$

where (4.52) follows from (4.6) and Claims 4.13 and 4.14. This concludes the proof of Lemma 4.12.

4.B.3 Further properties of a uniformly sampled batch

Proposition 4.15 *Consider a finite set $\mathcal{X}$ such that $|\mathcal{X}| \geq 2$, a distribution p_X on $\mathcal{X}$ such that $\forall x \in \mathcal{X}, p_X(x)>0$, a positive integer B, some $R > \log(B)/\log(|\mathcal{X}|)$, some $\Delta \in \mathbb{R}_+$, and a $p_X^{\otimes B}$-critic δ. Consider any conditional transition kernel $p_{Y|X}$ from $\mathcal{X}$ to $\mathcal{X}$ satisfying*

$$p_Y \equiv p_X, \quad \mathbb{E}_p[d(X,Y)] \leq \Delta. \tag{4.53}$$

Then, for any $\varepsilon \in (0, \Delta/2)$, and any $\gamma > 0$, there exists a family $\{y(m)\}_{m\in[\lfloor 2^{R_1}\rfloor]}$, denoted $\mathbf{c}$, of elements of $\mathcal{X}$, such that distribution

$$Q_{M,Y,X}\left(m,y,x\right) := \tfrac{1}{\lfloor 2^{R_1}\rfloor} \cdot \left(\mathbf{1}_{y=y(m)}\right) \cdot p_{X|Y=y(m)}(x) \tag{4.54}$$

satisfies

$$\|Q_X - p_X\|_{TV} \leq \frac{3\Delta}{\varepsilon}[p(\mathcal{A}_{R,\gamma}) + 2^{-\gamma\log(|\mathcal{X}|)/2}] \tag{4.55}$$

$$\mathbb{E}_{Q^{\otimes B}}[d(\mathbf{X}^{(B)}, \mathbf{Y}^{(B)})] \leq \Delta + \varepsilon \tag{4.56}$$

$$\mathbb{E}_{Q^{\otimes B}}[\delta(\mathbf{Y}^{(B)})] \leq \frac{3\Delta}{\varepsilon} \cdot \left(1 + \frac{B^2}{\lfloor 2^{R_1}\rfloor} \max_x B \log \frac{2}{p_X(x)}\right), \tag{4.57}$$

where $R_1 = R\log|X|$, and

$$\mathcal{A}_{R,\gamma} := \left\{(x,y) \in \mathcal{X}^2 \mid \log\left(\frac{p_{X,Y}(x,y)}{p_X(x)p_Y(y)}\right) - \log(\lfloor 2^{R_1}\rfloor) > -\gamma\log(|\mathcal{X}|)\right\}. \tag{4.58}$$

Proof: Fix some $R > \log(B)/\log(|\mathcal{X}|)$, some $\Delta > 0$, some $\varepsilon \in (0, \Delta/2)$, some $\gamma > 0$, and a conditional transition kernel $p_{Y|X}$ from $\mathcal{X}$ to $\mathcal{X}$ satisfying

$$p_Y \equiv p_X, \quad \mathbb{E}_p[d(X,Y)] \leq \Delta. \tag{4.59}$$

Define $R_1 = R \log |\mathcal{X}|$. We apply Lemma 4.12, and use the notation therein. Then, from Markov's inequality, we have

$$Q_{\mathcal{C}}\left(\mathbb{E}_Q[\delta_+(\mathbf{Y}^{(B)})|\mathcal{C}] \geq \frac{3\Delta}{\varepsilon}\left(1+\frac{B^2}{\lfloor 2^{R_1}\rfloor}\max_x B\log\frac{2}{p_X(x)}\right)\right) \leq \frac{\varepsilon}{3\Delta}. \tag{4.60}$$

We extend distribution Q as follows.

$$\begin{aligned} &Q_{\mathcal{C},\mathbf{M}^{(B)},\mathbf{Y}^{(B)},\mathbf{X}^{(B)}}\left(\{y(m')\}_{m'\in[\lfloor 2^{R_1}\rfloor]},\mathbf{m}^{(B)},\mathbf{y}^{(B)},\mathbf{x}^{(B)}\right) := \\ &Q_{\mathcal{C},\mathbf{M}^{(B)},\mathbf{Y}^{(B)}}\left(\{y(m')\}_{m'\in[\lfloor 2^{R_1}\rfloor]},\mathbf{m}^{(B)},\mathbf{y}^{(B)}\right)\cdot\prod_{k=1}^{B}p_{X|Y=y(m^{(k)})}(x^{(k)}). \end{aligned} \tag{4.61}$$

Distribution $Q_{\mathcal{C},M^{(1)},Y^{(1)},X^{(1)}}$ corresponds to the setting of a general version of the soft covering lemma [Cuf13, Theorem VII.1]. Since $p_Y \equiv p_X$, the latter lemma yields that for any $\tau \in \mathbb{R}$,

$$\mathbb{E}_{\mathcal{C}}\left[\|Q_{X^{(1)}|\mathcal{C}} - p_X\|_{TV}\right] \leq p(\mathcal{A}_\tau) + 2^{\tau/2}, \tag{4.62}$$

where

$$\mathcal{A}_\tau := \{(x,y) \mid \log(p_{Y|X=x}(y)/p_X(y)) - \log(\lfloor 2^{R_1}\rfloor) > \tau\}. \tag{4.63}$$

We choose $\tau = -\gamma \log|\mathcal{X}|$. Then, $\mathcal{A}_\tau = \mathcal{A}_{R,\gamma}$, with the notation of Proposition 4.15. Hence, from (4.62) and Markov's inequality, we have

$$Q_{\mathcal{C}}\Big(\|Q_{X^{(1)}|\mathcal{C}} - p_X\|_{TV} \geq \frac{3\Delta}{\varepsilon}[p(\mathcal{A}_{R,\gamma}) + 2^{-\gamma\log(|\mathcal{X}|)/2}]\Big) \leq \frac{\varepsilon}{3\Delta}. \tag{4.64}$$

By construction, we have $Q_{\mathbf{Y}^{(B)},\mathbf{X}^{(B)}} \equiv p_{Y,X}^{\otimes B}$. Therefore, from (4.59), and the additivity of d, we have

$$\mathbb{E}_Q[d(\mathbf{X}^{(B)},\mathbf{Y}^{(B)})] \leq \Delta. \tag{4.65}$$

Therefore, from Markov's inequality,

$$Q_{\mathcal{C}}\Big(\mathbb{E}_Q[d(\mathbf{X}^{(B)},\mathbf{Y}^{(B)})|\mathcal{C}] \geq \Delta+\varepsilon\Big) \leq \frac{\Delta}{\Delta+\varepsilon} = 1-\frac{\varepsilon}{\Delta}\cdot\frac{1}{1+\varepsilon/\Delta} < 1-\frac{2\varepsilon}{3\Delta}, \tag{4.66}$$

where we have used the fact that $\varepsilon \in (0,\Delta/2)$. From a union bound and (4.60), (4.64), and (4.66) there exists a realization c_* of $\mathcal{C}$ such that none of the corresponding events hold. Since, by construction,

$$Q_{\mathbf{M}^{(B)},\mathbf{Y}^{(B)},\mathbf{X}^{(B)}|\mathcal{C}=c*} \equiv Q^{\otimes B}_{M^{(1)},Y^{(1)},X^{(1)}|\mathcal{C}=c*},$$

and $\delta_+(\mathbf{Y}^{(B)}) \geq \delta(\mathbf{Y}^{(B)})$, this concludes the proof of Proposition 4.15. ■

4.B.4 Proof of Theorem 4.5

Fix some $R > \log(B)/\log(|\mathcal{X}|)$, some $\Delta > 0$, some $\varepsilon \in (0,\Delta/2)$, some $\gamma > 0$, and a conditional transition kernel $p_{Y|X}$ from $\mathcal{X}$ to $\mathcal{X}$ satisfying

$$p_Y \equiv p_X, \quad \mathbb{E}_p[d(X,Y)] \leq \Delta. \tag{4.67}$$

Define $R_1 = R\log|\mathcal{X}|$. Then, we can apply Proposition 4.15. We use the notation from the latter.

Compression scheme achieving close-to-uniform sampling

We define the following distribution $P_{X,Y,M}$, which differs from Q in having the correct marginal for X :

$$P_{X,M,Y} := p_X \cdot Q_{M,Y|X}. \tag{4.68}$$

Therefore, from (4.54), distribution P satisfies the Markov chain $X-M-Y$. Hence, it defines a $(R_1,0)$ code. From Lemma 1.9 (Appendix 1), comparing P with Q reduces to comparing

marginals, i.e. to (4.55) :

$$\begin{aligned}\left\|P_{M,X,Y}-Q_{M,X,Y}\right\|_{TV}&=\left\|P_X-Q_X\right\|_{TV}\\ &=\left\|p_X-Q_X\right\|_{TV}\leq\frac{3\Delta}{\varepsilon}[p(\mathcal{A}_{R,\gamma})+2^{-\gamma\log(|\mathcal{X}|)/2}].\end{aligned}\tag{4.69}$$

Since d is additive, we have

$$\mathbb{E}_{(P)^{\otimes B}}[d(\mathbf{X}^{(B)},\mathbf{Y}^{(B)})]=\mathbb{E}_P[d(X,Y)] \text{ and}$$

$$\mathbb{E}_{(Q)^{\otimes B}}[d(\mathbf{X}^{(B)},\mathbf{Y}^{(B)})]=\mathbb{E}_Q[d(X,Y)].$$

Since d is bounded, then we can apply Lemma 1.10 (Appendix 1). Then, from (4.69), and Lemma 1.8 (Appendix 1) with $W=(X,Y)$, we have

$$\begin{aligned}\mathbb{E}_{P^{\otimes B}}[d(\mathbf{X}^{(B)},\mathbf{Y}^{(B)})]&\leq\mathbb{E}_{Q^{\otimes B}}[d(\mathbf{X}^{(B)},\mathbf{Y}^{(B)})]+\frac{6\Delta}{\varepsilon}\max(d)[p(\mathcal{A}_{R,\gamma})+2^{-\gamma\log(|\mathcal{X}|)/2}]\\ &\leq\Delta+\varepsilon+\frac{6\Delta}{\varepsilon}\max(d)[p(\mathcal{A}_{R,\gamma})+2^{-\gamma\log(|\mathcal{X}|)/2}],\end{aligned}\tag{4.70}$$

where the last inequality follows from (4.56). Moving to the realism performance, we have the following property of the TVD; see Appendix 4.H:

Claim 4.16 *Given any two distributions P and Q on the same finite alphabet, we have, for any $B\in\mathbb{N}$,*

$$\left\|P^{\otimes B}-Q^{\otimes B}\right\|_{TV}\leq B\left\|P-Q\right\|_{TV}.$$

From Lemma 1.10 (Appendix 1), and from the combination of Claim 4.16, (4.69), and Lemma 1.8 (Appendix 1) with $W=\mathbf{Y}^{(B)}$, we have,

$$\begin{aligned}&\mathbb{E}_{P^{\otimes B}}[\delta(\mathbf{Y}^{(B)})]\leq\mathbb{E}_{Q^{\otimes B}}[\delta(\mathbf{Y}^{(B)})]+\frac{6B\Delta}{\varepsilon}\max(\delta)[p(\mathcal{A}_{R,\gamma})+2^{-\gamma\log(|\mathcal{X}|)/2}]\\ \leq\frac{3\Delta}{\varepsilon}\cdot&\left(1+\frac{B^2}{\lfloor 2^{R_1}\rfloor}\max_x B\log\frac{2}{p_X(x)}\right)+\frac{6B\Delta}{\varepsilon}[p(\mathcal{A}_{R,\gamma})+2^{-\gamma\log(|\mathcal{X}|)/2}]\cdot\max_x B\log\frac{2}{p_X(x)},\end{aligned}$$

where the last inequality follows from (4.57) and Claim 4.13. This concludes the proof.

4.C Achievability of Theorem 4.4

Consider some $\Delta \in \mathbb{R}_+$ such that $R^{(1)}(\Delta) < H_p(X)$, and a sequence $\{B_n\}_{n\geq 1}$ of positive integers such that

$$\log(B_n)/n \underset{n\to\infty}{\longrightarrow} 0. \tag{4.71}$$

We prove that

$$R(\Delta) \leq R_*(\Delta) \leq R^{(1)}(\Delta).$$

Fix $R \in (R^{(1)}(\Delta), H_p(X))$, $\varepsilon \in \Big(0, \min\big(R - R^{(1)}(\Delta), \Delta/2\big)\Big)$, and $\gamma \in (0, \varepsilon/\log(|\mathcal{X}|))$. Then, from (4.5), there exists $p_{Y|X}$ such that

$$p_Y \equiv p_X,\ \mathbb{E}_p[d(X,Y)] \leq \Delta,\ R \geq I_p(X;Y) + \varepsilon. \tag{4.72}$$

We use the powerful result of Theorem 4.10 regarding the existence of a so-called *universal critic* δ_0. From Definition 4.1, for every $n \in \mathbb{N}$, the restriction of δ_0 to $\mathcal{X}^{nB_n}$ is a $p_X^{\otimes nB_n}$-critic. Moreover, from (4.71), for large enough n, we have $nR > \log(B_n)$. Then, for large enough n, from the additivity of d, we can apply Theorem 4.5 for set $\mathcal{X}^n$, distribution $p_X^{\otimes n}$, transition kernel $\prod_{t=1}^n p_{Y|X}$, batch size B_n, critic δ_0, rate $nR/\log(|\mathcal{X}^n|)$, and constants $\Delta, \varepsilon, \gamma$. This gives that, for every n large enough, there is an $(n, R, 0)$ code, inducing a distribution $P^{(n)}$ such that

$$\mathbb{E}_{(P^{(n)})^{\otimes B_n}}\big[d(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)})\big] \leq \Delta + \varepsilon + \frac{6\Delta}{\varepsilon}\max(d)[p_{X,Y}^{\otimes n}(\mathcal{A}_{R,\gamma}^{(n)}) + 2^{-\gamma n \log(|\mathcal{X}|)/2}], \tag{4.73}$$

$$\begin{aligned}\mathbb{E}_{(P^{(n)})^{\otimes B_n}}\big[\delta_0(\mathbf{Y}^{(n,B_n)})\big] \leq \frac{3\Delta}{\varepsilon}\Bigg[1 + \frac{B_n^2}{\lfloor 2^{nR}\rfloor}\max_x\left(nB_n \log\frac{2}{p_X(x)}\right) \\ + 2B_n[p_{X,Y}^{\otimes n}(\mathcal{A}_{R,\gamma}^{(n)}) + 2^{-\gamma n\log(|\mathcal{X}|)/2}] \cdot \max_x\left(nB_n\log\frac{2}{p_X(x)}\right)\Bigg],\end{aligned} \tag{4.74}$$

where

$$\mathcal{A}^{(n)}_{R,\gamma} := \Big\{(x_{1:n}, y_{1:n}) \in (\mathcal{X}^n)^2 \mid \sum_{t=1}^{n} \log\Big(\frac{p_{X,Y}(x_t, y_t)}{p_X(x_t)p_Y(y_t)}\Big) - \log(\lfloor 2^{nR} \rfloor) > -\gamma n \log(|\mathcal{X}|)\Big\}, \quad (4.75)$$

with the convention $0/0 := 1$. From (4.72), $\log(\lfloor 2^{nR} \rfloor)/n - \gamma \log(|\mathcal{X}|) > I_p(X;Y)$ for large enough n. Then, since $\mathcal{X}$ is finite, we have, from Hoeffding's inequality,

$$p^{\otimes n}_{X,Y}(\mathcal{A}^{(n)}_{R,\gamma}) = O(e^{-\kappa n}), \quad (4.76)$$

for some $\kappa > 0$. Hence, from (4.71), (4.73), (4.74), and Theorem 4.10, we have

$$\limsup_{n\to\infty} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\big[d(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)})\big] \leq \Delta + \varepsilon, \quad (4.77)$$

and for any lower semi-computable $p^{\otimes *}_X$-critic δ,

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\big[\delta(\mathbf{Y}^{(n,B_n)})\big] < \infty. \quad (4.78)$$

From the proof of Theorem 4.5, we know that $P^{(n)}$ has a deterministic decoder. Hence, it only remains to de-randomize the encoder of $P^{(n)}$. We denote its decoder by $m \mapsto y_{1:n}(m)$. The following claim is a slight modification of Proposition 2.3 of Chapter 2. We provide details in Section 4.C.1.

Claim 4.17 *There exists a sequence of deterministic maps*

$$\begin{gathered} f^{(n)} : \mathcal{X}^n \to [2^{nR}], \quad \textit{such that} \\ \big\|\hat{\tilde{P}}^{(n)}_{\mathcal{X}^2}[X^n, y_{1:n}(M)] - \hat{P}^{(n)}_{\mathcal{X}^2}[X^n, y_{1:n}(M)]\big\|_{TV} \underset{n\to\infty}{\longrightarrow} 0, \\ \liminf_{n\to\infty} \frac{-1}{n} \log \big\|\tilde{P}^{(n)}_M - P^{(n)}_M\big\|_{TV} > 0, \quad \textit{where} \\ \tilde{P}^{(n)}_{X^n,M} := p^{\otimes n}_X \cdot \mathbf{1}_{M=f^{(n)}(X^n)}. \end{gathered} \quad (4.79)$$

Then, from (4.71) and Claim 4.16, we have

$$\liminf_{n\to\infty} \frac{-1}{n} \log \left\| (\tilde{P}^{(n)})_M^{\otimes B_n} - (P^{(n)})_M^{\otimes B_n} \right\|_{TV} > 0. \tag{4.80}$$

Thus, from Lemma 1.10 (Appendix 1) and (4.6), we have

$$\left| \mathbb{E}_{(\tilde{P}^{(n)})^{\otimes B_n}} \left[\delta(\mathbf{Y}^{(n,B_n)}) \right] - \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[\delta(\mathbf{Y}^{(n,B_n)}) \right] \right| \underset{n\to\infty}{\longrightarrow} 0. \tag{4.81}$$

Moreover, since d is bounded, then from Lemma 1.10 (Appendix 1), we obtain

$$\left| \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[d(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)}) \right] - \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \left[d(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)}) \right] \right| \underset{n\to\infty}{\longrightarrow} 0. \tag{4.82}$$

Since this analysis is valid for any $\varepsilon \in \Big(0, \min\big(R - R^{(1)}(\Delta), \Delta/2\big)\Big)$, then tuple $(R, \{B_n\}_{n\geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism with fully deterministic codes. This being true for any $R \in (R^{(1)}(\Delta), H_p(X))$, we have

$$R(\Delta) \leq R_*(\Delta) \leq R^{(1)}(\Delta),$$

as desired.

4.C.1 Encoder de-randomization

We show that Claim 4.17 follows from Proposition 2.3 of Chapter 2, and its proof. We can apply that result directly, since $R < H_p(X)$ and $\mathcal{X}$ is finite. This would give all properties in Claim 4.17, except for the exponential decay in (4.79). To obtain the latter, it is sufficient to adapt the proof of Proposition 2.3 (Chapter 2), by replacing the use of the law of large numbers with the use of Hoeffding's inequality, and using [Cuf13, Theorem VII.1] with $\tau = -n\gamma$, for small enough γ.

4.D Converse of Theorem 4.4

From standard information-theoretic arguments, we have the following result - see Appendix 4.F for a proof.

Lemma 4.18 *Consider a triplet (R, R_c, Δ) and a sequence of codes, the n-th being (n, R, R_c), inducing a sequence $\{P^{(n)}_{X_{1:n},J,M,Y_{1:n}}\}_{n\geq 1}$ of distributions such that*

$$\limsup_{n\to\infty} \mathbb{E}_{(P^{(n)})^{\otimes b_n}}\left[d(\mathbf{X}^{(n,b_n)}, \mathbf{Y}^{(n,b_n)})\right] \leq \Delta, \tag{4.83}$$

for some sequence $\{b_n\}_{n\geq 1}$ of positive integers. For every $n \geq 1$, let $T^{(n)}$ denote a uniform variable on $[nb_n]$ independent from all other random variables. Then, there exists a conditional distribution $p_{Y|X}$ and an increasing sequence $\{n_i\}_{i\geq 1}$ of positive integers such that

$$(P^{(n_i)})^{\otimes b_{n_i}}_{X_{T^{(n_i)}},Y_{T^{(n_i)}}} \underset{i\to\infty}{\longrightarrow} p_{X,Y} \tag{4.84}$$

$$\Delta \geq \mathbb{E}_p[d(X,Y)] \tag{4.85}$$

$$R \geq I_p(X;Y), \tag{4.86}$$

where $p_{X,Y}$ refers to $p_X \cdot p_{Y|X}$.

4.D.1 Converse proof

Consider some $\Delta \in \mathbb{R}_+$ such that $R^{(1)}(\Delta) < H_p(X)$, and a sequence $\{B_n\}_{n\geq 1}$ of positive integers such that

$$\log(B_n)/n \underset{n\to\infty}{\longrightarrow} 0. \tag{4.87}$$

We prove that $R(\Delta) \geq R^{(1)}(\Delta)$. Consider $(R, R_c) \in \mathbb{R}_+ \times \mathbb{R}_{\geq 0}$ such that $(R, R_c, \{B_n\}_{n\geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism. Fix $\varepsilon > 0$. Then, there exists a sequence of

codes, the n-th being (n, R, R_c), inducing a sequence $\{P^{(n)}_{X_{1:n},J,M,Y_{1:n}}\}_n$ of distributions such that

$$\limsup_{n\to\infty} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[d(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)})\right] \leq \Delta + \varepsilon, \tag{4.88}$$

and for any lower semi-computable $p_X^{\otimes *}$-critic δ,

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[\delta(\mathbf{Y}^{(n,B_n)})\right] < \infty. \tag{4.89}$$

Then, Lemma 4.18 applies, with $b_n = B_n$, for all n, with $R+\varepsilon$ instead of R, and $\Delta+\varepsilon$ instead of Δ. Then, there exists a conditional distribution $p_{Y|X}$ and an increasing sequence $\{n_i\}_{i\geq 1}$ of positive integers such that

$$(P^{(n_i)})^{\otimes b_{n_i}}_{X_{T^{(n_i)}},Y_{T^{(n_i)}}} \underset{i\to\infty}{\longrightarrow} p_{X,Y} \tag{4.90}$$

$$\Delta + \varepsilon \geq \mathbb{E}_p[d(X,Y)] \tag{4.91}$$

$$R + \varepsilon \geq I_p(X;Y), \tag{4.92}$$

where for any $n \in \mathbb{N}$, variable $T^{(n)}$ is uniformly distributed on $[nB_n]$, and independent from all other random variables. We prove that $p_Y \equiv p_X$. Fix $e_0 \in \mathcal{X}$. Consider the computable $p_X^{\otimes *}$-critic δ from Proposition 4.1, with q therein taken to be p_X. Then, from (4.89),

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[\delta(\mathbf{Y}^{(n,B_n)}) - 2\log(\delta(\mathbf{Y}^{(n,B_n)}) + 3)\right] < \infty. \tag{4.93}$$

Thus,

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[\delta(\mathbf{Y}^{(n,B_n)})\right] < \infty, \quad \text{and} \quad \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[\delta(\mathbf{Y}^{(n,B_n)}) - \frac{1}{2}\log(nB_n)\right] \underset{n\to\infty}{\longrightarrow} -\infty.$$

Thus, the frequency of e_0 in a batch of reconstructions converges in L_1 norm to $p_X(e_0)$. Hence, the expected frequencies converge to $p_X(e_0)$. This rewrites as

$$(P^{(n)})^{\otimes B_n}_{Y_{T^{(n)}}}(e_0) \to p_X(e_0). \tag{4.94}$$

This is true for any e_0 in $\mathcal{X}$. Thus, from (4.90), $p_Y \equiv p_X$. Hence, from (4.91) and (4.92), we have

$$R + \varepsilon \geq R^{(1)}(\Delta + \varepsilon). \tag{4.95}$$

This being true for any $\varepsilon > 0$, and since $R^{(1)}(\cdot)$ is convex -thus continuous- on $(0, \infty)$, we have

$$R \geq R^{(1)}(\Delta). \tag{4.96}$$

This being true for any $R \in \mathbb{R}_+$ such that there exists $R_c \in \mathbb{R}_{\geq 0}$ such that $(R, R_c, \{B_n\}_{n\geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism, we have

$$R(\Delta) \geq R^{(1)}(\Delta), \tag{4.97}$$

as desired.

4.E Proof of Theorem 4.6

Consider an increasing sequence $\{B_n\}_{n\geq 1}$ of positive integers such that

$$\frac{B_n}{|\mathcal{X}|^n} \to \infty, \tag{4.98}$$

some $R_c \in \mathbb{R}_{\geq 0}$, and some $(R, \Delta) \in (\mathbb{R}_+)^2$ such that tuple (R, R_c, Δ) is asymptotically achievable with near-perfect realism. From Theorem 1 in [Wag22], (R, R_c, Δ) achievable with *perfect realism* — in the presence of unlimited (decoder) private randomness — (Definition 2.3 in Chapter 2). Fix $\varepsilon > 0$, and a corresponding sequence of $(n, R + \varepsilon, R_c)$ codes[5]. Denote by $P^{(n)}$

[5]The fact that there is no need for an ε backoff in the common randomness rate follows from the proof of [Wag22, Theorem 1].

the distribution induces by the n-th code. Then, there exists an integer N_ε such that

$$\limsup_{n\to\infty} \mathbb{E}_{P^{(n)}}\left[d(X_{1:n}, Y_{1:n})\right] \leq \Delta + \varepsilon, \qquad (4.99)$$
$$\forall n \geq N_\varepsilon,\ P^{(n)}_{Y_{1:n}} \equiv p_X^{\otimes n}.$$

Thus,

$$\forall n \geq N_\varepsilon,\ (P^{(n)}_{Y_{1:n}})^{\otimes B_n} \equiv p_X^{\otimes nB_n}. \qquad (4.100)$$

From (4.99), (4.100), Claim 4.13, and the additivity of the distortion measure d, we have

$$\limsup_{n\to\infty} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[d(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)})\right] \leq \Delta + \varepsilon, \qquad (4.101)$$

and for any lower semi-computable $p_X^{\otimes *}$-critic δ,

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[\delta(\mathbf{Y}^{(n,B_n)})\right] < \infty. \qquad (4.102)$$

Since this analysis is valid for every $\varepsilon > 0$, then $(R, R_c, \{B_n\}_{n\geq 1}, \Delta)$ is asymptotically achievable with algorithmic realism. Moving to the converse, consider a computable increasing sequence $\{B_n\}_{n\geq 1}$ of positive integers such that

$$\frac{B_n}{|\mathcal{X}|^n} \to \infty, \qquad (4.103)$$

some $R_c \in \mathbb{R}_{\geq 0}$, and some $(R, \Delta) \in (\mathbb{R}_+)^2$ such that tuple (R, R_c, Δ) is asymptotically achievable with algorithmic realism. Fix $\varepsilon > 0$. Then, there exists a sequence of codes, the n-th being $(n, R + \varepsilon, R_c)$, such that the sequence $\{P^{(n)}\}_n$ of distributions induced by the codes satisfies

$$\limsup_{n\to\infty} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[d(\mathbf{X}^{(n,B_n)}, \mathbf{Y}^{(n,B_n)})\right] \leq \Delta + \varepsilon, \quad \text{and} \qquad (4.104)$$

and for any lower semi-computable $p_X^{\otimes *}$-critic δ,

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}}\left[\delta(\mathbf{Y}^{(n,B_n)})\right] < \infty. \tag{4.105}$$

Lemma 4.19 *[Can20] There exists a positive integer λ such that for any $k \in \mathbb{N}$, any distribution q on some finite set $\mathcal{W}$ of size k, any $\varepsilon, \eta > 0$, and any integer b satisfying*

$$b \geq \lambda \cdot \frac{k + \log(1/\eta)}{\varepsilon^2}, \tag{4.106}$$

we have

$$q^{\otimes b}\Big(\big\|\mathbb{P}_{\mathcal{W}}^{emp}[W^b] - q\big\|_{TV} \geq \varepsilon\Big) \leq \eta. \tag{4.107}$$

For every $n \in \mathbb{N}$, define

$$C_n := \left\lceil \left(\frac{B_n}{|\mathcal{X}|^n}\right)^{\frac{1}{3}} \right\rceil. \tag{4.108}$$

Since $\mathcal{X}$ is finite and $\{B_n\}_{n\geq 1}$ is a computable sequence of positive integers, the map $n \mapsto B_n/|\mathcal{X}|^n$ is a computable map from $\mathbb{N}$ to $\mathbb{Q}_{\geq 0}$ by Lemma 4.8(i). Hence $\{C_n\}_{n\geq 1}$ is a computable sequence of positive integers, by Lemma 4.8(iv) applied with $(p, q) = (1, 3)$. Moreover, from (4.103), we have

$$C_n \underset{n\to\infty}{\longrightarrow} \infty. \tag{4.109}$$

Choosing, for every $n \in \mathbb{N}$, $\eta = 1/3$ and $\varepsilon = 1/C_n$, then from Lemma 4.19 and (4.109) we have, for large enough n,

$$(P^{(n)})^{\otimes B_n}\left(\big\|\mathbb{P}_{\mathcal{X}^n}^{\mathrm{emp}}[\mathbf{Y}^{(n,B_n)}] - P_{Y_{1:n}}^{(n)}\big\|_{TV} \geq \frac{1}{C_n}\right) \leq \frac{1}{3}. \tag{4.110}$$

Consider the computable sequence of positive integers defined by

$$\forall n \in \mathbb{N},\ A_n := \Big\lceil \Big(\frac{B_n}{|\mathcal{X}|^n} \Big)^{\frac{4}{9}} \Big\rceil. \tag{4.111}$$

As for $\{C_n\}_{n\geq 1}$, the sequence $\{A_n\}_{n\geq 1}$ is a computable sequence of positive integers by Lemma 4.8(i) and (iv), the latter applied with $(p, q) = (4, 9)$. Since $\{B_n\}_{n\geq 1}$ is increasing, then for any $t \in \mathbb{N}$, there exists a unique integer $n \in \mathbb{N}_{\geq 0}$ such that

$$t \in [nB_n, (n+1)B_{n+1}),$$

with the definition $B_0 := 0$. We define $\phi : \cup_{t\in\mathbb{N}} \mathcal{X}^t \to \mathbb{N}_{\geq 0}$ as follows. For any integer $t \in [1, B_1)$, and any $x_{1:t} \in \mathcal{X}^t$, let $\phi(x) := 0$. For any $n \in \mathbb{N}$, any $t \in [nB_n, (n+1)B_{n+1})$, and any $x_{1:t} \in \mathcal{X}^t$, let

$$\phi(x_{1:t}) := \Big\lceil A_n \big\| \mathbb{P}^{\mathrm{emp}}_{\mathcal{X}^n}[x_{1:nB_n}] - p_X^{\otimes n} \big\|_{TV} \Big\rceil. \tag{4.112}$$

Claim 4.20 *From Lemma 4.19 and* (4.109)*, there exists a positive integer L such that $\phi - 2\log(\phi + 3) - L$ is a lower semi-computable $p_X^{\otimes *}$-critic.*

We provide a proof in Appendix 4.G.2. Then, we can apply (4.105) to critic $\phi - 2\log(\phi+3) - L$, and get,

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \big[\phi(\mathbf{Y}^{(n,B_n)}) - 2\log(\phi(\mathbf{Y}^{(n,B_n)}) + 3) - L \big] < \infty. \tag{4.113}$$

Thus,

$$\sup_{n\in\mathbb{N}} \mathbb{E}_{(P^{(n)})^{\otimes B_n}} \big[\phi(\mathbf{Y}^{(n,B_n)}) \big] < \infty, \quad \text{and} \quad (P^{(n)})^{\otimes B_n} \big(\phi(\mathbf{Y}^{(n,B_n)}) \geq C_n \big) \underset{n\to\infty}{\longrightarrow} 0,$$

because $\{C_n\}_{n\geq 1}$ tends to infinity. Combining this with (4.110) through a union bound, we obtain, from the triangle inequality for the TVD,

$$(P^{(n)})^{\otimes B_n}\big(\big\|P_{Y_{1:n}}^{(n)} - p_X^{\otimes n}\big\|_{TV} \leq \frac{C_n}{A_n} + \frac{1}{C_n}\big) > 0,$$

for large enough n. The above event does not depend on the random batch, hence the corresponding inequality is true, for large enough n. Since $\{C_n\}_{n\geq 1}$ tends to infinity and since from (4.103), (4.108), and (4.111), we have $C_n/A_n \to 0$, then we obtain

$$\big\|P_{Y_{1:n}}^{(n)} - p_X^{\otimes n}\big\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{4.114}$$

Hence, from (4.104) and the additivity of d, we have that (R, R_c, Δ) is asymptotically achievable with near-perfect realism. This concludes the proof.

4.F Standard converse arguments

Here, we provide a proof of Lemma 4.18 (Appendix 4.D). The sequence of probability distributions $(P^{(n)})_{X_T,Y_T}^{\otimes b_n}$ can be seen as a bounded sequence in $\mathbb{R}^{2^{2s}}$, thus it admits a converging subsequence:

$$(P^{(n_i)})_{X_T,Y_T}^{\otimes b_n} \underset{i\to\infty}{\longrightarrow} p_{X,Y}. \tag{4.115}$$

Since d is bounded, we have

$$\mathbb{E}_{(P^{(n_i)})^{\otimes b_n}}[d(X_T, Y_T)] \underset{i\to\infty}{\longrightarrow} \mathbb{E}_p[d(X, Y)]. \tag{4.116}$$

Since d is additive, we have, for any $n \in \mathbb{N}$,

$$\begin{aligned}&\mathbb{E}_{(P^{(n)})^{\otimes b_n}}\big[d(\mathbf{X}^{(n,b_n)}, \mathbf{Y}^{(n,b_n)})\big]\\ &\quad= \mathbb{E}_{(P^{(n)})^{\otimes b_n}}\big[d(X_T, Y_T)\big].\end{aligned} \tag{4.117}$$

From (4.83), (4.116) and (4.117), we have $\Delta \geq \mathbb{E}_p[d(X,Y)]$. Secondly, distribution $P^{(n)}$ satisfies

$$\begin{aligned}
nb_nR &\geq H(\{m^{(k)}\}_{k\in[b_n]}|\{J^{(k)}\}_{k\in[b_n]}) \\
&\geq I(\{m^{(k)}\}_{k\in[b_n]}; \mathbf{X}^{(n,b_n)}|\{J^{(k)}\}_{k\in[b_n]}) \\
&= I(\{m^{(k)}\}_{k\in[b_n]}, \{J^{(k)}\}_{k\in[b_n]}; \mathbf{X}^{(n,b_n)}) \\
&\geq I(\mathbf{Y}^{(n,b_n)}; \mathbf{X}^{(n,b_n)}) \\
&\geq \sum_{k=1}^{b_n}\sum_{t=1}^{n} I(Y_t^{(k)}; X_t^{(k)}) \\
&= nb_nI(Y_T; X_T|T) \\
&= nb_nI(T, Y_T; X_T) \\
&\geq nb_nI(Y_T; X_T).
\end{aligned}$$

Therefore, from (4.115), and by continuity of mutual information on the set of distributions on $(\{0,1\}^s)^2$, we have $R \geq I_p(X;Y)$.

4.G Frequency critics

4.G.1 Critic involving the frequency of a specific pattern

We provide a proof of Proposition 4.1.

ϕ is a computable integer-valued map

The function S is a computable integer-valued map. Since $q(e_0)$ and $1/n$ are computable, then the map inside the inner ceiling function is a computable real-valued map. Denote it by f. It is straightforward that if a real-valued map g is computable, and the set $\{x \mid g(x) \in \mathbb{Z}\}$ is computable, then $\lceil g \rceil \mathbf{1}_{g\neq 0}$ is a computable integer-valued function. The set $\{x \mid f(x) \in \mathbb{Z}\}$ is computable. Indeed, this set is empty if there does not exist integers a, b, a', b', c such that $q(e_0) = a/b - a'\sqrt{c}/b'$. Otherwise, there exists an algorithm having such a tuple (a, b, a', b', c)

in memory, which always terminates and determines whether $f(x)$ is rational by computing prime decompositions. Then, $\lceil f \rceil \mathbf{1}_{f\neq 0}$ is a computable integer-valued function. Thus, for the logarithm in base 2 (or any positive integer), $\{x \mid \mathbf{1}_{f(x)\neq 0} \log(\lceil f(x) \rceil) \in \mathbb{Z}\}$ is computable. Moreover, from point (ii) of Lemma 4.9, $\mathbf{1}_{f\neq 0} \log(\lceil f \rceil)$ is a real-valued computable map. Then, we can apply the same principle once more with $g = \mathbf{1}_{f\neq 0} \log(\lceil f \rceil)$. Therefore, ϕ is a computable integer-valued map.

Remainder of the proof of Proposition 4.1

This implies that ϕ, seen as a real-valued map, is computable. Therefore, from point (ii) of Lemma 4.9, $\phi - 2\log(\phi + 3)$ is a real-valued computable map. For any $(n, C) \in \mathbb{N}^2$, and any $x_{1:n} \in \mathcal{X}^n$, we have:

$$\begin{aligned}
\{\phi(x_{1:n}) \geq C\} &= \left\{ \left\lceil \log \left\lceil |S(x_{1:n}) - nq(e_0)| \;/\; \sqrt{n} \right\rceil \right\rceil \geq C \right\} \\
&= \left\{ \log \left\lceil |S(x_{1:n}) - nq(e_0)| \;/\; \sqrt{n} \right\rceil > C - 1 \right\} \\
&= \left\{ \left\lceil |S(x_{1:n}) - nq(e_0)| \;/\; \sqrt{n} \right\rceil > 2^{C-1} \right\} \\
&= \left\{ |S(x_{1:n}) - nq(e_0)| \;/\; \sqrt{n} > 2^{C-1} \right\}.
\end{aligned}$$

From this and Chebyshev inequality, we obtain:

$$\begin{aligned}
q^{\otimes n}(\phi(X_{1:n}) \geq C) &\leq \mathbb{E}_{q^{\otimes n}} \left[\left(S(X_{1:n}) - nq(e_0) \right)^2 / n \right] / 4^{C-1} \\
&= (q(e_0) - q(e_0)^2) / 4^{C-1} \qquad (4.118) \\
&\leq 4^{-C} \\
&\leq 2^{-C},
\end{aligned}$$

where (4.118) comes from the fact that $S(X_{1:n})$ follows a binomial distribution $B(n, q(e_0))$. Thus,

$$\mathbb{E}_{q^{\otimes n}}[\mathbf{1}_{\phi(X_{1:n})=C}] \leq 2^{-C}. \qquad (4.119)$$

Therefore,

$$\mathbb{E}_{q^{\otimes n}}[\mathbf{1}_{\phi(X_{1:n})=C} \cdot 2^{\phi(X_{1:n})-2\log(\phi(X_{1:n})+3)}] \leq \frac{1}{(C+3)^2}. \tag{4.120}$$

This also holds for $C = 0$. Summing over $C \in \mathbb{N}_{\geq 0}$ gives, for any $n \in \mathbb{N}$,

$$\sum_{x_{1:n} \in \mathcal{X}^n} q^{\otimes n}(x_{1:n}) \cdot 2^{\phi(x_{1:n})-2\log(\phi(x_{1:n})+3)} \leq 1. \tag{4.121}$$

Hence, we have that $\phi - 2\log(\phi+3)$ is a computable $q^{\otimes *}$-critic.

4.G.2 Critic involving an empirical distribution

We provide a proof of Claim 4.20.

Claim 4.21 *The map* $f : \cup_{t\in\mathbb{N}} \mathcal{X}^t \to \mathbb{R}$ *defined by* $\forall t \in [1, B_1) \cap \mathbb{N}, \forall x_{1:t} \in \mathcal{X}^t, f(x_{1:t}) := 0$, *and*

$$\forall n \in \mathbb{N}, \forall t \in [nB_n, (n+1)B_{n+1}) \cap \mathbb{N}, \forall x_{1:t} \in \mathcal{X}^t, \ f(x_{1:t}) := \left\| \mathbb{P}^{emp}_{\mathcal{X}^n}[x_{1:nB_n}] - p^{\otimes n} \right\|_{TV} \tag{4.122}$$

is computable.

Proof: Since there exists $s \in \mathbb{N}$ such that $\mathcal{X} \subseteq \{0,1\}^s$, then, given some $x \in \cup_{t\in\mathbb{N}} \mathcal{X}^t \to \mathbb{R}$, one can compute the unique corresponding t via a Turing machine. Moreover, since $\{B_n\}_{n\geq 1}$ is computable, one can further compute the unique n such that $t \in [nB_n, (n+1)B_{n+1})$ via a Turing machine, as well as the empirical probability appearing in (4.122). For any $k \in \mathbb{N}$, and any $x_0 \in \mathcal{X}$, one can call the rational-valued computable upper and lower approximations of p at point (x_0, k). Then, one can go over all $y_{1:n} \in \mathcal{X}^n$, and use the explicit constructions from the proof of Lemma 4.9 regarding the product, sum, and absolute value, yielding rational-valued computable upper and lower approximations of f. ■

We know that $\{A_n\}_{n\geq 1}$ is computable. From Lemma 4.9, the product of two computable functions is computable, thus lower semi-computable, and the ceiling function preserves semi-computability. Therefore, ϕ is lower semi-computable. Since $\phi - 2\log(\phi+3) = \log(2^\phi/(\phi+3)^2)$,

then by Lemma 4.9, for any positive integer L, function $\phi - 2\log(\phi+3) - L$ is lower semi-computable. It remains to prove that a certain choice of L yields a $p_X^{\otimes *}$-critic. From (4.103) and (4.111), there exists $N_0 \in \mathbb{N}$ such that

$$\forall n \geq N_0,\ B_n \geq \lambda(|\mathcal{X}|^n + 2)A_n^2. \tag{4.123}$$

For any $n \geq N_0$, any $C \geq 2$, any integer $t \in [nB_n, (n+1)B_{n+1})$, and any $x_{1:t} \in \mathcal{X}^t$, we have:

$$\begin{aligned}\{\phi(x_{1:t}) \geq C\} &= \left\{\left\lceil A_n \left\| \mathbb{P}^{\text{emp}}_{\mathcal{X}^n}[x_{1:nB_n}] - p^{\otimes n}\right\|_{TV}\right\rceil \geq C\right\} \\ &= \left\{A_n \left\| \mathbb{P}^{\text{emp}}_{\mathcal{X}^n}[x_{1:nB_n}] - p^{\otimes n}\right\|_{TV} > C-1\right\}.\end{aligned}$$

From this, (4.123), and Lemma 4.19, with distribution $p_X^{\otimes n}$, and parameters $b = B_n$, $\varepsilon = (C-1)/A_n, \eta = 2^{-C}$ we obtain,

$$\forall t \geq [N_0 B_{N_0}, \infty) \cap \mathbb{N}, \forall C \in \mathbb{N}_{\geq 2}, \quad p_X^{\otimes t}(\phi(X_{1:t}) \geq C) \leq 2^{-C}.$$

Therefore,

$$\forall t \geq [N_0 B_{N_0}, \infty) \cap \mathbb{N}, \forall C \in \mathbb{N}_{\geq 2}, \quad \mathbb{E}_{p_X^{\otimes t}}[\mathbf{1}_{\phi(X_{1:t})=C} \cdot 2^{\phi(X_{1:t}) - 2\log(\phi(X_{1:t})+3)}] \leq \frac{1}{(C+3)^2}.$$

This also holds for $C \in \{0,1\}$. Summing over $C \in \mathbb{N}_{\geq 0}$ gives,

$$\forall t \geq [N_0 B_{N_0}, \infty) \cap \mathbb{N}, \quad \sum_{x_{1:t} \in \mathcal{X}^t} p_X^{\otimes t}(x_{1:t}) \cdot 2^{\phi(x_{1:t}) - 2\log(\phi(x_{1:t})+3)} \leq 1. \tag{4.124}$$

In order to extend this to all positive integers t, it is sufficient to multiply by 2^{-L} for some L large enough. Therefore, there exists $L \in \mathbb{N}$ such that $\phi - 2\log(\phi+3) - L$ is a lower semi-computable $p_X^{\otimes *}$-critic. This concludes the proof.

4.H On the total variation distance

4.H.1 Proof of Claim 4.16

Let P and Q be any two distributions on the same alphabet. Fix a positive integer B. Then, we have, with the convention $\Pi \otimes \Gamma^{\otimes 0} \equiv \Pi$,

$$\begin{aligned}
\|P^B - Q^B\|_{TV} &= \|\sum_{k=1}^{B} (P^{\otimes(B-k+1)} \otimes Q^{\otimes(k-1)} - P^{\otimes(B-k)} \otimes Q^{\otimes k})\|_{TV} \\
&\leq \sum_{k=1}^{B} \|P^{\otimes(B-k+1)} \otimes Q^{\otimes(k-1)} - P^{\otimes(B-k)} \otimes Q^{\otimes k}\|_{TV} \qquad (4.125) \\
&\leq \sum_{k=1}^{B} \|P^{\otimes(B-k)} \otimes P \otimes Q^{\otimes(k-1)} - P^{\otimes(B-k)} \otimes Q \otimes Q^{\otimes(k-1)}\|_{TV} \\
&\leq \sum_{k=1}^{B} \|P - Q\|_{TV} = B\|P - Q\|_{TV}, \qquad (4.126)
\end{aligned}$$

where (4.125) follows from the triangle inequality for the TVD ; and (4.126) follows form Lemma 1.9 (Appendix 1), with $W = X_{B-k+1}$, $\Pi_W \equiv P$ and $\Gamma_W \equiv Q$.

4.I The birthday paradox

We provide a proof of Claim 4.14. We have

$$\begin{aligned}
(p^{\mathcal{U}}_{[\lfloor 2^{R_1} \rfloor]})^{\otimes B}\big(M^{(1)}, ..., M^{(B)} \text{ 2 by 2 distinct}\big) &= \prod_{k=1}^{B} \frac{\lfloor 2^{R_1} \rfloor - k + 1}{\lfloor 2^{R_1} \rfloor} \\
&\geq \frac{(\lfloor 2^{R_1} \rfloor - B + 1)^B}{\lfloor 2^{R_1} \rfloor^B} \\
&\geq \Big(1 - \frac{B-1}{\lfloor 2^{R_1} \rfloor}\Big)^B \\
&\geq 1 - \frac{B(B-1)}{\lfloor 2^{R_1} \rfloor} \qquad (4.127) \\
&\geq 1 - \frac{B^2}{\lfloor 2^{R_1} \rfloor},
\end{aligned}$$

where (4.127) follows from Bernoulli's inequality, since $R_1 > \log(B)$.

4.J Existence of a universal $p^{\otimes *}$-critic

We provide a proof of Theorem 4.10. From [LV19, Theorem 4.3.1], there exists a sequence $\{q_n\}_{n\geq 1}$ containing all lower semi-computable semi-measures on $\{0,1\}^*$, and a sequence $\{\pi_n\}_{n\geq 1}$ of (strictly) positive reals, such that the mixture defined by

$$\mathbf{m} := \sum_{n\geq 1} \pi_n q_n \tag{4.128}$$

is a lower semi-computable semi-measure on $\{0,1\}^*$. For every $n \in \mathbb{N}$, let $\mathbf{m}(\mathcal{X}^n)$ denote

$$\sum_{x_{1:n}\in\mathcal{X}^n} \mathbf{m}(x_{1:n}).$$

From (4.128), we have $\forall x \in \{0,1\}^*, \mathbf{m}(x) > 0$. Moreover, $\forall x_0 \in \mathcal{X},\ p(x_0) > 0$, thus $\forall x \in \cup_{n\in\mathbb{N}}\mathcal{X}^n$, $p^{\otimes *}(x) > 0$. Define function δ_0, by

$$\forall n \in \mathbb{N}, \forall x_{1:n} \in \mathcal{X}^n,\ \delta_0(x_{1:n}) := \log\Big(\frac{\mathbf{m}(x_{1:n})}{\mathbf{m}(\mathcal{X}^n)p^{\otimes n}(x_{1:n})}\Big). \tag{4.129}$$

Then, for any $n \in \mathbb{N}$, we have

$$\sum_{x\in\mathcal{X}^n} p^{\otimes n}(x)2^{\delta_0(x)} = \sum_{x\in\mathcal{X}^n} \frac{\mathbf{m}(x)}{\mathbf{m}(\mathcal{X}^n)} \leq 1. \tag{4.130}$$

Hence, δ_0 is a $p^{\otimes *}$-critic. Fix any lower semi-computable $p^{\otimes *}$-critic δ. Define map $q_\delta : \{0,1\}^* \to \mathbb{R}$ by

$$\forall x \in \cup_{n\in\mathbb{N}}\mathcal{X}^n, \quad q_\delta(x) := \mathbf{m}(\mathcal{X}^{l(x)})2^{\delta(x)}p^{\otimes *}(x),$$

and $x \mapsto 0$ elsewhere. From Lemma 4.9 (iii), the function which is null outside of $\cup_{n\in\mathbb{N}}\mathcal{X}^n$, and defined by $x \mapsto \mathbf{m}(\mathcal{X}^{l(x)})$ on $\cup_{n\in\mathbb{N}}\mathcal{X}^n$, is lower semi-computable. Moreover, $x \mapsto 2^{\delta(x)}$ and $x \mapsto p^{\otimes *}(x)$ are lower semi-computable by Lemma 4.9 (i) and (iii) respectively. Thus, q_δ

is the product three non-negative lower semi-computable functions. Hence, q_δ is lower semi-computable by Lemma 4.9 (i). Moreover, we have

$$\begin{aligned}\sum_{x\in\{0,1\}^*} q_\delta(x) &= \sum_{n\in\mathbb{N}} \mathbf{m}(\mathcal{X}^n) \sum_{x\in\mathcal{X}^n} 2^{\delta(x)} p^{\otimes n}(x) \\ &\leq \sum_{n\in\mathbb{N}} \mathbf{m}(\mathcal{X}^n) \qquad (4.131)\\ &\leq 1, \qquad (4.132)\end{aligned}$$

where (4.131) follows from the definition of a $p^{\otimes *}$-critic; and (4.132) follows from the fact that $\mathbf{m}$ is a semi-measure. Therefore, q_δ is a lower semi-computable semi-measure. Thus, from (4.128), we have $\mathbf{m} \geq \pi_{q_\delta} q_\delta$, for some positive real π_{q_δ}. In order to derive (4.38), fix $x \in \cup_{n\in\mathbb{N}} \mathcal{X}^n$, and denote $l(x)$ by n. From (4.128), we have $\mathbf{m}(x) > 0$. Therefore, since $\forall x_0 \in \mathcal{X}, p(x_0) > 0$, we have $q_\delta(x) > 0$. Thus, from (4.129), we have

$$\begin{aligned}\delta_0(x) &= \log\Big(\frac{\mathbf{m}(x_{1:n})}{\mathbf{m}(\mathcal{X}^n) p^{\otimes n}(x_{1:n})}\Big) \\ &\geq \log\Big(\frac{\pi_{q_\delta} q(x_{1:n})}{\mathbf{m}(\mathcal{X}^n) p^{\otimes n}(x_{1:n})}\Big) \\ &= \log(\pi_{q_\delta}) + \delta(x). \qquad (4.133)\end{aligned}$$

This is true for any lower semi-computable $p^{\otimes *}$-critic δ, and any $x \in \cup_{n\in\mathbb{N}} \mathcal{X}^n$. Since $\log(\pi_{q_\delta})$ does not depend on x, then property (4.38) holds. This concludes the proof.

Remark 4.22 *As detailed in Remark 4.3, our definition of a $p^{\otimes *}$-critic is an adaptation of classical definitions [LV19, Definitions 2.4.1 & 4.3.8]. In [LV19, Definition 4.3.8], a distribution $\{\pi(n)\}_n$ is assumed to be given. The universal critic under that definition is given by [LV19, Theorem 4.3.5]*

$$\forall n \in \mathbb{N}, \forall x_{1:n} \in \mathcal{X}^n,\ \delta_2(x_{1:n}) := \log\Big(\frac{\mathbf{m}(x_{1:n})}{\tilde{p}(x_{1:n})}\Big), \qquad (4.134)$$

where $\tilde{p}(x_{1:n}) = \pi(n) p^{\otimes n}(x_{1:n})$. One can verify that δ_2 dominates all functions δ which are $p^{\otimes *}$-critics in the sense of Definition 4.1, irrespective of the weights $\{\pi(n)\}_n$. However, δ_2 is not itself a critic according to Definition 4.1 for all choices of $\{\pi(n)\}_n$.

Remark 4.23 *The class of lower semi-computable $p^{\otimes *}$-critics is vast. Numerous examples of intuitive properties of randomness have been covered in Section 4.2. We provide a highly non-trivial example. It is unclear whether δ_0 is lower semi-computable. Lower semi-computability can be ensured by a slight modification of δ_0. Consider the function δ_1 defined by*

$$\forall n \in \mathbb{N}, \forall x_{1:n} \in \mathcal{X}^n,\ \delta_1(x_{1:n}) := \log\Big(\frac{\mathbf{m}(x_{1:n})}{p^{\otimes n}(x_{1:n})}\Big). \tag{4.135}$$

Since $\mathbf{m}$ is lower semi-computable and does not take value 0, then from Lemma 4.9, $\log(\mathbf{m})$ is lower semi-computable. Since $p^{\otimes *}$ is computable and does not take value 0, then from Lemma 4.9, $-\log(p^{\otimes *})$ is lower semi-computable. Thus, δ_1 is lower semi-computable. Moreover, from (4.129), and since $\mathbf{m}$ is a semi-measure, we have $\delta_1 \leq \delta_0$. Thus, δ_1 is a $p^{\otimes *}$-critic. It is not universal: while the restriction of δ_1 to $\mathcal{X}^n$ for a specific n, similarly to the critic in (4.20), would be an interesting, $p^{\otimes n}$-critic, obtaining a $p^{\otimes *}$-critic that dominates all other $p^{\otimes *}$-critics requires to add weights as in (4.129) and (4.134). The lower semi-computable $p^{\otimes *}$-critic δ_1 and the critic in (4.20) are of a fundamentally different nature compared to the other critics presented in Section 4.2, in part because the latter are computable real-valued maps. Indeed, if δ_1 were computable, then since $p^{\otimes *}$ is computable, the universal semi-measure $\mathbf{m}$ would be computable, which is false [LV19, Lemmas 4.3.1 & 4.3.2]. Moreover, the quantity δ_1 has close ties with the Kolmogorov complexity (see Remark 4.11), and thus is similar to the critic in (4.20). This may provide intuition on the nature of δ_1, and on the vastness of the set of lower semi-computable $p^{\otimes *}$-critics.

4.K Additional semi-computability arguments

4.K.1 Proof of Lemma 4.8

(i)–(ii). By Definition 4.6, there are Turing machines returning, on input x, triples (s, a, b) and (s', a', b') in $\{0,1\} \times \mathbb{N}_{\geq 0} \times \mathbb{N}$ with $f(x) = (2s-1)a/b$ and $g(x) = (2s'-1)a'/b'$. Each listed operation is realized by a Turing machine performing finitely many additions, multiplications, comparisons and Euclidean divisions of the non-negative integers a, b, a', b', which are exact operations on finite representations, and returning the result as such a triple. For instance, $f(x) \geq g(x)$ holds iff $(2s-1)ab' \geq (2s'-1)a'b$, a comparison of two integers, and $\lceil f(x) \rceil$ is read off the Euclidean division of a by b.

(iii). Set $\varphi(x,k) := f(x)$ for all $k \in \mathbb{N}$. Then φ is a computable map from $\mathcal{E} \times \mathbb{N}$ to $\mathbb{Q}$, is non-decreasing in k, and converges to $f(x)$. Hence f is lower semi-computable, and the same argument applies to $-f$.

(iv). For $y \geq 0$ and $m \in \mathbb{N}_{\geq 0}$, we have $m \geq y^{p/q}$ iff $m^q \geq y^p$, since $t \mapsto t^q$ is increasing on $\mathbb{R}_{\geq 0}$. Hence $\lceil f(x)^{p/q} \rceil = \min\{m \in \mathbb{N}_{\geq 0} : m^q \geq f(x)^p\}$, and by (i)–(ii) each test $m^q \geq f(x)^p$ is an exact comparison between two rationals; a Turing machine therefore determines this minimum by trying $m = 0, 1, 2, \ldots$, which terminates. The floor is obtained as $\max\{m \in \mathbb{N}_{\geq 0} : m^q \leq f(x)^p\}$.

4.K.2 Notation for the proof of Lemma 4.9

If f is lower semi-computable, we denote by $(x,k) \mapsto \varphi_{f,-}(x,k)$ a computable function from $\mathcal{E}$ to $\mathbb{Q}$, monotonically approaching f from below, in the sense of Definition 4.6. If f is upper semi-computable, then $\varphi_{f,+}(x,k)$ denotes a function of the form $\varphi_{-f,-}(x,k)$, which monotonically approaches f from above.

In all the constructions below, the approximants $\varphi_{\cdot,\pm}$ are computable maps into $\mathbb{Q}$ (Definition 4.6), and we manipulate them by finitely many additions, subtractions, products, quotients, absolute values, maxima, ceilings and comparisons, and by searches for the smallest integer satisfying such a comparison. Every map so obtained is again a computable map into $\mathbb{Q}$, and

every such comparison is decidable, by Lemma 4.8(i)–(ii); the integer-root search used for $f^{1/b}$ is licensed by Lemma 4.8(iv). We use this implicitly in what follows. It is essential that these operations are exact operations on *rationals*: the corresponding operations on the real values $f(x), g(x)$ are unavailable, a comparison such as "$f(x) \geq 0$" being undecidable.

4.K.3 Assume that f and g are computable

$f + g$

Function $\varphi_{f,-} + \varphi_{g,-}$ is a computable function from $\mathcal{E} \times \mathbb{N}$ to $\mathbb{Q}$, which monotonically approaches $f+g$ from below. Similarly, $\varphi_{f,+} + \varphi_{g,+}$ constitutes a computable rational upper approximation.

$|f|$

We construct $\varphi_{|f|,-}$ as follows. Let $x \in \mathcal{E}$ and $k \in \mathbb{N}$. If $\varphi_{f,-}(x,k) \geq 0$, return $|\varphi_{f,-}(x,k)|$. Otherwise, if $\varphi_{f,+}(x,k) \leq 0$, return $|\varphi_{f,+}(x,k)|$. Otherwise, return 0. We define $\varphi_{|f|,+}(x,k)$ as

$$\max\Big(|\varphi_{f,-}(x,k)|, |\varphi_{f,+}(x,k)|\Big).$$

Straightforwardly, this implies that $|f|$ is computable.

fg

Define $\varphi_{fg,-}(x,k)$ as follows. If $\varphi_{f,-}(x,k) \geq 0$ and $\varphi_{g,-}(x,k) \geq 0$, then return the product $\varphi_{f,-}(x,k)\varphi_{g,-}(x,k)$. Otherwise, if $\varphi_{f,+}(x,k) \leq 0$ and $\varphi_{g,+}(x,k) \leq 0$, then return the product $\varphi_{f,+}(x,k)\varphi_{g,+}(x,k)$. Otherwise, return

$$-\max\Big(|\varphi_{f,-}(x,k)|, |\varphi_{f,+}(x,k)|\Big) \max\Big(|\varphi_{g,-}(x,k)|, |\varphi_{g,+}(x,k)|\Big).$$

Define $\varphi_{fg,+}$ as $-\varphi_{(-f)g,-}$.

4.K.4 Suppose that f is computable and only takes positive values

$1/f$

Define $\varphi_{1/f,-}(x,k)$ as $1/\varphi_{f,+}(x,k)$. Compute $k_1(x)$, the smallest positive integer k such that $\varphi_{f,-}(x,k) > 0$. For all integers $k \in [1, k_1(x)]$, define $\varphi_{1/f,+}(x,k)$ as $1/\varphi_{f,-}(x,k_1(x))$. For all integers $k \in (k_1(x), \infty)$, define $\varphi_{1/f,+}(x,k)$ as $1/\varphi_{f,-}(x,k)$.

$\log(f)$

From the above paragraph, we know that f and $1/f$ are lower semi-computable. Since these functions only take positive values, and $-\log(f) = \log(1/f)$, then it is sufficient to show that for any lower semi-computable function g taking only positive values, the function $\log(g)$ is lower semi-computable. This property is part of point (iii) of Lemma 4.9, and is proved at the end of Appendix 4.K.6.

$f^{1/b}$

Compute $k_1(x)$, the smallest positive integer k such that $\varphi_{f,-}(x,k) > 0$. For all integers $k \in [1, k_1(x))$, define $\varphi_{f^{1/b},-}(x,k) := 0$ and $\varphi_{f^{1/b},+}(x,k) = \lceil \varphi_{f,+}(x,k) \rceil$. Consider an integer $k \in [k_1(x), \infty)$. Compute the greatest integer m such that $(m/2^k)^b \leq \varphi_{f,-}(x,k)$. Then, define $\varphi_{f^{1/b},-}(x,k) := m/2^k$. Therefore, we have

$$\forall k \geq k_1(x), \quad 0 \leq \varphi_{f,-}(x,k)^{1/b} - \varphi_{f^{1/b},-}(x,k) < \frac{1}{2^k}. \tag{4.136}$$

From (4.136), and since the b-th root function and $k \mapsto \varphi_{f,-}(x,k)$ are both non-decreasing, we have

$$\forall k \geq k_1(x) + 1, \quad \varphi_{f^{1/b},-}(x,k-1) \leq \varphi_{f,-}(x,k)^{1/b}. \tag{4.137}$$

Since $\varphi_{f^{1/b},-}(x,k-1)$ can also be written in the form $m'/2^k$, then, from the maximality of the integer m appearing in the construction of $\varphi_{f^{1/b},-}(x,k)$, we have

$$\forall k \geq k_1(x)+1, \quad \varphi_{f^{1/b},-}(x,k-1) \leq \varphi_{f^{1/b},-}(x,k). \tag{4.138}$$

This also holds for all integers $k \in [2, k_1(x)+1)$. Properties (4.136) and (4.138) imply that $f^{1/b}$ is lower semi-computable. We prove upper semi-computability similarly, using the smallest integer $\tilde{m}$ such that $(\tilde{m}/2^k)^b \geq \varphi_{f,+}(x,k)$, and setting $\varphi_{f^{1/b},+}(x,k) := \tilde{m}/2^k$.

4.K.5 Assume that f and g are lower semi-computable

$f+g$

Function $\varphi_{f,-}+\varphi_{g,-}$ is a computable function from $\mathcal{E}\times\mathbb{N}$ to $\mathbb{Q}$, which monotonically approaches $f+g$ from below.

$\lceil f \rceil$

Define $\varphi_{\lceil f \rceil,-}$ as $\lceil \varphi_{f,-} \rceil$.

2^f

Fix $x \in \mathcal{E}$ and $k \in \mathbb{N}$. Let $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ such that $\varphi_{f,-}(x,k) = a/b$. Compute the greatest integer m such that $(m/2^k)^b \leq 2^a$. Then, define $\varphi_{2^f,-}(x,k) := m/2^k$. Therefore, we have

$$0 \leq 2^{\varphi_{f,-}(x,k)} - \varphi_{2^f,-}(x,k) < \frac{1}{2^k}. \tag{4.139}$$

From (4.139), and since the exponential function and $k \mapsto \varphi_{f,-}(x,k)$ are both non-decreasing, we have

$$\forall k \geq 2, \ \varphi_{2^f,-}(x,k-1) \leq 2^{\varphi_{f,-}(x,k)}. \tag{4.140}$$

Since $\varphi_{2^f,-}(x,k-1)$ can also be written in the form $m'/2^k$, then, from the maximality of the integer m appearing in the construction of $\varphi_{2^f,-}(x,k)$, we have

$$\forall k \geq 2,\ \varphi_{2^f,-}(x,k-1) \leq \varphi_{2^f,-}(x,k). \tag{4.141}$$

Properties (4.139) and (4.141) imply that 2^f is lower semi-computable.

4.K.6 Assume that f and g are semi-computable and non-negative

fg

If $\varphi_{f,-}(x,k) \geq 0$ and $\varphi_{g,-}(x,k) \geq 0$, return $\varphi_{f,-}(x,k)\varphi_{g,-}(x,k)$. Otherwise, return 0.

$2^f/(3+f)^2$

There exists a real $\varepsilon \in (0,1)$ such that $u \mapsto 2^u/(3+u)^2$ is non-decreasing on $(-\varepsilon,\infty)$. Fix $x \in \mathcal{E}$. Compute $k_1(x)$, the smallest positive integer k such that $\varphi_{f,-}(x,k) > -\varepsilon$. For all integers $k \in [1,k_1(x))$, define $\varphi_{2^f/(3+f)^2,-}(x,k) := 0$. Fix an integer $k \geq k_1(x)$. Let $a \in \mathbb{Z}$ and $b \in \mathbb{N}$ such that $\varphi_{f,-}(x,k) = a/b$. Compute the greatest integer m such that $(m/2^k)^b \leq 2^a/(3+a/b)^{2b}$. Then, define $\varphi_{2^f/(3+f)^2,-}(x,k) := m/2^k$. Therefore, we have

$$\forall k \geq k_1(x),\quad 0 \leq \frac{2^{\varphi_{f,-}(x,k)}}{(3+\varphi_{f,-}(x,k))^2} - \varphi_{2^f/(3+f)^2,-}(x,k) < \frac{1}{2^k}. \tag{4.142}$$

From (4.142), and since $k \mapsto \varphi_{f,-}(x,k)$ is non-decreasing, and $u \mapsto 2^u/(3+u)^2$ is non-decreasing on $(-\varepsilon,\infty)$, we have

$$\forall k \geq k_1(x)+1,\quad \varphi_{2^f/(3+f)^2,-}(x,k-1) \leq \frac{2^{\varphi_{f,-}(x,k)}}{(3+\varphi_{f,-}(x,k))^2}. \tag{4.143}$$

Since $\varphi_{2^f/(3+f)^2,-}(x,k-1)$ can also be written in the form $m'/2^k$, then, from the maximality of the integer m appearing in the construction of $\varphi_{2^f/(3+f)^2,-}(x,k)$, we have

$$\forall k \geq k_1(x)+1, \quad \varphi_{2^f/(3+f)^2,-}(x,k-1) \leq \varphi_{2^f/(3+f)^2,-}(x,k). \tag{4.144}$$

This is also true for all integers $k \in [2, k_1(x)+1)$. Properties (4.142) and (4.144) imply that $2^f/(3+f)^2$ is lower semi-computable.

$\log(f)$

Assume that f only takes positive values. Fix $x \in \mathcal{E}$. Compute $k_1(x)$, the smallest positive integer k such that $\varphi_{f,-}(x,k) > 0$. Fix an integer $k \geq k_1(x)$. Compute the largest integer m such that $2^m \leq \varphi_{f,-}(x,k)^{2^k}$. Then, define $\varphi_{\log(f),-}(x,k) := m/2^k$. For all integers $k \in [1, k_1(x))$, define $\varphi_{\log(f),-}(x,k)$ as $\varphi_{\log(f),-}(x,k_1(x))$. Therefore, we have

$$\forall k \geq k_1(x), \quad 0 \leq \log(\varphi_{f,-}(x,k)) - \varphi_{\log(f),-}(x,k) < \frac{1}{2^k}. \tag{4.145}$$

From (4.145), and since the logarithm and $k \mapsto \varphi_{f,-}(x,k)$ are both non-decreasing, we have

$$\forall k \geq k_1(x)+1, \quad \varphi_{\log(f),-}(x,k-1) \leq \log(\varphi_{f,-}(x,k)). \tag{4.146}$$

Since $\varphi_{\log(f),-}(x,k-1)$ can also be written in the form $m'/2^k$, then, from the maximality of the integer m appearing in the construction of $\varphi_{\log(f),-}(x,k)$, we have

$$\forall k \geq k_1(x)+1, \quad \varphi_{\log(f),-}(x,k-1) \leq \varphi_{\log(f),-}(x,k). \tag{4.147}$$

This also holds for all integers $k \in [2, k_1(x)+1)$. Properties (4.145) and (4.147) imply that $\log(f)$ is lower semi-computable.

4.K.7 Functions of finite binary strings

Let $\mathcal{X}$ be a finite computable subset of $\{0,1\}^*$, and f be a lower semi-computable function from $\{0,1\}^*$ into $\mathbb{R}$.

Lemma 4.24 *$\cup_{n\in\mathbb{N}}\mathcal{X}^n$ is a computable set.*

Proof: By Definition 4.6, it is sufficient to construct a computable function τ from $\{0,1\}^*$ to $\{0,1\}$, which returns 1 if its input is in $\cup_{n\in\mathbb{N}}\mathcal{X}^n$, and 0 otherwise. Since $\mathcal{X}$ is computable, there exists a computable function τ_0 from $\{0,1\}^*$ to $\{0,1\}$, which returns 1 if its input is in $\mathcal{X}$, and 0 otherwise. Fix $x \in \{0,1\}^*$. Define $\tau(x)$ as follows. Enumerate all partitions of x into consecutive sub-strings. For each, call τ_0 on every sub-string. If for some partition, the output of τ_0 is 1 for every sub-string, then return 1. Otherwise, return 0. ■

Hereafter, we use the notation τ defined in the above proof.

Partial sums

Consider the function $\tilde{f} : \{0,1\}^* \to \mathbb{R}$ which is null outside of $\cup_{n\in\mathbb{N}}\mathcal{X}^n$, and is defined by

$$\forall x \in \cup_{n\in\mathbb{N}}\mathcal{X}^n, \quad \tilde{f}(x) = \sum_{y\in\mathcal{X}^{l(x)}} f(y).$$

Fix $x \in \{0,1\}^*$ and $k \in \mathbb{N}$. Define $\varphi_{\tilde{f},-}(x,k)$ as follows. Compute $\tau(x)$. If it is null, return 0. Otherwise: compute $l(x)$, and for each y in $\mathcal{X}^{l(x)}$, compute $\varphi_{f,-}(y,k)$, then return

$$\sum_{y\in\mathcal{X}^{l(x)}} \varphi_{f,-}(y,k).$$

Fix some $x \in \cup_{n\in\mathbb{N}}\mathcal{X}^n$. The (finite) set of indices of the above sum does not depend on k. Therefore, since for each $y \in \{0,1\}^*$ we have $\varphi_{f,-}(y,k) \underset{k\to\infty}{\to} f(y)$, we get

$$\forall x \in \cup_{n\in\mathbb{N}}\mathcal{X}^n, \quad \varphi_{\tilde{f},-}(x,k) \underset{k\to\infty}{\longrightarrow} \tilde{f}(x). \tag{4.148}$$

Similarly, since for any $y \in \{0,1\}^*$ and any $k \geq 1$, we have $\varphi_{f,-}(y,k) \leq \varphi_{f,-}(y,k+1)$, then we have

$$\forall x \in \cup_{n\in\mathbb{N}}\mathcal{X}^n, \forall k \in \mathbb{N}, \quad \varphi_{\tilde{f},-}(x,k) \leq \varphi_{\tilde{f},-}(x,k+1). \tag{4.149}$$

Properties (4.148) and (4.149) also hold for finite strings outside of $\cup_{n\in\mathbb{N}}\mathcal{X}^n$. Thus, $\tilde{f}$ is lower semi-computable.

Product distribution

Let p be a lower semi-computable probability measure on $\mathcal{X}$. Fix $x \in \{0,1\}^*$ and $k \in \mathbb{N}$. Define $\varphi_{p^{\otimes *},-}(x,k)$ as follows. Compute $\tau(x)$. If it is null, return 0. Otherwise, proceed as follows. Compute $l(x)$. We write x as $x_{1:l(x)}$, with $x_t \in \mathcal{X}$ for any integer t in $[1, l(x)]$. Compute and return

$$\prod_{t=1}^{l(x)} \varphi_{p,-}(x_t,k).$$

Fix some $x \in \cup_{n\in\mathbb{N}}\mathcal{X}^n$. The (finite) set of indices of the above product does not depend on k. Therefore, since for each $y \in \mathcal{X}$, we have $\varphi_{p,-}(y,k) \underset{k\to\infty}{\to} p(y)$, we get

$$\forall x \in \cup_{n\in\mathbb{N}}\mathcal{X}^n, \quad \varphi_{p^{\otimes *},-}(x,k) \underset{k\to\infty}{\longrightarrow} p^{\otimes *}(x). \tag{4.150}$$

Similarly, since for any $y \in \mathcal{X}$, and any $k \geq 1$, we have $\varphi_{p,-}(y,k) \leq \varphi_{p,-}(y,k+1)$, and since p is non-negative, then we have

$$\forall x \in \cup_{n\in\mathbb{N}}\mathcal{X}^n, \forall k \in \mathbb{N}, \quad \varphi_{p^{\otimes *},-}(x,k) \leq \varphi_{p^{\otimes *},-}(x,k+1). \tag{4.151}$$

Properties (4.150) and (4.151) also hold for finite strings outside of $\cup_{n\in\mathbb{N}}\mathcal{X}^n$. Thus, $p^{\otimes *}$ is lower semi-computable.

Chapter 5

Channel Synthesis Based Non-interactive Remote Coordination

5.1 Introduction

Achieving high-level coordination among distributed agents is essential for the agents to accomplish real-world tasks from policy making [H+24b], to distributed learning [C+21], cooperative game theory, and autonomous driving. However, in many applications, such as teams of robots or drones, agents to be coordinated are controlled remotely through rate-limited communication channels. In Section 1.3 of Chapter 1, we illustrated this through a soccer example: in a fast-paced match, a coach cannot perform lengthy communication with a given player, i.e., the communication is rate-limited. In the present chapter, we consider coordination among agents that do not interact directly. In the football example, players sometimes have to guess their teammates' next actions without communication, and act accordingly. Such coordination is made possible by a long history of training with teammates, and agreement upon a joint strategy. Instead of communication, each controller/encoder may have a sensor that provides observations of the environment — such as a coach and a team captain who each analyze the situation and give instructions to team members —, and sends information to its remote agent via the constrained communication link. Here, coordination is enabled through the correlation

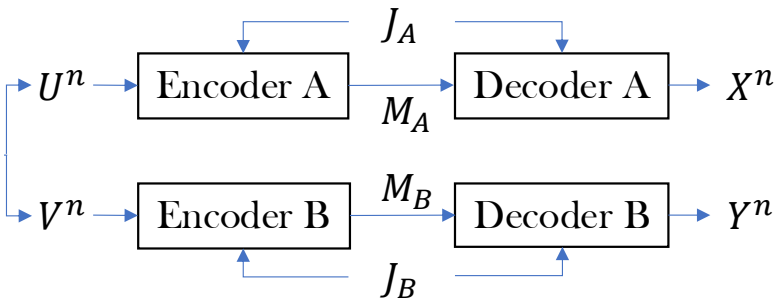


Figure 5.1: The system model.

among the sensor signals — e.g., what the coach and the captain observe are correlated. This case encompasses various machine learning problems, such as the generation of coordinated policies among distributed agents in multi-agent reinforcement learning problems, or neural based distributed compression of correlated images with the goal of realism , where the realism will require the reconstructed images to be realistic with respect to the joint distribution of the original images — similarly to the joint realism constraint studied in Chapter 3.

Our system model is illustrated in Figure 5.1. Here, two non-interacting controllers — e.g., the coach and the captain —, called Encoder A and Encoder B, observe two correlated source sequences, U^n and V^n, respectively. The goal of the encoders is to help two agents, called Decoder A and Decoder B, to achieve coordination of their respective actions X^n and Y^n by communicating with them over rate-limited channels. Encoder A maps input U^n to a message M_A of nR_A bits, which Decoder A uses to generate its output X^n. In parallel, Encoder B maps V^n into a message M_B of nR_B bits, used by Decoder B to output Y^n. Moreover, similarly to previous chapters, each encoder-decoder pair has access to a common randomness at a limited rate ($R_{c,A}$ and $R_{c,B}$, respectively), in the form of uniform random variables (J_A and J_B).

The problem of generating samples from a target joint distribution in a distributed manner, known as *distributed coordination*, has long been of interest [Wyn75, HV93, GK73]. Several notions of common information [Wyn75, GK73] have been shown to play a fundamental role in such problems. The general information theoretic formulation involves a network of terminals, connected by noisy or rate-limited noiseless channels — in our problem the channels are noiseless. Terminals may have access to correlated source sequences, and certain terminals

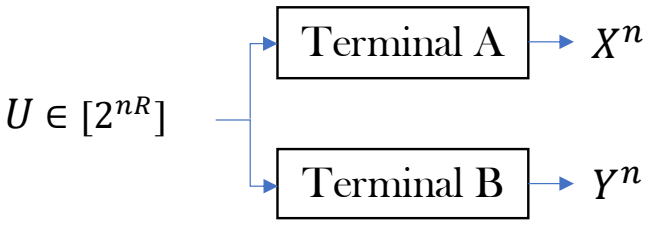


Figure 5.2: Wyner's setup [Wyn75].

are expected to output either a single symbol or a sequence of symbols. Coordination among these sequences involves their joint distribution P and a target joint distribution q. Different kinds of coordination constraints have been considered. Let $\mathcal{D}$ denote a divergence. As in the distribution matching formalism for realism studied in previous chapters, a constraint of the form

$$\mathcal{D}(P_{X^{(1)}_{1:n},\cdots,X^{(k)}_{1:n}}, q^{\otimes n}_{X^{(1)},\cdots,X^{(k)}}) \underset{n\to\infty}{\longrightarrow} 0 \tag{5.1}$$

with $k \geq 2$ is called a *strong coordination constraint*. An *empirical coordination constraint* is of the form

$$\mathcal{D}\Big(\mathbb{P}^{\text{emp}}_{\mathcal{X}^{(1)}\times\cdots\times\mathcal{X}^{(k)}}(X^{(1)}_{1:n},\cdots,X^{(k)}_{1:n}), q_{X^{(1)},\cdots,X^{(k)}}\Big) \underset{n\to\infty}{\xrightarrow{\mathcal{P}}} 0, \tag{5.2}$$

where for any $i \in [k]$, $X^{(i)}_{1:n}$ is a random sequence with alphabet denoted $\mathcal{X}^{(i)}$, and $\xrightarrow{\mathcal{P}}$ stands for convergence in probability[1]. Constraints of the form

$$\mathcal{D}(P_{X^{(1)},\cdots,X^{(k)}}, q_{X^{(1)},\cdots,X^{(k)}}) \underset{n\to\infty}{\longrightarrow} 0 \tag{5.3}$$

have also been considered, and are called *correlation distillation* constraints in the case where q satisfies $X^{(1)} = X^{(2)} = \cdots = X^{(k)}$. For each of the above constraints, *exact coordination* refers to the additional requirement that the divergence be null for large enough n — similarly to perfect realism (Definition 2.4 in Chapter 2). Constraints where the divergence is not required to vanish, but simply to be minimized, have also been considered [C+14].

[1]We say that $\{Y_n\}_{n\geq 1}$ converges to 0 in probability if for any $\varepsilon > 0$, we have $\mathbb{P}(|Y_n| \geq \varepsilon) \to 0$.

In the distributed coordination literature, a first line of work involves the coordination of sources with outputs [B$^+$02, C$^+$10, GA11, Cuf13, YGA15, C$^+$18, YT20, S$^+$20, A$^+$23, LL24, MP24, C$^+$24, SW24]. In the simplest formulation, an encoder having access to a source X^n sends a message to a decoder over a noiseless rate-limited channel, and the latter outputs Y^n, coordinated with X^n. It is also assumed that both terminals have access to common randomness independent form X^n. If strong coordination between X^n and Y^n is required, the problem is called *channel synthesis* or *channel synthesis* or *the reverse Shannon problem* [B$^+$02, Cuf13, B$^+$14, YT20, SW24] — the name *relative entropy coding* is also used for similar formulations in the case $n = 1$ [F$^+$20, HFHL24]. Note the similarity with the point-to-point setup of lossy compression with a near-perfect strong realism constraint studied in Chapter 2. The difference is that the coordination constraint involves the joint distribution of X^n and Y^n, instead of their marginal distributions. The problem of channel synthesis has been shown to be applicable to a broad range of lossy compression problems, as an alternative to simpler methods such as rounding to the nearest integer. In federated learning, channel synthesis has been shown to allow state-of-the-art communication costs while providing privacy guarantees [TRL21, HG24, H$^+$24a, SLL24, LCOL24]. In neural network based lossy source coding, it provides a differentiable alternative to standard rounding which has the merit of providing more control on the compression noise and of making use of any potential common randomness between the encoder and the decoder [H$^+$19, F$^+$20, AT20, TSHM22]. Channel synthesis is central to the non-interactive remote coordination schemes proposed in this chapter. We provide information on the role of common randomness in channel synthesis in Section 5.3.1. In Section 5.4, we study a problem which we coin remote channel synthesis. The latter is the key ingredient of the main scheme we propose. Please refer to [Li24] for a comprehensive overview of the channel synthesis problem and its applications.

Another line of work considers the coordination of decoder outputs, such as in our problem. A first subset of these involve a single source available at a single encoder, which sends a common message to several decoders [Wyn75, C$^+$10, GA11, Cuf13, K$^+$14, S$^+$20, K$^+$21]. Wyner [Wyn75] considered the distributed generation task depicted in Figure 5.2 — two terminals

with access to nR common i.i.d. fair bits independently generate correlated outputs X^n and Y^n, which are to satisfy a strong coordination constraint. The optimal rate R is known as the Wyner common information $C_{\text{Wyner}}(X,Y)$ between X and Y, which is a function of the target distribution $q_{X,Y}$. The attentive reader might notice the similarity between the setup in Figure 5.2 and in the soft covering lemma used in all the achievability proofs in previous chapters — see, e.g., Lemma 3.33 and Figure 3.3 in Chapter 3. Indeed, the results of Wyner [Wyn75] constitute the first instance of a soft covering lemma.

A second subset of works where the aim is to coordinate outputs involve non-interacting terminals taking as input distinct correlated sources — as is the case for Encoders A and B in our problem (Figure 5.1) —, and producing coordinated outputs [GK73, Wit75, AB07, KA16, S$^+$20, B$^+$21]. The first such problem, depicted in Figure 5.3, was studied by Gács and Körner [GK73]. Two non-interacting terminals have access to distinct source sequences U^n and V^n with joint distribution $q_{U,V}^{\otimes n}$ and each output $\lfloor nR \rfloor$ symbols, on which the strong coordination constraint

$$\| P_{X^{\lfloor nR \rfloor}, Y^{\lfloor nR \rfloor}} - q_{X,Y}^{\otimes \lfloor nR \rfloor} \|_{TV} \underset{n \to \infty}{\longrightarrow} 0$$

is imposed. The target distribution $q_{X,Y}$ satisfies $q(\{X = Y\}) = 1$ and $q_X \equiv Ber(1/2)$. The maximal achievable rate R is called the Gács-Körner (GK) common information $C_{GK}(U,V)$ of $q_{U,V}$ [GK73]. Further results with a correlation distillation constraint were obtained by Witsenhausen [Wit75]. Kamath and Anantharam [KA16] proposed a generalization to any target distribution q with a constraint in the form of (5.3), and mentioned the possibility of considering a strong coordination constraint such as in Figure 5.3. This was studied by Berg et. al. in the case $R = 1$, which is a special case of our problem (Figure 5.1) with $R_A = R_B = \infty$. For each of these types of coordination constraints, the problem of characterizing the achievable rates and target distributions remains open. Kamath and Anantharam proved necessary conditions of the form $\Gamma(U,V) \geq \Gamma(X,Y)$ with Γ being the maximal correlation — also used in [Wit75] — or a quantity related to the hypercontractivity ribbon [Bon68]. It was later noted that this

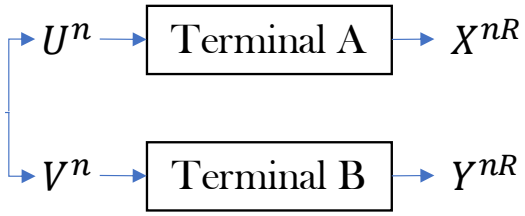


Figure 5.3: Non-interactive coordination problem, which subsumes those studies by Gács and Körner [GK73] and Berg et. al. [B$^+$21].

extends to a larger class of measures of correlation [BG18, B$^+$21, S$^+$20]. These necessary conditions also apply to the problem of Berg et. al. The latter designed non-trivial coding schemes by combining results on the Wyner common information and the GK common information — as discussed in Section 5.6 to follow.

The chapter is organised as follows. In Section 5.2, we provide a formal description of our problem. In Section 5.3, we present a first scheme based on channel synthesis, and discuss its sub-optimality. For the sake of developing a better scheme, we study a fundamental problem which we call *remote channel synthesis* in Section 5.4. We highlight the fact that the optimal schemes for that problem may be different to the optimal schemes for (standard) channel synthesis. We show that this is the case when the amount of available common randomness is small enough. In Section 5.5, we present our main scheme for the non-interactive remote coordination problem based on remote channel synthesis. In Section 5.6, we describe a promising direction to improve our scheme, based on GK common information. In Section 5.7, we derive relatively simple outer bounds by relating our problem to channel synthesis problems, and show that these are tight in a special case.

5.2 Problem formulation

Let $\mathcal{U}, \mathcal{V}, \mathcal{X}, \mathcal{Y}$ be four finite alphabets, $q_{U,V}$ a distribution on $\mathcal{U} \times \mathcal{V}$, and $q_{X,Y}$ a distribution on $\mathcal{X} \times \mathcal{Y}$. As shown in Figure 5.1, we consider two encoder-decoder pairs, with no communication across pairs, and with each pair having access to a rate-limited uniformly distributed common randomness. Each encoder can send a rate-limited message to the corresponding decoder. Pri-

vate randomness is unconstrained. As such, a coding scheme for this problem is a combination of two instances of a point-to-point coding scheme, as in Definition 2.3 in Chapter 2. This is formalized in the definition below. Similarly to Chapters 3 and 4, we write (n, R, R_c) code instead of (n, R, R_c, ∞) fixed-length code (Definition 2.3). More precisely, we write (n, R, R_c) code for $(\mathcal{X}, \mathcal{Y})$ when the alphabets of the input and output sequences are $\mathcal{X}$ and $\mathcal{Y}$, respectively — in all previous chapters the reconstruction sequences were required to have the same alphabet as the source sequences.

Definition 5.1 *Given a quadruplet $(R_A, R_{c,A}, R_B, R_{c,B}) \in \mathbb{R}^4_{\geq 0}$, an $(n, R_A, R_{c,A}, R_B, R_{c,B})$ code is a quadruplet $(F^{(A,n)}, G^{(A,n)}, F^{(B,n)}, G^{(B,n)})$ such that $(F^{(A,n)}, G^{(A,n)})$ (resp. $(F^{(B,n)}, G^{(B,n)})$) is an $(n, R_A, R_{c,A})$ (resp. $(n, R_B, R_{c,B})$) code for $(\mathcal{U}, \mathcal{X})$ (resp. $(\mathcal{V}, \mathcal{Y})$). The distribution induced by the latter code is given by*

$$P^{(n)}_{U^n,V^n,J_A,J_B,M_A,M_B,X^n,Y^n} := q^{\otimes n}_{U,V} \cdot p^{\mathcal{U}}_{[2^{\lfloor nR_{c,A} \rfloor}]} \cdot p^{\mathcal{U}}_{[2^{\lfloor nR_{c,B} \rfloor}]} \cdot F^{(A,n)}_{M_A|U^n,J_A} \cdot F^{(B,n)}_{M_B|V^n,J_B} \cdot G^{(A,n)}_{X^n|J_A,M_A} \cdot G^{(B,n)}_{Y^n|J_B,M_B}.$$

As formalized in the following definition, the only requirement for achievability is a strong coordination constraint. In particular, no distortion constraint is imposed — unlike in previous chapters.

Definition 5.2 *For a given target distribution $q_{X,Y}$, a tuple $(R_A, R_{c,A}, R_B, R_{c,B}) \in \mathbb{R}^4_{\geq 0}$ is said to be achievable if there exists a sequence of $(n, R_A, R_{c,A}, R_B, R_{c,B})$ codes with induced distributions $\{P^{(n)}\}_n$ such that*

$$\|P^{(n)}_{X^n,Y^n} - q^{\otimes n}_{X,Y}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{5.4}$$

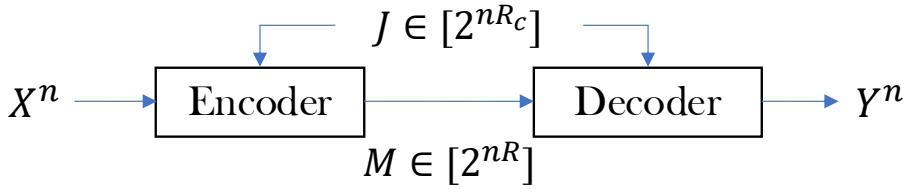


Figure 5.4: Block coding setup for channel synthesis, a.k.a. the reverse channel coding problem.

5.3 A natural scheme based on channel synthesis

5.3.1 The role of common randomness in channel synthesis

The problem of channel synthesis was briefly described in Section 5.1, and is as depicted in Figure 5.4. We only refer to the notation in the latter in the present paragraph. The goal in the channel synthesis problem is to determine the smallest compression rate R in bits per source symbol sufficient to satisfy the following strong coordination requirement

$$\mathcal{D}(P_{X^n,Y^n}, q_{X,Y}^{\otimes n}) \underset{n\to\infty}{\longrightarrow} 0,$$

where $q_{X,Y}$ is a target joint distribution. The answer has been shown to depend on the amount of available common randomness — as in Chapter 2, where we mentioned that common randomness is crucial under the strong near-perfect realism constraint, but is not necessary to achieve optimal performance under the per-symbol near-perfect realism constraint. The problem in the absence of common randomness was first studied in [Wyn75], under a slightly different formulation. As mentioned in Section 5.1, in that case, the optimal rate is the Wyner common information $C_q(X,Y)$. The case of unlimited common randomness was considered in [B$^+$02, Win02], where the optimal rate R is found to be the Shannon mutual information $I_q(X;Y)$. Bounds on the optimal average coding length in one-shot settings have been obtained in [H$^+$07, LEG18]. The role of common randomness is characterized in [BDH$^+$14, Cuf13] — see also [C$^+$10].

5.3.2 A simple scheme for non-interactive remote coordination

A naive approach would be to produce X^n (resp. Y^n) at Encoder A (resp. B) and send it losslessly to Decoder A (resp. B). If the communication rates R_A and R_B are not high enough, then one could think of performing lossy compression of U^n (resp. V^n). The problem is then: how to choose compression schemes so that as much correlation as possible can carry over from (U^n, V^n) to Decoders A and B? Note that lossy compression is a form of *noise addition*. The question is thus: how to choose the optimal compression noise? Channel synthesis appears as a natural tool, to implement certain choices of noise addition — i.e., specific target conditional distributions —, while respecting communication constraints. Encoder A and Decoder A can simulate a channel $p_{X|U}$, while Encoder B and Decoder B simulate a channel $p_{Y|V}$, with the following rates.

Proposition 5.1 *All tuples in the interior of the following region $\mathcal{S}^{(a)}$ are achievable.*

$$\left\{ \begin{array}{rcl} (R_A, R_{c,A}, R_B, R_{c,B}) & : & \exists\, p_{U,V,L_A,L_B,X,Y} \in \mathcal{D}^{(a)}, \\ R_A & \geq & I_p(U; L_A) \\ R_A + R_{c,A} & \geq & I_p(U, X; L_A) \\ R_B & \geq & I_p(V; L_B) \\ R_B + R_{c,B} & \geq & I_p(V, Y; L_B) \end{array} \right\},$$

with $\mathcal{D}^{(a)}$ defined as

$$\left\{ \begin{array}{c} p_{U,V,L_A,L_B,X,Y} : p_{X,Y} \equiv q_{X,Y},\ p_{U,V} \equiv q_{U,V}, \\ X - L_A - U - V - L_B - Y \\ |\mathcal{L}_A| \leq |\mathcal{X}||\mathcal{U}| + 1 \\ |\mathcal{L}_B| \leq |\mathcal{Y}||\mathcal{V}| + 1 \end{array} \right\}.$$

This follows directly from studies on the cost of channel synthesis, e.g., [Cuf13, Theorem II.1].

5.3.3 Sub-optimality

The channel synthesis based scheme employed to achieve the region in Proposition 5.1 exhibits certain weaknesses. Indeed, while the goal in the non-interactive remote coordination problem is only to produce jointly i.i.d. actions X^n and Y^n according to some target joint distribution, the channel synthesis based schemes are such that all four variables U^n, V^n, X^n, Y^n are jointly i.i.d.. This suggests that such schemes may be sub-optimal, and raises the following question: how to design schemes for which X^n and Y^n jointly i.i.d. but not all of U^n, V^n, X^n, Y^n? We provide an answer that involves a fundamental coordination problem which we coin *remote channel synthesis*, that we present and study in the following section.

5.4 Remote Channel Synthesis

5.4.1 Outline

As mentioned in Section 5.3.1, the optimal compression rate for channel synthesis depends on the amount of available common randomness. The role of the latter is studied in [Cuf13]. Therein, the minimum rate of common randomness that is necessary to reach the optimal compression rate $I_q(X;Y)$ was coined *necessary conditional entropy*. Notably, it is often equal to the Shannon conditional entropy $H_q(Y|X)$, and thus often larger than the optimal compression rate $I_q(X;Y)$. This poses a significant implementation challenge. Therefore, it is more practically relevant to assume that the rate of common randomness is limited. See also [F$^+$20, FMHL22, Fla23, FMHL23, FT23, HFHL24] for progress on the computational efficiency of algorithms for one-shot versions of the problem.

In this section, we consider the problem of *remote channel synthesis* (r.c.s.), depicted in Figure 5.5. It differs from the the standard channel synthesis problem (Figure 5.4) in that the encoder only has access to a partial or noisy version Z^n of the remote sequence X^n, such that (X^n, Z^n) follows an i.i.d. distribution $q_{X,Z}^{\otimes n}$. The goal is to achieve strong coordination between the remote source X^n and the decoder output Y^n. This problem bears a resemblance with the

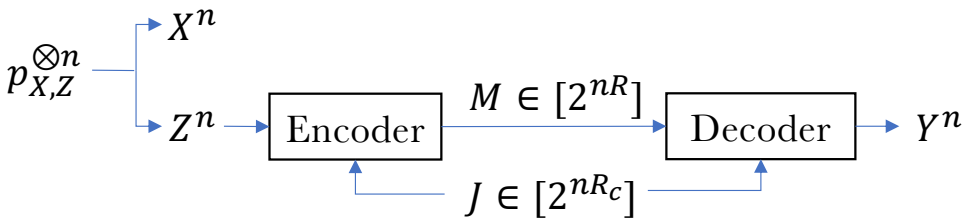


Figure 5.5: Our remote channel synthesis setup.

remote RD problem [DT62, WZ70], [Ber71, Chap. 3, Sec. 5], [Wit80], in which a constraint of the form

$$\mathbb{E}[d(X^n, Y^n)] \leq \Delta$$

where d is a measure of distortion, is imposed instead of strong coordination. We obtain a single-letter characterization (Theorem 5.3 to follow) of the optimal rate pairs (R, R_c), where R denotes the compression rate and R_c the common randomness rate. We find that in the more practically relevant setting with a small R_c, coordinating Z^n and Y^n such that

$$\mathcal{D}(P_{Z^n,Y^n}, q_{Z,Y}^{\otimes n}) \underset{n\to\infty}{\longrightarrow} 0 \tag{5.5}$$

for some $q_{Z,Y}$ is sub-optimal. In other words, in order to achieve the optimal rate R, it is necessary — in most cases — to employ a *vector scheme*, in the sense that the joint distribution of Z^n and Y^n cannot be arbitrarily close to an i.i.d. distribution $q_{Z,Y}^{\otimes n}$ — or more generally a product distribution (Proposition 5.6 to follow).

A channel coding problem with a name similar to remote channel synthesis has been studied in [COS20]. While that channel coding problem involves a coordination constraint similar to our remote coordination constraint, the problem is fundamentally different: the (channel coding) rate is a function of the target joint distribution, while our problem consists in characterizing the optimal (compression) rate for a given target joint distribution.

The problem formulation for remote channel synthesis is laid out in Section 5.4.2. We present our main results in Section 5.4.3. The proofs can be found in the Appendices 5.A, 5.B, 5.C, and 5.D. In Section 5.4.4, we provide additional remarks on the necessity of vector schemes to

reach the optimal compression rate in the case where the rate of common randomness is small.

5.4.2 Problem formulation

Throughout Section 5.4, the alphabet of the remote source is denoted by $\mathcal{X}$, that of the observed source is denoted by $\mathcal{Z}$, and that of the decoder output is denoted by $\mathcal{Y}$. We assume said alphabets to be finite. The following definition of a code is the same as in the definition of fixed-length codes with unlimited private randomness (Definition 2.3), with the exception of the induced distribution, which involves the remote source.

Definition 5.3 *Given a positive integer n, and non-negative reals R and R_c, an (n, R, R_c) code is a privately randomized encoder and decoder pair $(F^{(n)}, G^{(n)})$ consisting of a conditional probability kernel $F^{(n)}_{M|Z^n,J}$ from $\mathcal{Z}^n \times [2^{nR_c}]$ to $[2^{nR}]$, and a conditional probability kernel $G^{(n)}_{Y^n|J,M}$ from $[2^{nR_c}] \times [2^{nR}]$ to $\mathcal{Y}^n$. Given a distribution $q_{X,Z}$ on $\mathcal{X}\times\mathcal{Z}$, the following distribution*

$$P_{X^n,Z^n,J,M,Y^n} := q^{\otimes n}_{X,Z} \cdot p^{\mathcal{U}}_{[2^{nR_c}]} \cdot F^{(n)} \cdot G^{(n)}$$

is called the distribution induced by the code.

Definition 5.4 *Given a distribution $q_{X,Z}$ on $\mathcal{X} \times \mathcal{Z}$, and a distribution $q_{X,Y}$ on $\mathcal{X} \times \mathcal{Y}$, a rate pair (R, R_c) is $(q_{X,Z},q_{X,Y})$-achievable if there exists a sequence $\{(F^{(n)}, G^{(n)})\}_{n\in\mathbb{N}}$ of codes, the n-th being (n, R, R_c), such that*

$$\|P_{X^n,Y^n} - q^{\otimes n}_{X,Y}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{5.6}$$

Let $\mathcal{A}^{(r.c.s.)}$ denotee the set of $(q_{X,Z},q_{X,Y})$-achievable rate pairs (R, R_c), where r.c.s. stands for remote channel synthesis.

The following lemma is a straightforward reformulation of Definition 5.4.

Lemma 5.2 *Consider a joint distribution $q_{X,Z}$ on $\mathcal{X}\times\mathcal{Z}$, a joint distribution $q_{X,Y}$ on $\mathcal{X}\times\mathcal{Y}$, and a distribution P_{X^n,Z^n,J,M,Y^n} on $\mathcal{X}^n\times\mathcal{Z}^n\times\mathbb{N}\times\mathbb{N}\times\mathcal{Y}^n$. Suppose that there exist non-negative reals R and R_c such that $M\in[2^{nR}]$ and $J\in[2^{nR_c}]$, P-almost surely. Then, P coincides with the distribution induced by the (n,R,R_c) code $(P_{M|Z^n,J},P_{Y^n|J,M})$ if and only if P satisfies the following properties:*

- $P_{X^n,Z^n,J}\equiv q_{X,Z}^{\otimes n}\cdot p^{\mathcal{U}}_{[2^{nR_c}]}$
- $X^n-Z^n-(J,M)-Y^n$ *form a Markov chain*

5.4.3 Main results on the remote channel synthesis problem

The main result of Section 5.4 is stated in the following theorem.

Theorem 5.3 *Consider a joint distribution $q_{X,Z}$ on $\mathcal{X}\times\mathcal{Z}$, and a joint target distribution $q_{X,Y}$ on $\mathcal{X}\times\mathcal{Y}$. Define the region $\mathcal{S}^{(r.c.s.)}$ as*

$$\left\{\begin{array}{rcl}(R,R_c)\in\mathbb{R}^2_{\geq 0} & : & \exists\, p_{X,Z,W,Y}\in\mathcal{D}^{(r.c.s.)},\\ R & \geq & I_p(Z;W)\\ R+R_c & \geq & I_p(X,Y;W)\end{array}\right\},\tag{5.7}$$

with $\mathcal{D}^{(r.c.s.)}$ defined as

$$\left\{\begin{array}{c}p_{X,Z,W,Y}: p_{X,Z}\equiv q_{X,Z},\ p_{X,Y}\equiv q_{X,Y}\\ X-Z-W-Y\\ |\mathcal{W}|\leq|\mathcal{Y}||\mathcal{Z}|+1\end{array}\right\}.\tag{5.8}$$

Then, the closure of $\mathcal{A}^{(r.c.s.)}$ is equal to $\mathcal{S}^{(r.c.s.)}$.

The proof is very close to that in [Cuf13] for the standard channel synthesis problem, and can be found in Appendices 5.A and 5.B. Similarly to the previous chapters, the scheme uses the likelihood encoder [SCP16], which has been shown to be universal in that it is relevant

for numerous lossy source compression problems. In contrast, the traditional scheme for the similar problem of remote RD is not universal. Note that $\mathcal{S}^{(r.c.s.)}$ is non-empty if and only if $(q_{X,Z}, q_{X,Y})$ belongs to the set $\mathcal{F}^{(r.c.s.)}$ defined as

$$\{(q_{X,Z}, q_{X,Y}) \mid \exists q_{Y|Z},\ q_{X,Y} \equiv \sum_{z\in\mathcal{Z}} q_Z(z) q_{X|Z=z} q_{Y|Z=z}\}.$$

A pair $(q_{X,Z}, q_{X,Y})$ is said to be *feasible* if it belongs to $\mathcal{F}^{(r.c.s.)}$. When the condition in the latter's definition is satisfied by some conditional distribution $q_{Y|Z}$, we say that the resulting distribution $q_{X,Z} \cdot q_{Y|Z}$ is *compatible with* $(q_{X,Z}, q_{X,Y})$. A natural scheme for remote channel synthesis consists in synthesizing such a compatible channel $q_{Y|Z}$. We call this direct channel synthesis, abbreviated as d.c.s. This leads to the following achievable region.

Proposition 5.4 *(Direct channel synthesis)*
Consider a $(q_{X,Z}, q_{X,Y})$-achievable rate pair (R, R_c) and a distribution $q_{X,Z,Y}$ compatible with $(q_{X,Z}, q_{X,Y})$. Define the region $\mathcal{S}^{(d.c.s.)}(q_{X,Z,Y})$ as

$$\left\{ \begin{array}{rcl} (R, R_c) \in \mathbb{R}^2_{\geq 0} & : & \exists\ p_{X,Z,W,Y} \in \mathcal{D}^{(d.c.s.)}(q_{X,Z,Y}), \\ R & \geq & I_p(Z;W) \\ R + R_c & \geq & I_p(Z,Y;W) \end{array} \right\}, \tag{5.9}$$

with $\mathcal{D}^{(d.c.s.)}(q_{X,Z,Y})$ defined as

$$\left\{ \begin{array}{c} p_{X,Z,W,Y} : p_{X,Z,Y} \equiv q_{X,Z,Y} \\ X - Z - W - Y \\ |\mathcal{W}| \leq |\mathcal{Y}||\mathcal{Z}| + 1 \end{array} \right\}. \tag{5.10}$$

Then, $(R, R_c) \in \mathcal{S}^{(d.c.s.)}(q_{X,Z,Y})$ iff there exists $\varepsilon > 0$ and a sequence $\{(F^{(n)}, G^{(n)})\}_{n\in\mathbb{N}}$ of codes, the n-th being $(n, R+\varepsilon, R_c)$, such that

$$\|P_{Z^n,Y^n} - q_{Z,Y}^{\otimes n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{5.11}$$

When (5.11) holds, then, from the Markov chain relation $X - Z - W - Y$, and [Cuf13,

Lemma V.2], we have

$$\|P_{X^n,Z^n,Y^n} - q_{X,Z,Y}^{\otimes n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{5.12}$$

Hence, we always have $\mathcal{S}^{(d.c.s.)}(q_{X,Z,Y}) \subseteq \overline{\mathcal{A}}^{(r.c.s.)} = \mathcal{S}^{(r.c.s.)}$.

Proof: The single-letter characterization follows from the single-letter characterization [Cuf13, Theorem II.1] for the problem of channel synthesis. Indeed, the proof of the latter involves an arbitrarily small excess rate ε for the compression rate, but not for the common randomness rate. The convergence in (5.12) follows from the Markov chain relation $X - Z - W - Y$ and [Cuf13, Lemma V.2]. The equality $\overline{\mathcal{A}}^{(r.c.s.)} = \mathcal{S}^{(r.c.s.)}$ follows from Theorem 5.3. ■

For any distribution $p_{X,Z,W,Y}$ satisfying the Markov chain relation $X - Z - W - Y$, we have

$$\begin{aligned} I_p(Z,Y;W) &= I_p(X,Z,Y;W) \\ &\geq \max(I_p(Z;W), I_p(X,Y;W)). \end{aligned} \tag{5.13}$$

Therefore, in most non-trivial cases, for small enough common randomness rate R_c, direct channel synthesis is strictly sub-optimal. We show this by obtaining a lower bound on the sub-optimality gap when $R_c = 0$, depicted in Figure 5.6. In a setup with binary alphabets, we have the following formal statement.

Proposition 5.5 *Assume that* $(q_{X,Z}, q_{X,Y})$ *is feasible and*

$$\begin{aligned} q_X \equiv q_Z \equiv q_Y &\equiv Ber(1/2), \\ 0 < q_{X,Z}(X \neq Z) &< q_{X,Y}(X \neq Y) < 1/2. \end{aligned} \tag{5.14}$$

Then, there exists $r > 0$ *such that for any* $R_c \in [0, r)$,

$$\begin{aligned} &\min\{R \mid (R, R_c) \in \cup_{q_{X,Z,Y}} \mathcal{S}^{(d.c.s.)}(q_{X,Z,Y})\} \\ &\quad > \min\{R \mid (R, R_c) \in \mathcal{S}^{(r.c.s.)}\}, \end{aligned} \tag{5.15}$$

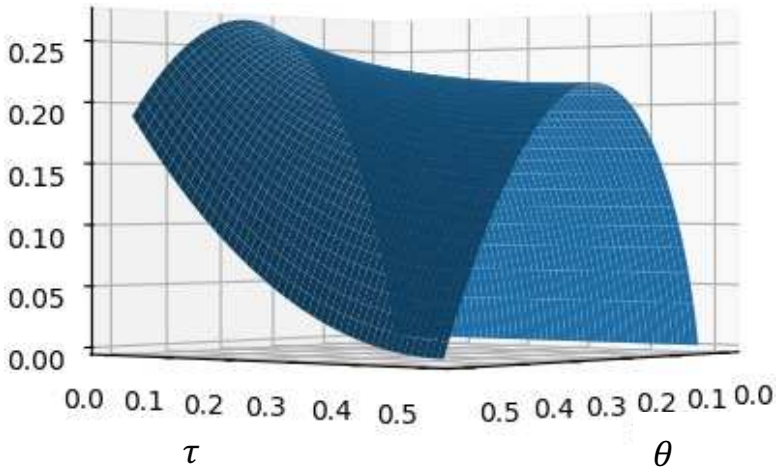


Figure 5.6: Lower bound on the gap between the optimal rates in (5.15) — see Appendix 5.C for details. Variables θ and τ denote $q_{Z,Y}(Y \neq Z)$ and $q_{X,Z}(X \neq Z)$, respectively, where $q_{X,Z,Y}$ is a distribution compatible with $(q_{X,Z}, q_{X,Y})$.

where the union is over all distributions $q_{X,Z,Y}$ *on* $\mathcal{X} \times \mathcal{Z} \times \mathcal{Y}$ *compatible with* $(q_{X,Z}, q_{X,Y})$.

The proof is provided in Appendix 5.C.

5.4.4 Scalar schemes v.s. vector schemes

In order to provide further intuitive understanding of Proposition 5.5, we consider a slightly different perspective, centered around the concept of *proper vector scheme*, defined below.

Definition 5.5 *Consider a source distribution* $q_{X,Z}$*, a target distribution* $q_{X,Y}$*, and reals* $R \geq 0$ *and* $R_c \geq 0$*. A* (R, R_c) *scheme for distribution* $q_{X,Y}$ *is a sequence* $(F^{(n)}, G^{(n)})$ *of codes, the* n*-th being* (n, R, R_c)*, such that the corresponding sequence* $\{P_{X^n,Z^n,J,M,Y^n}\}_{n\in\mathbb{N}}$ *of induced distributions satisfies*

$$\|P_{X^n,Y^n} - q_{X,Y}^{\otimes n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{5.16}$$

Definition 5.6 *Consider two non-negative reals* R *and* R_c*. A* (R, R_c) *scheme for distribution* $q_{X,Y}$ *is said to be a proper vector scheme if the corresponding sequence* $\{P_{X^n,Z^n,J,M,Y^n}\}_{n\in\mathbb{N}}$ *of induced distributions is such that there does not exists a sequence* $\{\rho_{Y|Z}^{(t)}\}_{t\in\mathbb{N}}$ *of single-letter*

distributions for which

$$\left\| P_{Z^n,Y^n} - \prod_{t=1}^{n} q_Z \cdot \rho_{Y|Z}^{(t)} \right\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{5.17}$$

From the results of Section 5.4.3, a proper vector scheme is necessary for optimal

Proposition 5.6 *Assume the setting of Proposition 5.5. Let* $R_{\min}$ *denote the right hand side of* (5.15) *in the case* $R_c = 0$, *i.e., the optimal rate for remote strong coordination in the absence of common randomness. Let* R_1 *denote the left hand side of* (5.15) *in the case* $R_c = 0$, *i.e., the optimal rate for direct strong coordination in the absence of common randomness. Then, for any* R *in* $[R_{\min}, R_1)$, *for small enough* R_c, *any* (R, R_c) *scheme for distribution* $q_{X,Y}$ *is a proper vector scheme.*

The proof is provided in Appendix 5.D.

5.5 Non-interactive remote coordination scheme based on remote channel synthesis

We return to the central object of this chapter, namely the problem of non-interactive remote coordination (Figure 5.1), and present our main result. As mentioned in Section 5.3.3, approaching the problem with the standard tool of channel synthesis achieves much stronger coordination than required. We propose a scheme based on remote channel synthesis, which induces a relatively weaker yet sufficient coordination.

5.5.1 Scheme description

Our scheme consists of the following steps.

1. Choose two single-letter conditional distributions $p_{X|U}$ and $p_{Y|V}$.

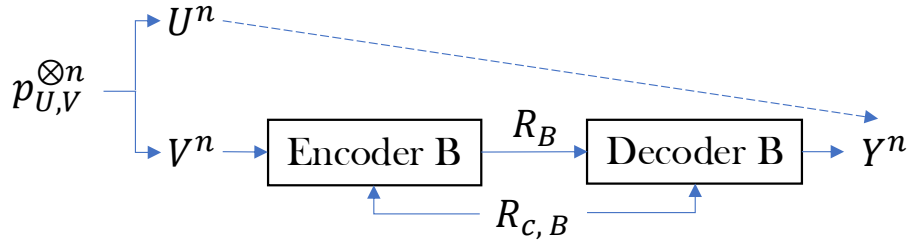


Figure 5.7: First step

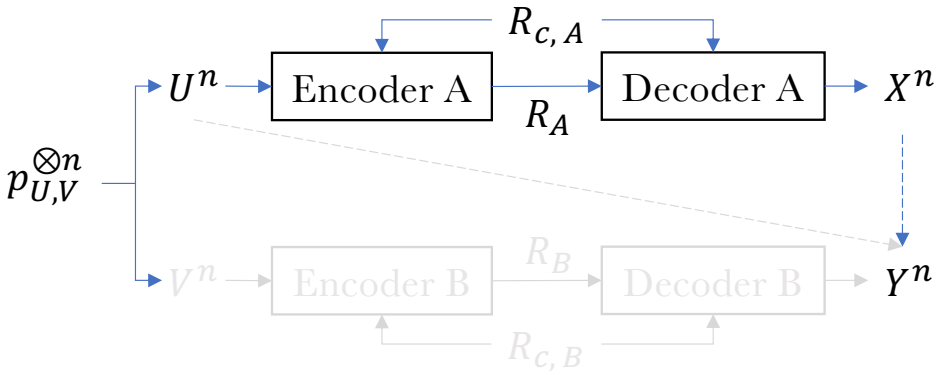


Figure 5.8: Second step

2. This induces a joint single-letter distribution $p_{U,V,X,Y}$.

3. Synthesize the memoryless channel $\prod p_{Y|U}$ using Encoder B and Decoder B via remote channel synthesis with remote source U^n. This is depicted in Figure 5.7.

4. Synthesize the memoryless channel $\prod p_{X|Y}$ using Encoder A and Decoder A via remote channel synthesis with remote source Y^n. This is depicted in Figure 5.8.

Step 4 is possible since Step 3 ensures that U^n and Y^n are approximately jointly i.i.d.. Indeed, although the problem of remote channel synthesis is defined for source sequences that are exactly jointly i.i.d., it is easy to show that a minor change in the distribution of the sources does not change the problem from an information theoretic point of view. This is due to the fact that the desired coordination between the remote source and the decoder output is only an approximate coordination measured via the TVD.

5.5.2 Achievable region

As a straightforward consequence of the single-letter characterization of remote channel synthesis in Theorem 5.3, we obtain the following achievable region for the non-interactive remote coordination problem. By symmetry, the roles of the encoder-decoder pairs can be switched, leading to an achievable region with two components.

Theorem 5.7 *All tuples in the interior of the following region $\mathcal{S}$ are achievable.*

$$\left\{ \begin{array}{rcl} (R_A, R_{c,A}, R_B, R_{c,B}) & : & \exists\, p_{U,V,L_A,L_B,X,Y} \in \mathcal{D}, \\ R_A & \geq & I_p(U; L_A) \\ R_A + R_{c,A} & \geq & I_p(X, Y; L_A) \\ R_B & \geq & I_p(V; L_B) \\ R_B + R_{c,B} & \geq & I_p(U, Y; L_B) \end{array} \right\} \bigcup$$

$$\left\{ \begin{array}{rcl} (R_A, R_{c,A}, R_B, R_{c,B}) & : & \exists\, p_{U,V,L_A,L_B,X,Y} \in \mathcal{D}, \\ R_B & \geq & I_p(V; L_B) \\ R_B + R_{c,B} & \geq & I_p(X, Y; L_B) \\ R_A & \geq & I_p(U; L_A) \\ R_A + R_{c,A} & \geq & I_p(V, X; L_A) \end{array} \right\},$$

with $\mathcal{D}$ defined as

$$\left\{ \begin{array}{c} p_{U,V,L_A,L_B,X,Y} : p_{X,Y} \equiv q_{X,Y},\ p_{U,V} \equiv q_{U,V}, \\ X - L_A - U - V - L_B - Y \\ |\mathcal{L}_A| \leq |\mathcal{U}||\mathcal{X}| + 1 \\ |\mathcal{L}_B| \leq |\mathcal{V}||\mathcal{Y}| + 1 \end{array} \right\}.$$

As detailed in Section 5.4, remote channel synthesis can be more efficient than direct channel synthesis if the rate of available common randomness is sufficiently low. When this is the case

for $R_{c,A}$ or $R_{c,B}$, then the achievable region in the above theorem is strictly larger than the one in Proposition 5.1. To make this more explicit, we provide an example below.

Example 5.8 *We exhibit an operating point achievable via Theorem 5.7 but not via Proposition 5.1, for a binary instance in the case $R_{c,B} = 0$. We show that in this case, the minimum rate R_B in the region of Proposition 5.1 is strictly larger than the one in the region of Theorem 5.7. For any $R_{c,B}$ smaller than this gap, the property remains true because $\mathcal{S}^{(a)}$ is closed under the operation $(R_B, R_{c,B}) \mapsto (R_B + R_{c,B}, 0)$. Throughout this example, we focus on exhibiting points in $\mathcal{S} \setminus \mathcal{S}^{(a)}$. Since $\mathcal{S}^{(a)}$ is closed (all alphabets have bounded cardinality), a small perturbation can then yield a point outside of $\mathcal{S}^{(a)}$ but in the interior of $\mathcal{S}$, and thus achievable by Theorem 5.7. For notation pertaining to binary variables and channel and related probability distributions, we refer the reader to Appendix 5.C. Let $\mathcal{U} = \mathcal{V} = \mathcal{X} = \mathcal{Y} = \{0, 1\}$ and $q_{U,V} = DSBS(\tau)$ with $\tau \in (0, 1/2)$ and $q_{X,Y} = DSBS(\eta)$ with $\eta \in (\tau, 1/2)$. This guarantees that $\mathcal{S}^{(a)}$ is non-empty — e.g., take $L_A = X, L_B = V$, and $p_{X|U} = BSC(a)$ with $a * \tau = \eta$. Let*

$$R_B = \min\{R \mid \exists (R_A, R_{c,A}) \text{ with } (R_A, R_{c,A}, R, 0) \in \mathcal{S}^{(a)}\},$$

where the minimum exists because $\mathcal{S}^{(a)}$ is closed as all alphabets have bounded cardinality. Let $(R_A, R_{c,A})$ be a corresponding pair, and $p \in \mathcal{D}^{(a)}$ be a corresponding single-letter distribution from the definition of $\mathcal{S}^{(a)}$. Note that since $\eta < 1/2$, then X and Y are not independent under p, thus, U and Y are not independent; hence, $p_{U,Y}(U \neq Y) \neq 1/2$ (a property of variables having marginal distribution $Ber(1/2)$). Without loss of generality, one can assume that $p_{U,Y}(U \neq Y) < 1/2$: otherwise, modify p by replacing X (resp. Y) with $1 - X$ (resp. $1 - Y$), which preserves $p_{X,Y}$ and mutual information quantities and the inequalities appearing in the definition of $\mathcal{S}^{(a)}$. Then, from the data processing inequality of mutual information, we have $q_{U,V}(U \neq V) \leq p_{U,Y}(U \neq Y)$. Moreover, by minimality we have $R_B < 1$ (easy to prove, e.g., [B+21]), thus $q_{U,V}(U \neq V) < p_{U,Y}(U \neq Y)$ (a property of variables that have marginal distribution $Ber(1/2)$).

Hence, the hypotheses of Proposition 5.5 are satisfied. Thus, we have $\rho_{r.c.s.} < \rho_{d.c.s.}$*, where*

$$\rho_{r.c.s.} := \min\{R : (R,0) \in \mathcal{S}^{(r.c.s.)}\},$$
$$\rho_{d.c.s.} := \min\{R : (R,0) \in \bigcup_{\tilde{p}} \mathcal{S}^{(d.c.s.)}(\tilde{p})\},$$

the pair-B minimum description rates under remote and direct channel synthesis, respectively, where the union is over all distributions $\tilde{p}_{U,V,Y}$ *on* $\{0,1\}^3$ *compatible with* $(q_{U,V}, p_{U,Y})$*. Since* p *is one such distribution, and* $R_B = I_p(V,Y;L_B)$ *by minimality of* R_B*, then, we have* $(R_B,0) \in \mathcal{S}^{(d.c.s.)}(p)$ *and thus* $\rho_{d.c.s.} \leq R_B$*. Let* $\tilde{p}_{L_B|V,Y}$ *be such that* $\tilde{p}_{U,V,L_B,Y} \in \mathcal{D}^{(r.c.s.)}$ *and attains* $\rho_{r.c.s.}$ *for pair B (definition of* $\mathcal{S}^{(r.c.s.)}$*). Then,* $\tilde{p}_{X,L_A,U,V,L_B,Y} \in \mathcal{D}$*. Choose*

$$R_A^\star := I_{\tilde{p}}(U;L_A), \qquad R_{c,A} \geq \max\big(0,\ I_{\tilde{p}}(X,Y;L_A) - R_A^\star\big), \qquad R_B^\star \in \big(\rho_{r.c.s.},\ \rho_{d.c.s.}\big).$$

Then $(R_A^\star, R_{c,A}, R_B^\star, 0)$ *lies in the region* $\mathcal{S}$ *of Theorem 5.7, witnessed by* $\tilde{p}$*: the pair-A inequalities* $R_A^\star \geq I_{\tilde{p}}(U;L_A)$ *and* $R_A^\star + R_{c,A} \geq I_{\tilde{p}}(X,Y;L_A)$ *hold by the choice of* $(R_A^\star, R_{c,A})$*, while* $R_B^\star > \rho_{r.c.s.}$ *ensures both pair-B inequalities* $R_B^\star \geq I_{\tilde{p}}(V;L_B)$ *and* $R_B^\star + 0 \geq I_{\tilde{p}}(U,Y;L_B)$*. It is* not *in the region* $\mathcal{S}^{(a)}$ *of Proposition 5.1 by minimality of* R_B*.*

In Section 5.6, we discuss promising directions to further improve our inner bound. In Section 5.7, we provide relatively simple channel synthesis based outer bounds, which imply that common randomness is necessary for optimal performance.

5.6 Leveraging Gács-Körner common information between the sources

5.6.1 Coordination schemes à la Wyner, improved with channel synthesis

General approach

In the distributed coordination literature, one of the main schemes to generate outputs satisfying a strong coordination constraint is the one proposed by Wyner [Wyn75], for the problem depicted in Figure 5.2. The latter consists in using the random number U as index of a codeword $w^n(U)$, then sampling X^n from a memoryless channel $\prod p_{X|W}$, and Y^n from a memoryless channel $\prod p_{Y|W}$. Going back to our non-interactive remote coordination problem, consider a scheme à la Wyner, that is, a scheme having the following characteristics.

1. Encoder A chooses a codeword $w_{1:n}^{(A)}$.
2. Encoder B chooses a codeword $w_{1:n}^{(B)}$.
3. Decoder A produces X^n by passing $w_{1:n}^{(A)}$ through a memoryless channel $\prod p_{X|W^{(A)}}$.
4. Decoder B produces Y^n by passing $w_{1:n}^{(B)}$ through a memoryless channel $\prod p_{Y|W^{(B)}}$.

Channel synthesis provides a natural alternative scheme: instead of sending codeword $w_{1:n}^{(A)}$ (resp. $w_{1:n}^{(A)}$), Encoder A and Decoder A can use a channel synthesis scheme to simulate the memoryless channel $\prod p_{X|W^{(A)}}$. If the rate $R_{c,A}$ of common randomness is high enough, then this scheme may require a smaller R_A then is needed for sending $w_{1:n}^{(A)}$ (losslessly).

An example

We provide an example where the channel synthesis scheme is more efficient. Consider a scheme satisfying conditions 1-4. Assume that $R_{c,A}, R_{c,B} \geq H(X,Y)$ and that codewords $w_{1:n}^{(A)}$ and $w_{1:n}^{(B)}$

are identical, similarly to the scheme of Wyner. Let $P^{(n)}$ denote the distribution induced by the scheme. The converse to Wyner's result [Wyn75, Section 5.2] implies that

$$\min(R_A, R_B) \geq \liminf_{n\to\infty} \frac{1}{n} H_{P^{(n)}}(W^n) \geq C_{\text{Wyner}}(X, Y), \tag{5.18}$$

with $C_{\text{Wyner}}(X, Y) = \min I_p(X, Y; W)$, where the minimum is over all p_W such that

$$\sum_w p_W(w) p_{X|W=w} p_{Y|W=w} \equiv q_{X,Y}. \tag{5.19}$$

There exists some p^*_W which achieves this minimum [Wyn75, Remark C]. Moreover, since the codewords are identical, the second inequality in (5.18) implies that $C_{GK}(U, V) \geq C_{\text{Wyner}}(X, Y)$ [GK73] and that Encoder A and Encoder B can agree on $nC_{\text{Wyner}}(X, Y)$ i.i.d. fair bits with probability going to 1 as $n \to \infty$. Therefore, $q_{U,V}$, R_A and R_B are such that it is possible to implement the scheme of Wyner [Wyn75] with codebook distribution p^*_W and channels $p_{X|W}$ and $p_{Y|W}$, where the common codeword $w_{1:n}$ is sent (losslessly) from Encoder A (resp. B) to Decoder A (resp. B). Let $\tilde{P}^{(n)}$ denote the distribution induced by said scheme. Then, it is known that

$$\forall \delta > 0, \quad \tilde{P}^{(n)}(W^n \notin \mathcal{T}^{(n)}_\delta(p^*_W)) \underset{n\to\infty}{\longrightarrow} 0, \tag{5.20}$$

where $\mathcal{T}^{(n)}_\delta(p^*_W))$ is the p^*_W-typical set of level δ [EGK11]. Since $R_{c,A}$ (resp. $R_{c,B}$) is large, then for any $\varepsilon > 0$, there exists a channel synthesis scheme of communication rate $I_{p^*_W \cdot p_{X|W}}(W; X) + \varepsilon$ which allows to simulate the memoryless channel $p_{X|W}$ with input W^n — see Appendix 5.E. This rate is strictly smaller than $C_W(X, Y) = I_{p^*_W \cdot p_{X|W} \cdot p_{Y|W}}(X, Y; W)$, which proves the superiority of performing channel synthesis. In this example, the availability of non-trivial GK common information between U^n and V^n is key. In the next section, we show that such GK common information can be leveraged in a more advanced scheme.

5.6.2 A hybrid scheme

In [AB07], a similar non-interactive coordination problem is considered, with $U = (A, W)$ and $V = (B, W)$ where W is independent from (A, B). The Markov chain relation $X - (A, W) - (B, W) - Y$ is studied and compared to the Markov chain relation $X - A - B - Y$. This suggests to explore more general channel synthesis based schemes for our non-interactive remote coordination problem, where channels of the form $\prod p_{X|U,W}$ and $\prod p_{Y|V,W}$ would be simulated, instead of channels of the form $\prod p_{X|U}$ and $\prod p_{Y|V}$. For example, it is relevant to combine channel synthesis with the scheme of Berg et. al. [B$^+$21], which also involves a Markov chain relation of the form $X - (U, W) - (V, W) - Y$. This yields a scheme which subsumes the ones presented in Section 5.3 and the second half of Section 5.6.1, with the following achievable rates.

Theorem 5.9 *All tuples in the interior of the following region $\mathcal{S}$ are achievable.*

$$\left\{ \begin{array}{rcl} (R_A, R_{c,A}, R_B, R_{c,B}) & : & \exists\ p_{U,V,K,W,L_A,L_B,X,Y} \in \mathcal{D}, \\ R_A & \geq & I_p(U, W; L_A) \\ R_A + R_{c,A} & \geq & I_p(U, W, X; L_A) \\ R_B & \geq & I_p(V, W; L_B) \\ R_B + R_{c,B} & \geq & I_p(V, W, Y; L_B) \end{array} \right\},$$

with $\mathcal{D}$ defined as

$$\left\{ \begin{array}{c} p_{U,V,K,W,L_A,L_B,X,Y} : p_{X,Y} \equiv q_{X,Y},\ p_{U,V} \equiv q_{U,V}, \\ W - K - (U, V),\ H(K|U) = H(K|V) = 0 \\ X - L_A - (U, W) - (V, W) - L_B - Y \\ H_p(K) \geq I_p(X, Y; K, W) \end{array} \right\}.$$

The corresponding scheme is depicted in Figure 5.9, and is closely related to the one in [B$^+$21]. Encoder A first extracts the GK common part K^n of U^n and V^n, then deterministically maps K^n to a string $w^n(K^n)$. Instead of sampling X^n from $\prod_{t=1}^n p_{X|U=u_t,W=w_t(K^n)}$ locally, the channel

is simulated in a distributed manner between Encoder A and Decoder A, via channel synthesis. The two other terminals proceed similarly, using the same map $w^n(\cdot)$, and simulate channel $\prod_{t=1}^n p_{Y|V=v_t,W=w_t(K^n)}$. A sketch of the proof is provided in Section 5.6.3, and a complete proof is provided in Appendix 5.E.

5.6.3 Main components of the proof of Theorem 5.9

The problem studied in [B$^+$21] is the special case of our problem, where $R_A = R_B = \infty$. Here the notation $R_A = R_B = \infty$ refers to any rates for which bijective (zero-error) encoding maps are possible, i.e. $R_A \geq \log|\mathcal{U}|, R_B \geq \log|\mathcal{V}|$. Then, Encoder A and Decoder A behave as a single terminal, which we shall call Terminal A, and Terminal B is defined similarly. Using the notation in the present chapter, we can reformulate [B$^+$21, Theorem 2] as follows.

Theorem 5.10 *Consider any $p \in \mathcal{D}$. Then, tuple $(\infty, 0, \infty, 0)$ is achieved by the following scheme.*

- *Terminal A passes its input U^n through channel $\prod p_{K|U}$ (which is a deterministic mapping), resulting in K^n (the GK common information of U^n and V^n).*
- *It then passes K^n through a well-chosen deterministic mapping $w^n(\cdot)$.*
- *Terminal B proceeds in the same manner, using the same mapping $w^n(\cdot)$, resulting in the same realizations of K^n and $w^n(K^n)$ as obtained at Terminal A.*
- *Terminal A samples X^n from $\prod_{t=1}^n p_{X|U=U_t,W=w_t(K^n)}$.*
- *Terminal B samples Y^n from $\prod_{t=1}^n p_{Y|V=V_t,W=w_t(K^n)}$.*

In order to prove Theorem 5.9, we use the same scheme at Encoders A and B, but replace the last two steps with a channel synthesis protocol: given U^n and $w^n(K^n)$ at Encoder A, the latter sends a message at rate R_A allowing the decoder to produce X^n distributed approximately according to $\prod_{t=1}^n p_{X|U=U_t,W=w_t(K^n)}$. Encoder B and Decoder B proceed similarly. One can

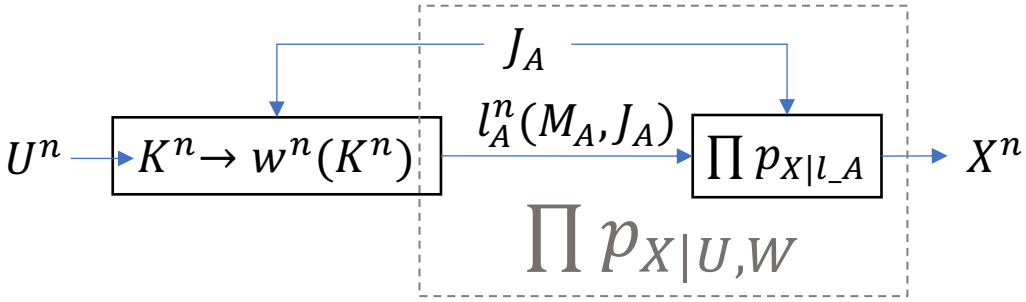


Figure 5.9: Illustration of the proposed coding scheme.

find a channel synthesis protocol of adequate communication and common randomness rates, provided that $(U^n, w^n(K^n))$ is a typical sequence, as formalized in the following result. We use the notion of typical sets defined in [EGK11], with the notation $\mathcal{T}_\varepsilon^{(n)}(p_Z)$ instead of the similar notation in [EGK11].

Proposition 5.11 *Let $\mathcal{Z}, \mathcal{L}, \mathcal{O}$ be three finite alphabets, and $p_{Z,L,O}$ be a distribution on their product, satisfying Markov chain $Z - L - O$. For any $\varepsilon > 0$, there exists a real $\delta > 0$ and an integer $N \geq 1$ such that, for any integer $n \geq N$, there exists an $(n, I_p(Z;L) + \varepsilon, I_p(Z,O;L) - I_p(Z;L) + \varepsilon)$ code for $(\mathcal{Z}, \mathcal{O})$, such that for any $z^n \in \mathcal{T}_\delta^{(n)}(p_Z)$,*

$$\left\| P_{O^n|Z^n=z^n}^{(n)} - \prod_{t=1}^n p_{O|Z=z_t} \right\|_{TV} \leq \varepsilon. \tag{5.21}$$

Proposition 5.11 is a generalization of the standard channel synthesis result in [Cuf13], which only pertains to the problem of identifying the rates for which it is possible to achieve a vanishing average of (5.21) over $Z^n \sim p_Z^{\otimes n}$. Proposition 5.11 is a corollary of the proofs from [BDH+14]. We provide the proofs of Proposition 5.11 and Theorem 5.9 in Appendix 5.E.

5.6.4 Can Gács-Körner common information be leveraged in a scheme based on remote channel synthesis?

In Section 5.6.2, we presented a hybrid scheme based on direct channel synthesis and the scheme in [B+21]. We expect that a similar hybrid scheme based on remote channel synthesis may be developed. This constitutes an interesting direction for future work. The main difficulty lies in

developing remote channel synthesis schemes for encoder inputs that are not i.i.d. sequences, which was crucial in the proof of Theorem 5.9. The method of types used for the latter — following [B$^+$14] — does not extend to remote channel synthesis in a straightforward manner. Nevertheless, we conjecture the following extension of Theorem 5.3.

Conjecture 5.12 *Let $\mathcal{X}, \mathcal{Z}, \mathcal{W}, \mathcal{Y}$ be finite alphabets, and $p_{X,Z,W,Y}$ be a distribution on their product, satisfying Markov chain $X - Z - W - Y$. For any $\varepsilon > 0$, there exists a real $\delta > 0$ and an integer $N \geq 1$ such that, for any integer $n \geq N$, there exists an $(n, I_p(Z;W) + \varepsilon, I_p(X,Y;W) - I_p(Z;W) + \varepsilon)$ code for $(\mathcal{Z}, \mathcal{Y})$ such that for any $(x^n, z^n) \in \mathcal{T}_\delta^{(n)}(p_{X,Z})$,*

$$\left\| P^{(n)}_{Y^n|X^n=x^n,Z^n=z^n} - \prod_{t=1}^n p_{Y|X=x_t} \right\|_{TV} \leq \varepsilon. \tag{5.22}$$

The next section is dedicated to outer bounds for the problem of non-interactive remote coordination.

5.7 Channel synthesis based outer bounds

In this section, we propose relatively simple outer bounds for the non-interactive remote coordination problem, based on links to channel synthesis. Given a joint distribution $q_{Z,O}$, we denote by $\mathcal{S}^{(c.s.)}_{Z->O}(q_{Z,O})$ the optimal rate region for the distributed synthesis of channel $q_{O|Z}$ with i.i.d. input of distribution q_Z [Cuf13]:

$$\left\{ \begin{array}{rcl} (R, R_c) & : & \exists\, p_{Z,L,O} \in \mathcal{D}^{(c.s.)}(q_{Z,O}), \\ R & \geq & I_p(Z;L) \\ R + R_c & \geq & I_p(Z,O;L) \end{array} \right\},$$

with $\mathcal{D}^{(c.s.)}(q_{Z,O})$ defined as

$$\left\{ \begin{array}{c} p_{Z,L,O} : p_{Z,O} \equiv q_{Z,O} \\ Z - L - O \\ |\mathcal{L}| \leq |\mathcal{Z}||\mathcal{O}| + 1 \end{array} \right\}.$$

We have the following outer bound for the non-interactive remote coordination problem.

Proposition 5.13 *If a tuple* $(R_A, R_{c,A}, R_B, R_{c,B}) \in \mathbb{R}^4_{\geq 0}$ *is achievable, then*

$$(R_A, R_{c,A}) \in cl(\mathcal{S}^{(c.s.)}_{Y->X}(q_{X,Y})),$$
$$(R_B, R_{c,B}) \in cl(\mathcal{S}^{(c.s.)}_{X->Y}(q_{X,Y})).$$

Moreover, if there exists at least one achievable tuple, then $q_{U,V}$ *and* $q_{X,Y}$ *satisfy*

$$C_{Wyner}(U,V) \geq C_{Wyner}(X,Y). \tag{5.23}$$

Proof: Consider an achievable tuple $(R_A, R_{c,A}, R_B, R_{c,B}) \in \mathbb{R}^4_{\geq 0}$, and a corresponding sequence of codes, the n-th being $(n, R_A, R_{c,A}, R_B, R_{c,B})$, with induced distribution $P^{(n)}$. Define

$$Q^{(n)}_{X^n,J_B,M_B,Y^n} := p_X^{\otimes n} \cdot p^{\mathcal{U}}_{[2^{nR_{c,B}}]} \cdot P^{(n)}_{M_B|X^n,J_B} \cdot P^{(n)}_{Y^n|M_B}.$$

Then, we have

$$Q^{(n)}_{X^n,J_B,M_B,Y^n} \equiv p_X^{\otimes n} \cdot P^{(n)}_{J_B,M_B,Y^n|X^n}.$$

Thus,

$$\|Q^{(n)}_{X^n,J_B,M_B,Y^n} - P^{(n)}_{X^n,J_B,M_B,Y^n}\|_{TV}$$
$$= \|p_X^{\otimes n} - P^{(n)}_{X^n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0.$$

Hence, from the triangle inequality for TVD, we have

$$\|Q^{(n)}_{X^n,Y^n} - p^{\otimes n}_{X,Y}\|_{TV} \to 0.$$

Therefore, by construction of $Q^{(n)}$, and from the standard channel synthesis result of Cuff [Cuf13, Theorem 1], we have

$$(R_B, R_{c,B}) \in cl(\mathcal{S}^{(c.s.)}_{X->Y}(q_{X,Y})).$$

The same proof can be used to show that

$$(R_A, R_{c,A}) \in cl(\mathcal{S}^{(c.s.)}_{Y->X}(q_{X,Y})).$$

Moreover, it is known that a joint distribution $q^{\otimes n}_{U,V}$ can be approximated with an input of distribution $q^{\otimes n}_U$ and a channel synthesis scheme of communication rate $C_{\text{Wyner}}(U,V)$ with no common randomness [Cuf13]. Replacing $q^{\otimes n}_{U,V}$ in $P^{(n)}$ with said approximation would preserve the strong coordination of X^n and Y^n. Therefore, the same argument as above yields

$$(C_{\text{Wyner}}(U,V), 0) \in cl(\mathcal{S}^{(c.s.)}_{X->Y}(q_{X,Y})),$$

which implies that $C_{\text{Wyner}}(U,V) \geq C_{\text{Wyner}}(X,Y)$ [Cuf13]. ■

The necessary condition (5.23) complements the ones mentioned in the introduction. From known properties of the channel synthesis rate region [Cuf13], Proposition 5.13 implies the following bounds.

Corollary 5.14 *If a tuple* $(R_A, 0, R_B, 0) \in \mathbb{R}^4_{\geq 0}$ *is achievable, then*

$$\min(R_A, R_B) \geq C_{Wyner}(X,Y).$$

If a tuple $(R_A, R_{c,A}, R_B, R_{c,B}) \in \mathbb{R}^4_{\geq 0}$ *is achievable, and*

$$\min(R_{c,A}, R_{c,B}) \geq H_q(X, Y),$$

then

$$\min(R_A, R_B) \geq I_q(X; Y).$$

This implies that the outer bound in Proposition 5.13 is tight in a special case.

Theorem 5.15 *If* $q_{U,V}(U = V) = 1$, *then for any* $(R_A, R_B) \in \mathbb{R}^2_{\geq 0}$, *tuple* $(R_A, 0, R_B, 0) \in \mathbb{R}^4_{\geq 0}$

- *Is achievable if*

$$\min(H(U), R_A, R_B) > C_{Wyner}(X, Y).$$

- *Is not achievable if*

$$\min(H(U), R_A, R_B) < C_{Wyner}(X, Y).$$

Proof: The first claim follows directly from using the scheme of Wyner [Wyn75]. Moving to the second claim, from the converse proof in [Wyn75], we have

$$H(U) \geq C_{\text{Wyner}}(X, Y).$$

The remaining inequalities follow from Corollary 5.14. ■

In Sections 5.3, 5.5 and 5.6, we showed that (possibly remote) channel synthesis and common randomness can be used to improve certain schemes, or develop new ones. This highlighted the benefits of common randomness for specific scheme designs. Theorem 5.15 implies that, in the special case $q_{U,V}(U = V) = 1$, common randomness is necessary for any scheme achieving operating points where one of the rates R_A and R_B is strictly lower than $C_{\text{Wyner}}(X, Y)$. Moreover, such schemes exist. Indeed, for any distribution $q_{X,Y}$, we have $C_{\text{Wyner}}(X, Y) \geq I(X; Y)$.

Consider a distribution $q_{X,Y}$ for which this inequality is strict. Let $q_{U,V}$ be such that $U = V$ and $q_U \equiv q_X$. Then, the following scheme is achievable. Encoder A sends U^n losslessly to Decoder A, which outputs it, and Encoder B and Decoder B simulate the memoryless channel $\prod p_{Y|X}$, with input distribution $p_V \equiv p_X$. If common randomness is available between Encoder B and Decoder B, then since $C_{\text{Wyner}}(X,Y) > I(X;Y)$, this channel synthesis can be performed at a communication rate R_B that is strictly lower than $C_{\text{Wyner}}(X,Y)$. This concludes the presentation of our outer bounds.

5.8 Conclusion

We have considered a non-interactive remote coordination problem, where two non-interacting controllers observe correlated sources, and can each communicate with a decoder, with the aim of helping the two decoders to generate coordinated actions. Two sources of common randomness, one for each pair, may be available. The coordination requirement is a distribution matching constraint similar to the near-perfect strong realism constraint studied in previous chapters - and specifically similar to the joint realism constraint studied in Chapter 3. Some of the proofs in the present chapter are very similar to the proofs in Chapters 2 and 3.

We have highlighted the benefits of common randomness for this problem. As mentioned in Chapter 4, owing to practical implementation challenges, the rate of common randomness is often relatively small — in Section 5.4.1, we explain that this is especially relevant for channel synthesis based schemes. We dedicated substantial attention to the regime of low common randomness rates ($R_{c,A}$ and $R_{c,B}$). We proposed several schemes and investigated directions for future improvement — contrary to most of the results in previous chapters, we do not have converse bounds matching our achievable regions in general.

We first presented a scheme based on a direct application of channel synthesis. For the purpose of developing schemes of superior performance, we studied a problem we coin remote channel synthesis: an encoder has access to a partial or noisy version Z^n of a remote source X^n, and communicates information to a decoder which aims to produce an output Y^n coordinated

with the remote source in the strong sense. Of auxiliary interest, we have obtained a single-letter characterization of the optimal compression and common randomness rate pairs for the remote channel synthesis problem. We find that when the rate of common randomness is small enough, the natural approach consisting in a direct application of channel synthesis is strictly sub-optimal for that problem. Similarly, for our initial problem of non-interactive remote coordination, we propose a scheme based on remote channel synthesis that achieves superior performance compared to our first scheme when the rates $R_{c,A}$ and $R_{c,B}$ of available common randomness are small enough.

We expect that the fundamental problem of remote channel synthesis can be used as a building block for more complex distributed coordination schemes. We have also shown that advanced non-interactive remote coordination schemes can be developed when the correlated sources share non-trivial GK common information. We conjecture that this can be combined with remote channel synthesis. Moreover, we expect that further channel synthesis based schemes and outer bounds can be uncovered.

This concludes the final technical chapter of this thesis. In the next chapter, we offer concluding remarks, highlight unifying themes, and discuss directions for future work.

Appendix

5.A Converse direction of Theorem 5.3

This proof closely tracks [Cuf13, Section VI].

Claim 5.16 *$\mathcal{S}^{(r.c.s.)}$ is closed in $\mathbb{R}^2$.*

Proof: Note that $\mathcal{S}^{(r.c.s.)}$ remains unchanged if variable W in $\mathcal{D}^{(r.c.s.)}$ is required to take values only in alphabet $\mathcal{W} = [|\mathcal{Y}||\mathcal{Z}| + 1]$. Hereafter, we assume that this requirement is added to the definition of $\mathcal{D}^{(r.c.s.)}$. Then, $\mathcal{D}^{(r.c.s.)}$ is a bounded closed subset of a finite-dimensional real vector space. Thus, it is compact. Since the inequalities in the definition of $\mathcal{S}^{(r.c.s.)}$ are not strict inequalities, and mutual information is continuous, and $\mathcal{D}^{(r.c.s.)}$ is closed and bounded, then for any $(R, R_c) \in \mathcal{S}^{(r.c.s.)}$, the set $\mathcal{D}^{(r.c.s.)}(R, R_c)$ of distributions $p \in \mathcal{D}^{(r.c.s.)}$ such that $R \geq I_p(Z;W)$ and $R + R_c \geq I_p(X,Y;W)$ is a closed and bounded subset of a finite-dimensional real vector space. Therefore, there exists an element $p^{(R,R_c)}$ of that set which is minimal in lexicographic order — so as not to have to use the axiom of choice later on. Consider a sequence $\{(R^{(n)}, R_c^{(n)})\}_{n\in\mathbb{N}}$ of elements of $\mathcal{S}^{(r.c.s.)}$ which converges in $\mathbb{R}^2$. Let (R^∞, R_c^∞) be the limit. From the compactness of $\mathcal{D}^{(r.c.s.)}$, there exists $p^* \in \mathcal{D}^{(r.c.s.)}$ and an increasing sequence $\{n_k\}_{k\in\mathbb{N}}$ of positive integers such that $p^{(R^{(n_k)}, R_c^{(n_k)})} \underset{k\to\infty}{\to} p^*$. Then, by continuity of the mutual information for finite alphabets, we have $R^\infty \geq I_{p^*}(Z;W)$ and $R^\infty + R_c^\infty \geq I_{p^*}(X,Y;W)$. Thus, $(R^\infty, R_c^\infty) \in \mathcal{S}^{(r.c.s.)}$. Hence, $\mathcal{S}^{(r.c.s.)}$ is closed. ■

We prove that $\overline{\mathcal{A}}^{(r.c.s.)} \subset \mathcal{S}^{(r.c.s.)}$ by proving that $\mathcal{A}^{(r.c.s.)} \subset \mathcal{S}^{(r.c.s.)}$. Define the region $\mathcal{S}_\varepsilon$ as

$$\left\{ \begin{array}{rcl} (R, R_c) \in \mathbb{R}^2_{\geq 0} & : & \exists\, p_{X,Z,W,Y} \in \mathcal{D}_\varepsilon, \\ R & \geq & I_p(Z;W) \\ R + R_c & \geq & I_p(X,Y;W) - 2g(\varepsilon) \end{array} \right\}, \tag{5.24}$$

with $\mathcal{D}_\varepsilon$ defined as

$$\left\{ \begin{array}{c} p_{X,Z,W,Y} : p_{X,Z} \equiv q_{X,Z}, \\ \|p_{X,Y} - q_{X,Y}\|_{TV} \leq \varepsilon \\ X - Z - W - Y \\ |\mathcal{W}| \leq |\mathcal{Y}||\mathcal{Z}| + 1 \end{array} \right\}, \tag{5.25}$$

and $g : (0, 1/2) \to \mathbb{R}$ defined as

$$g(\varepsilon) = 4\varepsilon\Big(\log|\mathcal{X}| + \log|\mathcal{Y}| + \log\frac{1}{\varepsilon}\Big). \tag{5.26}$$

Claim 5.17 *If a rate pair $(R, R_c) \in \mathbb{R}^2_{\geq 0}$ is $(q_{X,Z}, q_{X,Y})$-achievable, then, for any $\varepsilon \in (0, 1/4)$, we have $(R, R_c) \in \mathcal{S}_\varepsilon$.*

Proof: Let (R, R_c) be $(q_{X,Z}, q_{X,Y})$-achievable. Fix $\varepsilon \in (0, 1/4)$. Then, there exists $n \in \mathbb{N}$ and an (n, R, R_c) code inducing a joint distribution P such that

$$\|P_{X^n,Y^n} - q_{X,Y}^{\otimes n}\|_{TV} \leq \varepsilon. \tag{5.27}$$

Let T be a uniform random variable over $[n]$, independent of all other variables. Define $W{=}(M, J, T)$. We prove that $(R, R_c) \in \mathcal{S}_\varepsilon$ through Claims 5.18, 5.19, and 5.20.

Claim 5.18 *P_{X_T,Z_T,W,Y_T} satisfies all properties in the definition (5.25) of $\mathcal{D}_\varepsilon$ other than the cardinality bound.*

Proof: From (5.27) and Lemma 1.8 (Appendix 1), we have

$$\forall t \in [n],\ \|P_{X_t,Y_t} - q_{X,Y}\|_{TV} \leq \varepsilon.$$

Since P_{X_T,Z_T} is an arithmetic average of the P_{X_t,Y_t}, then from the triangle inequality for the TVD, we have

$$\|P_{X_T,Y_T} - q_{X,Y}\|_{TV} \leq \varepsilon.$$

Since $(X^n, Z^n) \sim q_{X,Z}^{\otimes n}$, the distribution of (X_T, Z_T) is $q_{X,Z}$ and (X_T, Z_T) is independent of T. Since for any $t \in [n]$, we have the Markov chain relation $X_t - Z_t - (M, J, Y_t)$, and T is independent of (X^n, Z^n, M, J, Y^n), then we have the relation $X_T - (Z_T, T) - (M, J, Y_T)$. Since (X_T, Z_T) is independent of T, then we have $X_T - Z_T - T$. Therefore, from the chain rule for mutual information, we have

$$X_T - Z_T - (M, J, T, Y_T). \tag{5.28}$$

Since for any $t \in [n]$, we have the Markov chain relation $Z_t - (M, J) - Y_t$, and T is independent of (Z^n, M, J, Y^n), then we have $Z_T - (M, J, T) - Y_T$. Combining this with (5.28) yields the desired Markov chain relation $X_T - Z_T - (M, J, T) - Y_T$. ■

Claim 5.19 (R, R_c) *satisfies the inequalities in the definition* (5.24) *of* $\mathcal{S}_\varepsilon$ *for distribution* P_{X_T,Z_T,W,Y_T}.

Proof: We have

$$\begin{aligned} nR &\geq H(M) \\ &\geq H(M|J) \\ &\geq I(M; Z^n|J) \\ &= I(M, J; Z^n) \\ &= \textstyle\sum_{t=1}^n I(M, J; Z_t|Z_{1:t-1}) \end{aligned} \tag{5.29}$$

$$
\begin{aligned}
&= \textstyle\sum_{t=1}^{n} I(M, J, Z_{1:t-1}; Z_t) \\
&\geq \textstyle\sum_{t=1}^{n} I(M, J; Z_t) \\
&= nI(M, J; Z_T|T) && (5.30) \\
&= nI(W; Z_T), && (5.31)
\end{aligned}
$$

where (5.29) follows from the independence between the common randomness and the sources; and (5.30) and (5.31) follow from the independence of T and all other variables and from the fact that variables in $\{(Z_t)\}_{t\in[n]}$ are i.i.d. Moreover, we have

$$
\begin{aligned}
n(R + R_c) &\geq I(M, J; X^n, Y^n) \\
&= \sum_{t=1}^{n} I(M, J; X_t, Y_t|X_{1:t-1}, Y_{1:t-1}) \\
&= \sum_{t=1}^{n} \big[I(M, J, X_{1:t-1}, Y_{1:t-1}; X_t, Y_t) \\
&\quad - I(X_{1:t-1}, Y_{1:t-1}; X_t, Y_t)\big] \\
&\geq -ng(\varepsilon) + \sum_{t=1}^{n} I(M, J; X_t, Y_t) && (5.32) \\
&= nI(M, J; X_T, Y_T|T) - ng(\varepsilon) && (5.33) \\
&= nI(T, M, J; X_T, Y_T) - nI(T; X_T, Y_T) - ng(\varepsilon) \\
&\geq nI(W; X_T, Y_T) - 2ng(\varepsilon), && (5.34)
\end{aligned}
$$

where (5.32) and (5.34) follow from (5.27), the fact that $\varepsilon \in (0, 1/4)$, and [Cuf13, Lemma VI.3]; and (5.33) follows from the independence of T from all other variables. ■

Claim 5.20 *There exists a distribution $p_{X,Z,W,Y}$ under which W satisfies the cardinality bound in the definition* (5.25) *of $\mathcal{D}_\varepsilon$, the Markov chain relation $X - Z - W - Y$, and*

$$
\begin{aligned}
p_{X,Z,Y} &\equiv P_{X_T, Z_T, Y_T}, && (5.35) \\
H_p(Z|W) &= H_P(Z_T|W), && (5.36) \\
H_p(X, Y|W) &= H_P(X_T, Y_T|W). && (5.37)
\end{aligned}
$$

Proof: For any finite alphabet $\mathcal{L}$, we denote by $\Delta_{\mathcal{L}}$ the set of probability distributions on $\mathcal{L}$. Consider the following subset of $\mathbb{R}^{|\mathcal{Z}||\mathcal{Y}|}$

$$\mathcal{E} := \{(Q_Z(z)Q_Y(y))_{z\in\mathcal{Z},y\in\mathcal{Y}} | Q_Z \in \Delta_{\mathcal{Z}}, Q_Y \in \Delta_{\mathcal{Y}}\}. \tag{5.38}$$

This set lies inside the following hyperplane of $\mathbb{R}^{|\mathcal{Z}||\mathcal{Y}|}$

$$\mathcal{H} := \{(v_{z,y})_{z\in\mathcal{Z},y\in\mathcal{Y}} | \sum_{z\in\mathcal{Z},y\in\mathcal{Y}} v_{z,y} = 1\}, \tag{5.39}$$

which has dimension $|\mathcal{Z}||\mathcal{Y}| - 1$. Define the linear map

$$\phi : (v_{z,y})_{z\in\mathcal{Z},y\in\mathcal{Y}} \mapsto (q_{X|Z=z}(x)v_{z,y})_{x\in\mathcal{X},z\in\mathcal{Z},y\in\mathcal{Y}}, \tag{5.40}$$

which is injective. Then, $\phi(\mathcal{E})$ lies inside of the affine subspace $\phi(\mathcal{H})$, which is of dimension at most $|\mathcal{Z}||\mathcal{Y}| - 1$. Define the following subset of $\phi(\mathcal{H}) \times \mathbb{R}^2$

$$\begin{aligned}\tilde{\mathcal{E}} := \{&(q_{X|Z=z}(x)Q_Z(z)Q_Y(y))_{x\in\mathcal{X},z\in\mathcal{Z},y\in\mathcal{Y}} \\ &\cup (H_{Q_Z}(Z), H_{Q_ZQ_Yq_{X|Z}}(X,Y)) | Q_Z \in \Delta_{\mathcal{Z}}, Q_Y \in \Delta_{\mathcal{Y}}\},\end{aligned}$$

where $\cup$ denotes the concatenation of vectors. Set $\tilde{\mathcal{E}}$ is compact and connected because it is the image of $\mathcal{E}$ by a continuous map, and $\mathcal{E}$ is compact and connected. Moreover, $\tilde{\mathcal{E}}$ lies in $\phi(\mathcal{H}) \times \mathbb{R}^2$, which is a strict affine subspace of $\phi(\mathbb{R}^{|\mathcal{Z}||\mathcal{Y}|}) \times \mathbb{R}^2$ by injectivity of ϕ. Therefore, by the Caratheodory theorem for connected sets — see the proof of [Cuf13, Lemma VI.1] — any point in the convex hull of $\tilde{\mathcal{E}}$ can be written as a convex combination of $|\mathcal{Z}||\mathcal{Y}|+1$ elements of $\tilde{\mathcal{E}}$. From Claim 5.18, P_{X_T,Z_T,W,Y_T} satisfies the Markov chain relation $X_T - Z_T - W - Y_T$. Therefore,

$$\begin{aligned}P_{X_T,Z_T,Y_T}(x,z,y) &= \sum_w P_W(w) q_{X|Z=z}(x) \\ &\qquad \cdot P_{Z_T|W=w}(z) P_{Y_T|W=w}(y), \\ H_P(Z_T|W) &= \sum_w P_W(w) H(P_{Z_T|W=w}),\end{aligned}$$

and

$$H_P(X_T,Y_T|W) = \sum_w P_W(w) \cdot H(\sum_z P_{Z_T|W=w}(z) q_{X|Z=z} P_{Y_T|W=w}).$$

Then, the following vector

$$(P_{X_T,Z_T,Y_T}(x,z,y))_{x\in\mathcal{X},z\in\mathcal{Z},y\in\mathcal{Y}} \cup (H_P(Z_T|W), H_P(X_T,Y_T|W))$$

belongs to the convex hull of $\tilde{\mathcal{E}}$. Therefore, it is a convex combination of $|\mathcal{Z}||\mathcal{Y}|+1$ points of the $\tilde{\mathcal{E}}$, which is equivalent to Claim 5.20. ■

Let $p_{X,Z,W,Y}$ be a distribution such as in Claim 5.20. From Claims 5.18 and 5.20, we have $p \in \mathcal{D}_\varepsilon$. Combining this with Claim 5.19 yields $(R, R_c) \in \mathcal{S}_\varepsilon$. This being true for any $\varepsilon \in (0, 1/4)$, Claim 5.17 is proved. ■

Similarly to [Cuf13, Lemma VI.5], we have the following.

Claim 5.21

$$\bigcap_{\varepsilon\in(0,1/4)} \mathcal{S}_\varepsilon = \mathcal{S}^{(r.c.s.)}. \tag{5.41}$$

Proof: The proof is very similar to that of Claim 5.16 (itself very similar to the proof of [Cuf13, Lemma VI.5]). ■

Combining Claims 5.17 and 5.21 yields $\mathcal{A}^{(r.c.s.)} \subset \mathcal{S}^{(r.c.s.)}$, as desired.

5.B Achievability direction of Theorem 5.3

This proof closely tracks [Cuf13, Section V]. Here, we prove that $\mathcal{S}^{(r.c.s.)} \subseteq \overline{\mathcal{A}}^{(r.c.s.)}$.

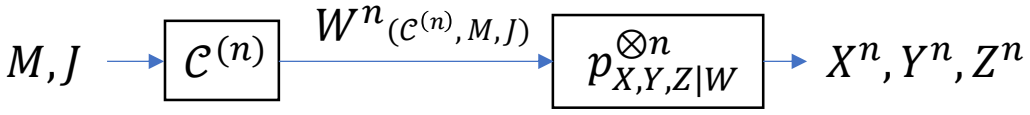


Figure 5.B.1: Graphical model for Q. See Equations (5.44) and (5.45).

5.B.1 Random codebook

Let (R, R_c) be a rate pair in $\mathcal{S}^{(r.c.s.)}$. Let $p_{X,Z,Y,W}$ be a corresponding distribution from the definition of $\mathcal{S}^{(r.c.s.)}$. Then, we have

$$R \geq I_p(Z;W) \tag{5.42}$$

$$R + R_c \geq I_p(X,Y;W) \tag{5.43}$$

Fix some $\varepsilon > 0$. For every $n \geq 1$, let $\mathcal{C}^{(n)}$ be a random codebook with $\lfloor 2^{n(R+\varepsilon)} \rfloor \times \lfloor 2^{nR_c} \rfloor$ i.i.d. codewords sampled from $p_W^{\otimes n}$. The codewords are indexed by pairs (m, j). We denote this random codebook distribution by $\mathbb{Q}_{\mathcal{C}^{(n)}}$. Given a realization $c^{(n)}$, the codeword of index (m, j) is denoted $w^n(c^{(n)}, m, j)$.

5.B.2 Distribution Q and soft-covering lemma

For every positive integer n, we define the distribution Q, described in Figure 5.B.1, as follows:

$$\begin{aligned} Q_{\mathcal{C}^{(n)},M,J,W^n,Z^n,Y^n,X^n} := \; & \mathbb{Q}_{\mathcal{C}^{(n)}} \cdot p^{\mathcal{U}}_{[2^{n(R+\varepsilon)}]} \cdot p^{\mathcal{U}}_{[2^{nR_c}]} \\ & \cdot \; \mathbf{1}_{W^n = w^n(c^{(n)},m,j)} \prod_{t=1}^{n} p_{Z|W=w_t} \\ & \cdot \; \prod_{t=1}^{n} p_{Y|W=w_t} \cdot \prod_{t=1}^{n} p_{X|Z=x_t}. \end{aligned} \tag{5.44}$$

By the definition of $\mathbb{Q}_{\mathcal{C}^{(n)}}$, we have $Q_{W^n} \equiv p_W^{\otimes n}$, and therefore, from (5.44) and the Markov chains in (5.8), we have

$$Q_{W^n,Z^n,Y^n,X^n} \equiv p_{W,Z,Y,X}^{\otimes n}. \tag{5.45}$$

We use the *soft covering lemma* [Cuf13, Corollary IV.1]. A nearly direct application gives, from

(5.42) and (5.43):

Claim 5.22

$$\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|Q_{X^n,Y^n|\mathcal{C}^{(n)}} - q_{X,Y}^{\otimes n}\|_{TV}\right] \underset{n\to\infty}{\longrightarrow} 0 \tag{5.46}$$

$$\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|Q_{J,Z^n|\mathcal{C}^{(n)}} - p^{\mathcal{U}}_{[2^{nR_c}]} q_Z^{\otimes n}\|_{TV}\right] \underset{n\to\infty}{\longrightarrow} 0. \tag{5.47}$$

The key difference with respect to the proof in [Cuf13] for standard channel synthesis is the following. We can introduce the remote source X via Lemma 1.9 (Appendix 1) with $L = X^n, U = (J, Z^n)$, and $\Pi_{L|U}$ the memoryless channel $\prod q_{X|Z}$: from (5.44) and (5.47), we get

$$\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|Q_{J,Z^n,X^n|\mathcal{C}^{(n)}} - p^{\mathcal{U}}_{[2^{nR_c}]} q_{Z,X}^{\otimes n}\|_{TV}\right] \underset{n\to\infty}{\longrightarrow} 0. \tag{5.48}$$

Proof of Claim 5.22:

Lemma 5.23 *[Cuf13, Corollary IV.1] Let $\mathcal{W}$ be a finite alphabet and ρ_W a distribution on the latter. Let R be a non-negative real number and let $(k_n)_{n\geq 1}$ be a sequence of positive integers satisfying $k_n/2^{nR} \underset{n\to\infty}{\to} 1$. For every positive integer n, let $\mathcal{E}^{(n)}$ be a randomly generated collection of k_n mutually independent sequences in $\mathcal{W}^n$ each drawn according to $\rho_W^{\otimes n}$. Its distribution is denoted $\Gamma_{\mathcal{E}^{(n)}}$. The sequences are indexed by some set $\mathcal{I}$ of size k_n and in a realization $e^{(n)}$ of $\mathcal{E}^{(n)}$ the sequence with index i is denoted by $w^n(e^{(n)}, i)$. A memoryless channel $(\rho_{U|W=w})_{w\in\mathcal{W}}$ induces an output distribution defined as*

$$\Gamma_{\mathcal{E}^{(n)},I,W^n,U^n} := \Gamma_{\mathcal{E}^{(n)}} \cdot p^{\mathcal{U}}_{[k_n]} \cdot \mathbf{1}_{w^n = w^n(e^{(n)},i)} \cdot \prod_{t=1}^{n} \rho^{(u_t)}_{U|W=w_t}. \tag{5.49}$$

If $R > I_\rho(W;U)$, then

$$\mathbb{E}_{\mathcal{E}^{(n)}}\left[\|\Gamma_{U^n|\mathcal{E}^{(n)}} - \rho_U^{\otimes n}\|_{TV}\right] \underset{n\to\infty}{\longrightarrow} 0,$$

where ρ_U is the marginal of $\rho_{W,U} = \rho_W \cdot \rho_{U|W}$.

We first use Lemma 5.23 with $\rho = p$, $\mathcal{E}^{(n)} = \mathcal{C}^{(n)}$, $k_n = \lfloor 2^{n(R+\varepsilon)} \rfloor \times \lfloor 2^{nR_c} \rfloor$, $I = (M, J)$, $U = (X,Y)$ and R replaced by $R + \varepsilon + R_c$. The corresponding output distribution is exactly $Q_{\mathcal{C}^{(n)},M,J,W^n,X^n,Y^n}$, by definition (5.44) of the latter. Indeed, from (5.44) distribution Q satisfies the Markov chain relation $(\mathcal{C}^{(n)}, M, J) - W^n - (X^n,Y^n)$, and from (5.45) we have $Q_{W^n,X^n,Y^n} \equiv p_{W,X,Y}^{\otimes n}$. Thus, from (5.43) and Lemma 5.23 together with the coordination property $p_{X,Y} \equiv q_{X,Y}$ from (5.8) we get (5.46).

Second, we use Lemma 5.23 as follows. For every $n \in \mathbb{N}$, we set $\mathcal{E}^{(n)}$ to be the sub-codebook of $\mathcal{C}^{(n)}$ corresponding to $j = 1$, denoted $\mathcal{C}_1^{(n)}$, and set $\rho = p$, $k_n = \lfloor 2^{n(R+\varepsilon)} \rfloor$. We further set $I = (M, 1)$, $U = Z$, and replace R in the lemma by $R+\varepsilon$. The corresponding output distribution is exactly $Q_{\mathcal{C}_1^{(n)},M,J,W^n,Z^n|J=1}$, by definition (5.44) of the latter. Indeed, from (5.44) distribution Q satisfies $(\mathcal{C}^{(n)}, M, J) - W^n - Z^n$, and from (5.45) we have $Q_{W^n,Z^n} \equiv p_{W,Z}^{\otimes n}$. From (5.3), we have $p_Z \equiv q_Z$. Thus, from (5.42) and Lemma 5.23 we have

$$\mathbb{E}_{\mathcal{C}_1^{(n)}} \left[\| Q_{Z^n|\mathcal{C}_1^{(n)},J=1} - q_Z^{\otimes n} \|_{TV} \right] \underset{n\to\infty}{\longrightarrow} 0. \tag{5.50}$$

From (5.44), we have $\forall j,\ Q_{Z^n|\mathcal{C}_j^{(n)},J=j} = \psi(\mathcal{C}_j^{(n)})$, where

$$\psi : \{w^n(m)\}_{m\in[k_n]} \in (\mathcal{W}^n)^{k_n} \mapsto \frac{1}{k_n} \sum_{m=1}^{k_n} \prod_{t=1}^{n} p_{Z|W=w_t(m)}$$

is a map which is defined independently of any index j. Therefore, $Q_{Z^n|\mathcal{C}^{(n)},J} \equiv Q_{Z^n|\mathcal{C}_J^{(n)},J}$. Moreover, since from (5.44) we have $Q_{\mathcal{C}^{(n)},J} \equiv \mathbb{Q}_{\mathcal{C}^{(n)}} p_{[2^{nR_c}]}^{\mathcal{U}}$, then, from Lemma 1.7 (Appendix 1),

$$\begin{aligned} &\| Q_{\mathcal{C}^{(n)},J,Z^n} - \mathbb{Q}_{\mathcal{C}^{(n)}} p_{[2^{nR_c}]}^{\mathcal{U}} p_Z^{\otimes n} \|_{TV} \\ &= \sum_{j=1}^{\lfloor 2^{nR_c} \rfloor} \frac{1}{\lfloor 2^{nR_c} \rfloor} \mathbb{E}_{\mathcal{C}_j^{(n)}} \left[\| Q_{Z^n|\mathcal{C}_j^{(n)},J=j} - q_Z^{\otimes n} \|_{TV} \right]. \end{aligned} \tag{5.51}$$

Moreover, by construction of $\mathcal{C}^{(n)}$, the distribution of sub-codebook $\mathcal{C}_j^{(n)} = (W^n(\mathcal{C}^{(n)}, m, j))_m$ is independent of j. Hence, this is also the case for $\psi(\mathcal{C}_j^{(n)})$, and all expectations in (5.51) are

identical. Using (5.50) and Lemma 1.7 (Appendix 1) we get (5.47). ■

5.B.3 Choosing a codebook

We know that L_1 convergence implies convergence in probability. Therefore, we get convergence in probability from (5.46) and (5.48). Then, for a certain N_0, for every integer $n \geq N_0$ there is a codebook $c_*^{(n)}$ such that

$$\|Q_{X^n,Y^n|\mathcal{C}^{(n)}=c_*^{(n)}} - q_{X,Y}^{\otimes n}\|_{TV} \leq \varepsilon, \tag{5.52}$$

$$\|Q_{J,Z^n,X^n|\mathcal{C}^{(n)}=c_*^{(n)}} - p_{[2^{nR_c}]}^{\mathcal{U}} q_{Z,X}^{\otimes n}\|_{TV} \leq \varepsilon. \tag{5.53}$$

5.B.4 Construction of a code

In this subsection, we show how the distribution $Q_{|\mathcal{C}^{(n)}=c_*^{(n)}}$ can be modified to lead to a code, with a negligible effect on the coordination performance. We define the distribution P which constitutes an $(n, R{+}\varepsilon, R_c)$ code. It differs from Q in having the correct marginal for (X^n,Z^n, J): let

$$\begin{aligned} P_{J,X^n,Z^n,M,Y^n} := p_{[2^{nR_c}]}^{\mathcal{U}} \cdot q_{X,Z}^{\otimes n} \cdot Q_{M|\mathcal{C}^{(n)}=c_*^{(n)},\, J=j,\, Z^n=z^n} \\ \cdot \;\; Q_{Y^n|\mathcal{C}^{(n)}=c_*^{(n)},\, M=m,\, J=j}, \end{aligned} \tag{5.54}$$

which from (5.44) can be written as

$$P_{J,X^n,Z^n,M,Y^n} = p_{[2^{nR_c}]}^{\mathcal{U}} \cdot q_{X,Z}^{\otimes n} \cdot Q_{M,\,Y^n|\mathcal{C}^{(n)}=c_*^{(n)},\, J=j,\, Z^n=z^n}. \tag{5.55}$$

From (5.54) distribution P satisfies

$$X^n - Z^n - (M, J) - Y^n. \tag{5.56}$$

The alphabet of M is $[2^{n(R+\varepsilon)}]$ throughout this proof by construction of the random codebook $\mathcal{C}^{(n)}$. Therefore, from (5.55)-(5.56) and Lemma 5.2, distribution P defines an $(n, R+\varepsilon, R_c)$ code. From (5.44), (5.55), and Lemma 1.9 (Appendix 1), comparing P with $Q_{|\mathcal{C}^{(n)}=c_*^{(n)}}$ reduces to comparing marginals, i.e., to (5.53):

$$\begin{aligned}&\left\|P_{J,X^n,Z^n,M,Y^n} - Q_{J,X^n,Z^n,M,Y^n|\mathcal{C}^{(n)}=c_*^{(n)}}\right\|_{TV}\\ &\quad= \left\|P_{J,X^n,Z^n} - Q_{J,X^n,Z^n|\mathcal{C}^{(n)}=c_*^{(n)}}\right\|_{TV} \leq \varepsilon. \end{aligned} \tag{5.57}$$

Therefore, from Lemma 1.8 (Appendix 1) with $U = (X^n,Y^n)$, (5.52), and the triangle inequality, we get

$$\left\|P_{X^n,Y^n} - q_{X,Y}^{\otimes n}\right\|_{TV} \leq 2\varepsilon. \tag{5.58}$$

This proves that $(R+\varepsilon, R_c)$ is $(q_{X,Z},q_{X,Y})$-achievable. This is true for every $\varepsilon > 0$. Therefore, we get $(R, R_c) \in \overline{\mathcal{A}}^{(r.c.s.)}$. This being true for any $(R, R_c) \in \mathcal{S}^{(r.c.s.)}$, we have $\mathcal{S}^{(r.c.s.)} \subseteq \overline{\mathcal{A}}^{(r.c.s.)}$, as desired.

5.C Proof of Proposition 5.5

Since the notion of compatibility with $(q_{X,Z}, q_{X,Y})$ consists of linear conditions, the set

$$\bigcup_{q_{X,Z,Y}} \mathcal{D}^{(d.c.s.)}(q_{X,Z,Y})$$

is closed. Since it is a subset of a finite-dimensional real vector space, then it is compact. Therefore, by continuity of the mutual information over finite alphabets, and due to non-strict inequalities, the set

$$\bigcup_{q_{X,Z,Y}} \mathcal{S}^{(d.c.s.)}(q_{X,Z,Y})$$

is closed. Similarly, $\mathcal{S}^{(r.c.s.)}$ is also closed. This justifies the existence of the minima in (5.15).

Claim 5.24 *It is sufficient to prove* (5.15) *for* $R_c = 0$.

Proof: Assume that (5.15) holds for $R_c = 0$. Denote by r the corresponding positive difference between both sides of (5.15), and by R_0 the rate achieving the minimum in the right hand side. Fix $R_c \in (0, r)$. For any $q_{X,Z,Y}$ and any $R \in \mathbb{R}_{\geq 0}$ such that $(R, R_c) \in \mathcal{S}^{(d.c.s.)}(q_{X,Z,Y})$, we have from the definition of $\mathcal{S}^{(d.c.s.)}(q_{X,Z,Y})$ (Eq. 5.9) that $(R+R_c, 0) \in \mathcal{S}^{(d.c.s.)}(q_{X,Z,Y})$, hence $R + R_c \geq R_0 + r > R_0 + R_c$. Then, (5.15) follows from the fact that $(R_0, R_c) \in \mathcal{S}^{(r.c.s.)}$ (since $(R_0, 0) \in \mathcal{S}^{(r.c.s.)}$). ∎

We consider $R_c = 0$ in the remainder of Appendix 5.C. Fix R as the rate attaining the first minimum, and consider a distribution $q_{X,Z,Y}$ compatible with $(q_{X,Z}, q_{X,Y})$ such that $(R, 0) \in \mathcal{S}^{(d.c.s.)}(q_{X,Z,Y})$. Then, there exists a distribution $p_{X,Z,W,Y} \in \mathcal{D}^{(d.c.s.)}(q_{X,Z,Y})$ such that $R \geq I_p(Z, Y; W)$. By minimality of R, we have $R = I_p(Z, Y; W)$. Therefore, $R \geq C_p(Z, Y)$, where $C_p(Z, Y)$ denotes the Wyner common information between Z and Y, defined as

$$C_p(Z, Y) := \min_{W \text{ s.t. } Z-W-Y} I(Z, Y; W). \tag{5.59}$$

We only have the following five cases.

Case 1: $R > C_p(Z, Y)$.

Let $\tilde{R} \in (C_p(Z, Y), R)$. Then, by definition of the Wyner common information, there exists $p_{\tilde{W}|Z,Y}$ such that $p_{Z,Y} \cdot p_{\tilde{W}|Z,Y}$ satisfies the Markov chain relation $Z - \tilde{W} - Y$ and $I_p(Z, Y; \tilde{W}) \leq \tilde{R}$. Therefore, we have $(\tilde{R}, 0)$ belongs to $\mathcal{S}^{(d.c.s.)}(q_{X,Z,Y})$, by definition of the latter. By construction, we have $\tilde{R} < R$. This contradicts the minimality of R.

Case 2: $R = C_p(Z, Y)$ and $q_{Z,Y}(Y \neq Z) = 0$.

Then, $q_{X,Y}(X \neq Y) = q_{X,Z}(X \neq Z)$, which contradicts the assumptions.

Case 3: $R = C_p(Z, Y)$ and $q_{Z,Y}(Y \neq Z) = 1$.

Then, $q_{X,Y}(X \neq Y) = 1 - q_{X,Z}(X \neq Z) > 1/2$, which contradicts the assumptions.

Case 4: $R = C_p(Z,Y)$ and $q_{Z,Y}(Y \neq Z) \in [1/2, 1)$.

Define $\theta = q_{Z,Y}(Y \neq Z)$ and $\tau = q_{X,Z}(X \neq Z)$. Then,

$$\begin{aligned} q_{X,Y}(X \neq Y) &= \tau(1-\theta) + (1-\tau)\theta \\ &= 1/2 + (\theta - 1/2)(1 - 2\tau) \geq 1/2. \end{aligned} \tag{5.60}$$

This contradicts the assumptions.

Case 5: $R = C_p(Z,Y)$ and $q_{Z,Y}(Y \neq Z) \in (0, 1/2)$.

Define $\theta = q_{Z,Y}(Y \neq Z)$. Since $q_Z \equiv q_Y \equiv Ber(1/2)$, then (Z,Y) is a *doubly binary symmetric source of parameter* θ [EGK11, Example 10.1], that is $q_{Z,Y}(0,1) = q_{Z,Y}(1,0) = \theta/2$ and $q_{Z,Y}(0,0) = q_{Z,Y}(1,1) = (1-\theta)/2$. This is abbreviated as $q_{Z,Y} \equiv DSBS(\theta)$. Therefore, as shown in [Wyn75, Section 3], one obtains $p_{Z,Y} \equiv q_{Z,Y}$ and $I_p(Z,Y;\tilde{W}) = C_p(Y,Z)$ when choosing $p_{Z,\tilde{W},Y}$ of the form $q_Z \cdot p_{\tilde{W}|Z} \cdot p_{Y|Z}$ with $p_{\tilde{W}|Z}$ and $p_{Y|\tilde{W}}$ identical BSC($\tilde{\theta}$), where

$$\tilde{\theta} = \frac{1}{2} - \frac{1}{2}\sqrt{1-2\theta}. \tag{5.61}$$

Define $p_{X,Z,\tilde{W},Y} := p_{X|Z} \cdot p_{Z,\tilde{W},Y}$ and

$$\tilde{R} := \max(I_p(Z;\tilde{W}), I_p(X,Y;\tilde{W})). \tag{5.62}$$

Then $(\tilde{R}, 0)$ belongs to $\mathcal{S}^{(r.c.s.)}$, by definition of the latter. It remains to show that $\tilde{R} < R$. Let

$h(\cdot)$ denote the binary entropy function. We have

$$\begin{aligned} R - I_p(Z;\tilde{W}) &= I_p(Z,Y;\tilde{W}) - I_p(Z;\tilde{W}) \\ &= I_p(Y;\tilde{W}|Z) \\ &= I_p(Y;\tilde{W},Z) - I_p(Z;Y) \\ &= I_p(Y;\tilde{W}) - I_p(Z;Y) \qquad (5.63) \\ &= 1 - h(\tilde{\theta}) - 1 + h(\theta) > 0, \qquad (5.64) \end{aligned}$$

where the last line follows from the fact that $\tilde{\theta} < \theta < 1/2$, which follows from the fact that $\sqrt{u} > u$ for $u \in (0,1)$. It remains to show that $R > I_p(X,Y;\tilde{W})$. By construction,

$$p_{Z,\tilde{W}} \equiv p_{\tilde{W},Y} \equiv DSBS(\tilde{\theta}). \qquad (5.65)$$

Moreover, with $\tau := q_{X,Z}(X \neq Z)$, we have

$$\begin{aligned} p_{X,\tilde{W}} &\equiv DSBS((1-\tau)\tilde{\theta} + \tau(1-\tilde{\theta})) \\ &\equiv DSBS(1/2 - (1/2-\tau)\sqrt{1-2\theta}), \qquad (5.66) \\ p_{X,Y} &\equiv DSBS((1-\tau)\theta + \tau(1-\theta)) \\ &\equiv DSBS(\tau + (1-2\tau)\theta). \end{aligned}$$

Then,

$$\begin{aligned} R - I_p(X,Y;\tilde{W}) &= I_p(Z,Y;\tilde{W}) - I_p(X,Y;\tilde{W}) \\ &= I_p(Z;\tilde{W}) + I_p(Y;\tilde{W}) - I_p(Z;Y) \\ &\quad - I_p(X;\tilde{W}) - I_p(Y;\tilde{W}) + I_p(X;Y) \qquad (5.67) \\ &= 1 - h(1/2 - (1/2)\sqrt{1-2\theta}) - 1 + h(\theta) \\ &\quad - 1 + h(1/2 - (1/2-\tau)\sqrt{1-2\theta}) \\ &\quad + 1 - h(\tau + (1-2\tau)\theta), \qquad (5.68) \end{aligned}$$

where each line in (5.67) follows from the chain of equalities leading to (5.63). A plot of the right hand side of (5.68) is provided in Figure 5.6. To conclude the proof, we use the following lemma.

Lemma 5.25 *Consider reals α, β such that $\alpha < \beta$, and a concave increasing function $f : (\alpha, \beta) \to \mathbb{R}$. Consider a, b, c, d in (α, β) such that*

$$a < \min(b, c),\ d > \max(b, c),\ b + c > a + d. \tag{5.69}$$

Then, $-f(a) + f(b) + f(c) - f(d) > 0$.

Proof: Denote $b' = \min(b, c)$ and $c' = \max(b, c)$. Quantity $-f(a) + f(b) + f(c) - f(d)$ rewrites as

$$(b' - a)\frac{f(b') - f(a)}{b' - a} - (d - c')\frac{f(d) - f(c')}{d - c'}. \tag{5.70}$$

By assumption, we have that $b' - a > d - c'$. By concavity of f, and since $\max(a, b') \leq \min(c', d)$, the slope between a and b' is greater or equal to the slope between c' and d. These slopes are (strictly) positive because f is increasing, which concludes. ■

It remains to check that we can apply the lemma, with $f = h$, to the right hand side of (5.68). First, h is concave and increasing on $(0, 1/2)$. It is straightforward to check that the arguments of h in (5.68) lie in $(0, 1/2)$, and that the first two conditions in (5.69) hold. Moreover,

$$\begin{aligned} &\theta + [1/2 - (1/2 - \tau)\sqrt{1 - 2\theta}] \\ &- [1/2 - 1/2\sqrt{1 - 2\theta}] - [\tau + (1 - 2\tau)\theta] \\ &= \tau\sqrt{1 - 2\theta} - \tau(1 - 2\theta) > 0. \end{aligned} \tag{5.71}$$

Thus, the third condition in (5.69) holds, and Lemma 5.25 can be applied. This concludes the proof in Case 5, which concludes the proof of Proposition 5.5.

5.D Proof of Proposition 5.6

Assume the setting of Proposition 5.5. Fix R in $[R_{\min}, R_1)$, and $R_c \in [0, R_1 - R)$ such that (R, R_c) is $(q_{X,Z}, q_{X,Y})$-achievable. For the sake of contradiction, assume that there exists a (R, R_c) scheme for $q_{X,Y}$ which is not a proper vector scheme. This scheme can be turned into a $(R + R_c, 0)$ scheme for $q_{X,Y}$ having the same induced P_{Z^n,Y^n}, by sampling variable J locally at the encoder, and taking (M, J) as message. In the remainder of the proof, we only study this new scheme. By assumption and by construction, the latter is not a proper vector scheme. Then, there exists a sequence $\{\rho^{(t)}_{Y|Z}\}_{t\in\mathbb{N}}$ of single-letter distributions for which

$$\left\| P_{Z^n,Y^n} - \prod_{t=1}^{n} q_Z \cdot \rho^{(t)}_{Y|Z} \right\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{5.72}$$

Since there are only a finite number of possible single-letter conditional distributions, then (e.g., from [Cuf13, Lemma V.1]) P_{Z^n,Y^n} is nearly i.i.d. *by part*. We show that there exits a $(R+R_c, 0)$ scheme for $q_{X,Y}$ which performs direct strong coordination by synthesizing one such conditional distribution. First, we have

Claim 5.26 *For any $t \in \mathbb{N}$, the conditional distribution $\rho^{(t)}_{Y|Z}$ is compatible with $(q_{X,Z}, q_{X,Y})$.*

Proof: From (5.72), from [Cuf13, Lemma V.2], we have

$$\left\| P_{X^n,Z^n,Y^n} - \prod_{t=1}^{n} q_{X,Z} \cdot \rho^{(t)}_{Y|Z} \right\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{5.73}$$

Thus, from [Cuf13, Lemma V.1], we have

$$\left\| P_{X^n,Y^n} - \prod_{t=1}^{n} \Big[\sum_z q_Z(z) q_{X|Z=z} \cdot \rho^{(t)}_{Y|Z=z} \Big] \right\|_{TV} \underset{n\to\infty}{\longrightarrow} 0.$$

Since the scheme is an (R, R_c) scheme for $q_{X,Y}$, then, from the triangle inequality for the TVD, we have

$$\left\| q^{\otimes n}_{X,Y} - \prod_{t=1}^{n} \Big[\sum_z q_Z(z) q_{X|Z=z} \cdot \rho^{(t)}_{Y|Z=z} \Big] \right\|_{TV} \underset{n\to\infty}{\longrightarrow} 0.$$

For any $t \in \mathbb{N}$, from [Cuf13, Lemma V.1] with $U = (X_t, Y_t)$ therein, we obtain

$$\left\| q_{X,Y} - \sum_z q_Z(z) q_{X|Z=z} \cdot \rho^{(t)}_{Y|Z=z} \right\|_{TV} \underset{n\to\infty}{\longrightarrow} 0.$$

Since this quantity does not depend on n, it is null. Hence, $\rho^{(t)}_{Y|Z}$ is compatible with $(q_{X,Z}, q_{X,Y})$.

■

Second, we have the following property.

Claim 5.27 *There exits $t \in \mathbb{N}$ such that*

$$(R + R_c, 0) \in \mathcal{S}^{(c.s.)}(\rho^{(t)}_{Y|Z}), \tag{5.74}$$

where, for a conditional distribution $\mu_{Y|Z}$, the set $\mathcal{S}^{(c.s.)}(\mu_{Y|Z})$ is defined as the achievable region for the standard channel synthesis problem [Cuf13, Eq 8 & 9] for channel $\mu_{Y|Z}$ and source distribution q_Z.

Proof: We present a sketch of the proof. The technical details are very similar to those of the converse proof of Theorem 5.3. Following steps similar to those in the converse proof of [Cuf13, Section VI] for the standard problem of channel synthesis, we prove the following. For every k, there exits t_k and n_k such that

$$\| P_{Z_T, Y_T} - q_Z \cdot \rho^{(t_k)}_{Y|Z} \|_{TV} \leq 1/k \tag{5.75}$$

and

$$R + R_c \geq I_P(W; Z_T, Y_T) - g(1/k), \tag{5.76}$$

where P denotes the distribution induced by the n_k-th code of the scheme, T is a certain random variable taking values in $[n_k]$, and $g : \mathbb{R}_+ \to \mathbb{R}$ is a function with limit 0 at 0. From a continuity argument similar to that in [Cuf13, Section VI], and from the fact that there exist only a finite number of different conditional distributions $\rho_{Y|Z}$, we obtain the desired conclusion. ■

The above two claims imply that $(R+R_c,0) \in \mathcal{S}^{(d.c.s.)}(q_{X,Z} \cdot \rho_{Y|Z}^{(t)})$ for a certain $t \in \mathbb{N}$. From (5.15), this contradicts the assumption that $R+R_c < R_1$, which concludes the proof.

5.E Continuation of the proof of Theorem 5.9

We prove a technical lemma in Section 5.E.1. Then, we prove Proposition 5.11 in Section 5.E.2, and Theorem 5.9 in Section 5.E.3.

5.E.1 On joint typicality in Theorem 5.10

Lemma 5.28 *In the scheme which yields Theorem 5.10, for any $\delta \in (0,1)$, the deterministic mappings $\{w^n(\cdot)\}_{n\geq 1}$ can be chosen such that the probability of the event*

$$(U^n, V^n, K^n, w^n(K^n)) \notin \mathcal{T}_\delta^{(n)}(p_{U,V,K,W}) \tag{5.77}$$

vanishes as n goes to infinity.

Proof: The proof of Theorem 5.10, i.e. that of [B$^+$21, Theorem 2], consists in showing the following. Let $P^{(n,w^n(\cdot))}$ denote the distribution induced by the scheme described in Theorem 5.10, with a choice of mapping $w^n(\cdot)$. Then,

$$\mathbb{E}_{\mathcal{C}^{(n)}}\left[\|P_{X^n,Y^n}^{(n,\mathcal{C}^{(n)})} - q_{X,Y}^{\otimes n}\|_{TV}\right] \underset{n\to\infty}{\longrightarrow} 0. \tag{5.78}$$

where $\mathcal{C}^{(n)}$ is a family $\{W_{k^n}^n\}_{k^n \in \mathcal{K}^n}$ of mutually independent variables, with the variable of index k^n being distributed according to $\prod_{t=1}^n p_{W|K=k_t}$. The latter proof is concluded by choosing any realization if $\mathcal{C}^{(n)}$ achieving a TVD below average. See the end of [B$^+$21, Section III-C] and [B$^+$21, Equation 45]. Therefore, to guarantee the existence of a realization of $\mathcal{C}^{(n)}$ which also satisfies (5.77), it only remains to show that the probability that $(U^n, V^n, K^n, W_{K^n}^n)$ is not

δ-typical vanishes as n goes to infinity. Fix $\delta \in (0,1)$, and $\delta' < \delta$. Then, we have

$$\begin{aligned}
&\mathbb{P}_{\mathcal{C}^{(n)},U^n,V^n,K^n}\Big((U^n,V^n,K^n,W^n_{K^n}) \notin \mathcal{T}^{(n)}_{\delta}(p)\Big) \\
&\leq p^{\otimes n}_{U,V,K}\Big((U^n,V^n,K^n) \notin \mathcal{T}^{(n)}_{\delta'}(p)\Big) + \sup_{(u^n,v^n,k^n)\in\mathcal{T}^{(n)}_{\delta'}} \\
&\Big(\prod_{t=1}^{n} p_{W|U=u_t,V=v_t,K=k_t}\Big)\Big((u^n,v^n,k^n,W^n_{k^n}) \notin \mathcal{T}^{(n)}_{\delta}(p)\Big).
\end{aligned}$$

The first probability in the right hand side vanishes as n goes to infinity, by definition of the typical set. We show that the second probability is upper bounded uniformly over $\mathcal{T}^{(n)}_{\delta'}$, by a term which vanishes as n goes to infinity. Fix $(u_0,v_0,k_0,w_0) \in \mathcal{U}\times\mathcal{V}\times\mathcal{K}\times\mathcal{W}$, such that $p_{Z,L,O}(u_0,v_0,k_0,w_0) > 0$. The probability according to $\prod_{t=1}^{n} p_{W|U=u_t,V=v_t,K=k_t}$ of the following event

$$\begin{aligned}
&\Big|\#\big\{i\in[n]|(u_i,v_i,k_i,W_i)=(u_0,v_0,k_0,w_0)\big\}/n \\
&\quad -p_{Z,L,O}(z_0,l_0,o_0)\Big| \leq \delta p_{Z,L,O}(z_0,l_0,o_0)
\end{aligned} \tag{5.79}$$

is upper bounded, in [EGK11, Appendix 2A], by a term of the form $\phi(a_{u^n,v^n,k^n})$. Map $\phi : \mathbb{N}_{\geq 0} \to [0,\infty)$ is a deterministic function which vanishes at infinity, and a_{u^n,v^n,k^n} is the number of occurrences of (u_0,v_0,k_0) in (u^n,v^n,k^n). We know that typicality entails $a_{u^n,v^n,k^n} \geq (1-\delta')np_{U,V,K}(u_0,v_0,k_0)$, for every (u^n,v^n,k^n) in $\mathcal{T}^{(n)}_{\delta'}(p_{U,V,K})$ [EGK11], and we have assumed $p_{Z,L,O}(u_0,v_0,k_0) > 0$. Therefore, the probability of the event in (5.79) tends to 0 as n goes to infinity, uniformly over all (u^n,v^n,k^n) in $\mathcal{T}^{(n)}_{\delta'}(p_{U,V,K})$. A union bound over all $(u_0,v_0,k_0,w_0) \in \mathcal{U}\times\mathcal{V}\times\mathcal{K}\times\mathcal{W}$ concludes the proof. ■

5.E.2 Proof of Proposition 5.11

We use the following result, from the main proof in [BDH$^+$14] - precisely the end of Section III.C therein. It pertains to the existence of a good channel simulation scheme tailored to a given joint type. To prove Proposition 5.11, we then use this only for typical types.

Proposition 5.29 *For every $\varepsilon>0$, there exists a positive integer N_ε, such that for all $n\geq N_\varepsilon$, there exists a family $\{(F^{(n,\tau)},G^{(n,\tau)})\}_{\tau\in\mathcal{T}^{(n)}(\mathcal{Z}\times\mathcal{L}\times\mathcal{O})}$ of codes, with the code of index τ being a $(n, I_\tau(Z;L)+\varepsilon, I_\tau(Z,O;L)-I_\tau(Z;L)+\varepsilon)$ code for $(\mathcal{Z},\mathcal{O})$, such that the following holds. Consider the following code for $(\mathcal{Z},\mathcal{O})$, where common randomness is assumed to be available at rate R_c satisfying*

$$\forall\tau\in\mathcal{T}^{(n)}(\mathcal{Z}\times\mathcal{L}\times\mathcal{O}),\ R_c\geq I_\tau(Z,O;L)-I_\tau(Z;L)+\varepsilon:=R_{c,\tau}.$$

- *From its input z^n, the encoder samples $(\tilde{L}^n,\tilde{O}^n)\sim\prod_{t=1}^n p_{L,O|Z=z_t}$, and sends the joint type $\tau:=t_{\mathcal{Z}\times\mathcal{L}\times\mathcal{O}}(z^n,\tilde{L}^n,\tilde{O}^n)$. There shall be no further use of $\tilde{L}^n,\tilde{O}^n$.*
- *Using $F^{(n,\tau)}$ and the integer corresponding to the first $\lfloor nR_{c,\tau}\rfloor$ bits of the common randomness' binary representation, it then encodes z^n and sends the resulting message.*
- *The decoder (which has received τ) uses $G^{(n,\tau)}$, the received message, and the integer corresponding to the first $\lfloor nR_{c,\tau}\rfloor$ bits of the common randomness' binary representation, to output a string O^n.*

Then, for any $n\geq N_\varepsilon$, and every z^n satisfying $p_Z^{\otimes n}(z^n)>0$, the above scheme results in a conditional distribution $P^{(n)}_{O^n|Z^n=z^n}$ such that

$$\left\|P^{(n)}_{O^n|Z^n=z^n}-\prod_{t=1}^n p_{O|Z=z_t}\right\|_{TV}\leq\varepsilon. \tag{5.80}$$

We proceed to proving Proposition 5.11. Let $p_{Z,L,O}$ be as in the latter. Fix some $\varepsilon>0$, and let $N_\varepsilon, R_c, \{P^{(n)}\}_{n\geq1}$ be the corresponding objects from Proposition 5.29. We use the scheme of the latter whenever $(z^n,\tilde{L}^n,\tilde{O}^n)\in\mathcal{T}_\delta^{(n)}(p_{Z,L,O})$, for some small enough $\delta>0$ to be specified later. Otherwise, the encoder sends an error message, and the decoder produces a default output. The remainder of this section is dedicated to analyzing this scheme. From Proposition 5.29, the definition of $\mathcal{T}_\delta^{(n)}(p_{Z,L,O})$ [EGK11], and the continuity of mutual information, there exists a choice of δ such that for any $n\geq N_\varepsilon$, our scheme defines an $(n, I_p(Z;L)+2\varepsilon, I_p(Z,O;L)-I_p(Z;L)+2\varepsilon)$ code for $(\mathcal{Z},\mathcal{O})$. Let $\tilde{P}^{(n)}_{Z^n,J,M,O^n,\tilde{L}^n,\tilde{O}^n}$ denote the distribution induced by this code,

and let $\tilde{R}_c := I_p(Z,O;L) - I_p(Z;L) + 2\varepsilon$. For any $j \in [2^{nR_c}]$, let $j_{\lfloor n\tilde{R}_c \rfloor}$ denote the sequence of first $\lfloor n\tilde{R}_c \rfloor$ bits of j. For any $n \geq N_\varepsilon$, any $\tilde{j} \in [2^{n\tilde{R}_c}]$, and any $(z^n, \tilde{l}^n, \tilde{o}^n) \in \mathcal{T}_\delta^{(n)}(p_{Z,L,O})$, we have

$$\begin{aligned} &\tilde{P}^{(n)}_{M,O^n|Z^n=z^n,\tilde{L}^n=\tilde{l}^n,\tilde{O}^n=\tilde{o}^n,J=\tilde{j}} \\ &\equiv P^{(n)}_{M,O^n|Z^n=z^n,\tilde{L}^n=\tilde{l}^n,\tilde{O}^n=\tilde{o}^n,J_{\lfloor n\tilde{R}_c \rfloor}=\tilde{j}}, \qquad (5.81) \\ &\text{and} \quad \tilde{P}^{(n)}(Z^n{=}z^n, \tilde{L}^n{=}\tilde{l}^n, \tilde{O}^n{=}\tilde{o}^n, J{=}\tilde{j}) \\ &= P^{(n)}(Z^n{=}z^n, \tilde{L}^n{=}\tilde{l}^n, \tilde{O}^n{=}\tilde{o}^n, J_{\lfloor n\tilde{R}_c \rfloor}{=}\tilde{j}). \qquad (5.82) \end{aligned}$$

Indeed, under both $\tilde{P}^{(n)}$ and $P^{(n)}$, the common randomness J is independent from $(Z^n, \tilde{L}^n, \tilde{O}^n)$. Thus, conditioned on a realization of the latter, for any $k \leq R_c$, the distribution of the first k bits of J is independent of R_c. From (5.81), (5.82), Proposition 5.29, and the triangle inequality for the TVD, for any $n \geq N_\varepsilon$ and any $z^n \in \mathcal{Z}^n$ such that $p_Z^{\otimes n}(z^n) > 0$, we have

$$\begin{aligned} &\left\| \tilde{P}^{(n)}_{O^n|Z^n=z^n} - \textstyle\prod_{t=1}^n p_{O|Z=z_t} \right\|_{TV} \leq \\ &\left\| P^{(n)}_{O^n|Z^n=z^n} - \textstyle\prod_{t=1}^n p_{O|Z=z_t} \right\|_{TV} + \left\| \tilde{P}^{(n)}_{O^n|Z^n=z^n} - P^{(n)}_{O^n|Z^n=z^n} \right\|_{TV} \\ &\leq \varepsilon + \Big(\textstyle\prod_{t=1}^n p_{L,O|Z=z_t} \Big) \Big((z^n, \tilde{L}^n, \tilde{O}^n) \notin \mathcal{T}_\delta^{(n)}(p_{Z,L,O}) \Big). \qquad (5.83) \end{aligned}$$

For any given $\delta > 0$, the probability in the right hand side of (5.83) is upper bounded uniformly over all z^n in $\mathcal{T}_{\delta/2}^{(n)}(p_Z)$, by a term which vanishes as n goes to infinity. This is shown, albeit with different notation, in the proof of Lemma 5.28. Since this analysis is valid for any ε, then Proposition 5.11 is proved.

5.E.3 Proof of Theorem 5.9

Scheme and relation to Theorem 5.10

Fix $(R_A, R_{c,A}, R_B, R_{c,B})$ in the interior of $\mathcal{S}$. Then, there exists some $\varepsilon^* > 0$ such that for any $\varepsilon < \varepsilon^*$, we have $(R_A - \varepsilon, R_{c,A} - \varepsilon, R_B - \varepsilon, R_{c,B} - \varepsilon) \in \mathcal{S}$. Consider $p \in \mathcal{D}$ corresponding to the latter tuple. Since p satisfies $(U,W) - L_A - X$, Proposition 5.11 applies. Let

$\delta_A, N_A, \{(F^{(A,n,1)}, G^{(A,n,1)})\}_{n\geq 1}$ be the objects from the latter, corresponding to ε. Note that for any $n \geq N_A$, $(F^{(A,n,1)}, G^{(A,n,1)})$ is an $(n, R_A, R_{c,A})$ code for $(\mathcal{U} \times \mathcal{W}, \mathcal{X})$. Define δ_B, N_B, and $\{(F^{(B,n,1)}, G^{(B,n,1)})\}_{n\geq 1}$ similarly. We apply Theorem 5.10, with the minor improvement of Lemma 5.28, for $\delta = \min(\delta_A, \delta_B)$. Let $P^{(n,2)}_{U^n,V^n,K^n,W^n,X^n,Y^n}$ denote the distribution induced by the corresponding scheme. Then, there exists $N_1 \in \mathbb{N}$ such that for any $n \geq N_1$, we have

$$\|P^{(n,2)}_{X^n,Y^n} - q^{\otimes n}_{X,Y}\|_{TV} \leq \varepsilon. \tag{5.84}$$

Let $N_2 := \max(N_1, N_A, N_B)$. For any $n \geq N_2$, consider the following $(n, R_A, R_{c,A}, R_B, R_{c,B})$ code, with induced distribution denoted $P^{(n,1)}$.

- Using its input U^n, Encoder A samples (K^n, W^n) using $P^{(n,2)}_{K^n,W^n|U^n}$.
- If $(U^n, W^n) \in \mathcal{T}^{(n)}_{\delta_A}(p_{U,W})$, then Encoder A chooses a message using $F^{(A,n,1)}$, then Decoder A uses the message and $G^{(A,n,1)}$ to produce an output string X^n.
- Otherwise, Encoder A sends an error message, and Decoder A outputs a default string.
- Encoder B and Decoder B follow similar steps.

From (5.21), for any $n \geq N_2$, for all $(u^n, w^n) \in \mathcal{T}^{(n)}_{\delta_A}(p_{U,W})$, we have

$$\left\| P^{(n,1)}_{X^n|U^n=u^n,W^n=w^n} - \prod_{t=1}^n p_{X|U=u_t,W=w_t} \right\|_{TV} \leq \varepsilon,$$

i.e.

$$\left\| P^{(n,1)}_{X^n|U^n=u^n,W^n=w^n} - P^{(n,2)}_{X^n|U^n=u^n,W^n=w^n} \right\|_{TV} \leq \varepsilon. \tag{5.85}$$

Then, since $P^{(n,1)}_{U^n,W^n} \equiv P^{(n,2)}_{U^n,W^n}$, we have

$$\begin{aligned}
&\left\| P^{(n,1)}_{U^n,X^n} - P^{(n,2)}_{U^n,X^n} \right\|_{TV} \\
&\leq \left\| P^{(n,1)}_{U^n,W^n,X^n} - P^{(n,2)}_{U^n,W^n,X^n} \right\|_{TV} && (5.86)\\
&\leq \varepsilon + P^{(n,2)}_{U^n,W^n}\big((U^n,W^n) \notin \mathcal{T}^{(n)}_{\delta_A}(p_{U,W})\big) && (5.87)\\
&\leq 2\varepsilon, && (5.88)
\end{aligned}$$

where (5.86) follows from Lemma 1.8 (Appendix 1); (5.87) follows form (5.85) and the triangle inequality for the TVD; and (5.88) holds for large enough n, and follows from (5.77) and the fact that $\delta \leq \delta_A$. The same inequality holds for variables (V^n, Y^n), for large enough n.

Proof conclusion

Hence, since $P^{(n,1)}$ and $P^{(n,2)}$ satisfy Markov chain $X^n - U^n - V^n - Y^n$, then we have

$$\begin{aligned}
&\left\| P^{(n,1)}_{U^n,V^n,X^n,Y^n} - P^{(n,2)}_{U^n,V^n,X^n,Y^n} \right\|_{TV} \\
&= \left\| P^{(n,1)}_{U^n,X^n} \cdot P^{(n,1)}_{V^n,Y^n|U^n} - P^{(n,2)}_{V^n,Y^n} \cdot P^{(n,2)}_{U^n,X^n|V^n} \right\|_{TV} \\
&= \Big\| P^{(n,1)}_{U^n,X^n} \cdot P^{(n,1)}_{V^n,Y^n|U^n} - P^{(n,2)}_{U^n,X^n} \cdot P^{(n,1)}_{V^n,Y^n|U^n} \\
&+ P^{(n,1)}_{V^n,Y^n} \cdot P^{(n,2)}_{U^n,X^n|V^n} - P^{(n,2)}_{V^n,Y^n} \cdot P^{(n,2)}_{U^n,X^n|V^n} \Big\|_{TV} \\
&\leq 2\varepsilon + 2\varepsilon, && (5.89)
\end{aligned}$$

where (5.89) follows from the triangle inequality for the TVD, Lemma 1.9 (Appendix 1), and (5.88). This yields, for large enough n,

$$\begin{aligned}
&\| P^{(n,1)}_{X^n,Y^n} - q^{\otimes n}_{X,Y} \|_{TV} \\
&\leq \left\| P^{(n,1)}_{X^n,Y^n} - P^{(n,2)}_{X^n,Y^n} \right\|_{TV} + \| P^{(n,2)}_{X^n,Y^n} - q^{\otimes n}_{X,Y} \|_{TV} && (5.90)\\
&\leq \left\| P^{(n,1)}_{U^n,V^n,X^n,Y^n} - P^{(n,2)}_{U^n,V^n,X^n,Y^n} \right\|_{TV} + \varepsilon \leq 5\varepsilon, && (5.91)
\end{aligned}$$

where (5.90) follows from the triangle inequality for the TVD; and (5.91) follows from Lemma 1.8 (Appendix 1), (5.84), and (5.89). Since this construction and analysis are valid for any $\varepsilon \in (0, \varepsilon^*)$, then $(R_A, R_{c,A}, R_B, R_{c,B})$ is achievable, which concludes the proof of Theorem 5.9.

Chapter 6

Conclusion

Due to the ever-growing volumes of data and communication in modern digital systems, it is crucial to adopt a lossy compression paradigm: data is a reservoir of information, and one can ingeniously discard parts of the latter that are not relevant for the target application. How do engineers develop masterful techniques to achieve that? Often, innovation arises after formally stating certain goals. This is not only useful for an individual engineer to structure an iterative process of investigation, but also allows for concerted efforts when multiple individuals tackle the same goals, and great progress can be achieved from incremental discoveries. Hence, formalization can be considered as a key step which unravels innovation. Specifically, the development and study of the formalisms for lossy compression with distortion and realism constraints or statistical coordination constraints constitute a key step towards the design of lossy compression schemes tailored to modern applications.

Through a systematic development, we have aimed to provide a comprehensive understanding of how distribution matching constraints — which can be used to formulate both realism and coordination requirements — fundamentally alter the nature of lossy compression, revealing new trade-offs, requiring new coding techniques, and ultimately suggesting a new framework for capturing the elusive notion of realism. The tractability of the mathematical models allowed an in-depth analysis that brought to light fundamental properties with significant practical implications.

In this chapter, we highlight the key components of the approaches adopted in this thesis, and our main results. Then, we mention directions for future work and give concluding remarks.

6.1 Lossy compression with distribution matching constraints

6.1.1 The role of randomization

As discussed in Chapter 1, the RDP trade-off with perfect realism constraints fundamentally differs from classical RD theory in that randomization becomes necessary to achieve optimal performance. This observation raises important questions about how randomization should be implemented in practical compression systems.

In Chapter 2, we formally defined the RDP trade-off with perfect realism constraints in the two most commonly studied forms: strong perfect realism $P_{Y^n} \equiv P_{X^n}$ and per-symbol perfect realism $\forall t \in [n], P_{Y_t} \equiv P_{X_t}$. We also defined the corresponding relaxations called near-perfect strong realism and near-perfect per-symbol realism, respectively:

$$\|P_{Y^n} - p_X^{\otimes n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0 \tag{6.1}$$

$$\max_{1\le t\le n} \|P_{Y_t} - p_X\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{6.2}$$

We characterized the five-way trade-off between compression rate, common randomness rate, encoder private randomness rate, decoder private randomness rate, and distortion under near-perfect strong realism and near-perfect per-symbol realism constraints — and under the corresponding perfect realism constraints when the source has infinite entropy. Our results reveal that encoder private randomness is not useful. This finding has significant practical implications. Indeed, motivated by studies on the benefits of common randomness, a common implementation strategy in the absence of common randomness is to sample a random number (e.g., that can serve as initialization for a pseudo-random number generator) at the encoder

and transmit it to the decoder. Our results imply that, in the limit of large blocklength and under a near-perfect (strong or per-symbol) realism constraint, such a scheme does not yield superior performance compared to schemes with deterministic encoders. Intuitively, either the seed has insufficient entropy — which is very common in standard implementations involving high entropy sources such as high-resolution images — and little impact, or it has relatively high entropy and its benefits do not outweigh the cost of sending it. Albeit pertaining to an idealized theoretical setting, our results raise important questions for codec design.

6.1.2 Side information and common randomness

Chapter 3 extended our investigation to compression scenarios aided by side information (denoted Z^n), which bears similarities to video compression where previous frames serve as references. We studied the Wyner-Ziv problem with the additional requirement of strong perfect or near-perfect realism on the reconstruction, and introduced a novel constraint of *joint realism* of the reconstruction and the side information:

$$\|P_{Y^n,Z^n} - p_{X,Z}^{\otimes n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0. \tag{6.3}$$

In contrast, we use the term *marginal realism* to refer to the standard near-perfect strong realism constraint.

Our characterization of the RD trade-off revealed the dual role of side information when available at both terminals: it serves simultaneously as a second look on the source and as a source of common randomness between encoder and decoder. Critically, this duality depends on whether the realism constraint involves the side information: under a joint realism constraint, common randomness cannot be extracted from the side information. This raises the question of the choice of realism requirement for practical applications. Under the marginal realism constraint, the (conditional) entropy term in our single-letter characterization quantifies the role of side information as a source of common randomness, providing valuable insight into video compression in the absence of other common randomness sources. For jointly Gaus-

sian sources with sufficient common randomness, we showed that the RD trade-off is identical whether side information is available only at the encoder or at both terminals—but this equivalence breaks down without common randomness, further highlighting the fundamental role of common randomness in compression with realism constraints.

6.2 Rethinking the realism paradigm

As outlined in Section 1.2, the distribution matching formalism for realism, while theoretically elegant, suffers from several fundamental limitations: it is mismatched to human perception (which operates on individual samples rather than ensembles), it requires common randomness in theory but not in practice (the "common randomness paradox"), and more generally imposes unnecessarily stringent requirements, which may consume extra compression rate.

Chapter 4 addressed these limitations by proposing a novel information-theoretic formalism based on distinguishing between computable tests (critics) of realism and incomputable ones — which is a standard paradigm of algorithmic information theory. Our specific formulation is inspired by a position paper by L. Theis [The24], where various problems involving realism are mentioned. By requiring that reconstructions satisfy all computable critics — a constraint we term algorithmic realism — we obtain a framework that aligns better with human perception and eliminates the artificial need for common randomness in practical scenarios. Under this framework, reconstructions must exhibit all the complex patterns and regularities present in natural source realizations, satisfying not only simple statistical tests but also subtle limit theorems like the Erdös-Rényi law of runs.

Our results reveal a fundamental insight about the role of randomization in lossy compression with realism constraints: the determining factor appears to be the number of samples that are jointly inspected by a critic. Theorem 4.4 establishes that common randomness does not improve the trade-off in all regimes where this number is not impractically large relative to the source dimension. Conversely, Theorem 4.6 shows that when it is extremely large, our formalism becomes equivalent to the distribution matching formalism, and common randomness becomes

useful. This finding suggests that the apparent paradox regarding common randomness in practical systems can be explained by considering the implicit numbers of samples jointly inspected in real-world scenarios.

6.3 Coordination problems in distributed systems

As introduced in Section 1.3 through our soccer team example, coordination problems naturally arise in numerous engineering applications, from multi-agent reinforcement learning to autonomous systems. We provided an example of how advanced soccer teams show an ability to achieve statistical coordination without direct communication — termed non-interactive coordination.

Chapter 5 studied a non-interactive remote coordination problem where two non-interacting controllers observe correlated sources and communicate with decoders to help them generate coordinated actions. A distribution matching formalism is used for coordination requirements, specifically a constraint similar to near-perfect strong realism:

$$\|P_{X^n,Y^n} - p_{X,Y}^{\otimes n}\|_{TV} \underset{n\to\infty}{\longrightarrow} 0, \tag{6.4}$$

where X^n and Y^n model actions to be coordinated according to a target joint distribution $q_{X,Y}$. We proposed schemes based on channel synthesis — which plays a central part in studies of similar problems — and highlighted the benefits of this technique and common randomness for the problem. Beyond the direct application of channel synthesis, we developed an extension which we termed remote channel synthesis, and proposed a scheme based on the latter. Our characterization of the optimal compression and common randomness rate pairs for remote channel synthesis revealed the following: when common randomness is sufficiently limited, bespoke remote channel synthesis schemes can yield superior performance for non-interactive remote coordination — and for the problem of remote channel synthesis itself — compared to direct applications of channel synthesis. This is of particular interest since in many practical implementations, the amount of available common randomness is quite limited — this applies

generally to compression and communication problems, in particular to lossy compression with realism constraints, as noted in Chapter 4.

6.4 Unifying themes and future directions

Several unifying themes emerge across our contributions:

The nuanced role of randomization: Throughout this thesis, we have seen that randomization plays a subtle and context-dependent role in information extraction and coordination. While common randomness can be crucial, encoder private randomness is not useful for lossy compression with near-perfect realism, even when used to share randomness with the decoder. What's more, the importance of common randomness depends critically on a number of factors: the presence of side information and its role in realism requirements, the formalism used for realism constraints (strong, per-symbol, algorithmic), the number of samples jointly inspected to assess realism, and whether we require that incomputable critics be satisfied, or simply computable ones.

The value of alternative formalizations: Our work demonstrates the importance of questioning established formalisms. The algorithmic realism framework overcomes fundamental limitations of the distribution matching formulation of realism, by shifting focus from ensemble properties to individual sample properties and raising questions of computability, while remote channel synthesis extends traditional channel synthesis to capture scenarios that are more distributed.

Raising questions pertaining to practical implementations: Several of our results address differences between theoretical predictions and approaches used in practice, such as the common randomness paradox and the use of encoder randomness in practical image compression algorithms, and the choice of realism metrics to assess the latter.

In light of the findings presented in this thesis, new questions, open problems, and unexplored ideas emerge. The main promising directions for future research are the following:

- Finite-blocklength and one-shot analysis: While the study of the role of randomness for long ergodic sequences allows for tractability, yielding valuable insight, practical compression systems are required to operate with much more general source distributions.

- Realism metrics: A systematic investigation of realism metrics with particular attention to the number of samples jointly inspected (batch size) could lead to a critical assessment of existing compression schemes and highlight specific strengths and weaknesses of current approaches.

- Randomness characterization: More precisely characterizing the amount of randomness needed as a function of batch size, beyond the two regimes considered in Chapter 4, is of theoretical interest.

- Extensions to distributed settings: Our algorithmic realism framework could be extended to compression with side information and other distributed settings, while the remote channel synthesis problem could serve as a building block for more complex distributed coordination schemes.

- Hybrid coordination schemes: In Section 5.6, we showed that, in the presence of non-trivial GK common information between the correlated sources, non-interactive remote coordination can be enhanced via a hybrid scheme that involves channel synthesis. We expect that similar hybrid schemes can be developed with remote channel synthesis.

6.5 Closing remarks

This thesis has contributed to our understanding of how to extract and communicate only the relevant information from signals — an exciting problem that still holds many mysteries. By characterizing fundamental trade-offs involving compression rate, distortion, realism, and various forms of randomness, we have provided both theoretical insights and raised questions that are highly relevant for the design of practical modern compression and coordination systems.

The digital age continues to generate ever-growing volumes of data, making efficient information extraction increasingly critical. As semantic communication and task-oriented communication paradigms gain popularity, the questions addressed in this thesis — how to compress while maintaining perceptual quality, how to coordinate without direct communication, and how to optimally use randomness —will only become more relevant. We hope that our contributions provide a foundation for continued progress in these vital areas of information theory and engineering.

Appendix 1

Mathematical background

1.1 Probability theory background

1.1.1 Conditional probabilities on general alphabets and sequential definitions

The family of Polish spaces includes all discrete spaces and all finite-dimensional real vector spaces, and any finite product of Polish spaces is a Polish space. Every measurable space we consider that is of the form $(\mathcal{W}, \mathcal{B})$ with $\mathcal{W}$ a Polish space satisfies that $\mathcal{B}$ is its Borel σ-algebra. A discrete alphabet is always endowed with its power set as topology, and hence its Borel σ-algebra is its power set. Finite-dimensional real vector spaces are endowed with the standard topology (induced by any norm). A product of the two measurable spaces will be implicitly endowed with the product σ-algebra. If the two spaces are Polish, then the product of (Borel) σ-algebras is equal to the Borel σ-algebra of the product Polish space. As is standard practice (see, e.g., [Bil12]), all conditional probability kernels we consider are defined for all elements of the input alphabet (not just almost everywhere):

Definition 1.1 *Given any two measurable spaces $(\mathcal{W}, \mathcal{B})$ and $(\mathcal{W}', \mathcal{B}')$, a regular channel ρ, also called a regular conditional probability kernel or regular transition kernel from the former*

to the latter is a function from $\mathcal{W} \times \mathcal{B}'$ *to* $[0,1]$ *such that for every* $w \in \mathcal{W}$*, function* $\rho(w,\cdot)$ *is a probability measure on* $(\mathcal{W}', \mathcal{B}')$ *and for every* $B' \in \mathcal{B}'$*, the function* $\rho(\cdot, B')$ *is* $\mathcal{B}$*-measurable.*

The well-known Fubini theorem generalizes beyond product measures as follows.

Theorem 1.1 *(Generalized Fubini theorem) [Nev70, Proposition III-2-1]*
Consider any two measurable spaces $(\mathcal{W}, \mathcal{B})$ *and* $(\mathcal{W}', \mathcal{B}')$*, a probability measure* μ *on the former and a regular transition kernel* $\rho(\cdot,\cdot)$ *from* $(\mathcal{W}, \mathcal{B})$ *to* $(\mathcal{W}', \mathcal{B}')$*. Then, there is a unique measure* Q *on the product space* $(\mathcal{W} \times \mathcal{W}', \mathcal{B} \otimes \mathcal{B}')$ *with*

$$\forall B \in \mathcal{B}, \forall B' \in \mathcal{B}',\ Q(B \times B') = \int_B \rho(w, B') d\mu(w). \tag{1.1}$$

We denote this distribution by $\mu \cdot \rho$*. Its marginal on* $\mathcal{W}$ *is* μ*. Moreover, for any non-negative measurable function* φ*, the map* $w \mapsto \int \varphi\, d\rho(w,\cdot)$ *is well-defined on all* $\mathcal{W}$ *and is* $\mathcal{B}$*-measurable, and we have*

$$\int \varphi\, dQ = \int \left(\int \varphi\, d\rho(w,\cdot) \right) d\mu(w).$$

When ρ *is constant in its first argument and equal to a distribution* ν *on* $\mathcal{W}'$*, the distribution* $\mu \cdot \rho$ *— or* $\mu \cdot \nu$ *— is the product distribution. Given a distribution* Q *on* $(\mathcal{W} \times \mathcal{W}', \mathcal{B} \otimes \mathcal{B}')$ *with marginal* μ *on* $\mathcal{W}$*, then any regular transition kernel* $\rho(\cdot,\cdot)$ *satisfying* (1.1) *is called a regular transition kernel of* Q *from* $\mathcal{W}$ *to* $\mathcal{W}'$*. If* (W, W') *is a random tuple with distribution* Q*, then we denote* μ *by* Q_W *and we denote* ρ *by* $Q_{W'|W}$ *— although such a transition kernel might not be unique. For any* $w \in \mathcal{W}$*, we denote the distribution* $\rho(w,\cdot)$ *by* $Q_{W'|W=w}$*. We also use, with abuse of notation:* $Q \equiv Q_W \cdot Q_{W'|W=w}$*, where* w *is a dummy variable.*

Remark 1.2 *Moreover, in the setting of Theorem 1.1, consider a distribution* Q *on* $(\mathcal{W} \times \mathcal{W}', \mathcal{B} \otimes \mathcal{B}')$ *and two regular transition kernels* $\rho(\cdot,\cdot)$ *and* $\tilde{\rho}(\cdot,\cdot)$ *of* Q *from* $\mathcal{W}$ *to* $\mathcal{W}'$*. Then, for any measurable set* B'*, the maps* $w \mapsto \rho(w, B')$ *and* $w \mapsto \tilde{\rho}(w, B')$ *are equal outside of a* μ*-null subset of* $\mathcal{W}$*, where* μ *is the marginal of* Q *on* $\mathcal{W}$*. Although there is no functional uniqueness of a regular transition kernel of* Q *from* $\mathcal{W}$ *to* $\mathcal{W}'$*, in many cases an integral involving such a kernel has a unique value.*

The claim in the above remark is proved in Section 1.3. Moreover, Theorem 1.1 implies the following corollary. We do not provide the proof, which is rather straightforward, for the sake of conciseness.

Corollary 1.3 *Consider four Polish measurable spaces* $(\mathcal{W}_1, \mathcal{B}_1), (\mathcal{W}_2, \mathcal{B}_2), (\mathcal{W}_3, \mathcal{B}_3), (\mathcal{W}_4, \mathcal{B}_4)$, *a random triplet* (W_1, W_2, W_3) *taking values in* $\mathcal{W}_1 \times \mathcal{W}_3 \times \mathcal{W}_3$ *and a transition kernel* $\rho(\cdot, \cdot)$ *from* $(\mathcal{W}_3, \mathcal{B}_3)$ *to* $(\mathcal{W}_4, \mathcal{B}_4)$. *Denote the distribution of* (W_1, W_2, W_3) *by* μ_{W_1,W_2,W_3}. *The following map*

$$\tilde{\rho} : \mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{W}_3 \times \mathcal{B}_4 \to [0, 1], \ (w_1, w_2, w_3, B_4) \mapsto \rho(w_3, B_4) \tag{1.2}$$

is a regular transition kernel from $(\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{W}_3, \mathcal{B}_1 \otimes \mathcal{B}_2 \otimes \mathcal{B}_3)$ *to* $(\mathcal{W}_4, \mathcal{B}_4)$. *Define*

$$Q_{W_1,W_2,W_3,W_4} := \mu_{W_1,W_2,W_3} \cdot \tilde{\rho}.$$

Then, ρ *is a regular transition kernel of* Q_{W_3,W_4} *from* $(\mathcal{W}_3, \mathcal{B}_3)$ *to* $(\mathcal{W}_4, \mathcal{B}_4)$, *and* $(w_2, w_3, B_4) \mapsto \rho(w_3, B_4)$ *is a regular transition kernel of of* Q_{W_2,W_3,W_4} *from* $(\mathcal{W}_2 \times \mathcal{W}_3, \mathcal{B}_2 \otimes \mathcal{B}_3)$ *to* $(\mathcal{W}_4, \mathcal{B}_4)$. *With some abuse of notation, we shall denote all the aforementioned maps by* ρ, *or* $Q_{W_4|W_3}$.

When given a joint distribution, such as from (3.13), we extensively use corresponding conditional distributions, which is justified if the alphabets are Polish spaces according to the following theorem, sometimes referred to as Jirina's theorem.

Theorem 1.4 *[Str10, Theorem 9.2.2] Let* $\mathcal{W}$ *and* $\mathcal{W}'$ *be Polish spaces and let* $\Omega = \mathcal{W} \times \mathcal{W}'$. *Let* Q *be a probability measure on* Ω. *Then, there exists a regular probability transition kernel* $\rho(\cdot, \cdot)$ *of* Q *from* $\mathcal{W}$ *to* $\mathcal{W}'$.

We will also use the following conditional version of Theorem 1.1, valid for Polish spaces, which allows us to build transition kernels sequentially. The proof is deferred to Section 1.3.

Proposition 1.5 *Consider three Polish measurable spaces* $(\mathcal{W}_1, \mathcal{B}_1), (\mathcal{W}_2, \mathcal{B}_2), (\mathcal{W}_3, \mathcal{B}_3)$, *a transition kernel* $\rho(\cdot, \cdot)$ *from* $(\mathcal{W}_1, \mathcal{B}_1)$ *to* $(\mathcal{W}_2, \mathcal{B}_2)$ *and a transition kernel* $\gamma(\cdot, \cdot, \cdot)$ *from* $(\mathcal{W}_1 \times$

$\mathcal{W}_2, \mathcal{B}_1 \otimes \mathcal{B}_2)$ *to* $(\mathcal{W}_3, \mathcal{B}_3)$. *Define* $\lambda : \mathcal{W}_1 \times (\mathcal{B}_2 \otimes \mathcal{B}_3) \to [0,1]$,

$$(w_1, B_{2,3}) \mapsto \Big(\rho(w_1, \cdot) \cdot \gamma(w_1, \cdot, \cdot)\Big)(B_{2,3}), \tag{1.3}$$

with the notation of Theorem 1.1. Then, λ *is a regular transition kernel from* $(\mathcal{W}_1, \mathcal{B}_1)$ *to* $(\mathcal{W}_2 \times \mathcal{W}_3, \mathcal{B}_2 \otimes \mathcal{B}_3)$. *We denote it* $\rho \cdot \gamma$. *Moreover, for any probability measure* μ *on* $(\mathcal{W}_1, \mathcal{B}_1)$ *we have associativity:*

$$(\mu \cdot \rho) \cdot \gamma \equiv \mu \cdot (\rho \cdot \gamma).$$

Consequently, for the resulting distribution, say Q, *kernel* ρ *is a transition kernel of* Q *from* $\mathcal{W}_1$ *to* $\mathcal{W}_2$, *kernel* γ *is a transition kernel of* Q *from* $\mathcal{W}_1 \times \mathcal{W}_2$ *to* $\mathcal{W}_3$ *and* λ *is a transition kernel of* Q *from* $\mathcal{W}_1$ *to* $\mathcal{W}_2 \times \mathcal{W}_3$. *We use the product notation* $\prod$ *to denote sequential composition.*

We provide a proof in Section 1.3.

1.1.2 The notion of Markov chain for general alphabets

We follow the definition of the notion of Markov chain for general alphabets given in [Gra11, page 196], which we state in a slightly different but equivalent form:

Definition 1.2 *Consider Polish spaces* $\mathcal{W}_1$, $\mathcal{W}_2$ *and* $\mathcal{W}_3$. *Let* (W_1, W_2, W_3) *be a random tuple taking values in* $\mathcal{W}_1 \times \mathcal{W}_2 \times \mathcal{W}_3$, *with distribution denoted* Q. *Then, we say that* Q *satisfies the Markov chain* $W_1 - W_2 - W_3$ *iff for some choice of transition kernel of* Q, *we have*

$$Q_{W_1,W_2,W_3} \equiv Q_{W_1,W_2} \cdot Q_{W_3|W_2=w_2}. \tag{1.4}$$

Moreover, this property does not depend on the choice of transition kernel. The notion extends to tuples of more than three variables. For any $n \geq 4$, *we say that a family* $(W_i)_{1 \leq i \leq n}$ *satisfies the Markov chain* $W_1 - W_2 - ... - W_n$ *iff for any* $k \in [n] \backslash \{1, n\}$, *we have the Markov chain* $W_{1:k-1} - W_k - W_{k+1:n}$.

The independence on the choice of kernels is a direct application of Remark 1.2. The notion of Markov chain is related to conditional mutual information, as stated in Lemma 1.15.

1.1.3 Density functions

When a joint distribution over Polish spaces is dominated by another, conditional densities exist and form a regular transition kernel, as stated in the following corollary of Theorems 1.1 and 1.4.

Corollary 1.6 *Consider Polish spaces $\mathcal{W}$ and $\mathcal{W}'$. Let P and Q be probability measures on $\mathcal{W} \times \mathcal{W}'$, with marginals on $\mathcal{W}$ denoted by μ and ν respectively. Assume that Q dominates P (then ν dominates μ). Fix a density function $\frac{dP}{dQ}(\cdot,\cdot)$ and a regular transition kernel $Q_{W'|W}$ of Q from $\mathcal{W}$ to $\mathcal{W}'$. Then,*

$$f : \mathcal{W} \to [0,\infty],\ w \mapsto \int \frac{dP}{dQ}(w,w')\ dQ_{W'|W=w}(w')$$

defines a density of μ with respect to ν. Define

$$g : \mathcal{W} \times \mathcal{W}' \to [0,\infty],\ (w,w') \mapsto \begin{cases} \frac{dP}{dQ}(w,w') \Big/ f(w) & \text{if } f(w) > 0, \\ 1 & \text{if } f(w) = 0. \end{cases}$$

Then, g is measurable and Q-integrable and the following map defines a regular probability transition kernel of P from $\mathcal{W}$ to $\mathcal{W}'$:

$$(w,B) \mapsto \int g(w,w') \cdot \mathbf{1}_B(w')\ dQ_{W'|W=w}(w').$$

The corresponding distribution for a given $w \in \mathcal{W}$ admits $g(w,\cdot)$ as a density function with respect to $Q_{W'|W=w}$.

The proof is provided in Section 1.3.

Definition 1.3 *The TVD between two distributions P, Q on general alphabets is defined by*

$$\sup_A \{P(A) - Q(A)\} = \frac{1}{2}\int |p - q| d\mu = \int (p - q)^+ d\mu, \tag{1.5}$$

for any common dominating μ and corresponding densities p, q, with x^+ being a notation for $\max(x, 0)$.

It satisfies the following properties, proved in Section 1.3.

Lemma 1.7 *Let Π and Γ be two distributions on the product of two Polish spaces $\mathcal{W}$ and $\mathcal{L}$, and let $\Pi_{L|W}, \Gamma_{L|W}$ be two channels. Then, we have*

$$\|\Pi_W \Pi_{L|W} - \Pi_W \Gamma_{L|W}\|_{TV} = \mathbb{E}_{\Pi_W}\big[\|\Pi_{L|W} - \Gamma_{L|W}\|_{TV}\big].$$

Lemma 1.8 *Let Π and Γ be two distributions on an alphabet $\mathcal{W} \times \mathcal{L}$. Then*

$$\|\Pi_W - \Gamma_W\|_{TV} \leq \|\Pi_{W,L} - \Gamma_{W,L}\|_{TV}.$$

Lemma 1.9 *Let Π and Γ be two distributions on an alphabet $\mathcal{W} \times \mathcal{L}$. Then when using the same channel $\Pi_{L|W}$ we have*

$$\|\Pi_W \Pi_{L|W} - \Gamma_W \Pi_{L|W}\|_{TV} = \|\Pi_W - \Gamma_W\|_{TV}.$$

Lemma 1.10 *[Cuf13, Eq. (29)] Let Π and Γ be two distributions on a finite set $\mathcal{W}$, and $f : \mathcal{W} \to \mathbb{R}$ be a bounded function. Then,*

$$|\ \mathbb{E}_\Pi[f] - \mathbb{E}_\Gamma[f]\ | \leq 2 \max|f| \cdot \|\Pi - \Gamma\|_{TV}.$$

1.2 Information quantities on general alphabets

Definition 1.4 *[Gra11, Section 7.1] Given a probability space $(\Omega, \mathcal{B}, P)$ and another probability measure Q on the same space, define the relative entropy or divergence of P with respect to Q*

as

$$D(P\|Q) = \sup_X \sum_{x\in\mathcal{X}} P_X(x) \log\Big(\frac{P_X(x)}{Q_X(x)}\Big), \tag{1.6}$$

where the supremum is over all random variables X defined from $(\Omega, \mathcal{B})$ into any finite-valued alphabet $\mathcal{X}$, and where we assume the convention $a/0 = +\infty$, for $a > 0$.

Lemma 1.11 *[Gra11, Lemma 7.1] For any distributions P and Q we have $D(P\|Q) \geq 0$, with equality iff $P \equiv Q$.*

Definition 1.5 *[Gra11, Equations 7.28, 7.29, 7.30] Let X, Y and Z denote random variables defined on the same probability space and taking values in some Polish spaces. Define the mutual information, entropy and conditional mutual information by*

$$\begin{aligned} I(X;Y) &= D(P_{X,Y}\|P_X \cdot P_Y) \\ H(X) &= I(X;X) \\ I(X;Y|Z) &= D(P_{X,Y,Z}\|P_Z \cdot P_{X|Z} \cdot P_{Y|Z}), \end{aligned}$$

where $P_{X,Y,Z}$ denotes the distribution of (X,Y,Z). The distribution $P_Z \cdot P_{X|Z} \cdot P_{Y|Z}$ does not depend on the choice of transition kernels $P_{X|Z}$ and $P_{Y|Z}$. Since these definitions depend only on distributions, we also denote $H(X)$ by $H(P_X)$.

These quantities satisfy the following properties.

Lemma 1.12 *[Gra11, Corollary 7.11, Lemma 7.20] Let X and Y denote random variables defined on the same probability space and taking values in some Polish spaces. Then,*

$$H(X) \geq I(X;Y). \tag{1.7}$$

If $I(X;Y) < \infty$, define $H(X|Y) := H(X) - I(X;Y)$. Let $P_{X,Y}$ denote the distribution of (X,Y).

If X is finite-valued, then for any regular transition kernel $P_{X|Y}$ of P, we have

$$H(X|Y) = \int H(P_{X|Y=y}) dP_Y(y). \tag{1.8}$$

Lemma 1.13 *[Gra11, Corollary 7.16] Let X and Y denote random variables defined on the same probability space and taking values in some Polish spaces. For any measurable functions f and g we have*

$$I(f(X); g(Y)) \leq I(X; Y).$$

Proposition 1.14 *[Gra11, Corollaries 7.14 & 7.15] Let X, Y, Z and W be finite-valued random variables. Then, we have*

$$I(X; Y, Z|W) = I(X; Y|W) + I(X; Z|Y, W). \tag{1.9}$$

Moreover, for random variables X', Y' and Z' taking values in Polish spaces, we have

$$I(X'; Y', Z') = I(X'; Y') + I(X'; Z'|Y'),$$

with the conventions $\infty + u = \infty$ for any (finite) real u and $\infty + \infty = \infty$.

Lemma 1.15 *[Gra11, Lemma 7.1 & Corollaries 7.14 & 7.15] Let X, Y, Z be random variable taking values in some Polish spaces. Then, X and Y are independent iff $I(X; Y) = 0$. Moreover, the Markov chain $X - Z - Y$ holds iff $I(X; Y|Z) = 0$. In that case, we have the data processing inequality:*

$$I(X; Z) \geq I(X; Y).$$

The conditional chain rule (1.9) generalizes to infinite alphabets under certain conditions, but in this thesis we prefer to perform mutual information computations with quantized variables

and take the limit. *A quantizer* is any finite-valued measurable map. Under the following conditions, information quantities of variables on general alphabets are limits of the corresponding quantities for quantized variables, and the sequence of quantizers can be chosen independently of the original variables' distribution.

Proposition 1.16 *[Gra11, Corollaries 7.14 & 7.15 & Lemma 7.18] Let X, Y and Z denote random variables defined on the same probability space and taking values in some Polish spaces $(\mathcal{W}_1, \mathcal{B}_1)$, $(\mathcal{W}_2, \mathcal{B}_2)$, and $(\mathcal{W}_3, \mathcal{B}_3)$, respectively. For each $i \in \{1,2,3\}$, let $(\kappa_i^{(n)})_{n\geq 1}$ be a sequence of quantizers such that the corresponding partitions asymptotically generate $\mathcal{B}_i$. If for some $i \in \{1,2,3\}$ the set $\mathcal{W}_i$ is finite, then $(\kappa_i^{(n)})_{n\geq 1}$ can be a constant sequence with $\kappa_i^{(1)}$ being the identity map. Then, whatever the distribution of (X,Y,Z) we have*

$$\begin{aligned} I(X;Y,Z) &= \lim_{n\to\infty} I\big(\kappa_1^{(n)}(X); \kappa_2^{(n)}(Y), \kappa_3^{(n)}(Z)\big) \\ H(X) &= \lim_{n\to\infty} H\big(\kappa_1^{(n)}(X)\big). \end{aligned}$$

Moreover, if $I(X;Z) < \infty$ or $I(Y;Z) < \infty$, then

$$I(X;Y|Z) = \lim_{n\to\infty} I\big(\kappa_1^{(n)}(X); \kappa_2^{(n)}(Y) | \kappa_3^{(n)}(Z)\big) \tag{1.10}$$

Proof: The set of partitions of quantizer $(y,z) \mapsto (\kappa_2^{(n)}(Y), \kappa_3^{(n)}(Z))$ is the product of the respective sets of partitions of $\kappa_2^{(n)}$ and $\kappa_3^{(n)}$. Then, that set asymptotically generates the product σ-algebra $\mathcal{B}_2 \otimes \mathcal{B}_3$, which is the Borel σ-algebra of the Polish product space $\mathcal{Y} \times \mathcal{Z}$. Hence, the first two properties of Proposition 1.16 follow from [Gra11, Lemma 7.18]. Assume that $I(X;Z) < \infty$. Then, from Proposition 1.14,

$$\begin{aligned} \forall n \in \mathbb{N},\ I\big(\kappa_1^{(n)}(X); \kappa_2^{(n)}(Y) | \kappa_3^{(n)}(Z)\big) &= I\big(\kappa_1^{(n)}(X); \kappa_2^{(n)}(Y), \kappa_3^{(n)}(Z)\big) - I\big(\kappa_1^{(n)}(X); \kappa_3^{(n)}(Z)\big), \\ I(X;Y|Z) &= I(X;Y,Z) - I(X;Z). \end{aligned}$$

Then, (1.10) follows from the first property of Proposition 1.16. By symmetry, (1.10) also holds if $I(Y;Z) < \infty$. ∎

Moreover, we have the following fundamental link to the *information density*.

Proposition 1.17 *[Gra11, Equation 7.28, Lemma 7.4] Let (X, Y) be a couple of joint distribution $P_{X,Y}$. If the product distribution $P_X \cdot P_Y$ dominates $P_{X,Y}$, we define the information density*

$$i_{P_{X,Y}} = \log\left(\frac{dP_{X,Y}}{d(P_X \cdot P_Y)}\right),$$

with values in $[-\infty, +\infty]$. Then, $\min(0, i_{P_{X,Y}})$ is $P_{X,Y}$-integrable and we have

$$I(X;Y) = \int i_{P_{X,Y}}(x,y) dP_{X,Y}(x,y). \tag{1.11}$$

Moreover, $P_X \cdot P_Y$ dominates $P_{X,Y}$ whenever $I(X;Y) < \infty$, and in that case the information density is well-defined and $P_{X,Y}$-integrable.

1.3 Proofs of lemmas stated in Section 1.1

1.3.1 Proof of Remark 1.2

Fix a measurable set $B' \subset \mathcal{W}'$. For any measurable set $B \subset \mathcal{W}$, we have

$$Q(B \times B') = \int_B \rho(w, B') d\mu$$

by taking $\varphi = \mathbf{1}_{B \times B'}$. If $Q(\mathcal{W} \times B') = 0$, then $\rho(\cdot, B') = 0$ μ-almost everywhere. If $Q(\mathcal{W} \times B') > 0$, then $w \mapsto \rho(w, B')$ is a density of $Q(\cdot|B')$ w.r.t. μ up to a multiplicative factor. Hence, it is unique up to a μ-null set.

1.3.2 Proof of Proposition 1.5

First we justify that for any $w_1 \in \mathcal{W}_1$, the object $\rho(w_1, \cdot) \cdot \gamma(w_1, \cdot, \cdot)$ is well defined. We just need to prove that $\gamma(w_1, \cdot, \cdot)$ is a regular transition kernel. Only the measurability is (highly) non-

trivial. Let B_3 be in $\mathcal{B}_3$ and A be a real Borel set. Then, $\gamma(\cdot,\cdot,B_3)^{-1}(A) \in \mathcal{B}_1 \otimes \mathcal{B}_2$ by definition of γ. Since $\mathcal{W}_1$ is Polish, it is metrizable, therefore T4 (normal), hence T1: its singletons are closed. Since $\mathcal{B}_1$ is the Borel σ-algebra of $\mathcal{W}_1$, we have $\{w_1\} \in \mathcal{B}_1$, hence $\{w_1\} \times \mathcal{W}_2 \in \mathcal{B}_1 \otimes \mathcal{B}_2$. Therefore, $\gamma(\cdot,\cdot,B_3)^{-1}(A) \cap (\{w_1\} \times \mathcal{W}_2) \in \mathcal{B}_1 \otimes \mathcal{B}_2$. The pre-image of the latter by the map $\pi : w_2 \mapsto (w_1, w_2)$ gives the desired set $\gamma(w_1,\cdot,B_3)^{-1}(A)$. The canonical bijection π is continuous. Since $\mathcal{B}_1$ and $\mathcal{B}_2$ are the Borel σ-algebras of Polish spaces $\mathcal{W}_1$ and $\mathcal{W}_2$, then $\mathcal{B}_1 \otimes \mathcal{B}_2$ is the Borel σ-algebra of $\mathcal{W}_1 \times \mathcal{W}_2$, that is, the σ-algebra generated by its open sets. From the continuity of π, the pre-image of an open set by π is an open set of $\mathcal{W}_2$. Therefore, the pre-image of $\mathcal{B}_1 \otimes \mathcal{B}_2$ by π is the σ-algebra generated by the open sets of $\mathcal{W}_2$ (e.g. [Nev70, Section II.1]), i.e. $\mathcal{B}_2$. This proves that $\gamma(w_1,\cdot,B_3)^{-1}(A) \in \mathcal{B}_2$. This being true for any real Borel set A, then $w_2 \mapsto \gamma(w_1, w_2, B_3)$ is $\mathcal{B}_2$-measurable. This being true for any $B_3 \in \mathcal{B}_3$, then $\gamma(w_1,\cdot,\cdot)$ is a regular transition kernel from $\mathcal{W}_2$ to $\mathcal{W}_3$. Since $\rho(w_1,\cdot)$ is a distribution on $\mathcal{W}_2$, then the object $\rho(w_1,\cdot) \cdot \gamma(w_1,\cdot,\cdot)$ is well defined, as per Theorem 1.1. This is true for any $w_1 \in \mathcal{W}_1$.

Second, consider a probability measure μ on $(\mathcal{W}_1, \mathcal{B}_1)$ and define the distribution $Q = (\mu \cdot \rho) \cdot \gamma$. Then, by Theorem 1.1, for any non-negative and $\mathcal{B}_1 \otimes \mathcal{B}_2 \otimes \mathcal{B}_2$-measurable φ, the map $(w_1, w_2) \mapsto \int \varphi d\gamma(w_1, w_2, \cdot)$ is $\mathcal{B}_1 \otimes \mathcal{B}_2$-measurable and

$$\int \varphi dQ = \int \left(\int \varphi(w_1, w_2, \cdot) d\gamma(w_1, w_2, \cdot) \right) d(\mu \cdot \rho).$$

Hence, we can apply Theorem 1.1 on $Q' = \mu \cdot \rho$ and obtain that the map

$$w_1 \mapsto \int \left(\int \varphi(w_1, w_2, \cdot) d\gamma(w_1, w_2, \cdot) \right) d\rho(w_1, \cdot) \tag{1.12}$$

is $\mathcal{B}_1$-measurable and that $\int \varphi dQ$ rewrites as

$$\int \left[\int \left(\int \varphi(w_1, w_2, \cdot) d\gamma(w_1, w_2, \cdot) \right) d\rho(w_1, \cdot) \right] d\mu. \tag{1.13}$$

Fix some w_1 in $\mathcal{W}_1$. By Theorem 1.1 on the double integral in (1.12) with joint distribution

$\lambda(w_1, \cdot) = \rho(w_1, \cdot) \cdot \gamma(w_1, \cdot, \cdot)$, the measurable map in (1.12) rewrites as

$$w_1 \mapsto \int \varphi(w_1, \cdot, \cdot) d\lambda(w_1, \cdot).$$

This shows that λ is a regular transition kernel. Furthermore, equation (1.13) becomes

$$\int \varphi dQ = \int \left(\int \varphi(w_1, \cdot, \cdot) d\lambda(w_1, \cdot) \right) d\mu.$$

Since λ is a regular transition kernel, then by Theorem 1.1 this becomes:

$$\int \varphi dQ = \int \varphi d(\mu \cdot \lambda).$$

This being true for any non-negative $\mathcal{B}_1 \otimes \mathcal{B}_2 \otimes \mathcal{B}_2$-measurable φ, we obtain

$$(\mu \cdot \rho) \cdot \gamma \equiv Q \equiv \mu \cdot \lambda.$$

1.3.3 Proof of Corollary 1.6

The fact that f is a density of μ with respect to ν follows from a simple manipulations of integrals using Theorem 1.1. For any $a \in \mathbb{R}$, we have

$$\begin{aligned} &\left\{ (w, w') \in \mathcal{W} \times \mathcal{W}' \mid f(w) > 0 \text{ and } a \geq \frac{dP}{dQ}(w, w') \Big/ f(w) \right\} \\ = &\left\{ (w, w') \in \mathcal{W} \times \mathcal{W}' \mid f(w) > 0 \text{ and } 0 \geq \frac{dP}{dQ}(w, w') - af(w) \right\} \end{aligned}$$

Therefore, $\{g \in (-\infty, a]\}$ is measurable because dP/dQ and f are. Hence, g is measurable. Fix $w \in \mathcal{W}$. If $f(w) = 0$, then $\int g(w, \cdot) dQ_{W'|W=w} = 1$. This is also the case if $f(w) > 0$ by definition of f and g. Hence, g is Q-integrable, and for any $w \in \mathcal{W}$, the measure with density $g(w, \cdot)$ with respect to $Q_{W'|W=w}$ is a probability measure. Fix $B \in \mathcal{B}(\mathcal{W}')$. Since $(w, w') \mapsto g(w, w') \cdot \mathbf{1}_B(w')$

is measurable, then from Theorem 1.1, the map

$$w \mapsto \int g(w, w') \cdot \mathbf{1}_B(w') \; dQ_{W'|W=w}(w')$$

is measurable. Hence, the map

$$(w, B) \mapsto \int g(w, w') \cdot \mathbf{1}_B(w') \; dQ_{W'|W=w}(w')$$

defines a regular transition kernel.

1.3.4 Proof of Lemma 1.7

Define $\mu := \Pi_W(\Pi_{L|W} + \Gamma_{L|W})/2$, which is a probability measure dominating both $\Pi_W\Pi_{L|W}$ and $\Pi_W\Gamma_{L|W}$. Then, $(\Pi_{L|W} + \Gamma_{L|W})/2$ is a regular transition kernel of μ from $\mathcal{W}$ to $\mathcal{L}$, which we denote by $\mu_{L|W}$. Note that for any $w \in \mathcal{W}$, the distribution $\mu_{L|W=w}$ dominates $\Pi_{L|W=w}$ (resp. $\Gamma_{L|W=w}$). Moreover, we have $\mu_W \equiv \Pi_W$. Then, it can be checked through Theorem 1.1 that

$$(w, l) \mapsto \frac{d\Pi_{L|W=w}}{d\mu_{L|W=w}}(l) \quad \left(\text{resp. } \frac{d\Gamma_{L|W=w}}{d\mu_{L|W=w}}(l)\right) \tag{1.14}$$

is a density of $\Pi_W\Pi_{L|W}$ (resp. $\Pi_W\Gamma_{L|W}$) with respect to μ. Moreover, from Theorem 1.1 and the definition of the TVD (Eq. 1.5) we obtain

$$\begin{aligned}
\|\Pi_W\Pi_{L|W} - \Pi_W\Gamma_{L|W}\|_{TV} &= \int \left| \frac{d\Pi_{L|W=w}}{d\mu_{L|W=w}}(l) - \frac{d\Gamma_{L|W=w}}{d\mu_{L|W=w}}(l) \right| d\mu_{W,L}(w, l) \\
&= \int \left(\int \left| \frac{d\Pi_{L|W=w}}{d\mu_{L|W=w}}(l) - \frac{d\Gamma_{L|W=w}}{d\mu_{L|W=w}}(l) \right| d\mu_{L|W=w}(l) \right) d\Pi_W(w) \\
&= \mathbb{E}_{\Pi_W}[\|\Pi_{L|W} - \Gamma_{L|W}\|_{TV}], \qquad (1.15)
\end{aligned}$$

where (1.15) follows from the definition of the TVD (Eq. 1.5).

1.3.5 Proof of Lemma 1.8

Taking the supremum of the same quantity over a smaller set reduces the supremum. Therefore,

$$\begin{aligned}\|\Pi_{W,L} - \Gamma_{W,L}\|_{TV} = \sup_{A \subset \mathcal{W}\times\mathcal{L}} \{\Pi_{W,L}(A) - \Gamma_{W,L}(A)\} &\geq \sup_{B \subset \mathcal{W}} \{\Pi_{W,L}(B \times \mathcal{L}) - \Gamma_{W,L}(B \times \mathcal{L})\} \\ &= \sup_{B \subset \mathcal{W}} \{\Pi_W(B) - \Gamma_W(B)\} \\ &= \|\Pi_W - \Gamma_W\|_{TV}.\end{aligned}$$

1.3.6 Proof of Lemma 1.9

Define the probability measure μ with marginal $\mu_W := (\Pi_W + \Gamma_W)/2$ and regular transition kernel $\Pi_{L|W}$. Then, $\Pi_W\Pi_{L|W}$ and $\Gamma_W\Pi_{L|W}$ are dominated by μ, and μ_W dominates Π_W (resp. Γ_W). Moreover, μ, Π, Γ share the same transition kernel from $\mathcal{W}$ to $\mathcal{L}$. Then, it can be checked through Theorem 1.1 that

$$(w, l) \mapsto \frac{d\Pi_W}{d\mu_W}(w) \quad \left(\text{resp. } \frac{d\Gamma_W}{d\mu_W}(w)\right) \tag{1.16}$$

is a density of $\Pi_W\Pi_{L|W}$ (resp. $\Gamma_W\Pi_{L|W}$) with respect to μ. Then, from Theorem 1.1 and the definition of the TVD (Eq. 1.5) we obtain

$$\begin{aligned}\|\Pi_W\Pi_{L|W} - \Gamma_W\Pi_{L|W}\|_{TV} &= \int \left|\frac{d\Pi_W}{d\mu_W}(w) - \frac{d\Gamma_W}{d\mu_W}(w)\right| d\mu_{W,L}(w, l) \\ &= \int \left|\frac{d\Pi_W}{d\mu_W}(w) - \frac{d\Gamma_W}{d\mu_W}(w)\right| d\mu_W(w) \\ &= \|\Pi_W - \Gamma_W\|_{TV}, \qquad (1.17)\end{aligned}$$

where (1.17) follows from the definition of the TVD (Eq. 1.5).

Bibliography

[A+23] T. A. Atif et al. Source coding for synthesizing correlated randomness. *IEEE Transactions on Information Theory*, 69(1), 2023.

[AB07] V. Anantharam and V. Borkar. Common randomness and distributed control: A counterexample. *Systems & Control Letters*, 56(7-8), 2007.

[AT20] E. Agustsson and L. Theis. Universally Quantized Neural Compression. In *Annual Conference on Neural Information Processing Systems*, 2020.

[ATM+19] E. Agustsson, M. Tschannen, F. Mentzer, R. Timofte, and L. Van Gool. Generative Adversarial Networks for Extreme Learned Image Compression. In *IEEE/CVF International Conference on Computer Vision*, 2019.

[B+02] C. H. Bennett et al. Entanglement-Assisted Capacity of a Quantum Channel and the Reverse Shannon Theorem. *IEEE Transactions on Information Theory*, 48(10), 2002.

[B+14] C. H. Bennett et al. The Quantum Reverse Shannon Theorem and Resource Tradeoffs for Simulating Quantum Channels. *IEEE Transactions on Information Theory*, 60(5), 2014.

[B+21] T. Berg et al. Distributed Source Simulation With No Communication. *IEEE Transactions on Information Theory*, 67(2), 2021.

[B+25] J. Ballé et al. Good, Cheap, and Fast: Overfitted Image Compression with Wasserstein Distortion. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2025.

[BDH+14] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter. The Quantum Reverse Shannon Theorem and Resource Tradeoffs for Simulating Quantum Channels. *IEEE Transactions on Information Theory*, 60(5):2926–2959, 2014.

[Ber71] T. Berger. *Rate Distortion Theory: A Mathematical Basis for Data Compression*. Englewood Cliffs, NJ: Prentice Hall, 1971.

[BG18] S. Beigi and A. Gohari. Φ -Entropic Measures of Correlation. *IEEE Transactions on Information Theory*, 64(4), 2018.

[Bil12] P. Billingsley. *Probability and Measure*. Wiley, Hoboken, NJ (United States of America), 4th edition, 2012.

[BLS17] J. Ballé, V. Laparra, and E. P. Simoncelli. End-to-end Optimized Image Compression. In *International Conference on Learning Representations*, 2017.

[BM18] Y. Blau and T. Michaeli. The Perception-Distortion Tradeoff. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

[BM19] Y. Blau and T. Michaeli. Rethinking Lossy Compression: The Rate-Distortion-Perception Tradeoff. In *International Conference on Machine Learning*, 2019.

[Bon68] A. Bonami. Ensembles $\Lambda(p)$ dans le dual de D^∞. In *Annales de l'institut Fourier*, volume 18, 1968.

[Boy72] D.W. Boyd. Losing Runs in Bernoulli Trials. *unpublished manuscript*, 1972.

[C+10] P. Cuff et al. Coordination capacity. *IEEE Transactions on Information Theory*, 56(9), 2010.

[C+14] S. O. Chan et al. On Extracting Common Random Bits From Correlated Sources on Large Alphabets. *IEEE Transactions on Information Theory*, 60(3), 2014.

[C+18] G. Cervia et al. Strong Coordination over Noisy Channels with Strictly Causal Encoding. In *Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, 2018.

[C+21] M. Chen et al. Distributed Learning in Wireless Networks: Recent Progress and Future Challenges. *IEEE Journal on Selected Areas in Communications*, 39(12), 2021.

[C+24] M. X. Cao et al. Channel Simulation: Finite Blocklengths and Broadcast Channels. *IEEE Transactions on Information Theory*, 70(10), 2024.

[Can20] C. L. Canonne. A short note on learning discrete distributions. *arXiv e-prints*, 2020. arxiv:2002.11457.

[COS20] G. Cervia, T. J. Oechtering, and M. Skoglund. Remote Joint Strong Coordination and Reliable Communication. In *IEEE International Symposium on Information Theory*, 2020.

[CT06] T.M. Cover and J.A. Thomas. *Elements of Information Theory*. Wiley-Interscience. Wiley, 2006.

[Cuf13] P. Cuff. Distributed Channel Synthesis. *IEEE Transactions on Information Theory*, 59(11), 2013.

[CYW+22] J. Chen, L. Yu, J. Wang, W. Shi, Y. Ge, and W. Tong. On the Rate-Distortion-Perception Function. *IEEE Journal on Selected Areas in Information Theory*, 3(4), 2022.

[DCSMS19] A. Djelouah, J. Campos, S. Schaub-Meyer, and C. Schroers. Neural Inter-Frame Compression for Video Coding. In *IEEE/CVF International Conference on Computer Vision*, 2019.

[DM91] E. J. Delp and O. R. Mitchell. Moment Preserving Quantization (Signal Processing). *IEEE Transactions on Communications*, 39(11), 1991.

[DT62] R. Dobrushin and B. Tsybakov. Information Transmission With Additional Noise. *IRE Transactions on Information Theory*, 8(5), 1962.

[EGK11] A. El Gamal and Y.-H. Kim. *Network Information Theory*. Cambridge University Press, Cambridge (England), 2011.

[F+20] G. Flamich et al. Compressing Images by Encoding Their Latent Representations with Relative Entropy Coding. In *Annual Conference on Neural Information Processing Systems*, 2020.

[Fla23] G. Flamich. Greedy Poisson Rejection Sampling. In *Annual Conference on Neural Information Processing Systems*, 2023.

[FMHL22] G. Flamich, S. Markou, and J. M. Hernandez-Lobato. Fast Relative Entropy Coding with A* Coding. In *International Conference on Machine Learning*, 2022.

[FMHL23] G. Flamich, S. Markou, and J. M. Hernández-Lobato. Faster Relative Entropy Coding with Greedy Rejection Coding. In *Annual Conference on Neural Information Processing Systems*, 2023.

[FT23] G. Flamich and L. Theis. Adaptive Greedy Rejection Sampling. In *IEEE International Symposium on Information Theory*, 2023.

[G+23] D. Gündüz et al. Beyond Transmitting Bits: Context, Semantics, and Task-Oriented Communications. *IEEE Journal on Selected Areas in Communications*, 41(1), 2023.

[GA11] A. A. Gohari and V. Anantharam. Generating Dependent Random Variables Over Networks. In *IEEE Information Theory Workshop*, 2011.

[GK73] P. Gács and J. Korner. Common Information Is Far Less Than Mutual Information. *Problems of Control and Information Theory*, 2, 1973.

[GKJ+20] D. Gündüz, D. B. Kurka, M. Jankowski, M. M. Amiri, E. Ozfatura, and S. Sreekumar. Communicate to Learn at the Edge. *IEEE Communications Magazine*, 58(12), 2020.

[GO80] L. J. Guibas and A. M. Odlyzko. Long Repetitive Patterns in Random Sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 53(3), 1980.

[GPC24] H. M. Garmaroudi, S. S. Pradhan, and J. Chen. Rate-Limited Optimal Transport for Quantum Gaussian Observables. In *IEEE International Symposium on Information Theory*, 2024.

[Gra11] R. M. Gray. *Entropy and Information Theory*. Springer, New York, NY (United States of America), 2nd edition, 2011.

[GSW86] L. Gordon, M. F. Shilling, and M. S. Waterman. An Extreme Value Theory for Long Head Runs. *Prob. Th. Rel. Fields*, 72, 1986.

[H+07] P. Harsha et al. The Communication Complexity of Correlation. In *IEEE Conference on Computational Complexity*, 2007.

[H+19] M. Havasi et al. Minimal Random Code Learning: Getting Bits Back from Compressed Model Parameters. In *International Conference on Learning Representations*, 2019.

[H+24a] Hegazy et al. Compression with Exact Error Distribution for Federated Learning. In *International Conference on Artificial Intelligence and Statistics*, 2024.

[H+24b] R. Huang et al. Provable Benefits of Multi-task RL under Non-Markovian Decision Making Processes. In *International Conference on Learning Representations*, 2024.

[Han03] T. S. Han. *Information-Spectrum Methods in Information Theory*. Springer, Berlin, Heidelberg (Germany), English edition, 2003.

[HFHL24] J. He, G. Flamich, and J. M. Hernndez-Lobato. Accelerating Relative Entropy Coding with Space Partitioning. In *Annual Conference on Neural Information Processing Systems*, 2024.

[HG24] B. Hasırcıoğlu and D. Gündüz. Communication efficient private federated learning using dithering. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2024.

[HLX21] Z. Hu, G. Lu, and D. Xu. FVC: A New Framework towards Deep Video Compression in Feature Space. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[HRTC19] A. Habibian, T. V. Rozendaal, J. Tomczak, and T. Cohen. Video Compression With Rate-Distortion Autoencoders. In *IEEE/CVF International Conference on Computer Vision*, 2019.

[HV93] T.S. Han and S. Verdu. Approximation Theory of Output Statistics. *IEEE Transactions on Information Theory*, 39(3), 1993.

[K+14] G. R. Kumar et al. Exact Common Information. In *IEEE International Symposium on Information Theory*, 2014.

[K+21] G. R. Kurri et al. Coordination Through Shared Randomness. *IEEE Transactions on Information Theory*, 67(8), 2021.

[KA16] S. Kamath and V. Anantharam. On Non-Interactive Simulation of Joint Distributions. *IEEE Transactions on Information Theory*, 62(6), 2016.

[Kan18] S. Kankanahalli. End-To-End Optimized Speech Coding with Deep Neural Networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018.

[KLL+18] W. B. Kleijn, F. S. C. Lim, A. Luebs, J. Skoglund, F. Stimberg, Q. Wang, and T. C. Walters. Wavenet Based Low Rate Speech Coding. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2018.

[KZLK13] J. Klejsa, G. Zhang, M. Li, and W. B. Kleijn. Multiple Description Distribution Preserving Quantization. *IEEE Transactions on Signal Processing*, 61(24), 2013.

[LCLK21] B. Liu, Y. Chen, S. Liu, and H. S. Kim. Deep Learning in Latent Space for Video Prediction and Compression. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021.

[LCOL24] Y. Liu, W.-N. Chen, A. Özgür, and C. T. Li. Universal Exact Compression of Differentially Private Mechanisms. In *Annual Conference on Neural Information Processing Systems*, 2024.

[LEG18] C. T. Li and A. El Gamal. Strong Functional Representation Lemma and Applications to Coding Theorems. *IEEE Transactions on Information Theory*, 64(11), 2018.

[Li24] C. T. Li. Channel Simulation: Theory and Applications to Lossy Compression and Differential Privacy. *Foundations and Trends® in Communications and Information Theory*, 21(6), 2024.

[Lin92] T. Lindvall. *Lectures on the Coupling Method*. Wiley, New York, 1992.

[LKK10] M. Li, J. Klejsa, and W. B. Kleijn. Distribution Preserving Quantization With Dithering and Transformation. *IEEE Signal Processing Letters*, 17(12), 2010.

[LKK11] M. Li, J. Klejsa, and W. Kleijn. On Distribution Preserving Quantization. *arXiv e-prints*, 2011. arXiv:1108.3728.

[LKOK12] M. Li, J. Klejsa, A. Ozerov, and W. B. Kleijn. Audio Coding with Power Spectral Density Preserving Quantization. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2012.

[LL24] Y. Liu and C. T. Li. One-Shot Coding over General Noisy Networks. In *IEEE International Symposium on Information Theory*, 2024.

[LV19] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Texts in Computer Science. Springer International Publishing, Cham, 4th edition, 2019.

[Mat18a] R. Matsumoto. Introducing the Perception-Distortion Tradeoff into the Rate-Distortion Theory of General Information Sources. *IEICE Communications Express*, 7(11), 2018.

[MAT+18b] F. Mentzer, E. Agustsson, M. Tschannen, R. Timofte, and L. V. Gool. Conditional Probability Models for Deep Image Compression. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018.

[Mat19] R. Matsumoto. Rate-Distortion-Perception Tradeoff of Variable-length Source Coding for General Information Sources. *IEICE Communications Express*, 8(2), 2019.

[MP24] M. A. Managoli and V. M. Prabhakaran. Broadcast Channel Synthesis from Shared Randomness. In *IEEE International Symposium on Information Theory*, 2024.

[MS20] D. Minnen and S. Singh. Channel-Wise Autoregressive Entropy Models for Learned Image Compression. In *IEEE International Conference on Image Processing*, 2020.

[Nev70] J. Neveu. *Bases Mathématiques du Calcul des Probabilités*. Masson, Paris (France), 2nd edition, 1970.

[NGBH23] X. Niu, D. Gündüz, B. Bai, and W. Han. Conditional Rate-Distortion-Perception Trade-Off. In *IEEE International Symposium on Information Theory*, 2023.

[Nie13] F. Nielsen. Hypothesis Testing, Information Divergence and Computational Geometry. In *International Conference on Geometric Science of Information*, 2013.

[PBK21] D. Petermann, S. Beack, and M. Kim. Harp-Net: Hyper-Autoencoded Reconstruction Propagation for Scalable Neural Audio Coding. In *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2021.

[PS11] W. A. Pearlman and A. Said. *Digital Signal Compression: Principles and Practice*. Cambridge University Press, Cambridge (England), 2011.

[QCYX25] X. Qu, J. Chen, L. Yu, and X. Xu. Rate-Distortion-Perception Theory for the Quadratic Wasserstein Space. *IEEE Transactions on Information Theory*, 71, 2025.

[QW24] Y. Qiu and A. B. Wagner. Low-Rate, Low-Distortion Compression with Wasserstein Distortion. In *IEEE International Symposium on Information Theory*, 2024.

[QZCK22] J. Qian, G. Zhang, J. Chen, and A. Khisti. A Rate-Distortion-Perception Theory for Binary Sources. In *International Zurich Seminar on Information and Communication*, 2022.

[R78] P. Révész. Strong Theorems on Coin Tossing. In *Proc. Intl. Congress of Mathematicians*. Academia Scientiarum Fennica, 1978.

[RNL+19] O. Rippel, S. Nair, C. Lew, S. Branson, A. Anderson, and L. Bourdev. Learned Video Compression. In *IEEE/CVF International Conference on Computer Vision*, 2019.

[S+20] M. Sudan et al. Communication for Generating Correlation: A Unifying Survey. *IEEE Transactions on Information Theory*, 66(1):5–37, 2020.

[Say12] K. Sayood. *Introduction to Data Compression*. Morgan Kaufmann, Waltham, MA (United States of America), 4th edition, 2012.

[SB18] R.S. Sutton and A.G. Barto. *Reinforcement Learning: An Introduction*. MIT Press, 2nd edition, 2018.

[SBS18] S. Santurkar, D. Budden, and N. Shavit. Generative Compression. In *Picture Coding Symposium*, 2018.

[Sch90] M. F. Schilling. The Longest Run of Heads. *College Math. Journal*, 21(3), 1990.

[SCKY24] S. Salehkalaibar, J. Chen, A. Khisti, and W. Yu. Rate-Distortion-Perception Tradeoff for Lossy Compression Using Conditional Perception Measure. In *IEEE International Symposium on Information Theory*, 2024.

[SCP16] E. C. Song, P. Cuff, and H. V. Poor. The Likelihood Encoder for Lossy Compression. *IEEE Transactions on Information Theory*, 62(4), 2016.

[Sha48] C. E. Shannon. A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3):379–423, 1948.

[SLL24] A. M. Shahmiri, C. W. Ling, and C. T. Li. Communication-Efficient Laplace Mechanism for Differential Privacy via Random Quantization. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2024.

[SLY15a] N. Saldi, T. Linder, and S. Yüksel. Output Constrained Lossy Source Coding With Limited Common Randomness. *IEEE Transactions on Information Theory*, 61(9), 2015.

[SLY15b] N. Saldi, T. Linder, and S. Yüksel. Randomized Quantization and Source Coding With Constrained Output Distribution. *IEEE Transactions on Information Theory*, 61(1), 2015.

[SPC+23] S. Salehkalaibar, T. B. Phan, J. Chen, W. Yu, and A. Khisti. On the Choice of Perception Loss Function for Learned Video Compression. In *Annual Conference on Neural Information Processing Systems*, 2023.

[Str10] D. W. Stroock. *Probability Theory : An Analytic View*. Cambridge University Press, Cambridge (England), 2nd edition, 2010.

[SW24] S. M. Sriramu and A. B. Wagner. Optimal Redundancy in Exact Channel Synthesis. In *IEEE International Symposium on Information Theory*, 2024.

[TA21] L. Theis and E. Agustsson. On the Advantages of Stochastic Encoders. In *Neural Compression: From Information Theory to Applications – Workshop at the International Conference on Learning Representations*, 2021.

[TAL18] M. Tschannen, E. Agustsson, and M. Lucic. Deep Generative Models for Distribution-Preserving Lossy Compression. In *Annual Conference on Neural Information Processing Systems*, 2018.

[TCN25] C. Tian, J. Chen, and K. Narayanan. Source-Channel Separation Theorems for Distortion Perception Coding. In *IEEE International Symposium on Information Theory (ISIT)*, 2025.

[The24] L. Theis. Position: What makes an image realistic? In *Forty-first International Conference on Machine Learning*, 2024.

[TRL21] A. Triastcyn, M. Reisser, and C. Louizos. Dp-rec: Private & Communication-Efficient Federated Learning. *arXiv e-prints*, 2021. arXiv:2111.05454.

[TSHM22] L. Theis, T. Salimans, M. D. Hoffman, and F. Mentzer. Lossy Compression with Gaussian Diffusion. *arXiv e-prints*, 2022. arxiv:2206.08889.

[TW21] L. Theis and A. B. Wagner. A Coding Theorem for the Rate-Distortion-Perception Function. In *Neural Compression: From Information Theory to Applications – Workshop at the International Conference on Learning Representations*, 2021.

[Vil03] C. Villani. *Topics in Optimal Transportation*, volume 58 of *Graduate Studies in Mathematics*. American Mathematical Society, 2003.

[Wag22] A. B. Wagner. The Rate-Distortion-Perception Tradeoff: The Role of Common Randomness. *arXiv e-prints*, 2022. arXiv:2202.04147.

[Win02] A. Winter. Compression of Sources of Probability Distributions and Density Operators. *arXiv e-prints*, 2002. arXiv:quant-ph/0208131.

[Wit75] H. S. Witsenhausen. On Sequences of Pairs of Dependent Random Variables. *SIAM Journal on Applied Mathematics*, 28(1), 1975.

[Wit80] H. Witsenhausen. Indirect Rate Distortion Problems. *IEEE Transactions on Information Theory*, 26(5), 1980.

[WKA11] A. B. Wagner, B. G. Kelly, and Y. Altug. Distributed Rate-Distortion With Common Components. *IEEE Transactions on Information Theory*, 57(7), 2011.

[WSK18] C.-Y. Wu, N. Singhal, and P. Krähenbühl. Video Compression Through Image Interpolation. In *European Conference on Computer Vision*, 2018.

[Wyn75] A. Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2), 1975.

[WZ70] J. Wolf and J. Ziv. Transmission of Noisy Information to a Noisy Receiver With Minimum Distortion. *IEEE Transactions on Information Theory*, 16(4), 1970.

[WZ76] A. Wyner and J. Ziv. The Rate-Distortion Function for Source Coding with Side Information at the Decoder. *IEEE Transactions on Information Theory*, 22(1), 1976.

[XLCZ25] L. Xie, L. Li, J. Chen, and Z. Zhang. Output-Constrained Lossy Source Coding With Application to Rate-Distortion-Perception Theory. *IEEE Transactions on Communications*, 73(3), 2025.

[YGA15] M. H. Yassaee, A. Gohari, and M. R. Aref. Channel Simulation via Interactive Communications. *IEEE Transactions on Information Theory*, 61(6), 2015.

[YT20] L. Yu and V. Y. F. Tan. Exact Channel Synthesis. *IEEE Transactions on Information Theory*, 66(5), 2020.

[ZLO+22] N. Zeghidour, A. Luebs, A. Omran, J. Skoglund, and M. Tagliasacchi. SoundStream: An End-to-End Neural Audio Codec. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 30, 2022.

[ZQCK21] G. Zhang, J. Qian, J. Chen, and A. Khisti. Universal Rate-Distortion-Perception Representations for Lossy Compression. In *Annual Conference on Neural Information Processing Systems*, 2021.