Coding for Communications and Secrecy

THÈSE N^{O} 7751 (2017)

PRÉSENTÉE LE 9 JUIN 2017

À LA FACULTÉ INFORMATIQUE ET COMMUNICATIONS
LABORATOIRE DE THÉORIE DE L'INFORMATION
PROGRAMME DOCTORAL EN INFORMATIQUE ET COMMUNICATIONS

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Mani BASTANIPARIZI

acceptée sur proposition du jury:

Prof. B. Rimoldi, président du jury
Prof. E. Telatar, directeur de thèse
Prof. E. Arikan, rapporteur
Prof. Y. Polyanskiy, rapporteur
Prof. R. Urbanke, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2017

Thesis No. 7751 (March 2017)

Thesis presented to the faculty of computer and communication sciences for obtaining the degree of Docteur ès Sciences

Accepted by the jury:

Professor Emre Telatar
Thesis director

Professor Erdal Arıkan
Expert

Professor Yury Polyanskiy
Expert

Professor Rüdiger Urbanke
Expert

Professor Bixio Rimoldi
President of the jury

Ecole Polytechnique Fédérale de Lausanne, 2017

Dedicated to the memory my grandfather,
Dr. Mohammad Ebrahim Bastani Parizi,
(1925–2014)

Acknowledgments

I am very fortunate to have had the privilege of being advised by Professor Emre Telatar. I express my profound gratitude to Emre for his continual guidance and invaluable contributions, without which this work could not have been possible. Over the past years, I have extensively benefited from Emre's exceptional generosity and his knowledge on a limitless range of technical and non-technical subjects. Emre has always allowed me the room to work in my own way and inspired me to learn about as many topics as possible and to interact with people with different perspectives and interests. Beyond his priceless technical advice on research problems, Emre is a nice, caring, and understanding supervisor. I am indebted to him for all his care and support.

I thank my thesis committee members, Professors Erdal Arıkan (the father of polar codes), Yury Polyanskiy, and Rüdiger Urbanke, for their time and their detailed and insightful comments that improved the quality of the thesis; and Professor Bixio Rimoldi for presiding over the committee. I thank Dr. Olivier Lévêque for his help in writing the French abstract of the thesis, and Dr. Pascal Giard for further editorial comments on the French abstract. I thank Holly Cogliati-Bauereis for patiently proofreading and editing the thesis, and cheerful correction sessions during the stressful thesis-writing days.

In the course of my PhD work, I had the honor of working with brilliant and passionate researchers, from whom I learned a lot. I thank Professor Andreas Burg and Dr. Alexios Balatsoukas-Stimming for the fruitful collaboration on the hardware implementation of polar list decoder, and the pleasant time we spent together in conferences. I thank Professor Neri Merhav for giving me the impetus to study the exact random-coding secrecy exponents, for his invaluable contributions to the work, and for his persistence that was an indispensable source of motivation for working on such a challenging topic.

Working in EPFL's Information Processing Group was a wonderful experience for which I would like to thank all the past and present members of the group. I thank Muriel Bardet and Françoise Behn for their efficient administrative support; and Damir Laurenzi and Giovanni Cangiani for managing the IT infrastructure of the lab. I thank Rüdiger Urbanke for his encouragements, his influential advice on my research in numerous occasions, and his incredible sense of humor. I thank Michael Gastpar, Olivier Lévêque, Nicolas

Macris, and Bixio Rimoldi for many interesting lunch-time conversations on a variety of themes — from politics to theoretical physics. I thank Jean Barbier, Mohamad Dia, Serj Haddad, Marco Mondelli, and Rajai Nasser for their camaraderie and memorable conference trips; Rafah El-Khatib and Mohammad Karzand for making the office a fun place to be; and Emmanuel Abbe, Andrei Guirgiu, Seyed Hamed Hassani, and Eren Şaşoğlu for insightful discussions and career advice.

Warm thanks go to all my friends whose companionship made the past eight years an unforgettable stage of my life full of memorable moments. I am hugely indebted to Sousan Abolhassani, Massoud Dadras, and Dara Bayat who welcomed me into their home from the first moment of my arrival in Switzerland and, with their everlasting support, never let me feel separated from my family by thousands of kilometers. My special thanks go to Esther Kleinhage, Massimo Ravasi, Lisa Spotz, Kevin Koehler, Ugo Campiglio, Mahshid Chekini, Arash Arami, and Hossein Afshari, my most attentive and thoughtful friends, who always stood by me through the twists and turns of life.

Last but not least, my heartfelt gratitude goes to my parents, Dordaneh Dorreh and Hamid Bastani Parizi, for their unconditional love and support.

This thesis was supported by the Swiss National Science Foundation under grant number 200020_146832. This funding is greatly appreciated.

Abstract

Shannon, in his landmark 1948 paper, developed a framework for characterizing the fundamental limits of information transmission. Among other results, he showed that reliable communication over a channel is possible at any rate below its *capacity*. In 2008, Arıkan discovered *polar codes*; the only class of explicitly constructed low-complexity codes that achieve the capacity of any binary-input memoryless symmetric-output channel. Arıkan's *polar transform* turns independent copies of a noisy channel into a collection of synthetic almost-noiseless and almost-useless channels. Polar codes are realized by sending data bits over the almost-noiseless channels and recovering them by using a low-complexity successive-cancellation (SC) decoder, at the receiver.

In the first part of this thesis, we study polar codes for communications. When the underlying channel is an erasure channel, we show that almost all correlation coefficients between the erasure events of the synthetic channels decay rapidly. Hence, the sum of the erasure probabilities of the information-carrying channels is a tight estimate of the block-error probability of polar codes when used for communication over the erasure channel.

We study SC *list* (SCL) decoding, a method for boosting the performance of short polar codes. We prove that the method has a numerically stable formulation in log-likelihood ratios. In hardware, this formulation increases the decoding throughput by 53% *and* reduces the decoder's size about 33%. We present empirical results on the trade-off between the length of the CRC and the performance gains in a CRC-aided version of the list decoder. We also make numerical comparisons of the performance of long polar codes under SC decoding with that of short polar codes under SCL decoding.

Shannon's framework also quantifies the secrecy of communications. Wyner, in 1975, proposed a model for communications in the presence of an eavesdropper. It was shown that, at rates below the *secrecy capacity*, there exist reliable communication schemes in which the amount of information leaked to the eavesdropper decays exponentially in the block-length of the code. In the second part of this thesis, we study the rate of this decay.

We derive the exact exponential decay rate of the *ensemble-average* of the information leaked to the eavesdropper in Wyner's model when a randomly constructed code is used for secure communications. For codes sampled from the ensemble of *i.i.d. random codes*, we show that the previously known lower bound to the exponent is exact. Our ensemble-optimal exponent for random *constant-composition codes* improves the lower bound extant in the literature. Finally, we show that random *linear* codes have the same secrecy power as i.i.d. random codes.

The key to securing messages against an eavesdropper is to exploit the randomness of her communication channel so that the statistics of her observation resembles that of a pure noise process for any sent message. We study the effect of feedback on this approximation and show that it does not reduce the minimum entropy rate required to approximate a given process. However, we give examples where variable-length schemes achieve much larger *exponents* in this approximation in the presence of feedback than the exponents in systems without feedback. Upper-bounding the best exponent that block codes attain, we conclude that variable-length coding is necessary for achieving the improved exponents.

Keywords: capacity-achieving error-correction codes, polar codes, successive-cancellation decoding, successive-cancellation list decoding, wiretap channel, channel resolvability, random-coding secrecy exponents, random-coding resolvability exponents, feedback resolvability exponents, variable-length resolvability codes

Résumé

Dans son article fondateur de 1948, Shannon a développé un cadre pour caractériser les limites fondamentales de la transmission des données. Il a montré entre autres qu'il est toujours possible de communiquer sur un canal de manière fiable tant que le débit de communication choisi reste en deçà de la *capacité* du canal. En 2008, Arıkan, a découvert les codes polaires qui sont les seuls codes de basse complexité avec une construction explicite atteignant la capacité de tous les canaux symétriques à entrée binaire et sans mémoire. La *transformation polaire* d'Arıkan convertit des copies indépendantes d'un canal bruité en une collection de canaux synthétiques, tous presque sans bruit ou au contraire presque inutiles. Les codes polaires sont donc réalisés en envoyant les données via des canaux presque sans bruit et en les récupérant au récepteur avec un décodeur par annulations successives (SC) de basse complexité.

Dans la première partie de cette thèse, nous étudions les codes polaires pour les communications. Lorsque le canal sous-jacent est un canal d'effacement, nous montrons que presque tous les coefficients de corrélation entre les effacements des canaux synthétiques déclinent rapidement. De là, nous déduisons que la somme des probabilités d'effacement des canaux portant des données est une bonne évaluation de la probabilité d'erreur des codes polaires lorsqu'ils sont employés pour la communication via le canal d'effacement.

Nous étudions le décodage par annulations successives *de type liste* (SCL), une méthode pour améliorer la performance des codes polaires courts. Nous montrons que la méthode a une formulation numériquement stable en termes des rapports de log-vraisemblances. Une fois implémentée, cette formulation permet d'augmenter le débit du décodeur jusqu'à 53% *et* de réduire la taille du matériel utilisé d'environ 33%. Nous présentons les résultats empiriques illustrant le compromis entre la longueur du code CRC et le gain de performance pour une version du décodeur de type liste assisté par CRC. Nous comparons aussi la performance des codes polaires longs avec décodage SC et celle des codes polaires courts avec décodage SCL.

Le cadre mathématique de Shannon quantifie aussi la confidentialité de la communication. En 1975, Wyner a proposé un modèle de communication en présence d'un espion. Il a été démontré que tant que le débit de communication choisi reste en deçà de la *capacité de confidentialité* du canal, il existe des méthodes de communication fiables pour lesquelles la quantité d'information divulguée à l'espion décroît exponentiellement en fonction de la longueur du code. Dans la deuxième partie de cette thèse, nous étudions le taux de cette décroissance.

Nous dérivons le taux précis de la décroissance exponentielle de la quantité d'information *moyenne* divulguée à l'espion dans le modèle de Wyner lorsqu'un code construit aléatoirement est utilisé pour des communications confidentielles. Pour ceux tirés de l'ensemble des *codes aléatoires i.i.d.*, nous montrons que le minorant précédemment connu à l'exposant est exact. Notre exposant optimal pour l'ensemble des *codes aléatoires de composition constante* est plus grand que le minorant déjà connu dans la littérature. Enfin, nous montrons que les codes aléatoires *linéaires* possèdent la même puissance de confidentialité que les codes aléatoires i.i.d.

La clé pour protéger les messages contre l'espion est d'exploiter l'entropie de son canal, de telle façon à ce que la statistique de l'observation de celui-ci ressemble à celle d'un processus de bruit pur, quel que soit le message envoyé. Nous étudions l'effet du rétrocontrôle sur cette approximation et montrons qu'il ne réduit pas le taux minimum d'entropie requis pour approximer un processus donné. Cependant, nous donnons aussi des exemples qui montrent que, toujours dans le cadre de cette approximation, des codes à longueur variable atteignent des *exposants* beaucoup plus grands en présence du rétrocontrôle que dans les systèmes sans rétrocontrôle. En dérivant un majorant sur le meilleur exposant atteignable avec des codes de longueur fixe, nous concluons que le codage à longueur variable est nécessaire afin d'atteindre ces exposants.

Mots-clés: codes correcteurs d'erreurs, codes polaires, décodage par annulations successives, décodage par annulations successives de type liste, canal wiretap, résolubilité du canal, exposants de confidentialité, exposants de résolubilité, exposants de résolubilité en présence du rétrocontrôle, codes de résolubilité à longueur variable

Contents

List of Figures

Notation

$a \coloneqq b$	a is defined as b.
$\square$	End of a proof
$\mathcal{A}, \mathcal{B}, \dots$	Sets or events
$\mathbb{N}$	Set of natural numbers $\{1, 2, 3, \dots\}$
$\mathbb{Z}$	Set of integers $\{\dots, -2, -1, 0, 1, 2, \dots\}$
$\mathbb{R}$	Set of real numbers
$\mathcal{A} \times \mathcal{B}$	Cartesian product of two sets $\mathcal{A}$ and $\mathcal{B}$
$\mathcal{A}^n$	n^{th} Cartesian power of the set $\mathcal{A}$
$\lvert\mathcal{A}\rvert$	Cardinality of the finite set $\mathcal{A}$
$\mathcal{A}^c$	Complement of the set (or event) $\mathcal{A}$
x_i^j	If $i \le j$, the vector $(x_i, x_{i+1}, \dots, x_j)$, otherwise the null vector
x^n	Shorthand notation for $x_1^n = (x_1, x_2, \dots, x_n)$
$x_{\mathcal{I}}$ (for $\mathcal{I} \subset \mathbb{N}$)	Sub-vector $(x_i \colon i \in \mathcal{I})$
$\mathrm{w_H}(x^n)$	Hamming weight of x^n
$\mathrm{d_H}(x^n, y^n)$	Hamming distance between x^n and y^n
$[a]^+$	Positive clipping operation, $$[a]^+ \coloneqq \max\{a, 0\}$$
$\mathbb{1}\{\cdot\}$	Indicator function; equals 1 if the statement inside the braces is true and 0 otherwise.
$\mathbb{1}_{\mathcal{E}}$	Indicator function of the event $\mathcal{E}$
$\Pr\{\cdot\}$	Probability of the event inside the braces
$\Pr(\mathcal{E})$	Probability of the event $\mathcal{E}$
$\mathbb{E}[X]$	Expected value of X
$\mathrm{var}(X)$	Variance of X
$X \multimap\!\!\!\!\circ\, Y \multimap\!\!\!\!\circ\, Z$	Random variables X, Y, and Z form a Markov chain.
$H(X)$	Entropy of the random variable X
$I(X;Y)$	Mutual information between random variables X and Y
$\mathcal{P}(\mathcal{A})$	Set of distributions on the alphabet $\mathcal{A}$

- For a distribution $P \in \mathcal{P}(\mathcal{A})$:

  $\mathrm{supp}(P)$ Support of P,

$$\mathrm{supp}(P) \coloneqq \{a \in \mathcal{A} : P(a) > 0\}$$

  $H(P)$ Entropy of P,

$$H(P) \coloneqq \sum_{a\in\mathcal{A}} P(a) \log\Big[\frac{1}{P(a)}\Big]$$

- In this thesis, without essential loss of generality, the bases of $\log(\cdot)$ and $\exp(\cdot)$ are assumed to be 2.
- For two distributions $P \in \mathcal{P}(\mathcal{A})$ and $Q \in \mathcal{P}(\mathcal{A})$:

  $P \ll Q$ P is absolutely continuous with respect to Q, i.e., $\mathrm{supp}(P) \subseteq \mathrm{supp}(Q)$

  $|P - Q|$ ℓ_1 distance between P and Q,

$$|P - Q| \coloneqq \sum_{a\in\mathcal{A}} |P(a) - Q(a)|$$

  $D(P\|Q)$ Kullback–Leibler (KL) divergence between P and Q,

$$D(P\|Q) \coloneqq \sum_{a\in\mathcal{A}} P(a) \log\Big[\frac{P(a)}{Q(a)}\Big]$$

- For a joint distribution $Q \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$:

  Q_X (resp. Q_Y) x-marginal (resp. the y-marginal) of Q,

$$Q_X(x) \coloneqq \sum_{y\in\mathcal{Y}} Q(x, y), \quad Q_Y(y) \coloneqq \sum_{x\in\mathcal{X}} Q(x, y)$$

  $I(Q)$ Mutual information between random variables X and Y when $(X, Y) \sim Q$,

$$I(Q) \coloneqq D(Q\|Q_X \times Q_Y)$$

- For a distribution $P \in \mathcal{P}(\mathcal{X})$ and a stochastic matrix $W\colon \mathcal{X} \to \mathcal{Y}$:

P^n n-fold product distribution

$$P^n(x^n) \coloneqq \prod_{i=1}^{n} P(x_i), \quad \forall x^n \in \mathcal{X}^n$$

W^n n-fold product matrix $W^n : \mathcal{X}^n \to \mathcal{Y}^n$

$$W^n(y^n|x^n) = \prod_{i=1}^{n} W(y_i|x_i), \quad \forall x^n \in \mathcal{X}^n, y^n \in \mathcal{Y}^n$$

$P \times W$ Joint distribution $(P \times W) \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$

$$(P \times W)(x, y) \coloneqq P(x)W(y|x), \quad \forall (x, y) \in \mathcal{X} \times \mathcal{Y}$$

$P \circ W$ y-marginal of the joint distribution $P \times W$,

$$(P \circ W)(y) \coloneqq \sum_{x \in \mathcal{X}} P(x)W(y|x), \quad \forall y \in \mathcal{Y}$$

1 Introduction and Overview of the Results

Communication is the process of transmitting *information* from an information *source* to an information *sink*, via a *medium*. Here, transmission can be in space or in time. Many of the systems we conventionally know as communication systems are those that transmit information in space. For example, when we make a phone call or send a fax we transfer information in space — from one point to another point, with practically no delay. The transmission of information in time is data storage; for example, all kinds of recording media (from magnetic tapes to modern solid state storage devices) or even our DNA transmit information in time.

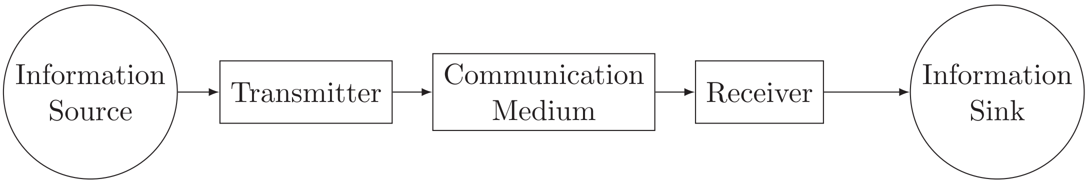


Figure 1.1: Transmission of Information from Source to Sink, via a Medium

The characteristic common to every communication medium available in the nature is noisiness: the information carried by the medium is subject to unpredictable, but statistically regular, alterations. Furthermore, in most circumstances, the distortion caused by the communication media is beyond tolerable levels. The principal task of communication system engineers is to design transceivers capable of transmitting information *reliably*, i.e., with low distortion, via a noisy communication medium.

Shannon, in his landmark paper [101] postulated a mathematical framework that models information transmission systems. In particular, he quantified the amount of information a medium can reliably carry as its *capacity*, and he showed that, rather surprisingly, as long as the *rate* of information to be communicated over the medium is below its capacity, transmission at arbitrarily low error probabilities is feasible.

Shannon used a non-constructive argument to demonstrate the existence of arbitrarily reliable communication schemes that operate at rates close to the capacity of the medium. Designing practical methods to exploit the communication resources close to their capacities has, ever since, been a challenge. About sixty years later, Arıkan [4, 5] proposed the first method, called *polar coding*, for explicitly constructing low-complexity schemes that provably allow reliable communications at rates arbitrarily close to the capacity. In the first part of this dissertation we consider polar codes for communications, analyze their performance, and discuss various aspects of the design and implementation of communication systems based on polar codes.

Shannon's breakthrough gave birth to *information theory*: the discipline of characterizing the fundamental limits of information transmission. The scope of information theory is not limited to point-to-point communications. In many circumstances more than two parties are involved in information transmission. Consider scenarios where sensitive information is communicated between two authorized parties, through a public communication media to which an unauthorized eavesdropper has access as well. In such settings, in addition to the reliability of information transmission, the *secrecy* of information is important. Shannon's framework also permits measuring the secrecy of information systems.

The foundations of *information-theoretic secrecy* were laid by Shannon [102] and later by Wyner [118]. In particular, Wyner [118] proposed a model, called the *wiretap* model, for communications in the presence of an eavesdropper through noisy media. He discovered that, as long as the eavesdropper's medium is noisier than that of the authorized receiver, reliable *and secure* communication, is feasible. In the second part of this thesis, we focus on Wyner's wiretap model and establish several results on the fundamental limits of secrecy guarantees in such a setting.

1.1 Channel Coding

Shannon's celebrated paper [101], answered the fundamental question of how much information can be transmitted reliably over a *noisy* communication channel. A communication channel is the mathematical model for the communication medium at hand. It can be seen as a 'black box' that accepts input symbols from an input *alphabet* and produces an output symbol from its output alphabet that is correlated to the input symbol, according to the *channel law*. The channel is assumed to be dictated by the nature and most often known to us.[1]

To transmit information through a channel, an *encoder* uses a *code* to map one of M different *messages* to a sequence of n channel input symbols and, at the receiver side, a *decoder* attempts to reproduce the communicated message upon observing the corresponding n channel output symbols (see Figure 1.2).

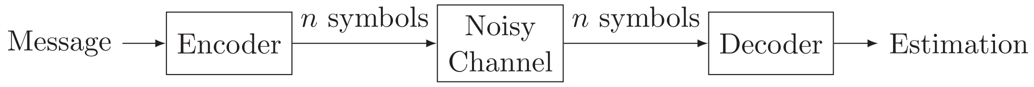


Figure 1.2: To communicate a message, the encoder maps it to n channel input symbols and the decoder estimates the sent message given noisy channel outputs.

A *code* is characterized by three parameters: (i) its block-length n, (ii) its *rate*, defined as

$$\frac{\log(M)}{n},$$

measuring how many messages can be transmitted via n channel uses, and (iii) its *block-error probability*, i.e., the probability of misestimating the sent message at the receiver. (The careful reader will notice that the block-error probability is not a well-defined concept without specifying a decoding algorithm and the channel over which the code is to be used for communication.)

In n uses of the channel, the number of messages that can reliably be transmitted can grow, at most, exponentially fast with n: there are exponentially many different combinations of channel input symbols, formally known as *codewords*, and two different messages should not be mapped to the same codeword. Using *all* possible codewords for information transmission typically results in a high block-error probability, because any small distortion from the channel can 'turn' the sent codeword to another one and confuse the decoder. Choosing only a subset of sufficiently 'separated' channel input sequences as codewords reduces the probability of error, because to shift one codeword to another, higher distortions from the channel would be necessary.

The common belief in the *pre-Shannon era* was that to improve the reliability of transmission it is always necessary to decrease the number of codewords, i.e., sacrifice the rate of information transmission. Shannon, contradicted this belief by introducing the notion of *channel capacity*. His *noisy channel coding theorem* asserts that, given any rate below the capacity of the channel and any desired level of reliability, there exists a code of that rate that meets the reliability requirement for communication over that channel, provided that we

[1] The latter might sound (and in practice is) a simplistic assumption. Usually the channel over which the communication system is used by the end-user is not exactly the one for which the system is designed. However, communication protocols often enable the transceivers to *estimate* the channel and accommodate to channel conditions by changing the transmission parameters accordingly. In addition, small changes in channel conditions would typically not cause radical changes in the performance of the transmission scheme. Therefore, when designing the system, the assumption about the knowledge of the channel is not unrealistic.

let the block-length of the code be sufficiently large. Conversely, for any rate above the capacity there is an upper bound to the reliability of any code of that rate, no matter how large its block-length is.

Instead of giving a recipe for constructing a capacity-achieving code, to prove his noisy channel coding theorem, Shannon relied on the probabilistic method. He mentions the following, just after the statement of [101, Theorem 11]:

> The method of proving the first part [direct part] of this theorem is not by exhibiting a coding method having the desired properties, but by showing that such a code must exist in a certain group of codes. In fact we average the frequency of errors over this group and show that average can be made less than ϵ. If the average of a set of numbers is less than ϵ there must exist at least one in the set which is less than ϵ. This will establish the desired result.

Shannon's *random-coding* method is an elegant approach for showing the existence of good codes and is the *de facto* method for almost all achievability proofs in information theory. However, employing the codes constructed via his random-selection approach is computationally and practically infeasible. Any reliable high-rate code must have a large block-length [42, Theorem 5.8.1]. Whereas, the memory and the computational power required for the coding scheme proposed by Shannon grows exponentially in block-length. Finding good *and efficiently implementable* codes (i.e., those with low block-error probability, efficient encoding and decoding algorithms, and a rate close to the capacity) was left as a challenging exercise to *coding theorists.*

Coding theory was inaugurated by the appearance of the perfect codes of Hamming and Golay [45, 46] and, later on, Reed–Muller codes [83, 91]. In the first decades of coding theory, the objective of coding theorists was to find linear codes[2] with a large number of codewords and a high *minimum distance.* In other words, to 'pack' as many codewords as possible into the space of channel input sequences and ensure that every two codewords differ at sufficiently many coordinates [73]. Some of the remarkable fruits of those efforts are the Bose–Chaudhuri–Hocquenghem (BCH) codes and Reed–Solomon (RS) codes [22, 55, 92].

Despite being very powerful and efficient (for the computing technology of time), the early codes failed to achieve Shannon's limits. Linearity is an important property for the computational efficiency of the code. In fact, it is well-known that over a large class of channels (virtually all channels we would deal with in practice) *random linear codes* perform essentially as well as Shannon's random codes that, in turn, perform as well as best codes at

[2] The component-wise sum of any two codewords in a linear code is a codeword. Thus, a linear code is a linear subspace of the space of channel input sequences. (Here we are assuming the channel input alphabet has a group property.) Such a code can be specified by (any) basis of the subspace it defines. Linearity significantly reduces the encoding complexity.

rates close to the capacity [42]. Therefore, by restricting the attention to linear codes, we are not compromising the performance. However, it turns out that maximizing the minimum distance is not the appropriate criteria for designing capacity-achieving codes. In fact, in the early days of coding theory, Elias's convolutional codes [38] were known to be powerful codes with efficient decoding algorithms [75, 112, 116] which were *not* designed based on the minimum distance criteria.

After the introduction of *turbo codes* [19] in the early 1990s, the attention in coding theory turned into *codes on graphs* that are decoded with *iterative decoding* algorithms. Such an approach was originally proposed by Gallager in [41], where he devised *low-density parity check* (LDPC) codes. In short, instead of finding a specific good code, the goal became finding an *ensemble* of codes such that the *distance profile* (as opposed to the minimum distance) of a randomly chosen code from the ensemble resembles that of Shannon's random codes. In addition, it is required that the graph representing the relations among the coordinates of the codewords has a simple structure to keep the decoding computationally efficient (e.g., the graph representing an LDPC code is assumed to be sparse).

Codes on graphs appeared very promising for approaching Shannon's limits. Substantial progress was made in the development and analysis of such codes and they were incorporated in many modern communication standards. It was proven recently that, combined with *spatial coupling* (a construction inspired by methods from statistical physics), such codes can achieve Shannon's capacity under iterative decoding [66]. We refer the reader interested in the history and key contributions in coding theory to the excellent survey of [26].

Arıkan [4, 5] postulated an information-theoretic perspective to channel coding. He proposed a transform that, when applied to n consecutive uses of any binary-input channel, 'polarizes' them into n *nearly extremal* channels; these channels are either almost noiseless or almost useless. Moreover, the fraction of noiseless channels equals the capacity of the underlying channel. As a result, capacity-achieving *polar codes* are constructed by transmitting information over the almost noiseless channels. Polar codes are very closely related to old Reed–Muller codes. The basis vectors for the subspace they define are chosen from the same matrix (the binary Walsh–Hadamard matrix). The key difference is that, instead of choosing the basis to optimize the distance profile of the code (i.e., the Reed–Muller choice), polar codes are constructed by choosing the coordinates that 'see' an almost-noiseless effective channel to the transmitter. This makes polar codes capacity-achieving under the low-complexity *successive-cancellation* (SC) decoding.[3]

Unlike randomly constructed codes on graphs, *a* polar code of length n and rate R for reliable communication over a given channel is uniquely de-

[3] Reed–Muller codes were for a long time conjectured to be capacity-achieving under optimal decoding. Very recently, this conjecture was proven when the transmission takes place over the erasure channel [65].

fined. Moreover, the successive-cancellation decoding algorithm proposed by Arıkan is a single-pass procedure that decodes the sent codeword in $O(n \log n)$ complexity (as opposed to iterative decoding methods for LDPC-like codes). To date, polar codes are the only class of *explicitly constructed* error-correction codes with a low-complexity encoding *and decoding* algorithm that provably achieve the capacity of *any* channel. As such, polar codes are the first practical answer to Shannon's challenge.

Although at the time of its introduction, constructing codes based on channel polarization was more an alluring theoretical method for designing computationally efficient capacity-achieving error-correction schemes, the extensive progress in improving their decoding algorithms (e.g., using successive-cancellation *list* decoding [108]), as well as in the VLSI implementation of polar codes, e.g. [2, 40, 67, 68, 80, 86, 90, 95, 96, 123–125], made them attractive for practical applications. Despite their infancy, polar codes have been recently included in the next-generation (5G) standard for communication systems [1].

Contribution of this Thesis — Part I

We study polar codes for communications in the first part of this thesis. We start by an introduction to Arıkan's channel polarization method in **Chapter 2**, see how this leads to the construction of capacity-achieving codes, and analyze the performance of polar codes under successive-cancellation (SC) decoding. We also review the methods for the construction of polar codes.

As we will see, yet another distinguishing characteristic of polar codes is that the performance guarantees of polar codes are not based on simulations. In other words, given a polar code and the channel over which this code is to be used, we can *compute* upper and lower bounds on the block-error probability of the code. This property makes them attractive for applications where very low error-probability guarantees (say, as small as 10^{-16}) are required. Even though both upper and lower bounds on the block-error probability of polar codes decay very fast (roughly like $2^{-\sqrt{n}}$ [6]), they still differ by orders of magnitude, especially when the block-length is relatively large. Hence, to avoid a conservative design based on an overestimation of the block-error probability, it is desirable to have better estimates of the true block-error probability of polar codes. In **Chapter 3**, we study the tightness of the upper bound (which is simply the union bound) on the block-error probability of polar codes by analyzing the correlation between the error events of the information-carrying channels, when transmission takes place over an erasure channel. We prove that the upper bound is indeed a tight estimate of the block-error probability of polar codes (Theorem 3.11) and also provide formulae for computing a tighter lower bound on the block-error probability of polar codes, when transmission takes place over an erasure channel (Lemma 3.5). For the typical block-lengths considered in Chapter 3, this lower bound practically matches the existing upper bound (see Figure 3.1).

Shortly after the publication of Arıkan's low-complexity SC decoding algorithm [5], an improved variant of it, called *successive-cancellation list* (SCL) decoding was proposed as a method for boosting the performance of polar codes while keeping the complexity of decoding algorithm low [108]. The empirical results of [108] suggest that the performance of a SCL decoder is very close to that of an optimal (but computationally complex) decoder for polar codes. In addition, to reduce their block-error probability further, using SCL decoder (instead of SC decoder) enables one to *concatenate* polar codes with CRC codes [85, 110]. The latter makes them competitive with existing commercially used error-correction codes of the same length. The original formulation of the successive-cancellation list-decoding algorithm is in terms of likelihoods that makes the decoder prone to underflow floating-point errors. In practice, the numerical stability of computations has to be guaranteed by implementing the algorithm in a numerically stable domain. In **Chapter 4**, we tackle the problem of design and implementation of a communication system using polar codes and SCL decoder. In Theorem 4.1, we show that successive-cancellation list decoding can be formulated exclusively in terms of *log-likelihood ratios*. Log-likelihood ratios are numerically stable and lead to a more efficient implementation of the decoder compared to the existing implementations of [10, 71, 72, 122, 126]. In addition, in § 4.3.1, we evaluate the performance of CRC-concatenated polar codes and highlight the importance of carefully tuning the length of the outer CRC code to the error-correction capabilities of the inner polar code (which depends on the choice of decoder). In the design of a communication system using polar codes, depending on the application, we can have the choice between adopting a long polar code to be decoded with the conventional SC decoder or a shorter polar code to be decoded with the more complex SCL decoder in order to guarantee a target block-error probability. In § 4.3.2, we compare the performance of short CRC-concatenated polar codes under SCL decoding with that of longer polar codes under SC decoding and show that, roughly speaking, at the same block-error probability and under the same decoding complexity, successive-cancellation list decoding and CRC concatenation enable us to reduce the block-length (and hence the decoding latency) by a factor of eight.

1.2 Information Theoretic Secrecy

In his subsequent work [102], Shannon used his mathematical framework for the analysis of communication systems, to analyze the secrecy of information transmission systems. He considered a model, today called the Shannon cipher system, where two legitimate parties communicate messages encrypted by using a securely shared *key* over a public communication channel. Based on his measure of information, i.e., the *entropy*, he proposed to assess the secrecy of the system via the amount of information that an *eavesdropper*, with access to publicly transmitted cryptograms (see Figure 1.3), would learn about the messages.

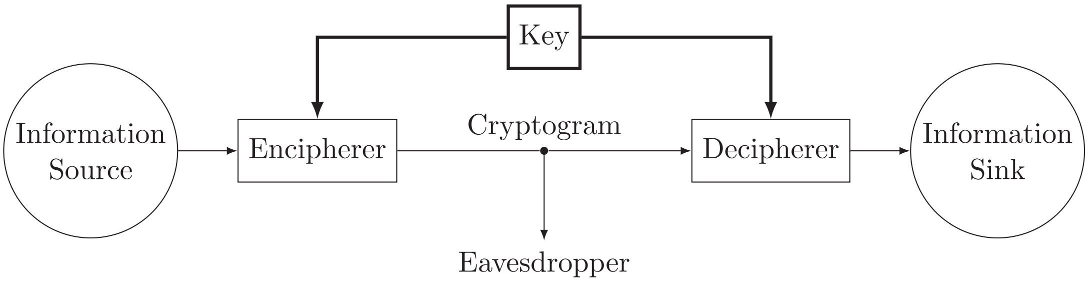


Figure 1.3: Shannon Cipher System

Shannon showed that, rather discouragingly, in order for the cipher system to be perfectly secure, i.e., to *leak zero information* to the eavesdropper, the entropy of the key must be at least as large as that of the information source. Such a requirement is typically hard to fulfill. Apart from specific situations[4], if a secure channel capable of communicating that much entropy (the one denoted by thick lines in Figure 1.3) is available, it could be directly used for the transmission of secret information (instead of the public channel).

At this point the reader might notice that Shannon's model (Figure 1.3) describes the operation of *symmetric-key cryptography* algorithms that are commonly used nowadays. For example to establish a secure connection when we connect to a WiFi network, we share a password (key) securely between our device and the access point. The entropy of such a key is definitely lower than the vast amount of information our device encrypts and sends to the access point or vice versa. Does Shannon's conclusion imply that such systems are insecure?

The answer is yes and no! Most of these algorithms are not *information-theoretically* secure. In fact, the only encryption technique providing perfect secrecy in Shannon's sense is *One Time Pad* that uses truly pre-shared random keys for enciphering text. The encryption methods used in modern systems have a similar basis, with the difference that the key is generated locally using a pair of *pseudorandom-number* generators at the transmitter and the receiver side, synchronized using the pre-shared password. It is easy to show that such a pseudorandom key does not have sufficient entropy for providing perfect secrecy. However, 'secrecy' from a cryptographic point of view usually relies on the boundedness of the eavesdropper's computational power. Even though, in most current cryptographic systems, the cryptogram is informative about the secret information it encodes, according to the current computational models and power, it would be computationally hard for an eavesdropper to deduce the secret information upon observing the cryptogram.

[4] e.g., when such a key can be pre-shared securely at the time when the information is not available yet or when the channel between the key source and the legitimate parties is one-way

Shannon's notion of the secrecy is, in a sense, the most stringent measure. Such a guarantee implies that an eavesdropper observing the cryptogram is as informed about the secret message as an entity who generates a fake cryptogram independently knowing only the statistics of the cryptogram. Although in some views *information-theoretic secrecy* was regarded as too strict to be attainable in practice, in 1975, Wyner [118] showed that information-theoretic secrecy can be guaranteed if we incorporate the noise of communication media into the model.

In Wyner's *wiretap channel* model (see Figure 1.4), the advantage of the legitimate receiver over the eavesdropper is in having a *cleaner* communication channel from the transmitter. The wiretap model is arguably a more realistic model for the actual information transmission scenarios in the presence of an eavesdropper. For example, your neighbour wiretapping your WiFi router's signals gets a noisier version of what you receive, because there are more obstacles in the signal path between the router and her than in the path between the router and you. Furthermore, in his model, the secrecy and reliability of information transmission is provided simultaneously by the channel encoder and decoder (this is why encipherer and decipherer are replaced by encoder and decoder in the diagram).[5]

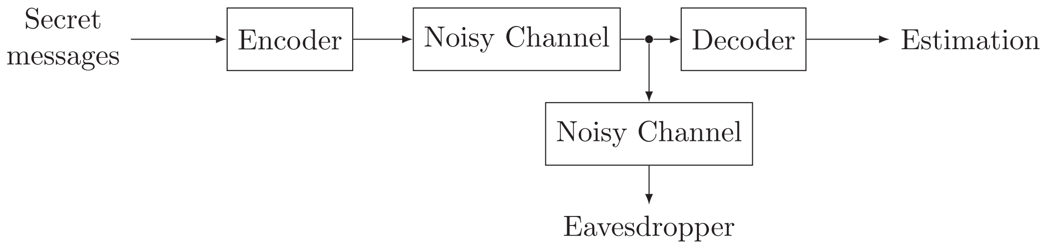


Figure 1.4: Wyner's wiretap model

Wyner [118] showed that there exist asymptotically *reliable* and *secure* communications schemes that enable the transmission of information at *positive rates* in his model. In other words, with sufficiently large n, we can find an encoder that, given one of M secret messages (where M is exponentially large in n), produces a sequence of n channel inputs, and a decoder that maps back the noisy versions of these symbols at the output of the channel to the sent message with arbitrarily low probability of error. Meanwhile the encoding method guarantees that the information that the eavesdropper learns about the secret message through her observations, normalized by the block-length n, is arbitrarily small. Indeed, Wyner characterized the *secrecy capacity* of the

[5] The reader also notices that the shared secret key is missing in Wyner's model. It turns out that sharing a key only increases the secure information transmission capacity of the system by an amount equal to the entropy of the key.

wiretap channel: the highest rate at which secure and reliable communications is possible.

Wyner's results were extended by Csiszár and Körner [29] to the cases where *more noisiness* of the eavesdropper's channel is not necessarily due to the concatenation of two physical channels, but the advantage of the legitimate receiver over the eavesdropper only follows from the mathematical descriptions of their channels. Csiszár and Körner also studied the scenarios where part of the information is to be broadcast to both parties (i.e., the eavesdropper also has to receive them reliably) while the rest is private to the legitimate receiver.

The principle in securing the messages against the eavesdropper is to *exploit* the random noise of her communication channel. To communicate a secret message over the wiretap channel, the encoder maps it to a *randomly* chosen codeword among a particular set of codewords. When randomness in the encoding operation and the intrinsic randomness of the eavesdropper's channel are combined, her observations appear like pure noise. Therefore, she would learn very little about the secret message.

As their ages suggest, the problem of secure communications in the presence of an eavesdropper is studied less, compared to point-to-point communications. Although some fundamental limits in such scenarios were discovered following the work of Wyner (see, for example, [29, 69]), and some structured coding schemes for the wiretap channel were proposed (e.g., [74, 100]), many fundamental questions concerning the model are still open. For example, given the asymptotic nature of the secrecy guarantees, we wonder how long should the block-length be in order to guarantee a certain level of information leakage to the eavesdropper? The analogous question in error correction was answered in the 1960s when it was found that the best code would guarantee exponentially small error probability. But, to our knowledge, it was only in 1996 when Csiszár [27] showed that the information leaked to the eavesdropper also vanishes exponentially fast in the block-length. Nevertheless, to date, the best exponential decay rate that we could hope to attain is unknown.

Contribution of this Thesis — Part II

In the second part of this dissertation, we focus on the exponential decay rate of the information leaked to the eavesdropper in Wyner's wiretap channel setting. As we characterize certain fundamental limits, we do not limit ourselves to a specific class of computationally efficient communication schemes (as opposed to the first part of the work where we specifically focus on polar codes for communications). In particular, we rely on the probabilistic method and take advantage of the convenience of random-coding arguments in our achievability proofs.

We formally introduce Wyner's wiretap channel model in **Chapter 5** and review the construction of codes for secure and reliable communications via random selection. We also discuss the concept of *channel resolvability* — the encoding method that, when used at the input of a channel, makes its output look like plain noise. We show how a resolvability-based approach to the wiretap channel leads to the construction of secure codes and study the relation between *secrecy* and *resolvability exponents*. The former is the exponential decay rate of the information leaked to the eavesdropper in the wiretap model, whereas the latter is the exponential decay rate of the distance between the artificial noise, simulated by a resolvability encoder, at the output of a channel and the 'target' noise process.

As we mentioned above, the exponential decay of the information leaked to the eavesdropper in the block-length, in the wiretap model, was first noted by Csiszár [27]. A sequence of recent works [47,51–54,58] study the achievable secrecy exponents over the wiretap channel, the best of which are reported in [47,52,54]. In these works the probabilistic method is employed to show the achievability of these exponents. Specifically, an exponentially decaying upper bound on the average amount of information leaked to the eavesdropper over an ensemble of randomly constructed codes is derived to conclude that there must exist a code guaranteeing such a tiny information leakage. The optimality of those exponents is, though, not known. In **Chapter 6**, we present a method for deriving *ensemble-optimal* secrecy exponents for the randomly constructed codes. That is to say, we derive exponentially tight upper *and lower* bounds on the average information leaked to the eavesdropper (see Theorem 6.3). The exponent of our bounds matches that previously reported for the ensemble of *i.i.d. random codes*, i.e., those constructed by independently drawing the letters of each codeword according to a given distribution on channel input alphabet (independent of other codewords). Furthermore, our exponent for the ensemble of *constant-composition random codes* — those for which codewords are obtained by randomly permuting the letters of a given word (independently for each codeword) — improves upon what was previously known to be achievable using such codes (see Lemma 6.5). We also show that, due to its generality, our simple analysis method is applicable to the ensemble of randomly constructed *linear* codes, as well. Similarly to the error-correction problem, Theorem 6.8 shows that random linear codes perform exactly the same as i.i.d. random codes for secure communications.

As we discussed briefly (and we shall see more rigorously in Chapter 5), the core concept in constructing codes for secure communications is channel resolvability. In a sense, resolvability is the counterpart of the error correction. In the latter the aim is to *combat* the channel noise and make it easy for the receiver to distinguish the sent codeword from the others despite being corrupted by the channel. Whereas, to simulate a random sequence at the output of the channel, in resolvability, we *exploit* the channel noise. The problem of communications in the presence of *feedback* is well studied. It is known that feedback does not increase the capacity of a communication channel [42, Exercise 4.6], but it enables us to achieve lower error probabilities compared to systems without feedback [23]. In **Chapter 7**, we consider the problem of resolvability in the presence of feedback. In Theorem 7.1, we show that feedback does not reduce the minimum amount of randomness the encoder needs to accurately simulate a given random process at the output of the channel. However, there are instances where, by using variable-length coding and the feedback signal, a much more accurate simulation, compared to what is attainable in the systems without feedback, is feasible (see Theorems 7.2 and 7.4). More importantly, we will prove (in Lemmas 7.3 and 7.5) that both employing variable-length codes and making use of the feedback are necessary to achieve such an accurate simulation of the desired random process — even the best block code results in a lower-quality approximation.

Part I

Communications

Arıkan's Polar Coding Paradigm 2

Shannon, in his seminal work [101], characterized the fundamental limits of data transmission through a noisy channel. Given the mathematical description of the communication medium, he derived an expression for the *capacity* of the channel — the highest rate at which reliable communication is possible. A fascinating aspect of Shannon's work is his proof of the feasibility of reliable communication at rates below capacity without an explicit prescription of a communication scheme. Instead of constructing a specific communication scheme, he showed that associating each message with a codeword built by choosing its letters *randomly* and independently from the input alphabet of the channel, with high probability leads to a reliable transmission scheme. Despite its elegance from a mathematical perspective, random coding does not lead to practical implementable codes: the computation power and memory required for decoding a randomly constructed code grows exponentially fast with the block-length.[1]

In 2008, Arıkan discovered the first (to date the only) method for explicitly constructing capacity-achieving codes with low-complexity encoding *and decoding* algorithms for any binary-input memoryless symmetric-output (BMS): this method is channel called *polar coding* [4, 5].[2] In this chapter, we give an overview of polar coding in order to set up the notation and foundations for the discussions in the following two chapters.

[1] Using the conventional form of random codes, the memory required for encoding a message into a codeword also increases exponentially fast with the block-length. However, this can be alleviated by using random linear codes [42, Chapter 6].

[2] Even though in the original work of Arıkan [4, 5] polar coding was introduced as a method for achieving the capacity of BMS channels, it was soon extended to channels with larger alphabets and non-symmetric channels [56, 81, 82, 84, 87, 94, 97–99, 106].

2.1 Preliminaries

Let $W : \mathbb{F}_2 \to \mathcal{Y}$ denote a binary-input memoryless channel described by the pair of conditional probabilities $W(\cdot|x) \in \mathcal{P}(\mathcal{Y})$, $x \in \mathbb{F}_2$. We further assume W is symmetric; i.e., that there exists a permutation π on $\mathcal{Y}$ such that (i) $\pi^{-1} = \pi$ and (ii) $\forall y \in \mathcal{Y}$, $W(y|1) = W(\pi(y)|0)$. Examples of binary-input memoryless symmetric (BMS) channels are:

Binary Symmetric Channel (BSC) whose output alphabet is $\{0, 1\}$ and has transition probabilities

$$W(0|0) = W(1|1) = 1 - p \quad \text{and} \quad W(0|1) = W(1|0) = p,$$

where $p \in [0 : 1/2]$ is called the *crossover probability* of the channel. We denote a BSC with crossover probability p as $\mathsf{BSC}(p)$.

Binary Erasure Channel (BEC) with output alphabet $\{0, 1, ?\}$ and transition probabilities

$$\begin{aligned} &W(0|0) = W(1|1) = 1 - p, \\ &W(0|1) = W(1|0) = 0, \quad \text{and} \\ &W(?|0) = W(?|1) = p, \end{aligned}$$

where p is called the *erasure probability* of the channel. We denote a BEC with erasure probability p as $\mathsf{BEC}(p)$.

Binary-Input Additive White Gaussian Noise Channel (BI-AWGNC) whose output alphabet is $\mathcal{Y} = \mathbb{R}$ and has transition probabilities

$$W(y|x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{[y-(-1)^x]^2}{2\sigma^2}}$$

where $\sigma^2 > 0$ is the noise variance.

Remark. It is well-known that for estimating the input of a binary-input memoryless channel, the likelihood ratio

$$\Lambda(y) \coloneqq \frac{W(y|0)}{W(y|1)}$$

is a sufficient statistic [70, Chapter 2]. In view of Lemma A.1, we can always merge the output symbols with equal likelihood and reduce the size of the output alphabet.

The capacity of a BMS channel $W : \mathbb{F}_2 \to \mathcal{Y}$, i.e., the highest rate at which reliable communication is feasible through that channel, is given by the mutual information between its input and its output, $I(X;Y)$ when X is uniformly distributed on $\mathbb{F}_2$, that is,

$$I(W) \coloneqq \sum_{x\in\mathbb{F}_2} \frac{1}{2} \sum_{y\in\mathcal{Y}} W(y|x) \log\Big[\frac{W(y|x)}{\frac{1}{2}W(y|0)+\frac{1}{2}W(y|1)}\Big]. \tag{2.1}$$

The capacity is measured in units of bits and (for a binary-input channel) is bounded between 0 and 1.

The other measures of noisiness of a BMS channel $W : \mathbb{F}_2 \to \mathcal{Y}$ are its bit-error probability, i.e., the probability of error of the ML decision on its input, upon observing the output of a single use of the channel,

$$P_{\mathrm{e}}(W) \coloneqq \frac{1}{2} \sum_{y\in\mathcal{Y}} \min\{W(y|0), W(y|1)\} \tag{2.2}$$

that lies in $[0, 1/2]$ and the Bhattacharyya parameter

$$Z(W) \coloneqq \sum_{y\in\mathcal{Y}} \sqrt{W(y|0)W(y|1)}, \tag{2.3}$$

that takes values in $[0, 1]$.

These measures of noisiness are related via the bounds summarized in the following lemma:

Lemma 2.1. *Let $W : \mathbb{F}_2 \to \mathcal{Y}$ be a BMS channel and $I(W)$, $P_{\mathrm{e}}(W)$, and $Z(W)$ its capacity, bit-error probability, and Bhattacharyya parameter, defined in* (2.1)*,* (2.2)*, and* (2.3)*, respectively. Then,*

(i) $I(W) + Z(W) \geq 1$ and $I(W)^2 + Z(W)^2 \leq 1$ [5, Propostions 1 and 11].

(ii) $2P_{\mathrm{e}}(W) \leq Z(W) \leq 2\sqrt{P_{\mathrm{e}}(W)(1-P_{\mathrm{e}}(W))}$ [93, Lemma 4.64].

A binary-input memoryless channel W is called *extremal* if $I(W) \in \{0, 1\}$. (According to Lemma 2.1, $I(W) \in \{0, 1\}$ implies $Z(W) \in \{0, 1\}$ and $P_{\mathrm{e}}(W) \in \{0, 1/2\}$.) Achieving the capacity of an extremal channel is trivial: a capacity-1 channel can be used to communicate uncoded data bits with zero error probability (and achieving the capacity of zero-capacity channel is even easier!).[3] Unfortunately, communication systems engineers usually have to deal with *mediocre*, hence not easy-to-use channels. *Channel polarization* [5] is a method for 'forging' asymptotically extremal channels out of (infinitely many) independent uses of a mediocre channel *while preserving the total capacity.*

[3] Note that for input alphabets larger than binary, *extremality* implies the existence of trivial codes to achieve the capacity but it is not necessary. For example, a 4-ary channel obtained as the product of two binary symmetric channels with total capacity of 1 bit per channel use (not necessarily the same) has capacity strictly less than log(4) (and hence is not extremal). But we can trivially communicate through such a channel at a rate of 1 bit per channel use with zero error probability.

Having forged such channels, constructing a capacity-achieving communication scheme would be easy. As the total capacity is preserved, the fraction of noiseless (i.e., capacity-1) channels must be equal to the capacity of the original (mediocre) channel. Using these channels to transfer uncoded data bits leads to a reliable communication scheme that operates at a rate arbitrarily close to the capacity.

2.2 Channel Polarization

Let us start with two copies of our mediocre BMS channel $W : \mathbb{F}_2 \to \mathcal{Y}$ and combine their inputs as shown in Figure 2.1.

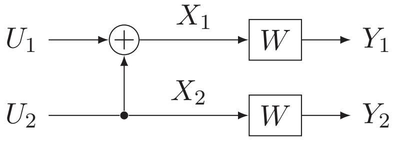


Figure 2.1: Singe-step Polar Transform

If (U_1, U_2) are independent Bernoulli(1/2) random variables, so are (X_1, X_2), hence we have

$$2I(W) = I(X_1, X_2; Y_1, Y_2) = I(U_1, U_2; Y_1, Y_2) \tag{2.4}$$

where the second equality follows as

$$(U_1, U_2) \mapsto (X_1, X_2) = (U_1 \oplus U_2, U_2) \tag{2.5}$$

is an invertible transform. Expanding $I(U_1, U_2; Y_1, Y_2)$ by using the chain rule of mutual information and the independence of U_1 and U_2, we get

$$\begin{aligned} 2I(W) &= I(U_1, U_2; Y_1, Y_2) &(2.6)\\ &= I(U_1; Y_1, Y_2) + I(U_2; Y_1, Y_2 | U_1) &(2.7)\\ &= I(U_1; Y_1, Y_2) + I(U_2; Y_1, Y_2, U_1). &(2.8) \end{aligned}$$

Both terms on the right-hand side of the above are mutual information between a uniformly distributed binary random variable (U_1 or U_2) and some other collection of random objects (Y_1, Y_2 or Y_1, Y_2, U_1, respectively). Indeed, they define the mutual information developed across the effective channels seen between U_1 and Y_1, Y_2, and U_2 and Y_1, Y_2, U_1, respectively. Moreover, the second term in (2.8) is lower-bounded as

$$\begin{aligned} I(U_2; Y_1, Y_2, U_1) &= I(U_2; Y_2) + I(U_2; Y_1, U_1 | Y_2) &(2.9)\\ &\geq I(U_2; Y_2) &(2.10)\\ &= I(X_2; Y_2) = I(W). &(2.11) \end{aligned}$$

The above inequality can be shown to be strict, i.e., $I(U_2; Y_1, U_1|Y_2) > 0$, unless $I(W) \in \{0, 1\}$ [5, Appendix C]. Consequently,

$$I(U_1; Y_1, Y_2) \leq I(W) \leq I(U_2; Y_1, Y_2, U_1) \tag{2.12}$$

(with equality iff $I(W) \in \{0, 1\}$).

By the virtue of (2.8) and (2.12), we can define two *synthetic* binary-input channels; $W^- : \mathbb{F}_2 \to \mathcal{Y}^2$ with transition probabilities

$$W^-(y_1, y_2|u_1) = \sum_{u_2 \in \mathbb{F}_2} \frac{1}{2} W(y_1|u_1 \oplus u_2) W(y_2|u_2), \tag{2.13}$$

and $W^+ : \mathbb{F}_2 \to \mathcal{Y}^2 \times \mathbb{F}_2$ with transition probabilities[4]

$$W^+(y_1, y_2, u_1|u_2) = \frac{1}{2} W(y_1|u_1 \oplus u_2) W(y_2|u_2). \tag{2.14}$$

Moreover

$$I(W^-) + I(W^+) = 2I(W), \tag{2.15}$$

and

$$I(W^-) \leq I(W) \leq I(W^+). \tag{2.16}$$

(Both channels W^- and W^+ can be shown to be symmetric [5, Proposition 13] hence, the mutual information $I(U_1; Y_1, Y_2)$ and $I(U_2; Y_1, Y_2, U_1)$ are indeed equal to the capacity of the synthetic channels W^- and W^+, respectively.)

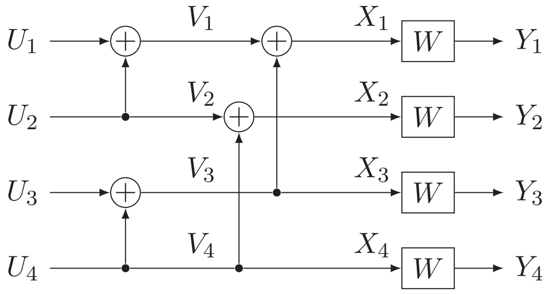
In short, out of two independent copies of a mediocre BMS channel W, we synthesize two *unequal* BMS channels: W^-, the channel seen from U_1 to Y_1, Y_2 treating U_2 as an internal noise component, which is *worse* than W; and W^+, the channel seen from U_2 to Y_1, Y_2, *and* U_1, which is *better* than W. Moreover, the total capacity is preserved during this transformation. This procedure is called *polar transform*.

As applying the polar transform on two independent copies of a BMS channel W results in a pair of BMS channels, we can repeatedly apply the transform to (independent copies of) each of these channels. For example, applying the polar transform once more, synthesizes four channels W^{--}, W^{-+}, W^{+-}, and W^{++} from two independent copies of W^- and two independent copies of W^+ (which are themselves obtained from four independent copies of W) as shown in Figure 2.2a. These channels have input and output as follows:

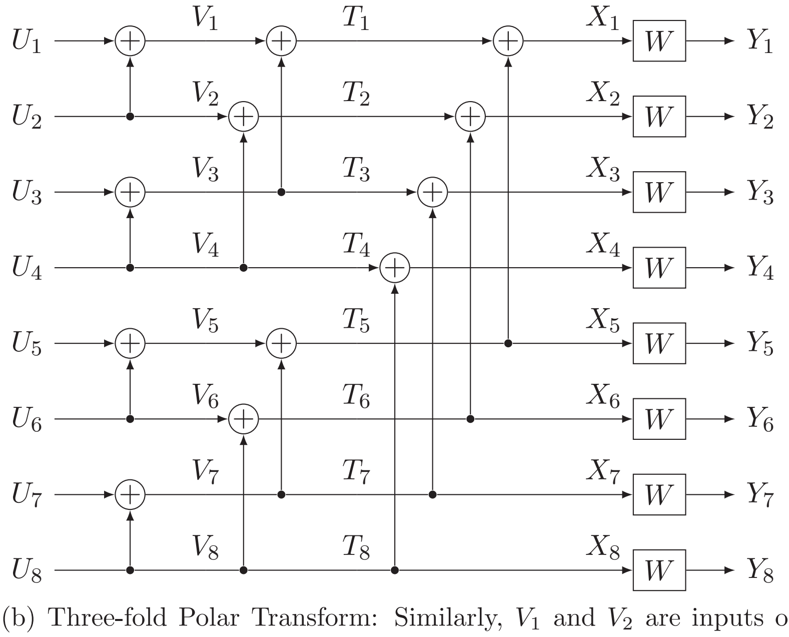
$$\begin{aligned} &W^{--} : \mathbb{F}_2 \to \mathcal{Y}^4 && U_1 \mapsto Y^4 \\ &W^{-+} : \mathbb{F}_2 \to \mathcal{Y}^4 \times \mathbb{F}_2 && U_2 \mapsto (Y^4, U_1) \\ &W^{+-} : \mathbb{F}_2 \to \mathcal{Y}^4 \times \mathbb{F}_2^2 && U_3 \mapsto (Y^4, V^2) \equiv (Y^4, U^2) \\ &W^{++} : \mathbb{F}_2 \to \mathcal{Y}^4 \times \mathbb{F}_2^3 && U_4 \mapsto (Y^4, V^2, U_3) \equiv (Y^4, U^3) \end{aligned} \tag{2.17}$$

[4] The reader might worry why we define a channel whose output is U_1; because U_1 is a value set at the encoder and, in principle, unknown to the decoder. As we will see in § 2.3, we impose a particular decoding order at the receiver and ask it to decode U_1 before decoding U_2. Therefore we can legitimately assume U_1 is available to the decoder at the time of decoding U_2.

We can again duplicate the structure of Figure 2.2a and obtain eight synthetic channels $W^{---}, \ldots, W^{+++}$ by applying the single-step polar transform to each of the four synthetic channels (cf. Figure 2.2b).



(a) Two-fold Polar Transform: Here V_1 and V_2 are inputs of two independent copies of W^- and V_3 and V_4 are inputs of independent copies of W^+.



(b) Three-fold Polar Transform: Similarly, V_1 and V_2 are inputs of two independnet copies of W^{--}, V_3 and V_4 are inputs of independnet copies of W^{-+}, V_5 and V_6 are inputs of independnet copies of W^{+-}, and V_7 and V_8 are inputs of independnet copies of W^{++}. Moreover, $T_1, \ldots, T_4$ are inputs to four independent copies of W^- and $T_5, \ldots, T_8$ are inputs to independent copies of W^+.

Figure 2.2: Two- and Three-fold Polar Transform

In general, the result of the m-fold application of the polar transform to $n = 2^m$ independent copies of a BMS channel W is a set of 2^m synthetic channels indexed by sign sequences of length m, W^{s^m}, $\mathsf{s}^m \in \{-,+\}^m$. The synthetic channel $W^{\mathsf{s}^m} : \mathbb{F}_2 \to \mathcal{Y}^n \times \mathbb{F}_2^{i-1}$, $\mathsf{s}^m \in \{-,+\}^m$ has input U_i, the i^{th} element of the synthetic channels input vector $U^n \in \mathbb{F}_2^n$; and output Y^n, U^{i-1}

(treating U_{i+1}^n as internal noise), where $i \in \{1, 2, \dots, n\}$ equals

$$i = 1 + \sum_{j=1}^{m} 2^{m-j} \cdot \mathbb{1}\{\mathsf{s}_j = +\}. \tag{2.18}$$

With a slight abuse of notation we can write

$$i = (\overline{\mathsf{s}_1 \mathsf{s}_2 \cdots \mathsf{s}_m})_2 + 1. \tag{2.19}$$

The physical channel input vector X^n is related to U^n via

$$X^n = G_n U^n \qquad \text{where} \qquad G_n = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}^{\otimes m}. \tag{2.20}$$

In the above $A^{\otimes m}$ denotes the m^{th} Kronecker power of A. Consequently, the transition probabilities of W^{s^m} are

$$W^{\mathsf{s}^m}(y^n, u^{i-1} | u_i) = \frac{1}{2^{n-1}} \sum_{u_{i+1}^n \in \mathbb{F}_2^{n-i}} W^n\big(y^n \big| G_n u^n\big) \tag{2.21}$$

(with $i = (\overline{\mathsf{s}_1 \mathsf{s}_2 \cdots \mathsf{s}_m})_2 + 1$).

Also, as the single-step polar transform is capacity-preserving,

$$2^m I(W) = \sum_{\mathsf{s}^m \in \{-,+\}^m} I(W^{\mathsf{s}^m}). \tag{2.22}$$

More importantly, as intuitively expected by looking at the fixed points of the single-step transform (which are the extremal channels), as $m \to \infty$ almost all synthetic channels become extremal. Since the total capacity is preserved by the polar transform, the fraction of capacity-1 channels will be equal to the capacity of the underlying channel W.

Theorem 2.2 ([5, Theorem 1])**.** *The synthetic channels W^{s^m}, $\mathsf{s}^m \in \{-,+\}^m$, obtained from m-fold application of the polar transform* polarize. *That is, $\forall a, b$ such that $0 < a < b < 1$,*

$$\lim_{m \to \infty} \frac{1}{2^m} \big|\big\{\mathsf{s}^m \in \{-,+\}^m \colon I\big(W^{\mathsf{s}^m}\big) \le a\big\}\big| = 1 - I(W), \tag{2.23}$$

$$\lim_{m \to \infty} \frac{1}{2^m} \big|\big\{\mathsf{s}^m \in \{-,+\}^m \colon I\big(W^{\mathsf{s}^m}\big) \in (a, b)\big\}\big| = 0, \tag{2.24}$$

$$\lim_{m \to \infty} \frac{1}{2^m} \big|\big\{\mathsf{s}^m \in \{-,+\}^m \colon I\big(W^{\mathsf{s}^m}\big) \ge b\big\}\big| = I(W). \tag{2.25}$$

Arıkan in [5] proved Theorem 2.2 using a martingale argument. In Appendix 2.A, we review a simpler proof of the theorem given in [3].

2.3 Polar Coding

As Theorem 2.2 suggests, for any rate $R < I(W)$, if m is large enough, there exists $\lceil 2^m R \rceil$ 'almost noiseless' synthetic channels, i.e., synthetic channels with capacity arbitrarily close to 1. Using these channels for communicating uncoded data bits, we obtain a sequence of capacity-achieving codes.

More precisely, given a rate R and a block-length $n = 2^m$, an $(n, \lceil nR \rceil)$ polar code[5] is defined as follows: Let $\mathcal{A} \subseteq \{-,+\}^m$ be the indices of $k = \lceil nR \rceil$ best synthetic channels and $\mathcal{I}$ the integer indices corresponding to the sign sequences in $\mathcal{A}$, that is $\mathcal{I} \coloneqq \{(\overline{\mathsf{s}_1\mathsf{s}_2\cdots\mathsf{s}_m})_2 + 1 \colon \mathsf{s}^m \in \mathcal{A}\}$. The encoder sets $u_{\mathcal{I}}$ to k *information* bits (to be communicated to the receiver) and $u_{\mathcal{F}}$, where $\mathcal{F} = \{1, 2, \cdots, n\} \setminus \mathcal{I}$, to some known-to-receiver *frozen* bits.[6] The vector u^n is then encoded to the codeword x^n through (2.20). It can be easily seen that encoding requires $\Theta(n \log n)$ binary additions [5, Section VII]. The codeword x^n is then transmitted via n independent uses of the channel to the receiver. Given $\mathcal{I}$ and $u_{\mathcal{F}}$, the encoding map $u_{\mathcal{I}} \mapsto x^n$ is

$$x^n = G_n[\mathcal{I}]u_{\mathcal{I}} + G_n[\mathcal{F}]u_{\mathcal{F}}, \tag{2.26}$$

(where $G_n[\mathcal{I}]$ denotes the $n \times k$ sub-matrix of G_n obtained by keeping only the columns with index in $\mathcal{I}$; similarly $G_n[\mathcal{F}]$ is the $n \times (n-k)$ sub-matrix of G_n obtained by keeping only the columns with index in $\mathcal{F}$).

The receiver observes the channel output vector y^n and estimates the elements of the $u_{\mathcal{I}}$ *successively* as follows: Suppose the information indices are ordered as $\mathcal{I} = \{i_1, i_2, \ldots, i_k\}$ (where $i_j < i_{j+1}$). Furthermore, for any s^m and $i = (\overline{\mathsf{s}_1\mathsf{s}_2\ldots\mathsf{s}_m})_2 + 1$, let $W_n^{(i)}$ be a synonym for W^{s^m} (to keep the notation simple). Having the channel output, the receiver has all the required information to decode the input of the synthetic channel $W_n^{(i_1)}$ as $\hat{u}_{i_1}$, since, in particular, u^{i_1-1} is a part of the known sub-vector $u_{\mathcal{F}}$. Hopefully, this estimation is correct and the decoder can subsequently proceed to index i_2, as the information required for decoding the input of $W_n^{(i_2)}$ is now available. As detailed in Algorithm 1, this process — known formally as *successive-cancellation* (SC) decoding — is continued until all the information bits have been estimated.

2.3.1 Performance of Polar Codes under SC Decoding

To show that polar codes are capacity-achieving, we analyze their block-error probability under successive-cancellation decoding and prove that it vanishes as the block-length increases, when the code rate R is below $I(W)$. To this end, there are two caveats to be addressed: First, we define the synthetic channel

[5] Following the convention in coding theory, we denote an affine code of block-length n and dimension k (hence, rate k/n) as an (n, k) code.

[6] In [5, Corollary 1] it has been shown that, as long as the channel W is symmetric, the choice of frozen bits does not affect the performance of the code. Hence, in order to make the code linear, $u_{\mathcal{F}}$ is typically chosen to be the all-zero sequence.

Algorithm 1: Successive-Cancellation (SC) Decoding [5].

```
1 for $i = 1, 2, \dots, n$ do
2     if $i \notin \mathcal{I}$ then                                  // frozen bits
3         $\hat{u}_i \leftarrow u_i$;
4     else                                                        // information bits
5         $\hat{u}_i \leftarrow \arg\max_{u_i \in \mathbb{F}_2} W_n^{(i)}(y^n, \hat{u}^{i-1} | u_i)$;
6 return $\hat{u}^n$ ;
```

W^{s^m} (or equivalently $W_n^{(i)}$, $i = (\overline{\mathsf{s}_1 \mathsf{s}_2 \dots \mathsf{s}_m})_2 + 1$) as the channel whose input is U_i and its output is Y^n and U^{i-1}. In successive-cancellation decoding we use an estimate of U^{i-1} (which could be wrong) instead, to decode U_i. Therefore, due to error propagation, it is not a priori clear how the performance of the decoder can be related to the quality of the synthetic channels. Second, Theorem 2.2 guarantees that, by taking m large enough, we can ensure that individual channels used for the transmission of information bits have a capacity as high as desired (hence, the probability of erroneous transmission through each channel is as small as desired). However, since the number of these channels increases linearly with n, this does not necessarily mean that the probability of the intersection of all "error-free transmission" events is close to 1 (thus proving that the block-error probability under successive-cancellation decoding is small).

To address the first point, let us consider a *genie-aided* version of the decoder by replacing the estimation of line 5 of Algorithm 1 by

$$\hat{u}_i = \arg\max_{u_i \in \mathbb{F}_2} W_n^{(i)}(y^n, u^{i-1} | u_i). \tag{2.27}$$

(Note that to implement (2.27) the decoder needs to ask for the correct value of the previous bits u^{i-1} from a genie.)

Upon observing a particular channel output sequence y^n, if the genie-aided SC decoder estimates all the information bits $u_{\mathcal{I}}$ correctly, then the plain SC decoder of Algorithm 1 does the same. Conversely, once the genie-aided decoder makes the *first* mistake, the same incorrect decision is taken by the plain SC decoder; and both decoders commit a block-error event — albeit from that point on the decoders can behave completely differently and their final outputs are possibly different. Consequently, we conclude that the *block-error* events of both decoders are the same. The genie-aided decoder makes an error if and only if decoding of the input any of the synthetic channels is erroneous.

Therefore, if we denote by $\mathcal{E}(\mathcal{A})$ the block-error event of the polar code defined by the set of information indices $\mathcal{A} \subseteq \{-,+\}^m$ under successive-cancellation decoding and, for $\forall \mathsf{s}^m \in \{-,+\}^m$, by $\mathcal{E}^{\mathsf{s}^m}$ the bit-error event of the synthetic channel W^{s^m} (in the genie-aided decoder), we have

$$\mathcal{E}(\mathcal{A}) \coloneqq \bigcup_{\mathsf{s}^m \in \mathcal{A}} \mathcal{E}^{\mathsf{s}^m}. \tag{2.28}$$

Lemma 2.3. *Let $P_{\mathrm{e}}(\mathcal{A})$ denote the block-error probability of a polar code defined defined by the set of information indices $\mathcal{A} \subseteq \{-,+\}^m$, under successive-cancellation decoding. Then:*

$$\max_{\mathsf{s}^m \in \mathcal{A}} P_{\mathrm{e}}(W^{\mathsf{s}^m}) \leq P_{\mathrm{e}}(\mathcal{A}) \leq \sum_{\mathsf{s}^m \in \mathcal{A}} P_{\mathrm{e}}(W^{\mathsf{s}^m}), \tag{2.29}$$

where $P_{\mathrm{e}}(W^{\mathsf{s}^m}) = \Pr(\mathcal{E}^{\mathsf{s}^m})$ is the bit-error probability of the synthetic channel W^{s^m}.

Proof. The result follows trivially from (2.28). The upper bound follows from the union bound. The lower bound follows because $\mathcal{E}^{\mathsf{s}^m} \subseteq \mathcal{E}$ for $\forall \mathsf{s}^m \in \mathcal{A}$. □

The last step in showing that polar codes are capacity-achieving is to show that the bit-error probability of individual synthetic channels decay sufficiently fast so that the upper bound of (2.29) vanishes as the block-length increases. This is guaranteed by the following theorem, from [6], which we present without proof:

Theorem 2.4 ([6, Theorem 1]). *For any BMS channel W, any rate $R < I(W)$, and $\forall \beta < 1/2$, there exists a sequence of information indices ($\mathcal{A}_m \subseteq \{-,+\}$, $m \in \mathbb{N}$) such that*

(i) $|\mathcal{A}_m| \geq \lceil 2^m R \rceil$

(ii) For all $\mathsf{s}^m \in \mathcal{A}_m$, $P_{\mathrm{e}}(W^{\mathsf{s}^m}) \in O\big(2^{-2^{\beta m}}\big)$ (where $P_{\mathrm{e}}(W^{\mathsf{s}^m}) = \Pr(\mathcal{E}^{\mathsf{s}^m})$ is the bit-error probability of the synthetic channel W^{s^m}, obtained by the m-fold application of the polar transform to W).

Corollary 2.5. *For any BMS channel W, any rate $R < I(W)$, and $\forall \beta < 1/2$, there exists a sequence of polar codes of block-length $n = 2^m$ and rate R such that their block-error probability satisfies*

$$P_{\mathrm{e}}(\mathcal{A}_m) \in O\big(2^{-n^{\beta}}\big). \tag{2.30}$$

Consequently, polar codes are capacity-achieving.

Remark 1. The bound of Equation (2.30) is essentially exponentially tight: From [6, Theorem 3] it follows that the fraction of synthetic channels whose bit-error probability decay faster than $2^{-2^{m\beta}}$ for $\beta > 1/2$ vanishes.[7] Thus, by virtue of the lower bound of (2.29), we conclude that, unless the code's rate vanishes, the block-error probability of polar codes under successive-cancellation decoding cannot decay faster than $2^{-\sqrt{n}}$.

[7] In view of (ii) in Lemma 2.1, if $P_{\mathrm{e}}(W^{\mathsf{s}^m}) \leq 2^{-2^{m\beta}}$ for some $\beta > \frac{1}{2}$, then for large enough m, $Z(W^{\mathsf{s}^m}) \leq 2^{-2^{m\beta'}}$ for any $\beta' \in (\frac{1}{2}, \beta)$. According to [6, Theorem 3], the fraction of synthetic channels for which $Z(W^{\mathsf{s}^m}) \leq 2^{-2^{m\beta'}}$ for $\beta' > 1/2$ vanishes as $m \to \infty$.

Remark 2. Successive-cancellation decoding is obviously sub-optimal: To decide on the value of the information bit u_i, $i \in \mathcal{I}$, the decoder pretends that all the following bits u_{i+1}^n are i.i.d. coin-flips (the exact definition of the corresponding synthetic channel), whereas the subsequent frozen bits put parity constraints on the value of the current bit. These constraints are not taken into account by that marginalization. In return, as we shall see in § 2.3.2, the decoding algorithm is efficient. Despite this sub-optimality, SC decoding has a good asymptotic performance (exponentially small block-error probability).

Remark 3. In the beginning of this section, we discussed a *rate-based* construction of polar codes, in other words, the method based on picking sufficient information indices to have as many codewords as desired in the code. By the virtue of the upper bound of (2.29), we can instead fix a target block-error probability and construct the code by increasing the rate, as long as the upper bound is below that target value.

2.3.2 Complexity of Successive-Cancellation Decoding

The computational task of the SC decoder is to calculate the pairs of likelihoods $W_n^{(i)}(y^n, \hat{u}^{i-1}|u_i)$, $u_i \in \mathbb{F}_2$ needed for the decisions in line 5 of Algorithm 1. Since the decisions are binary, it is sufficient to compute the *decision log-likelihood ratios (LLRs)*,

$$\lambda_m^{(i)} := \log\Big[\frac{W_n^{(i)}(y^n, \hat{u}^{i-1}|0)}{W_n^{(i)}(y^n, \hat{u}^{i-1}|1)}\Big]. \tag{2.31}$$

Due to the recursive nature of polar transform, it follows straightforwardly (see [5, Section VII] and [68]) that the decision LLRs (2.31) can be computed via the recursions,

$$\lambda_s^{(i)} = \begin{cases} f_-\big(\lambda_{s-1}^{(i)}, \lambda_{s-1}^{(i+2^{m-s})}\big) & \text{if } \lfloor (i-1)/2^{m-s} \rfloor \text{ is even} \\ f_+\big(\lambda_{s-1}^{(i-2^{m-s})}, \lambda_{s-1}^{(i)}, v_s^{(i-2^{m-s})}\big) & \text{if } \lfloor (i-1)/2^{m-s} \rfloor \text{ is odd} \end{cases} \tag{2.32}$$

for $i = 1, 2, \ldots, n$ and $s = m, m-1, \ldots, 1$ where $f_- \colon \mathbb{R}^2 \to \mathbb{R}$ and $f_+ \colon \mathbb{R}^2 \times \mathbb{F}_2 \to \mathbb{R}$ are defined as

$$f_-(a, b) := \log\Big[\frac{\exp(a+b)+1}{\exp(a)+\exp(b)}\Big], \tag{2.33a}$$

$$f_+(a, b, v) := (-1)^v a + b, \tag{2.33b}$$

respectively. The recursions terminate at $s = 0$ where

$$\lambda_0^{(i)} := \log\Big[\frac{W(y_i|0)}{W(y_i|1)}\Big], \qquad \forall i = 1, 2, \ldots, n, \tag{2.34}$$

are *channel LLRs*. The *partial sums* $v_s^{(i)}$, $s = m-1, \ldots, 1$, $i = 1, 2, \ldots, n$ are computed starting from $v_m^{(i)} := \hat{u}_i$ for $i = 1, 2, \ldots, n$, and setting

$$v_{s-1}^{(i)} = \begin{cases} v_s^{(i)} \oplus v_s^{(i+2^{m-s})} & \text{if } \lfloor (i-1)/2^{m-s} \rfloor \text{ is even,} \\ v_s^{(i)} & \text{if } \lfloor (i-1)/2^{m-s} \rfloor \text{ is odd.} \end{cases} \tag{2.35}$$

Therefore, the entire set of $m \times n$ LLRs $\lambda_s^{(i)}$ $s = 1, 2, \ldots, m$, $i = 1, 2, \ldots, n$ can be computed using $O(n \log n)$ *updates*: from each pair of LLRs at *stage* $s - 1$ a pair of LLRs at stage s is calculated using the update rules of (2.32) (see Figure 2.3). Additionally the decoder must keep track of $(m - 1) \times n$ partial sums $v_s^{(i)}$, $s = 1, 2, \ldots, m - 1$, $i = 1, 2, \ldots, n$ and update them after decoding each bit $\hat{u}_i$.

In terms of memory requirements, an elementary implementation of the decoder would require $O(n \log n)$ memory elements to store the intermediate LLRs (cf. Figure 2.3) and partial sums. This can further be improved to $O(n)$ elements by noting that, at the time of decoding bit i, only the intermediate LLR values corresponding to the *tree* rooted at $\lambda_m^{(i)}$ in the computational graph of the decoder needs to be stored (as opposed to entire $m \times n$ nodes of the graph). In other words, at any time instant, at each stage $s = 1, 2, \ldots, m$, only 2^{m-s} intermediate LLR values are required for computing the desired decision LLR. A similar observation holds for the partial sums, which eventually leads to a *space-efficient* implementation of the decoder [108, Section III].

Note also that the decoder, in principle, has to compute all the decision LLR values $\lambda_m^{(i)}$, $i = 1, 2, \ldots, n$, to make sure that all intermediate LLRs are computed and stored for the next stages — despite the fact that some indices correspond to frozen bits and the corresponding LLRs are not ultimately needed for a decision. In practice, depending on the structure of frozen and information bits, some computation time and power can be saved by skipping the LLR calculations for some clusters of frozen bits [2].

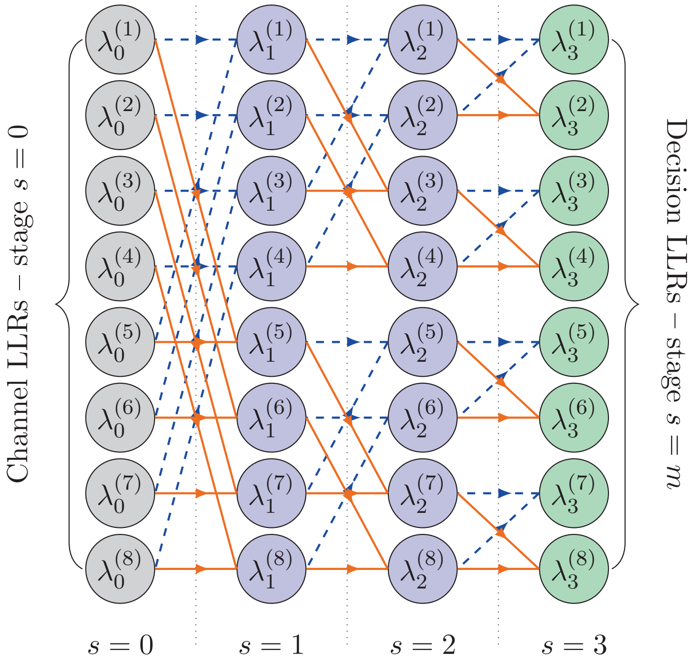


Figure 2.3: The butterfly computational structure of the SC decoder for $m = 3$; blue dashed and orange solid arrows show f_- and f_+ updates respectively.

Remark. It can easily be checked that

$$f_-(a,b) = \operatorname{sign}(a)\operatorname{sign}(b)\Big\{\min\{|a|,|b|\} + \log\Big[\frac{1+\exp(-(a+b))}{1+\exp(-|a-b|)}\Big]\Big\} \tag{2.36}$$

$$\approx \operatorname{sign}(a)\operatorname{sign}(b)\min\{|a|,|b|\} =: \tilde{f}_-(a,b). \tag{2.37}$$

In fact, it can be verified that $|f_-(a,b) - \tilde{f}_-(a,b)| \leq \log(2)$ [11, Lemma 1.8]. $\tilde{f}_-$ is a 'hardware-friendly' approximation of f_- as it involves only the easy-to-implement $\min\{\cdot,\cdot\}$ operation (compared to f_- which involves exponentiations and logarithms). For a hardware implementation of the SC decoder the update rule f_- is replaced by $\tilde{f}_-$. Given f_+, such an approximation is called the *min-sum approximation* of the decoder.

2.4 Construction of Polar Codes

In § 2.3 we saw that a polar code of block-length $n = 2^m$ and rate R for communication over the BMS channel W is constructed by choosing the 'best' $\lceil nR \rceil$ synthetic channels among 2^m channels W^{s^m}, $\mathsf{s}^m \in \{-,+\}^m$ for communicating information bits. Theorem 2.2 shows that for any rate $R < I(W)$ there exist $\lceil nR \rceil$ 'good' synthetic channels (provided that m is large enough); and Theorem 2.4 and its corollary show the bit-error probability of these channels decay like 2^{-n^β} (for any $\beta < 1/2$). But they do not tell us which sequences $\mathsf{s}^m \in \{-,+\}^m$ index those good channels. In principle, due to the recursive construction of the synthetic channels, we can compute the transition probabilities of W^{s^m} (for any $\mathsf{s}^m \in \{-,+\}^m$), hence fully characterize this channel and, accordingly, rank the synthetic channels. However, as the cardinally of the output alphabet of the these channels grows exponentially with $n = 2^m$, the computations will soon become intractable. Therefore, it is important to find computationally efficient methods for constructing polar codes.

Arıkan observed that when W is a $\mathsf{BEC}(p)$ both W^- and W^+ are *equivalent to* binary erasure channels. More specifically, despite the growth of their output alphabet, by merging the symbols with the same likelihood ratio (see Lemma A.1), both channels W^- and W^+ reduce to binary erasure channels. Moreover, W^- 'erases' if either independent copies of W erase , whereas W^+ erases if both copies of W erase. Therefore, W^- is a $\mathsf{BEC}(2p-p^2)$ and W^+ is a $\mathsf{BEC}(p^2)$ [5, Proposition 6]. Consequently, to construct a polar code for communication over a binary erasure channel, we can compute the erasure probabilities of each of 2^m binary erasure channels, denoted by Z^{s^m}, $\mathsf{s}^m \in \{-,+\}^m$, via the recursions

$$Z^{\mathsf{s}^m} = \begin{cases} 2Z^{\mathsf{s}^{m-1}} - (Z^{\mathsf{s}^{m-1}})^2 & \text{if } \mathsf{s}_m = - \\ (Z^{\mathsf{s}^{m-1}})^2 & \text{if } \mathsf{s}_m = +. \end{cases} \tag{2.38}$$

The recursions end at $Z^{\emptyset} = p$. Note that the entire set of $n = 2^m$ erasure probabilities can be computed in $O(n)$ operations by using $O(\log n)$ memory elements to store intermediate calculation results and n memory elements to store the final values (see Figure 2.4).

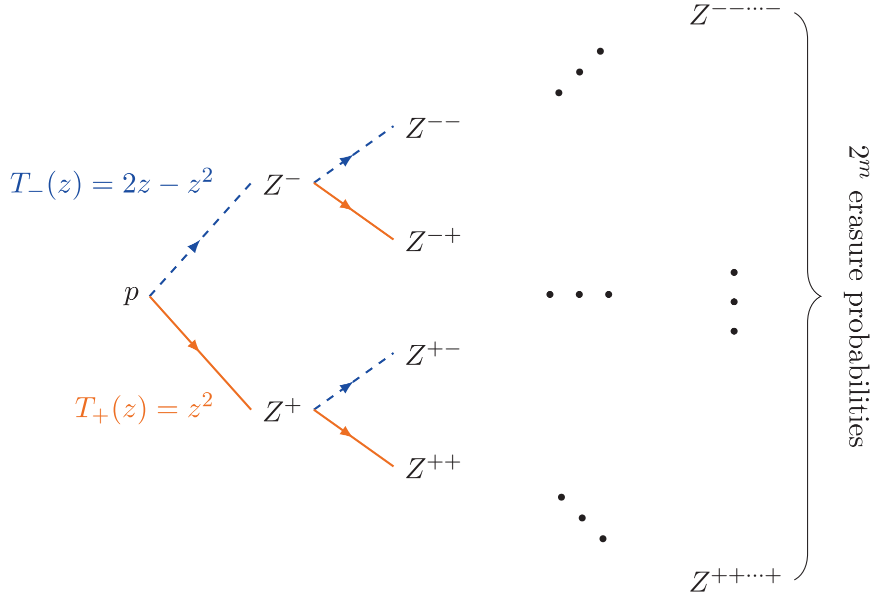


Figure 2.4: The entire set of $n = 2^m$ erasure probabilities of the synthetic BECs, obtained after m-fold application of the polar transform to a BEC(p), can be computed in $O(n)$ operations.

The set of binary erasure channels is the only known class of BMS channels that can be described with a finite number of parameters and is stable under the polar transform. Starting with any BMS, other than a BEC, the cardinality of the output alphabet of the synthetic channels, obtained after the m-fold application of Arıkan's polar transform (after merging symbols with equal likelihood ratio), grows exponentially in $n = 2^m$. Therefore, it is computationally infeasible to precisely keep track of the transition probabilities of the synthetic channels, when the underlying channel is not a BEC.

To rank the synthetic channels obtained from the repeated application of the polar transform to an arbitrary BMS channel W, Arıkan proposes a Monte Carlo estimation method, in [5, Section IX], as follows: Suppose we want to rank the synthetic channels based on their capacity.[8] One way to *estimate* the capacity of a BMS channel $V : \mathbb{F}_2 \to \mathcal{Y}$ is to note that

$$I(V) = 1 - \sum_{y\in\mathcal{Y}} V(y|0)\log\Big[1 + \frac{V(y|1)}{V(y|0)}\Big] \tag{2.39}$$

$$= 1 - \mathbb{E}_{Y\sim V(y|0)}\Big[\log\Big[1 + \frac{V(Y|1)}{V(Y|0)}\Big]\Big]. \tag{2.40}$$

[8] We can use, instead of capacity, any other measure of quality, for example the Bhattacharyya parameter that is used in [5, Section IX] or the bit-error probability.

If we transmit 0 through the channel, the output Y will have distribution $V(y|0)$. Hence, due to the law of large numbers, if we simulate the response of the channel to input 0 in independent trials and generate the samples $y_1, y_2, \dots, y_t$ independently from $V(y|0)$, the empirical average

$$\frac{1}{t}\sum_{i=1}^{t} \log\Big[1 + \frac{V(y_i|1)}{V(y_i|0)}\Big], \tag{2.41}$$

will converge to the expectation in (2.40) as $t \to \infty$. Consequently, we can estimate the capacity by simulating the channel sufficient times (and by using the empirical average instead of the expectation in (2.40)).

Suppose we run the successive-cancellation decoder for a polar code of block-length $n = 2^m$, defined by the set of information indices $\mathcal{A} = \emptyset$ and an arbitrary values for frozen bits, fed with the LLR values corresponding to independent uses of the channel W. Recall that, as we discussed in § 2.3.2, the decoder computes the decision LLRs

$$\lambda_n^{(i)} = \log\Big[\frac{W_n^{(i)}(y^n, \hat{u}^{i-1}|0)}{W_n^{(i)}(y^n, \hat{u}^{i-1}|1)}\Big], \qquad i = 1, 2, \dots, n, \tag{2.42}$$

in $O(n \log n)$ time. Since all indices $i \in \{1, 2, \dots, n\}$ are frozen, it, indeed, computes

$$\lambda_n^{(i)} = \log\Big[\frac{W_n^{(i)}(y^n, u^{i-1}|0)}{W_n^{(i)}(y^n, u^{i-1}|1)}\Big], \qquad i = 1, 2, \dots, n, \tag{2.43}$$

that are the LLRs of the outputs of the synthetic channel $W_n^{(i)}$, $i = 1, 2, \dots, n$. Running this experiment t times independently (for some large t), we can estimate the capacities of all synthetic channels $I(W_n^{(i)})$, $i = 1, 2, \dots, n$ using independent samples of their output LLRs.

An alternative method for constructing polar codes is to *approximate* the capacity (or any other measure of quality) of synthetic channels W^{s^m} as described in [109]: The key idea is that polar transform and channel degradation commute (see Theorem A.2 in Appendix A). If $V \preceq_{\mathrm{d}} W$, in other words, V is degraded with respect to W, (respectively $V \succeq_{\mathrm{d}} W$, i.e., V is upgraded with respect to W), then $V^{\mathsf{s}^m} \preceq_{\mathrm{d}} W^{\mathsf{s}^m}$ (resp. $V^{\mathsf{s}^m} \succeq_{\mathrm{d}} W^{\mathsf{s}^m}$) for any sequence $\mathsf{s}^m \in \{-,+\}^m$.

Consider the procedure described in Algorithm 2 where $V = \texttt{merge}(W, \kappa)$ is a BMS channel that is (i) degraded with respect to W, i.e., $V \preceq_{\mathrm{d}} W$, and (ii) has at most κ output symbols.[9] It follows straightforwardly that for $\forall l \in \{1, 2, \dots, m\}$, $V_l \preceq_{\mathrm{d}} W^{\mathsf{s}^l}$. Consequently, in particular, $I(V_m) \le I(W^{\mathsf{s}^m})$. Now, if we replace merge with a procedure that returns an *upgraded* version of W whose output alphabet has cardinality at most κ, and we run the algorithm

[9] Any reasonable merging function will return the channel W itself if its output alphabet has cardinality less than or equal to κ.

again, we obtain a second sequence of channels, call them $V_1', V_2', \dots, V_m'$, that satisfy $\forall l \in \{1, 2, \dots, m\}$, $V_l' \succeq_{\mathrm{d}} W^{\mathsf{s}^l}$. Thus, in particular,

$$I(V_m) \leq I(W^{\mathsf{s}^m}) \leq I(V_m'). \tag{2.44}$$

More importantly, to compute $I(V_m)$ and $I(V_m')$, we need to keep track of BMS channels whose output cardinality is at most κ (as opposed to the sequence of channels $W^{\mathsf{s}_1}, W^{\mathsf{s}_1\mathsf{s}_2}, \dots W^{\mathsf{s}^m}$ whose output alphabet grows doubly exponentially in m).

Algorithm 2: Algorithm to Approximate the Synthetic Channels [109].

```
Input: $W$ a BMS channel, $\mathsf{s}^m \in \{-,+\}^m$, $\kappa \in \mathbb{N}$
Output: A sequence of BMS channels $V_1, V_2, \dots, V_m$
$V_0 \leftarrow W$;
for $s = 1$ to $m$ do
    $V_l \leftarrow$ merge($V_{l-1}^{\mathsf{s}_l}, \kappa$);
return $V_1, V_2, \dots, V_m$
```

By designing a pair of good merging functions and an appropriate choice of κ, we can ensure that the upper and lower bounds of (2.44) are close enough, hence $I(W^{\mathsf{s}^m})$ is well-approximated by $I(V_m)$. In [88, 109] different merging methods are proposed and it is shown that by letting κ to scale as m^2 the approximation error vanishes as m grows large.

2.5 Summary

As we have seen in this chapter, *polar coding* is a computationally efficient method for communicating data at high rates reliably via noisy channels. A successive-cancellation decoder, in particular, is attractive from an implementation perspective, due to its very well-structured nature. Relatively soon after the publication of Arıkan's original work [5], several hardware architectures [40, 67, 68, 80, 86, 90, 95, 124] and simplifications of the original algorithm (to improve the decoding throughput) [2, 96, 123, 125] were proposed in the literature.

In addition to having low-complexity encoding *and decoding* algorithms, an important characteristic of polar codes is that we can compute bounds on their block-error probability (see Lemma 2.3) by using computationally efficient methods without the need for simulating the code. Given any BMS channel W, a block length $n = 2^m$ and a rate R, we can accurately compute the bit-error probability of the 2^m synthetic channels W^{s^m}, $\mathsf{s}^m \in \{-,+\}^m$. Subsequently, we can compute upper and lower bounds on the block-error probability of a polar code of rate R and block-length n, when the code is used for communication over the channel W. This makes polar codes a suitable candidate for applications like optical communications or storage, where the error probability guarantees of the order 10^{-16} are required. However, the upper and the lower bounds of (2.29) differ by $O(n)$, which, for practical block-lengths, translates to orders of magnitude of difference. To be able to optimize the code rate, given a target block-error probability, it is important to make sure that the union bound does not overestimate the block-error probability. In Chapter 3, we show that, at least when communication takes place over the erasure channel, the upper bound of (2.29) is essentially tight. Consequently, as a rule of thumb, we can safely design the code by pretending that the union bound is indeed equal to the actual block-error probability of the code, under successive-cancellation decoding.

Despite their very good asymptotic performance, short-to-moderate-length polar codes, under successive-cancellation decoding, have a relatively high block-error probability compared to competing coding schemes. For example, to achieve a target block-error probability of 10^{-5}, over a BI-AWGNC, a system using a rate-1/2 polar code of length $n = 1024$ would require about 1 dB energy per bit more than one based on the rate-1/2 length $n = 1296$ LDPC code of IEEE 802.11n (WiFi) standard. This deficiency is partly due to the sub-optimality of the successive-cancellation decoding; specifically, because the decoder is not allowed to go back and correct potentially wrong decisions upon observing a conflict between the decisions and the following frozen bits. In Chapter 4 we discuss an improved decoding algorithm, called *successive-cancellation list decoding* [108]: it decreases the block-error probability of short polar codes and makes them competitive with existing error correction schemes that are used in practice.

2.A Proof of Theorem 2.2

Let

$$\alpha_m(a) \coloneqq \frac{1}{2^m}\left|\left\{\mathsf{s}^m \in \{-,+\}^m : I\big(W^{\mathsf{s}^m}\big) \le a\right\}\right|, \tag{2.45}$$

$$\beta_m(b) \coloneqq \frac{1}{2^m}\left|\left\{\mathsf{s}^m \in \{-,+\}^m : I\big(W^{\mathsf{s}^m}\big) \ge b\right\}\right| \tag{2.46}$$

$$\gamma_m(a,b) \coloneqq \frac{1}{2^m}\left|\left\{\mathsf{s}^m \in \{-,+\}^m : I\big(W^{\mathsf{s}^m}\big) \in (a,b)\right\}\right|. \tag{2.47}$$

and define, for $m \in \mathbb{N}$,

$$\mu_m \coloneqq \frac{1}{2^m} \sum_{\mathsf{s}^m \in \{-,+\}^m} I\big(W^{\mathsf{s}^m}\big) \tag{2.48}$$

$$\nu_m \coloneqq \frac{1}{2^m} \sum_{\mathsf{s}^m \in \{-,+\}^m} I\big(W^{\mathsf{s}^m}\big)^2 \tag{2.49}$$

Equation (2.22) implies $\forall m \in \mathbb{N}$, $I(W) = \mu_m$. Moreover, using the identity $A^2 + B^2 = \frac{1}{2}[A+B]^2 + \frac{1}{2}[A-B]^2$, for any BMS channel W, we have

$$\begin{aligned} I(W^+)^2 + I(W^-)^2 &= \frac{1}{2}[I(W^+) + I(W^-)]^2 + \frac{1}{2}[I(W^+) - I(W^-)]^2 \\ &= 2I(W) + \frac{1}{2}[I(W^+) - I(W^-)]^2 && (2.50) \\ &\geq 2I(W). && (2.51) \end{aligned}$$

This shows that ν_m is an increasing sequence. Furthermore, since ν_m is bounded in $[0, 1]$ it must be convergent. Therefore,

$$\lim_{m\to\infty} [\nu_{m+1} - \nu_m] = 0. \tag{2.52}$$

As a consequence of Mrs. Gerber's Lemma [119], it is shown in [97, Lemma 2.1] that if $I(W) \in (a, b)$ (for any $0 < a < b < 1$) then

$$I(W^+) - I(W^-) \geq \eta(a, b) > 0 \tag{2.53}$$

Therefore, for all $\mathsf{s}^m \in \{-,+\}^m$ such that $I(W^{\mathsf{s}^m}) \in (a, b)$, (2.50) yields,

$$\frac{1}{2}[I(W^{\mathsf{s}^m+})^2 + I(W^{\mathsf{s}^m-})]^2 \geq I(W^{\mathsf{s}^m}) + \frac{1}{4}\eta(a,b)^2. \tag{2.54}$$

Hence,

$$\nu_{m+1} \geq \nu_m + \gamma_m(a, b)\frac{1}{4}\eta(a, b)^2. \tag{2.55}$$

Thus,

$$0 \leq \gamma_m(a, b) \leq \frac{4}{\eta(a, b)^2}[\nu_{m+1} - \nu_m]. \tag{2.56}$$

Using (2.52) we can immediately conclude that

$$\lim_{m\to\infty} \gamma_m(a, b) = 0 \tag{2.57}$$

which proves (2.24).

To prove (2.25) and (2.23), on one hand we have, $\forall m \in \mathbb{N}$,

$$\begin{aligned} I(W) = \mu_m &\leq a\alpha_m(a) + b\gamma_m(a, b) + \beta_m(b) && (2.58) \\ &\overset{(*)}{=} a + (b - a)\gamma_m(a, b) + (1 - a)\beta_m(b), && (2.59) \end{aligned}$$

where $(*)$ follows since $\alpha_m(a) + \beta_m(b) + \gamma_m(a, b) = 1$. Equation (2.59) implies

$$(1 - a) \liminf_{m\to\infty} \beta_m(b) + a \geq I(W). \tag{2.60}$$

Since, the above holds for any $a \in (0, b)$, we must have

$$\liminf_{m\to\infty} \beta_m(b) \geq I(W). \tag{2.61}$$

Similarly, $\forall m \in \mathbb{N}$,

$$\begin{aligned} I(W) = \mu_m &\geq a\gamma_m(a,b) + b\beta_m(b) \\ &\overset{(*)}{=} b - b\alpha_m(a) - (b-a)\gamma_m(a,b), \end{aligned} \tag{2.62}$$

which implies

$$1 - I(W) \leq (1-b) + b \liminf_{m\to\infty} \alpha_m(a). \tag{2.63}$$

Once again, since the above holds for any $b \in (a,1)$,

$$1 - I(W) \leq \liminf_{m\to\infty} \alpha_m(b). \tag{2.64}$$

On the other hand, $\alpha_m(a) + \beta_m(b) \leq 1$. Hence, using (2.61) and (2.64), we have

$$1 \leq \liminf_{m\to\infty} \alpha_m(a) + \liminf_{m\to\infty} \beta_m(b) \leq \limsup_{m\to\infty} \alpha_m(a) + \limsup_{m\to\infty} \beta_m(b) \leq 1. \tag{2.65}$$

The above implies both inequalities in (2.64) and (2.61) must be equality. Therefore,

$$\lim_{m\to\infty} \beta_m(b) = I(W), \qquad \text{and} \tag{2.66}$$

$$\lim_{m\to\infty} \alpha_m(a) = 1 - I(W), \tag{2.67}$$

which prove (2.25) and (2.23), respectively. □

3 Correlation between Synthetic BECs

We have seen in § 2.3.1 that the sum of the error probabilities of the 'good' synthetic channels — obtained by applying the union bound to the block-error event of polar codes — is used as a proxy on the block-error probability of polar codes, under successive-cancellation decoding. The union-bound estimate is sufficiently tight for the analysis of the exponential decay rate of the block-error probability; because the upper bound of (2.29) is at most n times larger than the lower bound of (2.29). However, especially for relatively large block-lengths, the bounds still differ by orders of magnitude. Hence, it is natural to ponder whether the union bound on the block-error probability of polar codes is tight?

The union bound overestimates, at most by a factor of 2, the probability of the union of pairwise independent events. For events $\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_k$ satisfying

$$\Pr(\mathcal{E}_i \cap \mathcal{E}_j) = \Pr(\mathcal{E}_i)\Pr(\mathcal{E}_j), \qquad i \neq j, \tag{3.1}$$

by using the inclusion–exclusion principle (see Lemma 3.1) it trivially follows that, when $\sum_{j=1}^{k} \Pr(\mathcal{E}_j) \leq 1$,

$$\frac{1}{2}\sum_{i=1}^{k} \Pr(\mathcal{E}_j) \leq \Pr\left(\bigcup_{j=1}^{k} \mathcal{E}_j\right) \leq \sum_{i=1}^{k} \Pr(\mathcal{E}_j). \tag{3.2}$$

For data transmission using polar codes, the error events in the synthetic channels become only *asymptotically* pairwise independent: Consider the sequence of polar codes of rate $R < I(W)$ defined by the sequence of information indices $\mathcal{A}_m \subset \{-,+\}^m$, $m \in \mathbb{N}$ for communication over the BMS channel W. For every $\mathsf{s}^m \in \{-,+\}^m$,

$$\max_{\substack{\mathsf{t}^m \in \{-,+\}^m:\\ \mathsf{t}^m \neq \mathsf{s}^m}} \left|\Pr\bigl(\mathcal{E}^{\mathsf{s}^m} \cap \mathcal{E}^{\mathsf{t}^m}\bigr) - \Pr\bigl(\mathcal{E}^{\mathsf{s}^m}\bigr)\Pr\bigl(\mathcal{E}^{\mathsf{t}^m}\bigr)\right| \leq \min\{\Pr(\mathcal{E}^{\mathsf{s}^m}), 1 - \Pr(\mathcal{E}^{\mathsf{s}^m})\}, \tag{3.3}$$

where $\mathcal{E}^{\mathsf{s}^m}$ and $\mathcal{E}^{\mathsf{t}^m}$ are the bit-error events of the synthetic channels W^{s^m} and W^{t^m} respectively. According to Theorem 2.4, $\forall \mathsf{s}^m \in \mathcal{A}_m \Pr(\mathcal{E}^{\mathsf{s}^m}) = P_{\mathrm{e}}(W^{\mathsf{s}^m}) \in O(2^{-2^{m\beta}})$ (for any $\beta < 1/2$). Therefore, for any $\delta > 0$, there exists $m_0(\delta)$ such that $\forall m \geq m_0$, and $\forall \mathsf{s}^m \in \mathcal{A}_m$,

$$\max_{\substack{\mathsf{t}^m \in \{-,+\}^m: \\ \mathsf{t}^m \neq \mathsf{s}^m}} \left|\Pr\bigl(\mathcal{E}^{\mathsf{s}^m} \cap \mathcal{E}^{\mathsf{t}^m}\bigr) - \Pr\bigl(\mathcal{E}^{\mathsf{s}^m}\bigr) \Pr\bigl(\mathcal{E}^{\mathsf{t}^m}\bigr)\right| \leq \delta \tag{3.4}$$

However, for asymptotically pairwise independent events the union bound could well overestimate the probability of their union.

Example 3.1. Suppose we have k events $\mathcal{E}_1, \dots, \mathcal{E}_k$ satisfying

$$\mathcal{E}_1 = \mathcal{E}_2 = \cdots = \mathcal{E}_k \qquad \text{and} \qquad \Pr(\mathcal{E}_1) = \frac{1}{k^2}. \tag{3.5}$$

These events are asymptotically pairwise independent; for $i \neq j$,

$$\bigl|\Pr(\mathcal{E}_i \cap \mathcal{E}_j) - \Pr(\mathcal{E}_i)\Pr(\mathcal{E}_j)\bigr| = \frac{k^2-1}{k^4} \tag{3.6}$$

which can be made as small as desired by choosing k large enough. The probability of the union of the events is

$$\Pr\left(\bigcup_{j=1}^{k} \mathcal{E}_j\right) = \frac{1}{k^2}. \tag{3.7}$$

But the union-bound estimate evaluates to $1/k$.

The above example shows that measuring the independence of the pair of events $\mathcal{E}_i$ and $\mathcal{E}_j$ via their *covariance*

$$\operatorname{cov}(\mathcal{E}_i, \mathcal{E}_j) \coloneqq \Pr(\mathcal{E}_i \cap \mathcal{E}_j) - \Pr(\mathcal{E}_i)\Pr(\mathcal{E}_j) \tag{3.8}$$

is not good enough for assessing the tightness of the union bound. In fact, for the union bound to be tight, this difference should be small compared to the probability of the individual events. It is, instead, better to measure the dependence between the events via pairwise *correlation coefficients*:

$$\rho(i,j) \coloneqq \frac{\Pr(\mathcal{E}_i \cap \mathcal{E}_j) - \Pr(\mathcal{E}_i)\Pr(\mathcal{E}_j)}{\sqrt{\Pr(\mathcal{E}_i)\bigl(1-\Pr(\mathcal{E}_i)\bigr)}\sqrt{\Pr(\mathcal{E}_j)\bigl(1-\Pr(\mathcal{E}_j)\bigr)}}. \tag{3.9}$$

Note that $\rho(i,j)$ as defined above is the Pearson correlation coefficient between the pair of indicator random variables of the events $\mathcal{E}_i$ and $\mathcal{E}_j$, i.e., $\bigl(\mathbb{1}_{\mathcal{E}_i}, \mathbb{1}_{\mathcal{E}_j}\bigr)$.

Lemma 3.1. *Let $\mathcal{E}_1, \dots, \mathcal{E}_k$ be k arbitrary events,*

$$\mathcal{E} \coloneqq \bigcup_{j=1}^{k} \mathcal{E}_j, \tag{3.10}$$

and

$$P_{\mathrm{UB}}(\mathcal{E}_1, \dots, \mathcal{E}_k) \coloneqq \sum_{j=1}^{k} \Pr(\mathcal{E}_j). \tag{3.11}$$

Then if $P_{\mathrm{UB}}(\mathcal{E}_1, \dots, \mathcal{E}_k) \leq 1$,

$$\frac{1}{2}[1 - k \max_{i \neq j} \rho(i,j)] P_{\mathrm{UB}}(\mathcal{E}_1, \dots, \mathcal{E}_k) \leq \Pr(\mathcal{E}) \leq P_{\mathrm{UB}}(\mathcal{E}_1, \dots, \mathcal{E}_k) \tag{3.12}$$

(with $\rho(i,j)$ *defined as in* (3.9)*).*

In other words, if all pairwise correlation coefficients between the events are much smaller than $1/k$ (with k being the number of events), the union bound estimates the probability of the union of the events as good as it would if the events were pairwise independent. Lemma 3.1 is an immediate consequence of the inclusion–exclusion principle. (See the proof in Appendix 3.A.)

In view of Lemma 3.1, we can conclude that if the pairwise correlation coefficients between the error events of the synthetic channels (obtained by the m-fold application of the polar transform to a channel), decay faster than 2^{-m}, the union bound on the block-error probability of a polar code used for transmission over that channel is essentially tight.

In this chapter, we consider the problem of communication over a binary erasure channel using polar codes and successive-cancellation decoding. The set of binary erasure channels is stable under the polar transform [5, Proposition 6]: if W is a BEC, then both W^- and W^+ are equivalent to BECs. Moreover when the transmission takes place over an erasure channel, the successive-cancellation decoder either correctly decodes an information bit or sees an 'erasure' of that information bit. With a pessimistic presumption on the decoder, we assume that a decoding failure occurs if any of the information bits are erased.[1] Therefore, to assess the tightness of the union bound on the block-error probability of the code, we study the correlation coefficients between the pairs of error (in this case *erasure*) events of the synthetic channels. The stability of the set of binary erasure channels under the polar transform facilitates the computation of those correlation coefficients. We give recursive formulae for computing these correlation coefficients (see § 3.1); and show numerical evidence that the union-bound estimate on the block-error probability is indeed tight (in § 3.2). We, then, prove that as m grows large *almost all* pairwise correlation coefficients decay faster than 2^{-m}, hence the union bound precisely estimates the block-error probability of polar codes, when used for communication over the erasure channel (see § 3.3).

The results of this chapter was published in part in [12].

[1] A practical decoder can flip a coin and decide on the value of an erased information bit correctly with 50% of chance. An analysis analogous to that of this chapter applies to such a decoder.

3.1 Recursive Computation of Correlation Coefficients

As Arıkan showed in [5, Propositon 6], if W is a binary erasure channel, both W^- and W^+ are also equivalent to binary erasure channels; W^- erases if either independent copies of W erase, whereas W^+ erases if both independent copies of W erase. Accordingly, if $\mathcal{E}_1$ and $\mathcal{E}_2$ denote the erasure events of two independent copies of a BEC, W, the erasure events of W^- and W^+ will, respectively, be

$$\mathcal{E}^- = \mathcal{E}_1 \cup \mathcal{E}_2 \qquad \Longleftrightarrow \qquad \mathcal{E}^{-c} = \mathcal{E}_1^c \cap \mathcal{E}_2^c \tag{3.13a}$$

$$\mathcal{E}^+ = \mathcal{E}_1 \cap \mathcal{E}_2. \tag{3.13b}$$

Obviously, although the underlying BECs (and hence their erasure events) are independent, $\mathcal{E}^-$ and $\mathcal{E}^+$ are heavily correlated: An erasure in the 'plus' channel implies that in the 'minus' channel.

Arıkan has already shown that the erasure probabilities of the synthetic BECs W^{s^m}, denoted as Z^{s^m}, for $\mathsf{s}^m \in \{-,+\}^m$, can be computed via single-step recursions [5, Proposition 6]. Indeed, the independence of $\mathcal{E}_1$ and $\mathcal{E}_2$, together with (3.13), shows that $\forall \mathsf{s}^m \in \{-,+\}^m$,

$$Z^{\mathsf{s}^m -} := \Pr(\mathcal{E}^{\mathsf{s}^m -}) = 2Z^{\mathsf{s}^m} - (Z^{\mathsf{s}^m})^2 \qquad \Longleftrightarrow \qquad \overline{Z^{\mathsf{s}^m -}} = (\overline{Z^{\mathsf{s}^m}})^2, \tag{3.14a}$$

$$Z^{\mathsf{s}^m +} := \Pr(\mathcal{E}^{\mathsf{s}^m +}) = (Z^{\mathsf{s}^m})^2, \tag{3.14b}$$

where $Z^{\mathsf{s}^m} := \Pr(\mathcal{E}^{\mathsf{s}^m})$, which is nothing but (2.38). (In this chapter, we use the shorthand notation $\overline{x} := 1 - x$, for $x \in [0,1]$.)

Let us index the *covariance* and *correlation coefficients* between the error events of pairs of synthetic channels by sign sequences. For sign sequences s^m and t^m, let

$$C(\mathsf{s}^m, \mathsf{t}^m) := \Pr(\mathcal{E}^{\mathsf{s}^m} \cap \mathcal{E}^{\mathsf{t}^m}) - \Pr(\mathcal{E}^{\mathsf{s}^m}) \Pr(\mathcal{E}^{\mathsf{t}^m}) \tag{3.15}$$

be the covariance between the error events of W^{s^m} and W^{t^m} and

$$\rho(\mathsf{s}^m, \mathsf{t}^m) := \frac{\Pr(\mathcal{E}^{\mathsf{s}^m} \cap \mathcal{E}^{\mathsf{t}^m}) - \Pr(\mathcal{E}^{\mathsf{s}^m}) \Pr(\mathcal{E}^{\mathsf{t}^m})}{\sqrt{\Pr(\mathcal{E}^{\mathsf{s}^m})[1 - \Pr(\mathcal{E}^{\mathsf{s}^m})]}\sqrt{\Pr(\mathcal{E}^{\mathsf{t}^m})[1 - \Pr(\mathcal{E}^{\mathsf{t}^m})]}}, \tag{3.16}$$

be the correlation coefficient between those events. Interestingly, the covariances $C(\mathsf{s}^m, \mathsf{t}^m)$ and, accordingly, correlation coefficients $\rho(\mathsf{s}^m, \mathsf{t}^m)$ can also be computed via single-step recursions.

Lemma 3.2. *The covariances between the erasure events of the synthetic channels, after $m+1$ steps of Arıkan's polar transform, are related to those of the synthetic channels, after m steps of the polar transform via*

$$C(\mathsf{s}^m -, \mathsf{t}^m -) = 2\overline{Z^{\mathsf{s}^m}}\,\overline{Z^{\mathsf{t}^m}}\, C(\mathsf{s}^m, \mathsf{t}^m) + C(\mathsf{s}^m, \mathsf{t}^m)^2, \tag{3.17a}$$

$$C(\mathsf{s}^m -, \mathsf{t}^m +) = 2\overline{Z^{\mathsf{s}^m}}\, Z^{\mathsf{t}^m}\, C(\mathsf{s}^m, \mathsf{t}^m) - C(\mathsf{s}^m, \mathsf{t}^m)^2, \tag{3.17b}$$

$$C(\mathsf{s}^m +, \mathsf{t}^m -) = 2Z^{\mathsf{s}^m}\, \overline{Z^{\mathsf{t}^m}}\, C(\mathsf{s}^m, \mathsf{t}^m) - C(\mathsf{s}^m, \mathsf{t}^m)^2, \tag{3.17c}$$

$$C(\mathsf{s}^m +, \mathsf{t}^m +) = 2Z^{\mathsf{s}^m}\, Z^{\mathsf{t}^m}\, C(\mathsf{s}^m, \mathsf{t}^m) + C(\mathsf{s}^m, \mathsf{t}^m)^2, \tag{3.17d}$$

The recursions terminate at $m = 0$ where $C(\emptyset, \emptyset) = p\overline{p}$ with p being the erasure probability of the underlying BEC.

Corollary 3.3. *The correlation coefficients between the erasure events of the synthetic channels (defined as in* (3.9)*) after $m+1$ steps of Arıkan's polar transform are related to those of the synthetic channels after m steps of the polar transform via:*

$$\begin{aligned}\rho(\mathsf{s}^m-, \mathsf{t}^m-) = {} & 2\sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+\overline{Z^{\mathsf{s}^m}}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+\overline{Z^{\mathsf{t}^m}}}}\rho(\mathsf{s}^m, \mathsf{t}^m) \\ & + \sqrt{\frac{Z^{\mathsf{s}^m}}{1+\overline{Z^{\mathsf{s}^m}}}}\sqrt{\frac{Z^{\mathsf{t}^m}}{1+\overline{Z^{\mathsf{t}^m}}}}\rho(\mathsf{s}^m, \mathsf{t}^m)^2, \end{aligned} \tag{3.18a}$$

$$\begin{aligned}\rho(\mathsf{s}^m-, \mathsf{t}^m+) = {} & 2\sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+\overline{Z^{\mathsf{s}^m}}}}\sqrt{\frac{Z^{\mathsf{t}^m}}{1+Z^{\mathsf{t}^m}}}\rho(\mathsf{s}^m, \mathsf{t}^m) \\ & - \sqrt{\frac{Z^{\mathsf{s}^m}}{1+\overline{Z^{\mathsf{s}^m}}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}}}\rho(\mathsf{s}^m, \mathsf{t}^m)^2, \end{aligned} \tag{3.18b}$$

$$\begin{aligned}\rho(\mathsf{s}^m+, \mathsf{t}^m-) = {} & 2\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+\overline{Z^{\mathsf{t}^m}}}}\rho(\mathsf{s}^m, \mathsf{t}^m) \\ & - \sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{Z^{\mathsf{t}^m}}{1+\overline{Z^{\mathsf{t}^m}}}}\rho(\mathsf{s}^m, \mathsf{t}^m)^2, \end{aligned} \tag{3.18c}$$

$$\begin{aligned}\rho(\mathsf{s}^m+, \mathsf{t}^m+) = {} & 2\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{Z^{\mathsf{t}^m}}{1+Z^{\mathsf{t}^m}}}\rho(\mathsf{s}^m, \mathsf{t}^m) \\ & + \sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}}}\rho(\mathsf{s}^m, \mathsf{t}^m)^2. \end{aligned} \tag{3.18d}$$

The recursions end at $m = 0$ with $\rho(\emptyset, \emptyset) = 1$.

Proofs of Lemma 3.2 and Corollary 3.3 are straightforward but tedious. We relegate them to Appendices 3.B.1 and 3.B.2, respectively.

It follows that we can compute the covariances (and correlation coefficients) between pairs of erasure events of synthetic channels via single-step recursions because the probabilities of intersections of pairs of error events, after $m+1$ levels of polarization, are computable in terms of the joint probabilities after m levels of polarization. The property of being computable via single-step recursions generalizes to higher order statistics.

Lemma 3.4. *In general the probabilities of the intersections of any number k of erasure events are computable via single-step recursions.*

Proof. Let $\mathsf{s}^{m+1}(i) \in \{-,+\}^{m+1}$, $i = 1, 2, \ldots, k$ be a collection of (not necessarily distinct) sign sequences of length $m+1$. We know that

$$\Pr\left(\bigcap_{i=1}^{k} \mathcal{E}^{\mathsf{s}^{m+1}(i)}\right) = \mathbb{E}\left[\prod_{i=1}^{k} \mathbb{1}_{\mathcal{E}^{\mathsf{s}^{m+1}(i)}}\right]. \tag{3.19}$$

Using (3.13) we have

$$\mathbb{1}_{\mathcal{E}^{\mathsf{s}^{m+1}}} = \begin{cases} \mathbb{1}_{\mathcal{E}_1^{\mathsf{s}^m}} + \mathbb{1}_{\mathcal{E}_2^{\mathsf{s}^m}} - \mathbb{1}_{\mathcal{E}_1^{\mathsf{s}^m}} \mathbb{1}_{\mathcal{E}_2^{\mathsf{s}^m}} & \text{if } \mathsf{s}_{m+1} = - \\ \mathbb{1}_{\mathcal{E}_1^{\mathsf{s}^m}} \mathbb{1}_{\mathcal{E}_2^{\mathsf{s}^m}} & \text{if } \mathsf{s}_{m+1} = + \end{cases} \tag{3.20}$$

(where $\mathcal{E}_1^{\mathsf{s}^m}$ and $\mathcal{E}_2^{\mathsf{s}^m}$ are independent events with identical probability). Therefore,

$$\prod_{i=1}^{k} \mathbb{1}_{\mathcal{E}^{\mathsf{s}^{m+1}}} \tag{3.21}$$

can be decomposed into the summation of terms in the form of

$$\prod_{j=1}^{l_1} \mathbb{1}_{\mathcal{E}_1^{\mathsf{s}^m(i_j)}} \times \prod_{j=1}^{l_2} \mathbb{1}_{\mathcal{E}_2^{\mathsf{s}^m(i_j)}} \tag{3.22}$$

for some $l_1 \leq k$ and $l_2 \leq k$. The independence of the indicator variables of the events with subscript 1 and those with subscript 2 implies the expectation of such a product will be the product of the expectations, each of which is the probability of intersection of (at most) k erasure events after m levels of polarization. □

3.2 A Tight Lower Bound on the Block-Error Probability

Recall Equation (2.29): In general, the block-error probability of a polar code defined by information set $\mathcal{A}_m \in \{-,+\}^m$, when used for communication over the channel W, is lower-bounded as

$$P_{\mathrm{e}}(\mathcal{A}_m) \geq \max_{\mathsf{s}^m \in \mathcal{A}_m} \Pr(\mathcal{E}^{\mathsf{s}^m}),$$

(where $\Pr(\mathcal{E}^{\mathsf{s}^m})$ is the bit-error probability of synthetic channel W^{s^m}).

For the case of erasure channel, since we can compute the second order statistics of the error events (see Lemma 3.2), we can compute the inclusion–exclusion lower bound on the block-error probability:

Lemma 3.5. *The block-error probability of a polar code of block length 2^m, defined by information indices $\mathcal{A}_m \in \{-,+\}^m$, when used for communicating* *over a* $\mathsf{BEC}(p)$ *is lower-bounded as*

$$P_{\mathrm{e}}(\mathcal{A}_m) \geq \sum_{\mathsf{s}^m \in \mathcal{A}_m} Z^{\mathsf{s}^m} - \frac{1}{2} \sum_{\substack{(\mathsf{s}^m, \mathsf{t}^m) \in \mathcal{A}_m^2 : \\ \mathsf{s}^m \neq \mathsf{t}^m}} \Big[Z^{\mathsf{s}^m} Z^{\mathsf{t}^m} + C(\mathsf{s}^m, \mathsf{t}^m) \Big] \tag{3.23}$$

where the values of Z^{s^m}, Z^{t^m} *and* $C(\mathsf{s}^m, \mathsf{t}^m)$ *can be computed via the single-step recursions of* (3.14) *and* (3.17) *(starting with* $Z^{\emptyset} = p$ *and* $C(\emptyset, \emptyset) = p\overline{p}$*).*

Remark 1. We will see later in Equation (3.27) that the covariances $C(\mathsf{s}^m, \mathsf{t}^m)$ will become tiny as m gets large. Thus, for numerical stability, it is safer to compute the normalized correlation coefficients, $\rho(\mathsf{s}^m, \mathsf{t}^m)$, via the recursions of (3.18) instead (and compute the inclusion–exclusion lower bound in terms of the correlation coefficients).

Remark 2. In § 2.4 we have seen that the erasure probabilities of synthetic channels (after the m-fold application of the polar transform to a binary erasure channel) can be computed in $O(n)$ operations with $O(\log n)$ internal and $O(n)$ external memory elements to store the values. Similar considerations show that the covariances $C(\mathsf{s}^m, \mathsf{t}^m)$, $(\mathsf{s}^m, \mathsf{t}^m) \in \{-, +\}^m \times \{-, +\}^m$ can be computed in $O(n^2)$ operations, with $O(\log n)$ internal memory and $O(n^2)$ external memory (for storing the values).

The quadratic growth of the computational complexity is a bottleneck for computing the inclusion–exclusion lower bound on the block-error probability of long polar codes. For example, consider a (relatively long) polar code of block-length $n = 2^{20}$. Assuming each erasure probability is stored as an 8-byte double-precision floating point number (as in the IEEE standard for floating point arithmetics), storing all erasure probabilities (hence computing the union bound on the block-error probability) requires only eight megabytes of memory whereas if the same data-types are used for storing the covariances, eight *terabytes* of storage space would be required to compute the inclusion–exclusion lower bound.

In Table 3.1 and Figure 3.1, we present the upper bound, the trivial lower bound of (2.29), and the inclusion–exclusion lower bound of (3.23) on the block-error probability of polar codes of different rates used for communication over a binary erasure channel with erasure probability 1/2. Note that the inclusion–exclusion lower bound of (3.23) is so tight that, except for high rates, it is visually indistinguishable from the union bound in the curves of Figure 3.1. (The careful reader will notice that we have intentionally chosen the channel and the block-lengths in the same way as in [5, Fig. 7] — with the difference that we were unable to compute the lower bound of (3.23) for the block length of $n = 2^{20}$, for the reasons discussed above.)

$\lvert\mathcal{A}_n\rvert$	Union Bound (2.29)	Lower Bound of (2.29)	Lower Bound of (3.23)
128	6.362×10^{-17}	1.388×10^{-17}	6.362×10^{-17}
160	1.894×10^{-13}	2.894×10^{-14}	1.894×10^{-13}
192	1.488×10^{-10}	3.297×10^{-11}	1.488×10^{-10}
224	7.425×10^{-8}	1.299×10^{-8}	7.424×10^{-8}
256	5.685×10^{-6}	6.330×10^{-7}	5.684×10^{-6}
288	1.496×10^{-4}	1.328×10^{-5}	1.494×10^{-4}
320	2.482×10^{-3}	2.141×10^{-4}	2.464×10^{-3}
352	2.598×10^{-2}	1.649×10^{-3}	2.500×10^{-2}
384	0.186	1.034×10^{-2}	0.152
416	0.934	4.207×10^{-2}	0.261

(a) $n = 2^{10}$

$\lvert\mathcal{A}_n\rvert$	Union Bound (2.29)	Lower Bound of (2.29)	Lower Bound of (3.23)
8,192	2.172×10^{-31}	5.000×10^{-33}	2.172×10^{-31}
8,704	9.366×10^{-27}	1.699×10^{-28}	9.366×10^{-27}
9,216	5.503×10^{-23}	8.120×10^{-25}	5.503×10^{-23}
9,728	1.420×10^{-19}	2.173×10^{-21}	1.420×10^{-19}
10,240	2.073×10^{-16}	2.777×10^{-18}	2.073×10^{-16}
10,752	1.508×10^{-13}	1.894×10^{-15}	1.508×10^{-13}
11,264	5.008×10^{-11}	5.035×10^{-13}	5.008×10^{-11}
11,776	7.502×10^{-9}	6.836×10^{-11}	7.502×10^{-9}
12,288	5.801×10^{-7}	4.745×10^{-9}	5.801×10^{-7}
12,800	3.147×10^{-5}	2.221×10^{-7}	3.146×10^{-5}
13,312	1.074×10^{-3}	6.668×10^{-6}	1.073×10^{-3}
13,824	2.314×10^{-2}	1.306×10^{-4}	2.271×10^{-2}
14,336	0.319	1.471×10^{-3}	0.254

(b) $n = 2^{15}$

Table 3.1: Bounds on the Block-Error Probability of Polar Code on BEC$(1/2)$

3.3 Decay of Correlations

The numerical examples show that the inclusion–exclusion lower bound on the block-error probability of polar codes is extremely close to the upper bound that the union bound gives. In the remainder of this chapter, we will prove that the correlation between the synthetic erasure channels decay very rapidly, hence the union bound is a tight estimate of the block-error probability of polar codes when used for communication over an erasure channel.

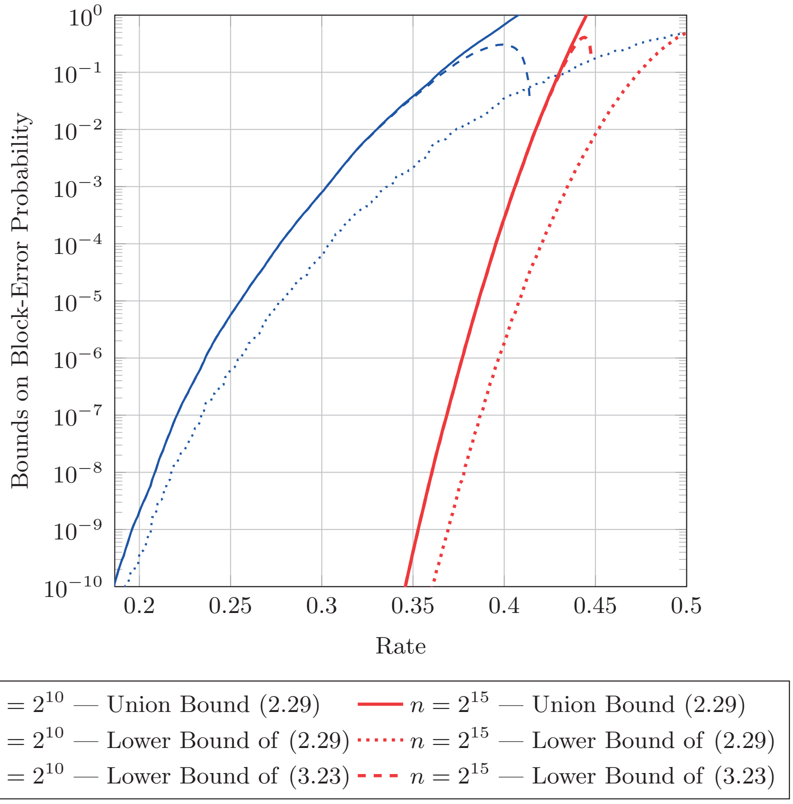


Figure 3.1: Bounds on the Block-Error Probability of Polar Code on BEC$(1/2)$

3.3.1 Basic Properties

Using the recursions of Corollary 3.3, we can derive the properties summarized in the following lemma on the correlation coefficients.

Lemma 3.6. *The following properties hold for the correlation coefficients:*

(i) $\forall \mathsf{s}^m, \mathsf{t}^m \in \{-,+\}^m$,

$$0 \le \rho(\mathsf{s}^m, \mathsf{t}^m) \le \min\left\{\sqrt{\frac{\overline{Z^{\mathsf{s}^m}}\, Z^{\mathsf{t}^m}}{Z^{\mathsf{s}^m}\, \overline{Z^{\mathsf{t}^m}}}}, \sqrt{\frac{Z^{\mathsf{s}^m}\, \overline{Z^{\mathsf{t}^m}}}{\overline{Z^{\mathsf{s}^m}}\, Z^{\mathsf{t}^m}}}\right\}. \tag{3.24}$$

(ii) $\forall \mathsf{s}^{m+1}, \mathsf{t}^{m+1} \in \{-,+\}^{m+1}$

$$\rho(\mathsf{s}^{m+1}, \mathsf{t}^{m+1}) \le \rho(\mathsf{s}^m, \mathsf{t}^m), \tag{3.25}$$

with equality if and only if

(a) $\rho(\mathsf{s}^m, \mathsf{t}^m) = 0$*; or*

(b) $\mathsf{s}_{m+1} = \mathsf{t}_{m+1}$*,* $Z^{\mathsf{s}^m} = Z^{\mathsf{t}^m}$*, and* $\rho(\mathsf{s}^m, \mathsf{t}^m) = 1$*; or*

(c) $Z^{\mathsf{s}^m} = \mathbb{1}\{\mathsf{s}_{m+1} = +\}$ *and* $Z^{\mathsf{t}^m} = \mathbb{1}\{\mathsf{t}_{m+1} = +\}$.

(iii) $\forall \mathsf{s}^m \neq \mathsf{t}^m$,

$$\rho(\mathsf{s}^m, \mathsf{t}^m) \leq \frac{1}{3}. \tag{3.26}$$

Proof.

(i) As $\Pr(\mathcal{E}^{\mathsf{s}^m} \cap \mathcal{E}^{\mathsf{t}^m}) \leq \min\{\Pr(\mathcal{E}^{\mathsf{s}^m}), \Pr(\mathcal{E}^{\mathsf{t}^m})\}$, and $\Pr(\mathcal{E}^{\mathsf{s}^m}) = Z^{\mathsf{s}^m}$ (respectively $\Pr(\mathcal{E}^{\mathsf{t}^m}) = Z^{\mathsf{t}^m}$),

$$C(\mathsf{s}^m, \mathsf{t}^m) \leq \min\{Z^{\mathsf{s}^m}\overline{Z^{\mathsf{t}^m}}, \overline{Z^{\mathsf{s}^m}}Z^{\mathsf{t}^m}\}. \tag{3.27}$$

Hence, the upper bound on correlation coefficients follows by definition (3.9). The positivity of the correlation coefficients follows by induction on m: at $m = 0$ the only covariance value $C(\emptyset, \emptyset) = p\overline{p}$ is positive. Equations (3.17a) and (3.17d) map positive $C(\mathsf{s}^m, \mathsf{t}^m)$ to positive covariance values at level $m + 1$; equations (3.17b) and (3.17c) do so as well, because of the upper bound of (3.27).

(ii) When $\rho(\mathsf{s}^m, \mathsf{t}^m) = 0$ the claim is trivial. Thus, we assume $\rho(\mathsf{s}^m, \mathsf{t}^m) \neq 0$ and divide the left-hand sides of (3.18) by $\rho(\mathsf{s}^m, \mathsf{t}^m)$ to get,

$$\begin{aligned}\frac{\rho(\mathsf{s}^m-, \mathsf{t}^m-)}{\rho(\mathsf{s}^m, \mathsf{t}^m)} &= 2\sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+\overline{Z^{\mathsf{s}^m}}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+\overline{Z^{\mathsf{t}^m}}}} \\ &\quad + \sqrt{\frac{Z^{\mathsf{s}^m}}{1+\overline{Z^{\mathsf{s}^m}}}}\sqrt{\frac{Z^{\mathsf{t}^m}}{1+\overline{Z^{\mathsf{t}^m}}}}\rho(\mathsf{s}^m, \mathsf{t}^m),\end{aligned} \tag{3.28a}$$

$$\begin{aligned}\frac{\rho(\mathsf{s}^m-, \mathsf{t}^m+)}{\rho(\mathsf{s}^m, \mathsf{t}^m)} &= 2\sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+\overline{Z^{\mathsf{s}^m}}}}\sqrt{\frac{Z^{\mathsf{t}^m}}{1+Z^{\mathsf{t}^m}}} \\ &\quad - \sqrt{\frac{Z^{\mathsf{s}^m}}{1+\overline{Z^{\mathsf{s}^m}}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}}}\rho(\mathsf{s}^m, \mathsf{t}^m),\end{aligned} \tag{3.28b}$$

$$\begin{aligned}\frac{\rho(\mathsf{s}^m+, \mathsf{t}^m-)}{\rho(\mathsf{s}^m, \mathsf{t}^m)} &= 2\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+\overline{Z^{\mathsf{t}^m}}}} \\ &\quad - \sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{Z^{\mathsf{t}^m}}{1+\overline{Z^{\mathsf{t}^m}}}}\rho(\mathsf{s}^m, \mathsf{t}^m),\end{aligned} \tag{3.28c}$$

$$\begin{aligned}\frac{\rho(\mathsf{s}^m+, \mathsf{t}^m+)}{\rho(\mathsf{s}^m, \mathsf{t}^m)} &= 2\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{Z^{\mathsf{t}^m}}{1+Z^{\mathsf{t}^m}}} \\ &\quad + \sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}}}\rho(\mathsf{s}^m, \mathsf{t}^m).\end{aligned} \tag{3.28d}$$

The right-hand side of (3.28d) is upper-bounded as

$$\begin{aligned}
&2\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{Z^{\mathsf{t}^m}}{1+Z^{\mathsf{t}^m}}}+\sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}}}\rho(\mathsf{s}^m,\mathsf{t}^m)\\
&\quad\overset{\text{(a)}}{\leq}\sqrt{\frac{2Z^{\mathsf{s}^m}+\rho(\mathsf{s}^m,\mathsf{t}^m)\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{2Z^{\mathsf{t}^m}+\rho(\mathsf{s}^m,\mathsf{t}^m)\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}}}\\
&\quad=\sqrt{1-\frac{1-(Z^{\mathsf{s}^m}+\rho(\mathsf{s}^m,\mathsf{t}^m)\overline{Z^{\mathsf{s}^m}})}{1+Z^{\mathsf{s}^m}}}\sqrt{1-\frac{1-(Z^{\mathsf{t}^m}+\rho(\mathsf{s}^m,\mathsf{t}^m)\overline{Z^{\mathsf{t}^m}})}{1+Z^{\mathsf{t}^m}}}\\
&\quad\overset{\text{(b)}}{\leq}1.
\end{aligned}\tag{3.29}$$

In the above (a) follows from Cauchy–Schwarz inequality (and is strict unless $Z^{\mathsf{s}^m}=Z^{\mathsf{t}^m}$) and (b) since $\rho(\mathsf{s}^m,\mathsf{t}^m)\leq 1$ (and is strict unless $\rho(\mathsf{s}^m,\mathsf{t}^m)=1$). The same argument applies to (3.28a).

The right-hand side of (3.28c) is upper-bounded as

$$\begin{aligned}
&2\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+\overline{Z^{\mathsf{t}^m}}}}-\sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{Z^{\mathsf{t}^m}}{1+\overline{Z^{\mathsf{t}^m}}}}\rho(\mathsf{s}^m,\mathsf{t}^m)\\
&\qquad\qquad\overset{\text{(a)}}{\leq}2\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+\overline{Z^{\mathsf{t}^m}}}}\overset{\text{(b)}}{\leq}1,
\end{aligned}\tag{3.30}$$

where (a) follows since $\rho(\mathsf{s}^m,\mathsf{t}^m)\geq 0$ (because of (3.24)) and (b) as $\sqrt{\alpha/(1+\alpha)}\leq 1/\sqrt{2}$ for $\alpha\in[0,1]$. Both inequalities are strict unless $Z^{\mathsf{s}^m}=1-Z^{\mathsf{t}^m}=1$. The same argument applies to (3.28b).

(iii) Consider a pair of sign sequences s^m and t^m that differ only in their last component, say $\mathsf{s}_m=+$ and $\mathsf{t}_m=-$ (and agree on the first $m-1$ components). It is sufficient to prove the result for such a pair. The result for general case follows from property (ii).

Since $\rho(\mathsf{s}^{m-1},\mathsf{t}^{m-1})=1$, (3.18c) implies,

$$\begin{aligned}
\rho(\mathsf{s}^m,\mathsf{t}^m)&=2\sqrt{\frac{Z^{\mathsf{s}^{m-1}}}{1+Z^{\mathsf{s}^{m-1}}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^{m-1}}}}{1+\overline{Z^{\mathsf{t}^{m-1}}}}}-\sqrt{\frac{\overline{Z^{\mathsf{s}^{m-1}}}}{1+Z^{\mathsf{s}^{m-1}}}}\sqrt{\frac{Z^{\mathsf{t}^{m-1}}}{1+\overline{Z^{\mathsf{t}^{m-1}}}}}\\
&\overset{\text{(a)}}{=}2\sqrt{\frac{Z^{\mathsf{s}^{m-1}}}{1+Z^{\mathsf{s}^{m-1}}}}\sqrt{\frac{\overline{Z^{\mathsf{s}^{m-1}}}}{1+\overline{Z^{\mathsf{s}^{m-1}}}}}-\sqrt{\frac{\overline{Z^{\mathsf{s}^{m-1}}}}{1+Z^{\mathsf{s}^{m-1}}}}\sqrt{\frac{Z^{\mathsf{s}^{m-1}}}{1+\overline{Z^{\mathsf{s}^{m-1}}}}}\\
&=\sqrt{\frac{Z^{\mathsf{s}^{m-1}}}{1+Z^{\mathsf{s}^{m-1}}}}\sqrt{\frac{\overline{Z^{\mathsf{s}^{m-1}}}}{1+\overline{Z^{\mathsf{s}^{m-1}}}}}\\
&=\sqrt{\frac{Z^{\mathsf{s}^{m-1}}\overline{Z^{\mathsf{s}^{m-1}}}}{2+Z^{\mathsf{s}^{m-1}}\overline{Z^{\mathsf{s}^{m-1}}}}}\overset{\text{(b)}}{\leq}\frac{1}{3}
\end{aligned}\tag{3.31}$$

where (a) follows as $\mathsf{s}^{m-1} = \mathsf{t}^{m-1}$ and (b) as $\alpha/(2+\alpha) \le 1/9$ for $\alpha \in [0, 1/4]$. □

Property (ii) in Lemma 3.6 shows that the correlation coefficients are decreasing and properties (i) and (iii) show that the *off-diagonal* correlation coefficients (i.e., $\rho(\mathsf{s}^m, \mathsf{t}^m)$ for $\mathsf{s}^m \neq \mathsf{t}^m$) lie in $[0, 1/3]$. Before analyzing how the collection of $2^{2m} - 2^m$ off-diagonal correlation coefficients decay, let us make sure that they converge to 0 in a point-wise manner. In other words, we verify that applying the recursions of (3.18) to an off-diagonal correlation coefficient repeatedly will eventually drive the result to 0.

Lemma 3.7. *Let* s^∞ *and* t^∞ *be two unequal infinite sign sequences, and* s^m *and* t^m *be the sub-sequences corresponding to their first* m *elements, respectively. Then*

$$\lim_{m\to\infty} \rho(\mathsf{s}^m, \mathsf{t}^m) = 0. \tag{3.32}$$

Proof. Let ℓ be the length of the common prefix of s^∞ and t^∞. For $m > \ell$, by (ii) and (iii) in Lemma 3.6 we know $\rho(\mathsf{s}^m, \mathsf{t}^m) \in [0, 1/3]$ and is decreasing. Hence, $(\rho(\mathsf{s}^m, \mathsf{t}^m), m \in \mathbb{N})$ is a convergent sequence. Suppose its limit is $\rho^\star > 0$. This implies for every $\epsilon > 0$, $\exists m_0(\epsilon)$ such that $\forall m > m_0$,

$$\frac{\rho(\mathsf{s}^{m+1}, \mathsf{t}^{m+1})}{\rho(\mathsf{s}^m, \mathsf{t}^m)} \ge 1 - \epsilon \tag{3.33}$$

and

$$\rho(\mathsf{s}^m, \mathsf{t}^m) \in [\epsilon, 1 - \epsilon]. \tag{3.34}$$

By the continuity of (3.28), we must have $\forall m \ge m_0$,

$$|Z^{\mathsf{s}^m} - \mathbb{1}\{\mathsf{s}_m = +\}| \le \delta \quad \text{and} \quad |Z^{\mathsf{t}^m} - \mathbb{1}\{\mathsf{t}_m = +\}| \le \delta \tag{3.35}$$

where δ is a quantity that approaches 0 as $\epsilon \to 0$. Since the evolutions of Z^{s^m} and Z^{t^m} do not allow jumps from one extreme to the other, the latter requires both s and t sequences to be constant, i.e., $\mathsf{s}_m = \mathsf{s}_\star$ and $\mathsf{t}_m = \mathsf{t}_\star$ for $m \ge m_0$. Without loss of generality, assume $\mathsf{s}_* = +$ which, in turn, requires $Z^{\mathsf{s}^m} \ge 1 - \delta$. We now have an incompatible situation: s_m being a ‘plus’ sign will drive Z^{s^m} to zero as m grows. Consequently, we conclude that $\rho(\mathsf{s}^m, \mathsf{t}^m)$ cannot converge to a non-zero value. □

3.3.2 Exponential Decay of Correlation Coefficients

Lemma 3.7 gives evidence that the recursions of (3.18) will eventually push the correlation coefficients down to 0. Now, we look at the sequence of collections of 2^{2m} correlation coefficients and prove that they decay sufficiently fast with m. Let us first consider the average of this collection.

Lemma 3.8. *For all* $\mathsf{s}^m, \mathsf{t}^m \in \{-,+\}^m$,

$$\frac{1}{4} \sum_{(\mathsf{s}_{m+1},\mathsf{t}_{m+1})\in\{-,+\}^2} \rho(\mathsf{s}^{m+1},\mathsf{t}^{m+1}) \leq \frac{2}{3}\rho(\mathsf{s}^m,\mathsf{t}^m). \tag{3.36}$$

Proof. Using (3.18), it is easy to see that

$$\begin{aligned}
&\sum_{(\mathsf{s}_{m+1},\mathsf{t}_{m+1})\in\{-,+\}^2} \rho(\mathsf{s}^{m+1},\mathsf{t}^{m+1}) \\
&= 2\left[\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}} + \sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+\overline{Z^{\mathsf{s}^m}}}}\right]\left[\sqrt{\frac{Z^{\mathsf{t}^m}}{1+Z^{\mathsf{t}^m}}} + \sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+\overline{Z^{\mathsf{t}^m}}}}\right]\rho(\mathsf{s}^m,\mathsf{t}^m) \\
&\quad + \left[\sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}} - \sqrt{\frac{Z^{\mathsf{s}^m}}{1+\overline{Z^{\mathsf{s}^m}}}}\right]\left[\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}}} - \sqrt{\frac{Z^{\mathsf{t}^m}}{1+\overline{Z^{\mathsf{t}^m}}}}\right]\rho(\mathsf{s}^m,\mathsf{t}^m)^2
\end{aligned} \tag{3.37}$$

$$\begin{aligned}
&= \rho(\mathsf{s}^m,\mathsf{t}^m)\left\{2\left[\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}} + \sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+\overline{Z^{\mathsf{s}^m}}}}\right]\left[\sqrt{\frac{Z^{\mathsf{t}^m}}{1+Z^{\mathsf{t}^m}}} + \sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+\overline{Z^{\mathsf{t}^m}}}}\right]\right. \\
&\quad \left. + \left[\sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}} - \sqrt{\frac{Z^{\mathsf{s}^m}}{1+\overline{Z^{\mathsf{s}^m}}}}\right]\left[\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}}} - \sqrt{\frac{Z^{\mathsf{t}^m}}{1+\overline{Z^{\mathsf{t}^m}}}}\right]\rho(\mathsf{s}^m,\mathsf{t}^m)\right\}.
\end{aligned} \tag{3.38}$$

Both sides of the above are positive, and the term inside the curly brackets can be upper-bounded as

$$\begin{aligned}
&\left\{2\left[\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}} + \sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+\overline{Z^{\mathsf{s}^m}}}}\right]\left[\sqrt{\frac{Z^{\mathsf{t}^m}}{1+Z^{\mathsf{t}^m}}} + \sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+\overline{Z^{\mathsf{t}^m}}}}\right]\right. \\
&\quad \left. + \left[\sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}} - \sqrt{\frac{Z^{\mathsf{s}^m}}{1+\overline{Z^{\mathsf{s}^m}}}}\right]\left[\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}}} - \sqrt{\frac{Z^{\mathsf{t}^m}}{1+\overline{Z^{\mathsf{t}^m}}}}\right]\rho(\mathsf{s}^m,\mathsf{t}^m)\right\}^2 \\
&\overset{\text{(a)}}{\leq} \left\{2\left[\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}} + \sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+\overline{Z^{\mathsf{s}^m}}}}\right]^2 + \left[\sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}} - \sqrt{\frac{Z^{\mathsf{s}^m}}{1+\overline{Z^{\mathsf{s}^m}}}}\right]^2\rho(\mathsf{s}^m,\mathsf{t}^m)\right\} \\
&\quad \cdot \left\{2\left[\sqrt{\frac{Z^{\mathsf{t}^m}}{1+Z^{\mathsf{t}^m}}} + \sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+\overline{Z^{\mathsf{t}^m}}}}\right]^2 + \left[\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}}} - \sqrt{\frac{Z^{\mathsf{t}^m}}{1+\overline{Z^{\mathsf{t}^m}}}}\right]^2\rho(\mathsf{s}^m,\mathsf{t}^m)\right\}
\end{aligned} \tag{3.39}$$

$$\begin{aligned}
&\leq \left\{2\left[\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}} + \sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+\overline{Z^{\mathsf{s}^m}}}}\right]^2 + \left[\sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}} - \sqrt{\frac{Z^{\mathsf{s}^m}}{1+\overline{Z^{\mathsf{s}^m}}}}\right]^2\right\} \\
&\quad \cdot \left\{2\left[\sqrt{\frac{Z^{\mathsf{t}^m}}{1+Z^{\mathsf{t}^m}}} + \sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+\overline{Z^{\mathsf{t}^m}}}}\right]^2 + \left[\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}}} - \sqrt{\frac{Z^{\mathsf{t}^m}}{1+\overline{Z^{\mathsf{t}^m}}}}\right]^2\right\}
\end{aligned} \tag{3.40}$$

$$= 2\left\{1 + \sqrt{\frac{Z^{\mathsf{s}^m}\overline{Z^{\mathsf{s}^m}}}{2 + Z^{\mathsf{s}^m}\overline{Z^{\mathsf{s}^m}}}}\right\} \cdot 2\left\{1 + \sqrt{\frac{Z^{\mathsf{t}^m}\overline{Z^{\mathsf{t}^m}}}{2 + Z^{\mathsf{t}^m}\overline{Z^{\mathsf{t}^m}}}}\right\} \tag{3.41}$$

$$\overset{\text{(b)}}{\leq} \left(\frac{8}{3}\right)^2. \tag{3.42}$$

In the above (a) follows by Cauchy–Schwarz inequality and (b) since $\alpha/(2+\alpha)$ is increasing for $\alpha \in [0, 1/4]$. Using the above in (3.38) establishes (3.36). □

Corollary 3.9. *The average of the correlation coefficients decays exponentially fast in m according to*

$$\frac{1}{4^m} \sum_{\mathsf{s}^m, \mathsf{t}^m \in \{-,+\}^m} \rho(\mathsf{s}^m, \mathsf{t}^m) \leq \left(\frac{2}{3}\right)^m. \tag{3.43}$$

Corollary 3.9 implies that for large enough m, almost all of *off-diagonal* correlation coefficients, that is $\rho(\mathsf{s}^m, \mathsf{t}^m)$ for $\mathsf{s}^m \neq \mathsf{t}^m$, are small. However, the bound it gives is not strong enough to show the asymptotic tightness of the union bound on the block-error probability of polar codes. According to Lemma 3.1, in order to prove this tightness, we must show (i) that the correlations decay as fast as $2^{-(1+\alpha)m}$ for some $\alpha > 0$, and (ii) that this bound applies to $\max_{\mathsf{s}^m \neq \mathsf{t}^m} \rho(\mathsf{s}^m, \mathsf{t}^m)$ over the sign sequences s^m and t^m that index the information channels, as opposed to the average of correlation coefficients.

We can indeed show that for any $\alpha > 0$, $\max_{\mathsf{t}^m \neq \mathsf{s}^m} \rho(\mathsf{s}^m, \mathsf{t}^m) \leq 2^{-(1+\alpha)m}$, for almost all $\mathsf{s}^m \in \{-,+\}^m$, provided that m is large enough:

Theorem 3.10. *For any $\alpha > 0$ and $\delta > 0$, there exists $m_0(\alpha, \delta)$ such that $\forall m \geq m_0$,*

$$\frac{1}{2^m}\left|\left\{\mathsf{s}^m \in \{-,+\}^m \colon \max_{\mathsf{t}^m \neq \mathsf{s}^m} \rho(\mathsf{s}^m, \mathsf{t}^m) \leq 2^{-m(1+\alpha)}\right\}\right| \geq 1 - \delta. \tag{3.44}$$

Having proven Theorem 3.10, we can immediately conclude that the union bound is asymptotically tight:

Theorem 3.11. *Let $P_{\mathrm{UB}}(n, R, p)$ be the sum of $\lceil nR \rceil$, $n = 2^m$, smallest erasure probabilities of the synthetic BECs, obtained by the m-fold application of polar transform to a $\mathsf{BEC}(p)$ — that is, the sum of the $\lceil nR \rceil$ smallest values among $n = 2^m$ elements of the set $\{Z^{\mathsf{s}^m} : \mathsf{s}^m \in \{-,+\}^m\}$ where the values are computed according to the single-step recursions of* (3.14). *Then, if $P_{\mathrm{e}}(n, R, p)$ denotes the block-error probability of a polar code of block-length $n = 2^m$ and rate $R < 1 - p$, designed for communicating over a $\mathsf{BEC}(p)$, for every $\delta > 0$, there exists $m_0(\delta)$, such that $\forall m \geq m_0(\delta)$, with $n = 2^m$,*

$$(1 - \delta) P_{\mathrm{UB}}\big(n, R - \delta, p\big) \leq P_{\mathrm{e}}(n, R, p) \leq P_{\mathrm{UB}}(n, R, p). \tag{3.45}$$

Proof. The upper bound is already known and we only need to prove the lower bound. For each $m \in \mathbb{N}$, let $\mathcal{A}_m$ be the indices of best $\lceil 2^m R \rceil$ synthetic channels and

$$\mathcal{D}_m := \Big\{ \mathsf{s}^m \in \{-,+\}^m : \max_{\mathsf{t}^m \neq \mathsf{s}^m} \rho(\mathsf{s}^m, \mathsf{t}^m) \leq \delta 2^{-m} \Big\}. \tag{3.46}$$

In view of Theorem 3.10, choose $m \geq m_1(\delta)$ such that

$$\frac{|\mathcal{D}_m|}{2^m} \geq 1 - \delta. \tag{3.47}$$

Consider the sequence polar codes[2] defined by the set of information indices $\mathcal{A}'_m = \mathcal{A}_m \cap \mathcal{D}'_m$. The rate of this sequence of codes is

$$R' := \frac{|\mathcal{A}_m \cap \mathcal{D}'_m|}{2^m} = \frac{|\mathcal{A}_m|}{2^m} + \frac{|\mathcal{D}'_m|}{2^m} - \frac{|\mathcal{A}_m \cup \mathcal{D}'_m|}{2^m} \geq R - \delta. \tag{3.48}$$

Moreover since $\mathcal{A}'_m \subseteq \mathcal{A}_m$,

$$P_{\mathrm{e}}(\mathcal{A}'_m) \leq P_{\mathrm{e}}(\mathcal{A}_m) = P_{\mathrm{e}}(n, R, p). \tag{3.49}$$

(Recall that $P_{\mathrm{e}}(\mathcal{A})$ denotes the block-error probability of the polar code defined by the set of information indices $\mathcal{A}$.) Let

$$P_{\mathrm{UB}}(\mathcal{A}'_m) = \sum_{\mathsf{s}^m \in \mathcal{A}'_m} \Pr(\mathcal{E}^{\mathsf{s}^m}), \tag{3.50}$$

be the union bound on the block-error probability of the code defined by $\mathcal{A}'_m$. Since $P_{\mathrm{UB}}(\mathcal{A}'_m) \leq P_{\mathrm{UB}}(m, R, p)$, by construction, and $R < 1 - p$, there exists m_2 such that for $m \geq m_2$, $P_{\mathrm{UB}}(\mathcal{A}'_m) \leq 1$.

For $m \geq m_0 := \max\{m_1, m_2\}$, the lower bound of (3.12) implies

$$P_{\mathrm{e}}(\mathcal{A}'_m) \geq \frac{1}{2}(1 - \delta) P_{\mathrm{UB}}(\mathcal{A}'_m). \tag{3.51}$$

Moreover, by definition

$$P_{\mathrm{UB}}(\mathcal{A}'_m) \geq P_{\mathrm{UB}}(m, R', p) \geq P_{\mathrm{UB}}(m, R - \delta, p). \tag{3.52}$$

Combining (3.49), (3.51), and (3.52), yields

$$P_{\mathrm{e}}(\mathcal{A}_m) \geq \frac{1}{2}(1 - \delta) P_{\mathrm{UB}}(m, R - \delta, p). \tag{3.53}$$

Finally, we note that the multiplicative factor in the above lower bound can easily be improved to $(1 - \delta)$: Since $P_{\mathrm{UB}}(\mathcal{A}'_m)$ decays with m as $2^{-2^{\beta m}}$ (for any $\beta < 1/2$), for large enough m, it is below δ. Using this upper bound (as opposed to 1) in (3.97) in the proof of Lemma 3.1 (see Appendix 3.A) yields the tighter lower bound of (3.45). □

[2] We slightly abuse the terminology here: these codes are not *real* polar codes as $\mathcal{A}'_m$ does not necessarily index the $|\mathcal{A}'_m|$ best synthetic channels.

We devote the rest of this section to proving Theorem 3.10. To this end, we establish a probabilistic framework similar to that used in [5] for proving the channel polarization theorem.

Let $\mathsf{S}_1, \mathsf{S}_2, \dots,$ be a sequence i.i.d. Bernoulli(1/2) random variables taking values in $\{-,+\}$, and let $\mathcal{F}_m \coloneqq \sigma(\mathsf{S}_1, \dots, \mathsf{S}_m)$ be the σ-algebra generated by the first m of them. We consider the sequence of random variables $(Z^{\mathsf{S}^m}, m \in \mathbb{N})$ and $(\rho(\mathsf{S}^m, \mathsf{t}^m), m \in \mathbb{N})$ (for an arbitrary infinite sign sequence t^∞, with t^m denoting its first m elements) which are all $\mathcal{F}_m$-measurable.

Define also the $\mathcal{F}_m$-measurable random process $(\rho_m, m \in \mathbb{N})$ as

$$\rho_m \coloneqq \max_{\mathsf{t}^m \neq \mathsf{S}^m} \rho(\mathsf{S}^m, \mathsf{t}^m), \tag{3.54}$$

and observe that (3.44) is equivalent to

$$\Pr\{\rho_m \leq 2^{-m(1+\alpha)}\} \geq 1 - \delta. \tag{3.55}$$

Note also that ρ_m is the maximum of two processes

$$\rho^{(1)}_{m,l_m} \coloneqq \max_{\substack{\mathsf{t}^m \neq \mathsf{S}^m: \\ \mathsf{t}^{l_m} = \mathsf{S}^{l_m}}} \rho(\mathsf{S}^m, \mathsf{t}^m) \tag{3.56}$$

$$\rho^{(2)}_{m,l_m} \coloneqq \max_{\substack{\mathsf{t}^m \neq \mathsf{S}^m: \\ \mathsf{t}^{l_m} \neq \mathsf{S}^{l_m}}} \rho(\mathsf{S}^m, \mathsf{t}^m) \tag{3.57}$$

for any sequence of integers $(l_m, m \in \mathbb{N})$ (such that $l_m \leq m$). In other words, $\rho^{(1)}_{m,l_m}$ is the largest correlation coefficient between the erasure events of the synthetic channels W^{S^m} and W^{t^m}, where t^m and S^m have a common prefix of length at least l_m (but differ at some position after the ${l_m}^{\text{th}}$) and $\rho^{(2)}_{m,l_m}$ is the maximum correlation coefficient between the erasure events W^{S^m} and the rest of synthetic channels.

Obviously, (3.55) follows if we can show that for $m \geq m_0(\alpha, \delta)$,

$$\Pr\{\rho^{(1)}_{m,l_m} \leq 2^{-m(1+\alpha)}\} \geq 1 - \delta/2, \quad \text{and} \tag{3.58}$$

$$\Pr\{\rho^{(2)}_{m,l_m} \leq 2^{-m(1+\alpha)}\} \geq 1 - \delta/2. \tag{3.59}$$

In the rest of this section we will establish (3.58) and (3.59) through a sequence of lemmas.

Closely related s^m and t^m

If two channels are *closely related*, i.e., they are indexed by a pair of sign sequences s^m and t^m, that share a long common prefix $\mathsf{s}^\ell = \mathsf{t}^\ell$, then with high probability, their parent channel, W^{s^ℓ}, is already *well-polarized*. In particular Z^{s^ℓ} is, with high probability, doubly exponentially small in ℓ. Recursions of (3.18) show that the correlation coefficient between the pair of erasure events of the children of an extremal channel is small. More precisely, if $\ell = \Omega(\log m)$, since Z^{s^ℓ} is exponentially small in m, the correlation coefficient between the children of W^{s^ℓ} will be exponentially small in m too.

Lemma 3.12. *Let* $l_m := \lceil 4\log\bigl(2(1+\alpha)m-1\bigr)\rceil$. *Then,* $\forall \alpha \geq 0$ *and* $\forall \delta > 0$, $\exists m_1(\alpha,\delta)$ *such that* $\forall m \geq m_1$,

$$\Pr\bigl\{\rho^{(1)}_{m,l_m} \leq 2^{-(1+\alpha)m}\bigr\} \geq 1-\delta/2.$$

Proof. Let $\mathsf{t}^m \neq \mathsf{S}^m$ be any sign sequence that shares a prefix of length $\ell \geq l_m$ with S^m. Note that S^ℓ is a uniformly chosen sign sequence in $\{-,+\}^\ell$. According to Lemma 3.6 (ii),

$$\rho(\mathsf{S}^m,\mathsf{t}^m) \leq \rho(\mathsf{S}^{\ell+1},\mathsf{t}^{\ell+1}) \tag{3.60}$$

Moreover, since $\rho(\mathsf{S}^\ell,\mathsf{t}^\ell)=1$,

$$\rho(\mathsf{S}^{\ell+1},\mathsf{t}^{\ell+1}) = \sqrt{\frac{Z^{\mathsf{S}^\ell}\overline{Z^{\mathsf{S}^\ell}}}{2+Z^{\mathsf{S}^\ell}\overline{Z^{\mathsf{S}^\ell}}}} \leq \sqrt{\frac{1}{2}\min\bigl\{Z^{\mathsf{S}^\ell},\overline{Z^{\mathsf{S}^\ell}}\bigr\}}. \tag{3.61}$$

In [6, Theorem 1] it has been shown that for any fixed $0<\beta<1/2$ and $\delta>0$ there exists $l_0(\beta,\delta)$ such that for $\ell \geq l_0$

$$\Pr\bigl\{Z^{\mathsf{S}^\ell} \in \bigl[2^{-2^{\ell\beta}}, 1-2^{-2^{\ell\beta}}\bigr]\bigr\} \leq \delta/2. \tag{3.62}$$

Note also that since $\ell \geq l_m$ (by assumption), $2^{-2^{\ell\beta}} \leq 2^{-2^{l_m\beta}}$, hence

$$\Pr\bigl\{Z^{\mathsf{S}^\ell} \in \bigl[2^{-2^{l_m\beta}}, 1-2^{-2^{l_m\beta}}\bigr]\bigr\} \leq \Pr\bigl\{Z^{\mathsf{S}^\ell} \in \bigl[2^{-2^{\ell\beta}}, 1-2^{-2^{\ell\beta}}\bigr]\bigr\}. \tag{3.63}$$

Thus, choosing $\beta = \frac{1}{4}$ and, accordingly, $m_1(\alpha,\delta)$ such that $m \geq m_1$ implies $l_m \geq l_0(\frac{1}{4},\delta)$, we will have (for $m \geq m_1$),

$$\Pr\bigl\{Z^{\mathsf{S}^\ell} \in \bigl[2^{-2^{l_m/4}}, 1-2^{-2^{l_m/4}}\bigr]\bigr\} \leq \delta/2, \tag{3.64}$$

which, together with the fact that $l_m \geq 4\log\bigl(2(1+\alpha)m-1\bigr)$, implies

$$\Pr\bigl\{Z^{\mathsf{S}^\ell} \in \bigl[2\cdot 2^{-2(1+\alpha)m}, 1-2\cdot 2^{-2(1+\alpha)m}\bigr]\bigr\} \leq \delta/2. \tag{3.65}$$

The above yields

$$\Pr\left\{\sqrt{\frac{1}{2}\min\bigl\{Z^{\mathsf{S}^\ell},\overline{Z^{\mathsf{S}^\ell}}\bigr\}} \geq 2^{-(1+\alpha)m}\right\} \leq \delta/2. \tag{3.66}$$

Due to (3.60) and (3.61), Equation (3.66) implies, with probability at least $1-\delta/2$ over the choice of S^ℓ,

$$\rho(\mathsf{S}^m,\mathsf{t}^m) \leq 2^{-m(1+\alpha)} \tag{3.67}$$

for all $\mathsf{t}^m \neq \mathsf{S}^m$ sharing a common prefix of length at least l_m with S^m. □

Distantly related s^m and t^m

For distantly related channels, at the separation point, the parent channel can still be mediocre. However, in this case, according to (3.28), the correlation coefficients decay somehow geometrically (that is, they are multiplied by some factor smaller than 1) in $m - l_m = \Theta(m)$ steps. Consequently, we can use a bootstrapping method to establish an exponentially decaying upper bound on $\rho^{(2)}_{m,l_m}$. More precisely, (3.28) implies

$$\frac{\rho(\mathsf{s}^{m+1},\mathsf{t}^{m+1})}{\rho(\mathsf{s}^m,\mathsf{t}^m)} \leq \begin{cases} \sqrt{\rho(\mathsf{s}^m,\mathsf{t}^m)^2 \dfrac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}} + \dfrac{2Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}} & \text{if } \mathsf{s}_{m+1} = +, \\ \sqrt{\rho(\mathsf{s}^m,\mathsf{t}^m)^2 \dfrac{Z^{\mathsf{s}^m}}{1+\overline{Z^{\mathsf{s}^m}}} + \dfrac{2\overline{Z^{\mathsf{s}^m}}}{1+\overline{Z^{\mathsf{s}^m}}}} & \text{if } \mathsf{s}_{m+1} = -. \end{cases} \tag{3.68}$$

(See the proof in Appendix 3.C.) Inspecting the right-hand side of (3.68), we see that if, for instance, the process Z^{S^m} is *sealed* outside the interval $[3\gamma^2/8, 1-3\gamma^2/8]$ and ρ_m below $\frac{1}{2}\gamma$, at each step, with probability half ρ_m is multiplied by at most γ, which results in an exponentially decaying bound on ρ_m.

Lemma 3.13. *For any sequence l_m such that $\lim_{m\to\infty} m - l_m = \infty$ and any $\gamma \geq 0$, there exits $m_s(\gamma,\delta)$ such that $\forall m \geq m_s$,*

$$\Pr\Big\{\rho^{(2)}_{m,l_m} \leq \frac{1}{2}\gamma\Big\} \geq 1-\delta \tag{3.69}$$

Proof. If $\gamma \geq \frac{2}{3}$ then $\Pr\{\rho^{(2)}_{m,l_m} \leq \frac{1}{2}\gamma\} = 1$ for $\forall m > l_m$ due to Lemma 3.6. Thus, in the rest of proof we assume $\gamma < \frac{2}{3}$.

We first observe that due to (ii) and (iii) in Lemma 3.6, $\rho^{(2)}_{l_m+1,l_m} \leq 1/3$ with probability 1. Now we will show that

$$\Pr\Big\{\rho^{(2)}_{m,l_m} \geq \frac{1}{2}\gamma\Big\} \leq \delta, \tag{3.70}$$

for sufficiently large m. Note that $\rho^{(2)}_{m,l_m} \geq \frac{1}{2}\gamma$ means $\exists \mathsf{t}^m \neq \mathsf{S}^m$ (in particular different in the first l_m positions) for which

$$\rho(\mathsf{t}^m,\mathsf{S}^m) \geq \frac{1}{2}\gamma. \tag{3.71}$$

Let $\mathsf{s}^m,\mathsf{t}^m$ be such a pair and consider the sequence of successive ratios

$$r_\ell := \frac{\rho(\mathsf{s}^\ell,\mathsf{t}^\ell)}{\rho(\mathsf{s}^{\ell-1},\mathsf{t}^{\ell-1})}, \qquad \ell = l_m+2,\ldots,m. \tag{3.72}$$

Let $q_m := m - l_m - 1$ and note that at most

$$\frac{\log(3\gamma/2)}{\log(1-1/\sqrt{q_m})} \leq \underbrace{-\log(3\gamma/2)}_{=:c_\gamma>0}\sqrt{q_m} \tag{3.73}$$

elements of the sequence r_ℓ (out of q_m) can be smaller than $1 - 1/\sqrt{q_m}$. Otherwise,

$$\rho(\mathsf{s}^m, \mathsf{t}^m) = \rho(\mathsf{s}^{l_m}, \mathsf{t}^{l_m}) \cdot \prod_{\ell=l_m+2}^{m} r_\ell < \frac{1}{3} \cdot \frac{3\gamma}{2} = \frac{1}{2}\gamma, \tag{3.74}$$

which contradicts (3.71). These elements partition the interval $\{l_m+2, \dots, m\}$ into at most $c_\gamma\sqrt{q_m}$ segments, one of which must have length at least $c_\gamma^{-1}\sqrt{q_m}$. The proof of Lemma 3.7 shows that on this segment the sign sequence s^m must be constant because, on this segment, $r_\ell > 1 - 1/\sqrt{q_m}$. The set of sequences of length q_m that have a run of the same sign for an interval of length $c_\gamma^{-1}\sqrt{q_m}$ has probability at most $2q_m 2^{-c_\gamma^{-1}\sqrt{q_m}}$. However, by assumption $q_m = m - l_m - 1$ goes to infinity as m gets large. Therefore, the probability of having such S^m sequences is arbitrarily small, provided that m is large enough. □

Lemma 3.14. *Let* $l_m := \lceil 4\log\bigl(2(1+\alpha)m - 1\bigr)\rceil$. *Then,* $\forall\alpha \geq 0$ *and* $\forall\delta > 0$, $\exists m_2(\alpha, \delta)$ *such that* $\forall m \geq m_2$,

$$\Pr\bigl\{\rho^{(2)}_{m,l_m} \leq 2^{-(1+\alpha)m}\bigr\} \geq 1 - \delta/2. \tag{3.75}$$

Proof. Fix $\gamma > 0$ (to be chosen later). With our choice of l_m, according to Lemma 3.13, $\exists m_s(\gamma, \delta)$, such that $\forall m \geq m_s$,

$$\Pr\bigl\{\rho^{(2)}_{m/2,l_m} \leq \gamma/2\bigr\} \geq 1 - \delta/6. \tag{3.76}$$

Moreover, the polarization of Z^{S^m} process (see [6, Lemma 2]) implies $\exists m_s'(\gamma, \delta)$ such that $\forall m \geq m_s'(\gamma, \delta)$,

$$\Pr\bigl\{\forall \ell \geq m/2\colon Z^{\mathsf{S}^\ell} \notin [3\gamma^2/8, 1 - 3\gamma^2/8]\bigr\} \geq 1 - \delta/6. \tag{3.77}$$

Finally, for $\ell = m/2+1, \dots, m$, let $D_\ell := \min\{Z^{\mathsf{S}^\ell}, \overline{Z^{\mathsf{S}^\ell}}\}$ for the sake of brevity, and

$$\tilde{\mathsf{S}}_\ell := \begin{cases} + & \text{if } Z^{\mathsf{S}^{\ell-1}} \leq \frac{1}{2}, \\ - & \text{if } Z^{\mathsf{S}^{\ell-1}} \geq \frac{1}{2} \end{cases} \tag{3.78}$$

Since $\tilde{\mathsf{S}}_\ell$ depends only on $\mathsf{S}^{\ell-1}$,

$$B_\ell := \mathbb{1}\{\tilde{\mathsf{S}}_\ell = \mathsf{S}_\ell\}, \qquad \ell = \frac{m}{2} + 1, \frac{m}{2} + 2, \dots, m. \tag{3.79}$$

are i.i.d. Bernoulli($\frac{1}{2}$) random variables. Consequently, by the weak law of large numbers, $\exists m_b(\delta)$ such that $\forall m \geq m_b$,

$$\Pr\biggl\{\frac{1}{m/2}\sum_{\ell=m/2+1}^{m} B_\ell \geq \frac{1}{4}\biggr\} \geq 1 - \delta/6. \tag{3.80}$$

Let $\mathsf{t}_\star^m = \mathsf{t}_\star^m(\mathsf{S}^m)$ be the sign sequence for which the maximum in (3.57) is attained. We have

$$\begin{aligned}
\log\bigl(\rho^{(2)}_{m,l_m}\bigr) &= \log\bigl[\rho(\mathsf{S}^{m/2}, \mathsf{t}_\star^{m/2})\bigr] + \sum_{\ell=m/2}^{m-1} \log\Bigl[\frac{\rho(\mathsf{S}^{\ell+1}, \mathsf{t}_\star^{\ell+1})}{\rho(\mathsf{S}^{\ell}, \mathsf{t}_\star^{\ell})}\Bigr] && (3.81)\\
&\overset{\text{(a)}}{\leq} \sum_{\ell=m/2}^{m-1} \log\Bigl[\frac{\rho(\mathsf{S}^{\ell+1}, \mathsf{t}_\star^{\ell+1})}{\rho(\mathsf{S}^{\ell}, \mathsf{t}_\star^{\ell})}\Bigr] && (3.82)\\
&\overset{\text{(b)}}{\leq} \sum_{\ell=m/2}^{m-1} \log\Bigl[\frac{\rho(\mathsf{S}^{\ell+1}, \mathsf{t}_\star^{\ell+1})}{\rho(\mathsf{S}^{\ell}, \mathsf{t}_\star^{\ell})}\Bigr] B_{\ell+1} && (3.83)
\end{aligned}$$

where (a) follows because $\rho(\mathsf{S}^{m/2}, \mathsf{t}_\star^{m/2}) \leq 1$ and (b) as $\rho(\mathsf{S}^{\ell+1}, \mathsf{t}_*^{\ell+1})/\rho(\mathsf{S}^{\ell}, \mathsf{t}_*^{\ell}) \leq 1$ (due to (ii) of Lemma 3.6), thus excluding some terms from the summation can only increase its value.

Inspecting (3.68), we see that if $B_\ell = 1$,

$$\begin{aligned}
\frac{\rho(\mathsf{S}^{\ell+1}, \mathsf{t}_*^{\ell+1})}{\rho(\mathsf{S}^{\ell}, \mathsf{t}_*^{\ell})} &\leq \sqrt{\rho(\mathsf{S}^{\ell}, \mathsf{t}_*^{\ell})^2 \frac{\overline{D_\ell}}{1+D_\ell} + \frac{2D_\ell}{1+D_\ell}} && (3.84)\\
&\leq \sqrt{\rho(\mathsf{S}^{\ell}, \mathsf{t}_*^{\ell})^2 + 2D_\ell} && (3.85)\\
&\leq \sqrt{\bigl(\rho^{(2)}_{m/2,l_m}\bigr)^2 + 2D_\ell}. && (3.86)
\end{aligned}$$

Using (3.76) and (3.77), we see that, for $m \geq \max\{m_s, m_s'\}$, with probability at least $1-\delta/3$,

$$\sqrt{\bigl(\rho^{(2)}_{m/2,l_m}\bigr)^2 + 2D_\ell} \leq \gamma. \tag{3.87}$$

Further, (3.80) shows that, for $m \geq m_b$ with probability at least $1-\delta/6$,

$$\sum_{\ell=m/2+1}^{m} B_\ell \geq \frac{m}{8}. \tag{3.88}$$

Therefore, continuing (3.83), we conclude that for $m \geq \max\{m_s, m_s', m_b\}$, with probability $1-\delta/2$,

$$\log \rho^{(2)}_{m,l_m} \leq -m\log(1/\gamma)/8. \tag{3.89}$$

Choosing $\gamma = 2^{-8(1+\alpha)}$ ensures that, for $m \geq m_1(\alpha, \delta) := \max\{m_s, m_s', m_b\}$, with probability at least $1-\delta/2$,

$$\rho^{(2)}_{m,\ell_m} \leq 2^{-m(1+\alpha)}.$$

□

3.4 Summary and Extensions

In this chapter, we have studied the correlations between the synthetic erasure channels and have shown that, even though after a single step of polar transform two highly correlated erasure channels are synthesized from two independent copies of an erasure channel, almost all correlation coefficients between the erasure events of 2^m erasure channels (after the m-fold application of Arıkan's polar transform) decay faster than any exponential in m. Consequently, by using the inclusion–exclusion lower bound, we conclude that the union bound on the block-error probability of polar codes under successive-cancellation decoding is a tight estimate of the actual block-error probability, when communication takes place over the binary erasure channel. Our numerical examples also confirm that the inclusion–exclusion lower bound on the block-error probability of polar codes of different rates and block-lengths is indeed very close to the union bound.

An important implication of our result is that the optimal choice for information indices is to select those indexing the synthetic channels with the lowest bit-error probability. Such a choice minimizes the upper bound of (2.29), but whether it minimizes the actual block-error probability of the code was not a priori clear. Our results show that, at least for communication over the erasure channel, such a design rule for the code is, indeed, optimal.

It is noteworthy that our results are immediately extensible to the case of communication with q-ary polar codes over a q-ary erasure channel, namely the channel $W : \mathbb{F}_q :\to \mathcal{Y}$ with transition probabilities

$$W(y|x) = \begin{cases} p & \text{if } y = ?, \\ 1-p & \text{if } y \neq ? \text{ and } y = x. \end{cases} \tag{3.90}$$

It can easily be checked that, exactly like the binary case, in this case both W^- and W^+ are *effectively* q-ary erasure channels; and W^- erases if either copy of W erase, whereas W^+ erases if both copies erase.[3] Thus, the same correlation structure between the erasure events of the synthetic q-ary erasure channels exists and our conclusions will remain valid for non-binary erasure channels as well.

It is important to extend the results of this chapter to communication over channels other than BEC. This is a challenging open problem. The main difficulty here is that, as we discussed in § 2.4, apart from the set of erasure channels, no other class of easily described channels is known to be stable under Arıkan's polar transform. Consequently, it would be impractical to keep track of the probability of joint error events after the repeated application of the

[3] Here we assumed that the polar transform is applied in the same way as the binary case, replacing the XOR operation the modulo-q addition. According to [99], for general q-ary channels using such a polar transform, polarization — in the sense of converging to capacity-0 and capacity-log(q) channels — happens only when q is prime. However, for the special case of erasure channels polarization does happen even if q is not prime.

polar transform to a channel other than BEC. In fact, to do so, we would need to calculate the joint distributions of pairs of log-likelihood ratios of synthetic channels recursively, which is computationally intractable, due to the exponential growth of the output alphabet of the synthetic channels (see § 2.4).

One tempting experiment would be to estimate the correlation coefficients between the pairs of error events via computer simulations. The problem there is that, especially when the block-length is large, most of the channels become well-polarized and the probability of joint bit-error events will be misestimated to be 0 (while it is actually a very small value). This renders the estimations meaningless.

An alternative approach and one perhaps more interesting from a practical perspective would be to apply the same method of [109] here: The inclusion–exclusion lower bound on the block-error probability of polar codes under successive-cancellation decoding involves the probability of individual error events and the probability of joint error events (and the latter appears with a negative sign in the expression). The approximation method proposed in [109] readily gives us tight lower bounds on individual bit-error probabilities of the synthetic channels. We still need to study how the probability of joint bit-error events of the synthetic channels change after their output symbols are merged and to develop computationally efficient methods to compute upper bounds on the probability of joint bit-error events of synthetic channels accordingly. While writing these paragraphs, it came to our attention that such an approach is pursued in [104] to compute lower bounds on the block-error probability of polar codes under successive-cancellation decoding for any BMS channel.

3.A Proof of Lemma 3.1

The upper bound of (3.12) is trivial. To prove the lower bound we use the inclusion–exclusion principle,

$$\Pr(\mathcal{E}) \geq \sum_{j=1}^{k} \Pr(\mathcal{E}_j) - \frac{1}{2}\sum_{i \neq j} \Pr(\mathcal{E}_i \cap \mathcal{E}_j). \tag{3.91}$$

Moreover, for $i \neq j$,

$$\Pr(\mathcal{E}_i \cap \mathcal{E}_j) = \Pr(\mathcal{E}_i)\Pr(\mathcal{E}_j) + \rho(i,j)\sqrt{\Pr(\mathcal{E}_i)\big(1-\Pr(\mathcal{E}_i)\big)}\sqrt{\Pr(\mathcal{E}_j)\big(1-\Pr(\mathcal{E}_j)\big)} \tag{3.92}$$

$$\leq \Pr(\mathcal{E}_i)\Pr(\mathcal{E}_j) + \rho_{\max}\sqrt{\Pr(\mathcal{E}_i)}\sqrt{\Pr(\mathcal{E}_j)} \tag{3.93}$$

where we have defined $\rho_{\max} = \max_{i \neq j} \rho(i,j)$ for notational brevity. Consequently,

$$\sum_{i \neq j} \Pr(\mathcal{E} \cap \mathcal{E}_j) \leq \sum_{i \neq j} \Pr(\mathcal{E}_i)\Pr(\mathcal{E}_j) + \rho_{\max}\sum_{i \neq j} \sqrt{\Pr(\mathcal{E}_i)}\sqrt{\Pr(\mathcal{E}_j)} \tag{3.94}$$

$$\overset{(a)}{\leq} \left(\sum_{j=1}^{k} \Pr(\mathcal{E}_j)\right)^2 + \rho_{\max}\left(\sum_{j=1}^{k} \sqrt{\Pr(\mathcal{E}_j)}\right)^2 \tag{3.95}$$

$$\overset{(b)}{\leq} \left(\sum_{j=1}^{k} \Pr(\mathcal{E}_j)\right)^2 + k\rho_{\max}\sum_{j=1}^{k} \Pr(\mathcal{E}_j) \tag{3.96}$$

where (a) follows by including the terms corresponding to $i = j$ in the double sums and (b) from convexity of $s \mapsto s^2$. Using (3.96) in (3.91), we have

$$\Pr(\mathcal{E}) \geq \left[1 - \frac{1}{2}k\rho_{\max}\right]\sum_{j=1}^{k} \Pr(\mathcal{E}_j) - \frac{1}{2}\left(\sum_{j=1}^{k} \Pr(\mathcal{E}_j)\right)^2, \tag{3.97}$$

which, for $\sum_{j=1}^{k} \Pr(\mathcal{E}_j) \leq 1$ can be further lower-bounded as

$$\begin{aligned}\Pr(\mathcal{E}) &\geq \left[1 - \frac{1}{2}k\rho_{\max}\right]\sum_{j=1}^{k} \Pr(\mathcal{E}_j) - \frac{1}{2}\sum_{j=1}^{k} \Pr(\mathcal{E}_j)\\ &= \frac{1}{2}[1 - k\rho_{\max}]\sum_{j=1}^{k} \Pr(\mathcal{E}_j)\end{aligned}$$

□

3.B Second Order Statistics of the Erasure Events

In this section we prove Lemma 3.2 and Corollary 3.3. To this end, the following results turn out to be useful:

Lemma 3.15. *For two arbitrary events $\mathcal{A}$ and $\mathcal{B}$, let* $\operatorname{cov}(\mathcal{A}, \mathcal{B}) := \Pr(\mathcal{A} \cap \mathcal{B}) - \Pr(\mathcal{A})\Pr(\mathcal{B})$. *Then,*

$$\operatorname{cov}(\mathcal{A}^c, \mathcal{B}^c) = \operatorname{cov}(\mathcal{A}, \mathcal{B}) \tag{3.98a}$$

$$\operatorname{cov}(\mathcal{A}, \mathcal{B}^c) = \operatorname{cov}(\mathcal{A}^c, \mathcal{B}) = -\operatorname{cov}(\mathcal{A}, \mathcal{B}) \tag{3.98b}$$

Proof. Obviously, it is sufficient to prove that

$$\operatorname{cov}(\mathcal{A}^c, \mathcal{B}) = -\operatorname{cov}(\mathcal{A}, \mathcal{B}). \tag{3.99}$$

The rest of the claims follow from (3.99) using the symmetry of $\operatorname{cov}(\cdot, \cdot)$. By the law of total probability,

$$\Pr(\mathcal{A}^c \cap \mathcal{B}) = \Pr(\mathcal{B}) - \Pr(\mathcal{A} \cap \mathcal{B}). \tag{3.100}$$

Therefore

$$\begin{aligned}\operatorname{cov}(\mathcal{A}^c, \mathcal{B}) &= \Pr(\mathcal{A}^c \cap \mathcal{B}) - \Pr(\mathcal{A}^c)\Pr(\mathcal{B}) &(3.101)\\ &= \Pr(\mathcal{B}) - \Pr(\mathcal{A} \cap \mathcal{B}) - [1 - \Pr(\mathcal{A})]\Pr(\mathcal{B}) &(3.102)\\ &= \Pr(\mathcal{A})\Pr(\mathcal{B}) - \Pr(\mathcal{A} \cap \mathcal{B}), &(3.103)\end{aligned}$$

which establishes (3.99). □

Corollary 3.16. *Let the correlation coefficient between pair of events $\mathcal{A}$ and $\mathcal{B}$ be defined as*

$$\rho(\mathcal{A},\mathcal{B}) \coloneqq \frac{\operatorname{cov}(\mathcal{A},\mathcal{B})}{\sqrt{\Pr(\mathcal{A})[1-\Pr(\mathcal{A})]}\sqrt{\Pr(\mathcal{B})[1-\Pr(\mathcal{B})]}}. \tag{3.104}$$

Then

$$\begin{aligned} \rho(\mathcal{A}^c,\mathcal{B}^c) &= \rho(\mathcal{A},\mathcal{B}) && (3.105\text{a}) \\ \rho(\mathcal{A},\mathcal{B}^c) = \rho(\mathcal{A}^c,\mathcal{B}) &= -\rho(\mathcal{A},\mathcal{B}). && (3.105\text{b}) \end{aligned}$$

3.B.1 Proof of Lemma 3.2

We first prove (3.17d) and then show how the rest of results easily follow by using Lemma 3.15.

Recall that $\mathcal{E}^{\mathsf{s}^m+} = \mathcal{E}_1^{\mathsf{s}^m} \cap \mathcal{E}_2^{\mathsf{s}^m}$ (and $\mathcal{E}^{\mathsf{t}^m+} = \mathcal{E}_1^{\mathsf{t}^m} \cap \mathcal{E}_2^{\mathsf{t}^m}$).

$$\begin{aligned} C(\mathsf{s}^m+,\mathsf{t}^m+) &= \Pr\bigl(\mathcal{E}_1^{\mathsf{s}^m} \cap \mathcal{E}_2^{\mathsf{s}^m} \cap \mathcal{E}_1^{\mathsf{t}^m} \cap \mathcal{E}_2^{\mathsf{t}^m}\bigr) - \Pr\bigl(\mathcal{E}_1^{\mathsf{s}^m} \cap \mathcal{E}_2^{\mathsf{s}^m}\bigr)\Pr\bigl(\mathcal{E}_1^{\mathsf{t}^m} \cap \mathcal{E}_2^{\mathsf{t}^m}\bigr) && (3.106) \\ &\overset{(*)}{=} \Pr(\mathcal{E}^{\mathsf{s}^m} \cap \mathcal{E}^{\mathsf{t}^m})^2 - \Pr(\mathcal{E}^{\mathsf{s}^m})^2 \Pr(\mathcal{E}^{\mathsf{t}^m})^2 && (3.107) \\ &= \bigl(\Pr(\mathcal{E}^{\mathsf{s}^m} \cap \mathcal{E}^{\mathsf{t}^m}) - Z^{\mathsf{s}^m} Z^{\mathsf{t}^m}\bigr)^2 \\ &\qquad + 2Z^{\mathsf{s}^m} Z^{\mathsf{t}^m}\bigl[\Pr(\mathcal{E}^{\mathsf{s}^m} \cap \mathcal{E}^{\mathsf{t}^m}) - Z^{\mathsf{s}^m} Z^{\mathsf{t}^m}\bigr] && (3.108) \\ &= C(\mathsf{s}^m,\mathsf{t}^m)^2 + 2Z^{\mathsf{s}^m} Z^{\mathsf{t}^m} C(\mathsf{s}^m,\mathsf{t}^m), && (3.109) \end{aligned}$$

where in $(*)$ we have used the fact that the $(\mathcal{E}_1^{\mathsf{s}^m}, \mathcal{E}_2^{\mathsf{s}^m})$ (respectively $(\mathcal{E}_1^{\mathsf{t}^m}, \mathcal{E}_2^{\mathsf{t}^m})$) are independent events with equal probability).

Now, observe that (3.98a) implies

$$C(\mathsf{s}^m-,\mathsf{t}^m-) = \operatorname{cov}\bigl(\mathcal{E}^{\mathsf{s}^m-}, \mathcal{E}^{\mathsf{t}^m-}\bigr) = \operatorname{cov}\bigl((\mathcal{E}^{\mathsf{s}^m-})^c, (\mathcal{E}^{\mathsf{t}^m-})^c\bigr). \tag{3.110}$$

Moreover, since $(\mathcal{E}^{\mathsf{s}^m-})^c = (\mathcal{E}_1^{\mathsf{s}^m})^c \cap (\mathcal{E}_2^{\mathsf{s}^m})^c$ (resp. $(\mathcal{E}^{\mathsf{t}^m-})^c = (\mathcal{E}_1^{\mathsf{t}^m})^c \cap (\mathcal{E}_2^{\mathsf{t}^m})^c$) due to (3.13a), and $\operatorname{cov}\bigl((\mathcal{E}^{\mathsf{s}^m})^c, (\mathcal{E}^{\mathsf{t}^m})^c\bigr) = C(\mathsf{s}^m,\mathsf{t}^m)$ because of (3.98a), following the same lines as (3.109), shows

$$\operatorname{cov}\bigl((\mathcal{E}^{\mathsf{s}^m-})^c, (\mathcal{E}^{\mathsf{t}^m-})^c\bigr) = C(\mathsf{s}^m,\mathsf{t}^m)^2 + 2\overline{Z^{\mathsf{s}^m} Z^{\mathsf{t}^m}} C(\mathsf{s}^m,\mathsf{t}^m). \tag{3.111}$$

The above, together with (3.110) establish (3.17a).

Similarly,

$$C(\mathsf{s}^m-,\mathsf{t}^m+) = \operatorname{cov}\bigl(\mathcal{E}^{\mathsf{s}^m-}, \mathcal{E}^{\mathsf{t}^m+}\bigr) = -\operatorname{cov}\bigl((\mathcal{E}^{\mathsf{s}^m-})^c, \mathcal{E}^{\mathsf{t}^m+}\bigr). \tag{3.112}$$

and

$$\operatorname{cov}\bigl((\mathcal{E}^{\mathsf{s}^m-})^c, \mathcal{E}^{\mathsf{t}^m+}\bigr) = C(\mathsf{s}^m,\mathsf{t}^m)^2 - 2\overline{Z^{\mathsf{s}^m}} Z^{\mathsf{t}^m} C(\mathsf{s}^m,\mathsf{t}^m) \tag{3.113}$$

(because $\operatorname{cov}\bigl(\mathcal{E}^{\mathsf{s}^m}, (\mathcal{E}^{\mathsf{t}^m})^c\bigr) = -C(\mathsf{s}^m,\mathsf{t}^m)$) which prove (3.17b). Equation (3.17c) can also be established in the same way. □

3.B.2 Proof of Corollary 3.3

Once again we first prove (3.18d). Setting

$$C(\mathsf{s}^m, \mathsf{t}^m) = \rho(\mathsf{s}^m, \mathsf{t}^m)\sqrt{Z^{\mathsf{s}^m}\overline{Z^{\mathsf{s}^m}}}\sqrt{Z^{\mathsf{t}^m}\overline{Z^{\mathsf{t}^m}}} \tag{3.114}$$

in both sides of (3.17d) and using the fact that $Z^{\mathsf{s}^m+} = (Z^{\mathsf{s}^m})^2$ (similarly $Z^{\mathsf{t}^m+} = (Z^{\mathsf{t}^m})^2$), we get:

$$\begin{aligned}\rho(\mathsf{s}^m+, \mathsf{t}^m+)&\sqrt{(Z^{\mathsf{s}^m})^2\overline{(Z^{\mathsf{s}^m})^2}(Z^{\mathsf{t}^m})^2\overline{(Z^{\mathsf{t}^m})^2}} \\ &= 2Z^{\mathsf{s}^m}Z^{\mathsf{t}^m}\sqrt{Z^{\mathsf{s}^m}\overline{Z^{\mathsf{s}^m}}Z^{\mathsf{t}^m}\overline{Z^{\mathsf{t}^m}}}\rho(\mathsf{s}^m, \mathsf{t}^m) + \left(Z^{\mathsf{s}^m}\overline{Z^{\mathsf{s}^m}}Z^{\mathsf{t}^m}\overline{Z^{\mathsf{t}^m}}\right)\rho(\mathsf{s}^m, \mathsf{t}^m)^2.\end{aligned} \tag{3.115}$$

Eliminating $Z^{\mathsf{s}^m}Z^{\mathsf{t}^m}$ from both sides of the above we get,

$$\begin{aligned}\rho(\mathsf{s}^m+, \mathsf{t}^m+)&\sqrt{\overline{(Z^{\mathsf{s}^m})^2}\cdot\overline{(Z^{\mathsf{t}^m})^2}} \\ &= 2\sqrt{Z^{\mathsf{s}^m}\overline{Z^{\mathsf{s}^m}}Z^{\mathsf{t}^m}\overline{Z^{\mathsf{t}^m}}}\rho(\mathsf{s}^m, \mathsf{t}^m) + \left(\overline{Z^{\mathsf{s}^m}}\,\overline{Z^{\mathsf{t}^m}}\right)\rho(\mathsf{s}^m, \mathsf{t}^m)^2,\end{aligned} \tag{3.116}$$

which yields (3.18d). Equations (3.18a)–(3.18c) follow using Corollary 3.16 and the symmetry between 'plus' and 'minus' transforms. □

3.C Proof of Equation (3.68)

An application of Cauchy–Schwarz inequality, slightly different than what we did in the proof of (ii) in Lemma 3.6, to the right-hand side of (3.28d) yields

$$\frac{\rho(\mathsf{s}^m+, \mathsf{t}^m+)}{\rho(\mathsf{s}^m, \mathsf{t}^m)} = 2\sqrt{\frac{Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{Z^{\mathsf{t}^m}}{1+Z^{\mathsf{t}^m}}} + \sqrt{\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}}}\sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}}}\rho(\mathsf{s}^m, \mathsf{t}^m) \tag{3.117}$$

$$\le \sqrt{\frac{\overline{Z^{\mathsf{t}^m}}}{1+Z^{\mathsf{t}^m}} + \frac{2Z^{\mathsf{t}^m}}{1+Z^{\mathsf{t}^m}}}\cdot\sqrt{\rho(\mathsf{s}^m, \mathsf{t}^m)^2\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}} + \frac{2Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}} \tag{3.118}$$

$$= \sqrt{\rho(\mathsf{s}^m, \mathsf{t}^m)^2\frac{\overline{Z^{\mathsf{s}^m}}}{1+Z^{\mathsf{s}^m}} + \frac{2Z^{\mathsf{s}^m}}{1+Z^{\mathsf{s}^m}}} \tag{3.119}$$

Likewise, using (3.28a), we get

$$\frac{\rho(\mathsf{s}^m-, \mathsf{t}^m-)}{\rho(\mathsf{s}^m, \mathsf{t}^m)} \le \sqrt{\rho(\mathsf{s}^m, \mathsf{t}^m)^2\frac{Z^{\mathsf{s}^m}}{1+\overline{Z^{\mathsf{s}^m}}} + \frac{2\overline{Z^{\mathsf{s}^m}}}{1+\overline{Z^{\mathsf{s}^m}}}}. \tag{3.120}$$

Also, the right-hand side of (3.28c) is upper-bounded by $\sqrt{2Z^{\mathsf{s}^m}/(1+Z^{\mathsf{s}^m})}$ that is in turn smaller than (3.119). Similarly, the right-hand side of (3.28b) is upper-bounded by $\sqrt{2\overline{Z^{\mathsf{s}^m}}/(1+\overline{Z^{\mathsf{s}^m}})}$ that is in turn smaller than (3.120). □

4 Efficient Implementation of Polar List Decoder

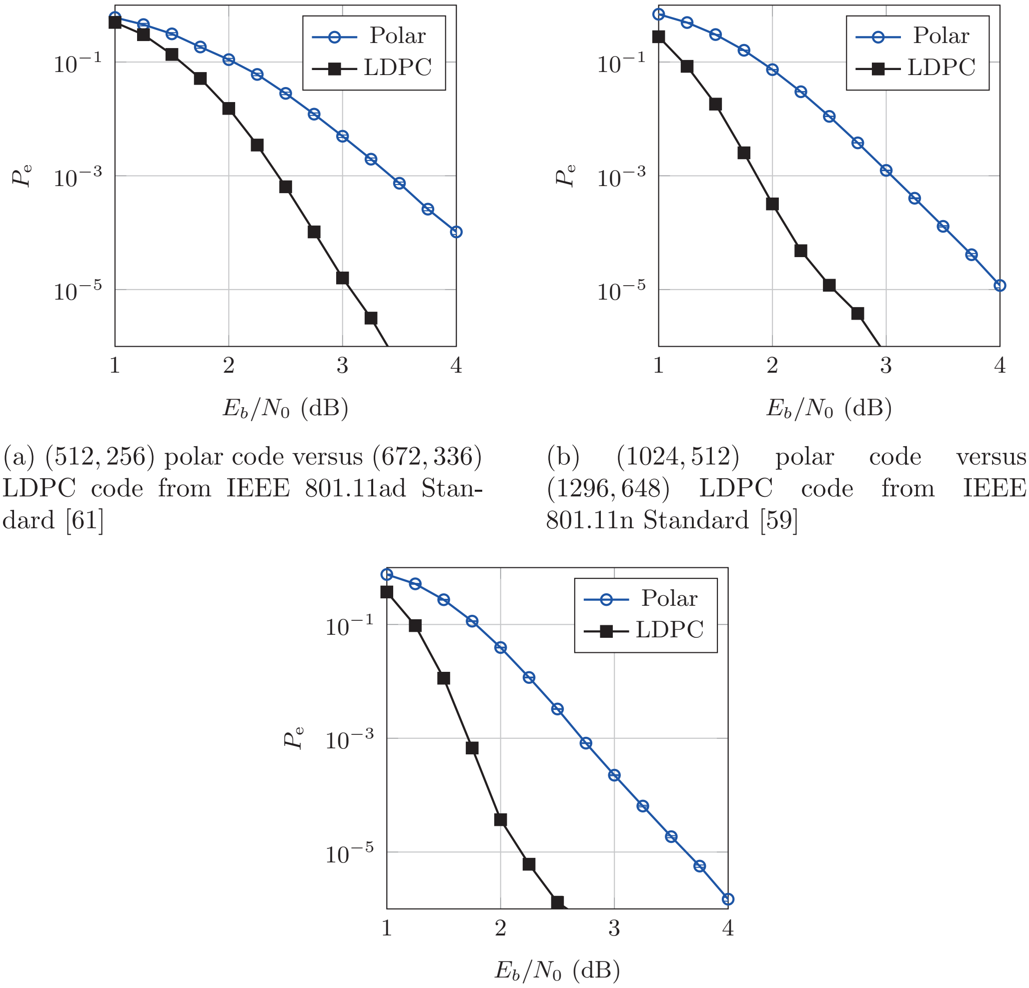
Even though the block-error probability of polar codes, when decoded with a low-complexity successive-cancellation decoder, scales roughly like $O(2^{-\sqrt{n}})$ (where $n = 2^m$ is the block-length of the code), short- to moderate-length polar codes do not perform very well compared to existing coding schemes. For example, in Figure 4.1 we compare the performance of polar codes of block-lengths $n = 512$, $n = 1024$, and $n = 2048$ and rate 1/2 with LDPC codes of similar block-length and same rates, when used for communication with BPSK signalling over AWGN channel.[1] We see that to guarantee a block-error probability of 10^{-3} polar codes require, roughly speaking, about 1 dB of higher energy per bit. This gap increases to more than a dB if we aim for a block-error probability of 10^{-5}.

This poor performance is due to two factors: the suboptimality of the successive-cancellation decoder (as we discussed in § 2.3.1), and the intrinsic weakness of short polar codes.[2]

Shortly after the introduction of polar codes, *successive-cancellation list* (SCL) decoding of polar codes emerged as an efficient solution for improving the performance of short polar codes [108]. As we will see in § 4.1, successive-cancellation list decoding has a complexity of $O(L \cdot n \log n)$, where L, the *list size,* is a parameter that enables the complexity–performance trade-off of the algorithm: with $L = 1$ the algorithm reduces to the conventional successive-cancellation decoding, and as L increases the algorithm improves upon the successive-cancellation decoding (in return for a slightly higher decoding com-

[1] We saw in Chapter 2 that polar codes are channel specific. In all examples of this chapter, the polar codes are constructed for a BI-AWGNC with noise variance $\sigma^2 = 0.63096$ using the Monte Carlo method of [5, Section IX]. For the code rate of 1/2 this corresponds to $E_b/N_0 = 2$ dB.

[2] As we will see later in this chapter, even under optimal MAP decoding, short- to moderate-length polar codes do not exhibit a good performance.



(a) $(512, 256)$ polar code versus $(672, 336)$ LDPC code from IEEE 801.11ad Standard [61]

(b) $(1024, 512)$ polar code versus $(1296, 648)$ LDPC code from IEEE 801.11n Standard [59]

(c) $(2048, 1024)$ polar code versus $(2304, 1152)$ LDPC code from IEEE 801.16 Standard [60]

Figure 4.1: Performance Comparison between Polar and LDPC Codes

plexity). Numerical evidence show that under successive-cancellation list decoding with a relatively small list size, the block-error probability of polar codes approaches that under the optimal MAP decoder. Hence successive-cancellation list decoding compensates for the suboptimality of the successive-cancellation decoder. Furthermore, successive-cancellation list decoding enables us to use *modified polar codes* by concatenating a polar code with a cyclic redundancy check (CRC) code as the outer code [85, 110]. Adding the CRC does not increase either the computational complexity of the encoder or that of the decoder by any notable amount, but it does significantly reduce the block-error probability. This makes polar codes competitive with existing codes (cf. § 4.1.2).

The original successive-cancellation list-decoding algorithm in [108] is formulated in terms of likelihood ratios, which makes the decoder's computations prone to underflow errors. Even if this issue is circumvented by normalizing the intermediate computation results (as proposed in [110]), the update rules in the likelihood domain involve multiplications and divisions that are expensive for a hardware implementation of the decoder (see § 2.3.2). We show in § 4.2 that the successive-cancellation list-decoding algorithm can — similarly to conventional SC decoding — be formulated exclusively in the *log-likelihood ratio* (LLR) domain. Such a formulation, in particular, results in a much more efficient hardware implementation of the decoder [8].

As we will discuss in § 4.1.2, in the design of CRC-concatenated polar codes, the length of CRC is a crucial design parameter that has to be chosen correctly. In § 4.3.1, we provide examples to illustrate the trade-off between the length of CRC and the performance of the code and show that, to obtain the best possible performance, the length of CRC should be carefully tuned based on the list size.

Another important question in the design of polar codes for error correction is the choice between a long polar code with the conventional low-complexity SC decoder or a short polar code with the successive-cancellation list decoder (which requires relatively more computational resources). In § 4.3.2, we compare the block-error probability of polar codes of block-lengths $n = 4096$, $n = 8192$, $n = 16384$, and $n = 32768$ under SC decoding to that of short modified polar codes (of block lengths $n = 512$, $n = 1024$, and $n = 2048$) that are decoded using a successive-cancellation list decoder with a list size chosen such that the computational complexity of all decoders are almost the same. From these examples, we conclude that to achieve a target block-error probability, and under the same per-codeword computational cost, successive-cancellation list decoding, together with CRC-concatenation, allows a decrease of the block-length by a factor of eight (hence reduces the per-codeword decoding latency).

Part of the material presented in this chapter is based on joint work with A. Balatsoukas-Stimming and A. P. Burg [7,8].

4.1 List Decoding of Polar Codes

To understand the successive-cancellation list-decoding algorithm, it is useful to picture the set of all 2^k, $k = |\mathcal{I}|$, possible source-plus-frozen bits

$$\mathcal{U}_n \coloneqq \{u^n \in \mathbb{F}_2^n \colon u_{\mathcal{F}} \text{ is fixed}\} \tag{4.1}$$

as 2^k paths from the root to the leaves of a full binary tree of height n (see Figure 4.2).

Any decoding algorithm is hence a tree-search algorithm that picks a path on the tree that corresponds to the set of all possible u^n sequences. Specifically, upon observing the channel output y^n, the optimal MAP decoder would associate each path with its *posterior probability*, $\Pr\{U^n = u^n | Y^n = y^n\}$, or

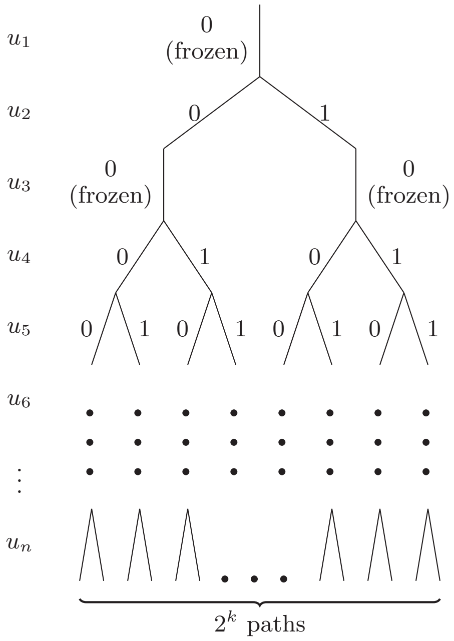


Figure 4.2: Each $u^n \in \mathcal{U}_n \subset \mathbb{F}_2^n$ corresponds to one of 2^k paths on a binary tree.

any other *score* that is a monotone function of its posterior probability, and chooses the path with the highest score:

$$\hat{u}^N_{\text{MAP}} = \underset{u^n \in \mathcal{U}_n}{\arg\max} \Pr\{U^n = u^n | Y^n = y^n\}. \tag{4.2}$$

Obviously, implementing the MAP decoder is infeasible as its computational complexity grows exponentially fast with the block-length.

In contrast, the low-complexity successive-cancellation decoder, finds a suboptimal solution by maximizing the likelihood via a *greedy* one-time pass through the tree: starting from the root, at each level $i \in \mathcal{I}$, the decoder extends the existing path by choosing the child that maximizes the *partial likelihood*:

$$\hat{u}_i = \underset{u_i \in \mathbb{F}_2}{\arg\max} W_n^{(i)}(y^n, \hat{u}_1^{i-1} | u_i). \tag{4.3}$$

(See Figure 4.3.)

Furthermore, observe that in the last step of SC decoding, the likelihood value $W_n^{(n-1)}(y^n, \hat{u}^{n-1}|\hat{u}_n)$ is proportional to the posterior probability $\Pr\{U^n = \hat{u}^n | Y^n = y^n\}$ as we have assumed the information bits are i.i.d. coin flips:

$$\begin{aligned} W_n^{(n-1)}(y^n, \hat{u}^{n-1}|\hat{u}_n) &= 2\Pr\{U^n = \hat{u}^n, Y^n = y^n\} && (4.4) \\ &= 2\Pr\{U^n = \hat{u}^n | Y^n = y^n\} \Pr\{Y^n = y^n\}. && (4.5) \end{aligned}$$

Consequently, in principle, one way to realize an optimal MAP decoder for polar codes would be to allow the decoder to *duplicate* in two *decoding threads*, whenever it reaches an information index, and to descend the tree in both paths

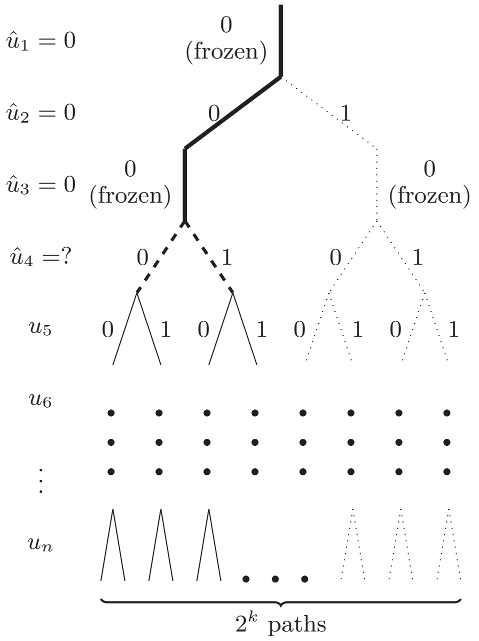


Figure 4.3: SC Decoder chooses a path on the tree in a greedy single pass through the tree: In this example $\hat{u}_4 = \arg\max_{u \in \mathbb{F}_2} W_n^{(4)}(y^n, \hat{u}^3|u)$.

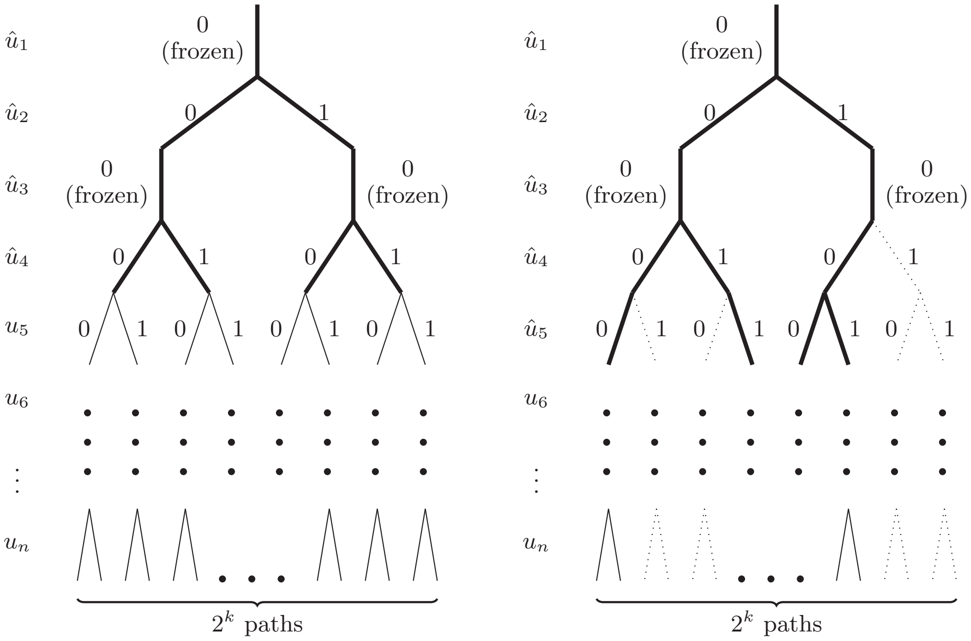
in parallel — instead of forcing the decoder to choose one direction to follow. This way, the decoder would eventually end up with 2^k decoding threads; each of them corresponding to one of the sequences in $\mathcal{U}_n$ and containing a score proportional to the posterior probability of that particular sequence.

Successive-cancellation list (SCL) decoding [108] interpolates between the suboptimal (but efficient) approach of the conventional successive-cancellation decoder and the optimal (but impractical) MAP decoder by letting the decoder descend the tree in L parallel paths: Starting from the root of the tree, at each level corresponding to an information index, the decoder is duplicated into two decoding threads that descend in either possible direction. However, in order to avoid the exponential growth of the number of decoding threads, as soon as the number of parallel threads reaches L, at each level, only the L threads corresponding to the L most likely paths (out of $2L$ tentative ones) are retained.[3] In other words, successive-cancellation list decoding converts the greedy one-time-pass search of SC decoding into a *breadth-first search* under a complexity constraint. The decoder will eventually end up with a list of L possible candidates for the sent u^n sequence, that we denote as $\{\hat{u}^n[\ell], \ell = 1, 2, \ldots, L\}$, among which the most likely one is declared as the final estimate (see Figure 4.4).[4] This procedure is formalized in Algorithm 3.

[3] Although it is not necessary, L is normally a power of 2

[4] Note that the output of successive-cancellation list decoder is a single codeword and not a list of codewords — unlike the conventional list decoder introduced by Elias and Wozencraft [39, 115].

At this point we should mention that, before the advent of polar codes, both successive-cancellation decoding and successive-cancellation list-decoding algorithms were proposed in [35, 36] for decoding the Reed–Muller codes.[5]



(a) For the first few steps the decoder only expands exploring all possibe subsequences {0000, 0001, 0100, 0101}

(b) Decoding proceeds by keeping 4 best extensions of exiting 4 paths (out of 8 possible): {00000, 00011, 01000, 01001}

Figure 4.4: Successive-cancellation list decoding descends the tree in L parallel paths. In this example $L = 4$.

As reported in [108] (and independently confirmed by our experiments, see § 4.1.1), with a relatively small list size, the performance of the successive-cancellation list decoder will be very close to that of the optimal MAP decoder for polar codes. It is also noteworthy that in the entire run of the successive-cancellation list decoding there are $\Theta(n)$ *copy* operations per decoding thread (see lines 8 and 18 of Algorithm 3). Each copy operation requires duplication of the internal data structures of size $\Omega(n)$ of a SC decoder. Consequently, a naïve implementation of the successive-cancellation list-decoding algorithm would have a complexity of $\Omega(L \cdot n^2)$. However, due to the structured nature of the SC decoder, using a *copy-on-write* mechanism, a successive-cancellation list decoder can indeed be implemented in $O(L \cdot n \log n)$ complexity [108].

[5] Polar and Reed–Muller codes [83, 91] are very similar. The generator matrices for both codes are sub-matrices of G_n (2.20). The generator matrix of an (n, k) Reed–Muller code is obtained by keeping the k columns of G_n with largest Hamming weight (a channel-independent choice). It is discussed in [5, Section X] that the Reed–Muller choice can lead to unreliable codes under SC decoding.

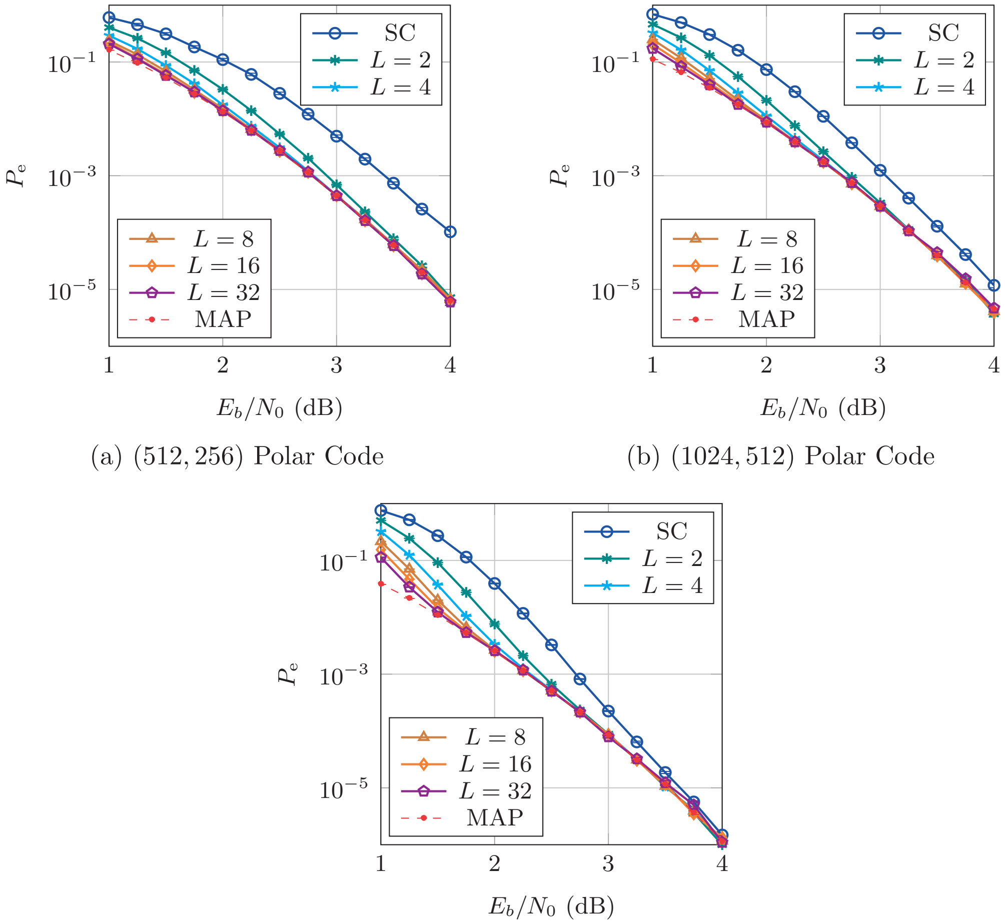
Algorithm 3: Successive-Cancellation List Decoding [108]

```
$\mathcal{L} \leftarrow \{1\}$ ;                    // start with a single active thread
for $i = 1$ to $n$ do
    if $i \notin \mathcal{I}$ then                              // frozen bits
        $\hat{u}_i[\ell] \leftarrow u_i$ for $\forall \ell \in \mathcal{L}$;
    else                                                   // information bits
        if $|\mathcal{L}| < L$ then                     // duplicate all the threads
            foreach $\ell \in \mathcal{L}$ do
                forkPath($\ell$);
        else                                        // choose $L$ best extensions
            $\forall \ell \in \mathcal{L}$ and $\forall u \in \mathbb{F}_2$, compute $P_{\ell,u} := W_n^{(i)}(y^n, \hat{u}^{i-1}[\ell]|u)$;
            $\tau \leftarrow$ the median of $2L$ likelihoods $P_{\ell,u}$;
            foreach $\ell \in \mathcal{L}$ such that $P_{\ell,0} < \tau$ and $P_{\ell,1} < \tau$ do
                Kill the thread $\ell$ and set $\mathcal{L} \leftarrow \mathcal{L} \setminus \{\ell\}$;
            for $\ell \in \mathcal{L}$ do
                if $P_{\ell,u} > \tau$ while $P_{\ell,u\oplus 1} < \tau$ then
                    $\hat{u}_i[\ell] \leftarrow u$;
                else                                  // $P_{\ell,0} \geq \tau$ and $P_{\ell,1} \geq \tau$
                    forkPath($\ell$);
$\ell^* \leftarrow \arg\max_{\ell \in \mathcal{L}} W_n^{(n-1)}(y^n, \hat{u}^{n-1}[\ell]|\hat{u}_n[\ell])$;
return $\hat{u}^n[\ell^*]$;
subroutine forkPath($\ell$)
    Copy the thread $\ell$ into a new thread $\ell' \notin \mathcal{L}$;
    $\mathcal{L} \leftarrow \mathcal{L} \cup \{\ell'\}$;
    $\hat{u}_i[\ell] \leftarrow 0$;
    $\hat{u}_i[\ell'] \leftarrow 1$;
```

4.1.1 Performance of Successive-Cancellation List Decoding

In Figure 4.5, we present the empirical performance of successive-cancellation list decoders with list sizes of $L = 2$, 4, 8, 16, and 32 for polar codes of rate 1/2 and block lengths $n = 512$, 1024, and 2048. We also plot a lower bound on the block-error probability of a MAP decoder. This bound is obtained as follows: During the simulations of the list decoder, whenever a decoding error occurs, we compute the posterior probability of the (mis)decoded codeword and compare it with that of the sent codeword. If the decoded codeword has a higher posterior probability, the MAP decoder would also make a wrong decision. Thus, by counting the frequency of such events, we can obtain a lower bound on the block-error probability of the MAP decoder.

As the plots show, for signal-to-noise ratios above 1.5 dB, a relatively small list size of $L = 32$ is sufficient to boost the performance of the suboptimal



(a) $(512, 256)$ Polar Code

(b) $(1024, 512)$ Polar Code

(c) $(2048, 1024)$ Polar Code

Figure 4.5: Successive-cancellation list decoder achieves close-to-optimal block-error probability.

successive-cancellation decoder up to that of the optimal MAP decoder. Moreover, as the list size is increased, we observe a *diminishing returns* phenomena. For example, for a $(512, 256)$ polar code the block-error probabilities at list sizes $L = 8$, 16 and 32 are practically indistinguishable (and they are all almost equal to that of the optimal MAP decoder) in the whole range of SNR. Similarly, for our $(1024, 512)$ polar code the block-error probabilities of list decoders with list sizes $L = 16$ and $L = 32$ are almost the same. Finally, we observe that as the SNR increases, say, above 3 dB, even a small list size of $L = 4$ is sufficient to achieve the optimal block-error probability.

4.1.2 CRC-Aided List Decoding

During the experiments with the successive-cancellation list decoder, we observe that when the decoder fails, in most of the cases the sent codeword is present in the final list but is not declared due to the presence of another more likely codeword in the list. (Note that in such circumstances the MAP decoder would have failed to decode too.) Consequently, if the decoder were 'assisted' for its final choice (line 19 of Algorithm 3) we could improve its performance. It turns out that such an assistance can easily be realized by *modifying* polar codes as follows: We can increase the number of information bits by r (i.e., increase the polar code rate to $(k+r)/n$) and set the last r information bits to an r-bit CRC of the first k information bits. Note that in this case the *effective* information rate of the code is unchanged. The successive-cancellation list decoder, in line 19, first discards the paths that do not pass the CRC test and then chooses the most likely path among the survivors [110]. As the empirical results of [110] show (we will also see shortly) a *CRC-aided* successive-cancellation list decoder has a significantly lower block-error probability and is competitive with existing error-correction codes. Before proceeding to the simulation results, a few remarks are in order:

1. In some works, e.g., [71,72,85] a modified polar code is realized by setting the last r information bits to the CRC of preceding $k-r$ information bits. This reduces the effective information rate, making the comparison between the performance of CRC-aided decoder and that of the unaided decoder unfair (as the codes subject to comparison do not have the same rate). According to [107], the simulation results of [110] are based on equi-rate comparisons.

2. In principle, a modified polar code can be constructed by using any $(k+r,k)$ code as the outer code and an $(n,k+r)$ polar code as the inner code. The concatenated code has a better distance profile and the list decoder, in line 19 of Algorithm 3, would first discard the candidates that are not codewords of the outer code and then declare the most likely one among the survivals (see for instance [113]). In particular, the advantage in using a cyclic code (such as a CRC code) as the outer code is that the code syndrome can be computed serially without the need for storing the decoded information bits per each path[6] [73, Chapter 7].

3. Another popular approach, to assist the decoder in declaring the path that corresponds to the (hopefully) correct codeword, is to partition the information bits into a few segments and append each segment with a very short CRC (as opposed to appending a longer CRC on the entire sequence of information bits) [50, 127]. Such a partitioning has the advantage that the decoder can prune the paths with an invalid CRC in early stages, which decreases the decoding complexity. Specifically, in a low-SNR operation regime, such an approach provides an *early stopping criteria* to the decoder: If at some point none of the paths pass the parity checks, the decoding stops (and a retransmission of that word will be

[6] This is essential for a space-efficient implementation of SC decoder which, in turn, leads to $O(L \cdot n \log n)$ implementation of the list decoder.

requested from higher layers) without wasting time to decode the entire word. This increases the decoding throughput.

4. As the careful reader might have noticed, in the design of concatenated polar codes, there is a trade-off between the length of the CRC (more generally the redundancy r of the outer $(k, k+r)$ code) and the performance improvements due to a CRC. A longer CRC helps the decoder to reject more incorrect codewords in line 19. Meanwhile, the longer CRC degrades the performance of the inner polar code (it requires a higher-rate code). In § 4.3.1, we show further empirical results that illustrate this trade-off.

In Figure 4.6, we present the simulated performance of CRC-aided successive-cancellation list decoding with list size $L = 32$ and the CRC-16 defined with generator polynomial

$$x^{16} + x^{15} + x^2 + 1,$$

We see, in these examples, that modified polar codes outperform the LDPC codes: they have both a better block-error probability and about 20% shorter block-length.

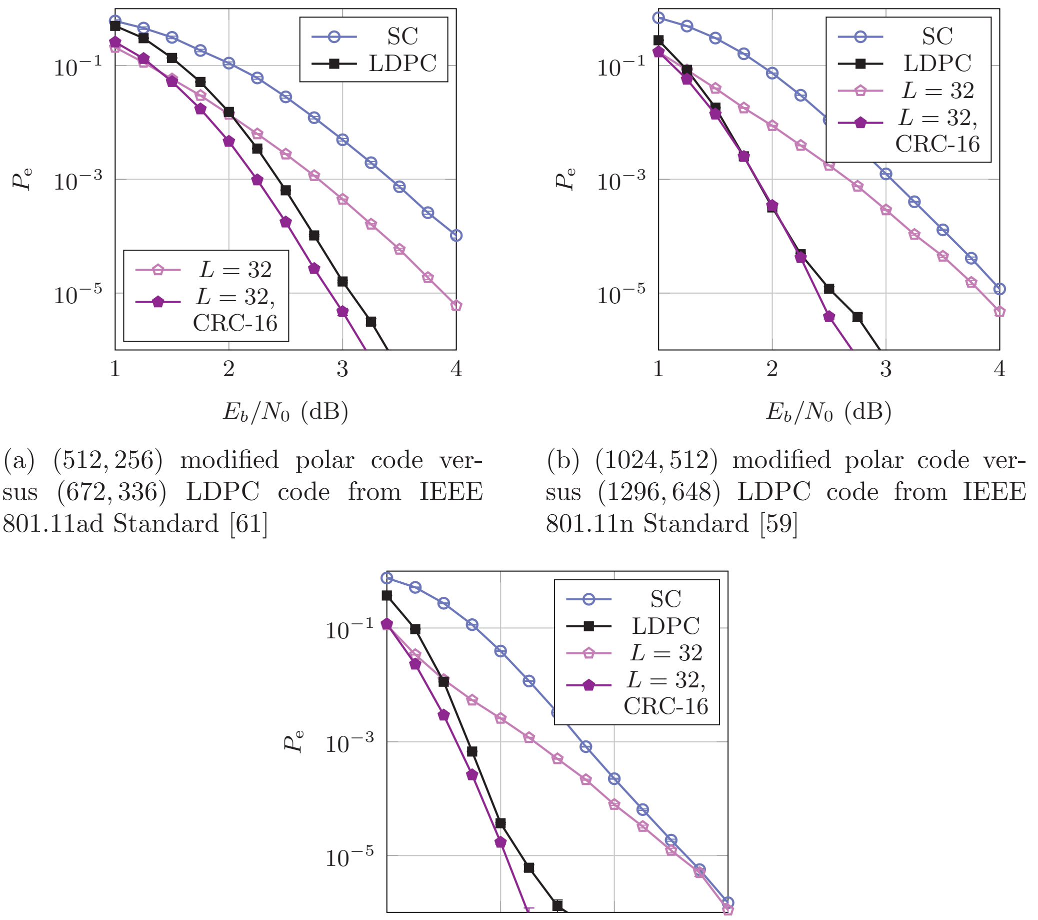
To highlight the superiority of modified polar codes under CRC-aided list decoding over the LDPC codes from IEEE standards, we compare the performance of all these codes using the uniform scale of [89] in Figure 4.7. The plot is obtained as follows. Using our simulation results, we find the SNR required by each code to achieve the target block-error probability of 10^{-4}. Given this SNR value and the code block-length, we compute an upper bound $R^\star$ on the rate of the best code (of that specific block-length) whose block-error probability does not exceed 10^{-4} using the bounds of [89]. Then, we plot the ratio of the code's rate (in our case always 1/2) to the optimal coding rate $R^\star$ versus the block-length.

4.2 LLR-Based Formulation of List Decoding Algorithm

Algorithm 3 is a valid high-level description of successive-cancellation list decoding. However, for implementing the algorithm, the stability of the computations is crucial. Algorithm 3 is described in terms of likelihoods that are *not* safe quantities to work with: a decoder implemented by using the likelihoods is prone to underflow errors because they quickly become tiny numbers as the block-length increases. As noted in [110], $\forall y^n \in \mathcal{Y}^n, u^i \in \mathbb{F}_2^i$, we trivially have

$$W_n^{(i)}(y^n, u_1^{i-1}|u_i) \overset{\text{(a)}}{=} 2\Pr\{Y^n = y^n, U^i = u^i\} \leq 2\Pr\{U^i = u^i\} \overset{\text{(b)}}{=} 2^{-(i-1)}. \quad (4.6)$$

(In the above (a) and (b) both follow from the assumption that information bits are independent, uniformly distributed, Bernoulli random variables.)



(a) $(512, 256)$ modified polar code versus $(672, 336)$ LDPC code from IEEE 802.11ad Standard [61]

(b) $(1024, 512)$ modified polar code versus $(1296, 648)$ LDPC code from IEEE 802.11n Standard [59]

(c) $(2048, 1024)$ polar code versus $(2304, 1152)$ LDPC code from IEEE 802.16 Standard [60]

Figure 4.6: Performance of CRC-Aided SC List Decoder

Consider the binary-tree picture that we provided in § 4.1. The decision LLRs

$$\lambda_m^{(i)} := \log\left[\frac{W_n^{(i)}(y^n, \hat{u}^{i-1}|0)}{W_n^{(i)}(y^n, \hat{u}^{i-1}|1)}\right] \tag{4.7}$$

summarize all the necessary information for choosing the most likely child among two children of the same parent at level i. In other words, decision LLRs are sufficient statistics for the SC decoder's decisions. In § 2.3.2 we have seen that having this type of decisions in the conventional SC decoder enables us to implement the computations in the LLR domain by using numerically stable operations. However the successive-cancellation list decoder, in lines 10–18 of Algorithm 3, has to choose the L most likely children out of $2L$ children of L different parents (cf. Figure 4.4). For these comparisons, the decision

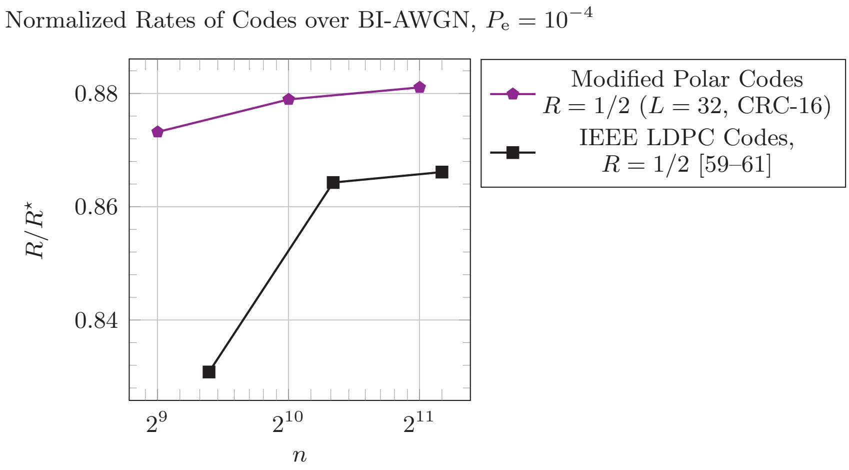


Figure 4.7: Modified polar codes outperform IEEE LDPC codes. The data points are computed using the numerical routines of [25].

log-likelihood *ratios*, $\lambda_m^{(i)}$, alone are not sufficient.

Consequently, the software implementation of the decoder in [108] implements the decoder in the likelihood domain by rewriting the recursions of § 2.3.2 for computing pairs of likelihoods $W_n^{(i)}(y^n, \hat{u}^{i-1}[\ell]|u_i)$, $u_i \in \mathbb{F}_2$ from pairs of channel likelihoods $W(y_i|x_i), x_i \in \mathbb{F}_2$, $i = 1, 2, \ldots, n$. To avoid underflows, at each intermediate step of updates, the likelihoods are scaled by a common factor such that $P_{\ell,u}$ in line 10 of Algorithm 3 is proportional to $W_n^{(i)}(y^n, \hat{u}^{i-1}|u_i)$ [110].

Alternatively, such a normalization step can be avoided by performing the computations in the log-likelihood (LL) domain, i.e., by computing the pairs

$$-\log\left[W_n^{(i)}(y^n, \hat{u}^{i-1}[\ell]|u_i)\right], \quad u_i \in \mathbb{F}_2 \tag{4.8}$$

as a function of channel log-likelihood pairs $\log[W(y_i|x_i)]$, $x_i \in \mathbb{F}_2$, $i = 1, 2, \ldots, n$ [10]. Log-likelihoods provide some numerical stability. Moreover, similar to log-likelihood ratios (see § 2.3.2), the recursive formulae for computing the log-likelihoods can be replaced by their min–sum approximations that involve only cheap-to-implement additions and comparisons. This leads to a suitable formulation for a hardware implementation of the decoder [10, 71, 72, 122, 126]. However, one important disadvantage of log-likelihoods is that they are all positive numbers being added throughout the recursive update rules. Consequently, as the log-likelihoods propagate from the channel level to the decision level, their dynamic range grows. Therefore, to avoid catastrophic overflows we need to increment by one the bit-width of memory elements in the decoder after each update, which leads to an irregular memory structure for the decoder. In particular, starting with a Q-bit quantization of channel log-likelihoods, the decision log-likelihoods need to be quantized using $Q + \log_2(n)$ bits. For example, in [10], the channel likelihoods are quantized as 4-bit unsigned integers which, for the block-length of $n = 1024$, leads to 14-bit decision log-likelihood ratios. This also means that all the processing elements of the decoder need to support 14-bit integers. Moreover, it is obvious that the number of memory elements per log-likelihood based SC decoder core would be twice that of a LLR-based SC decoder (a *pair* of log-likelihoods instead of each log-likelihood ratio has to be stored).

4.2.1 Ranking the Paths Based on the LLRs

Luckily, it turns out that the decoding paths can still be ordered according to their likelihoods by using all of the past decision LLRs $\lambda_m^{(j)}$, $j \in \{1, 2 \ldots, i\}$ and the trajectory of each path:

Theorem 4.1. *For each path ℓ and at each level $i \in \{1, 2, \ldots, n\}$ let the* path metric *be defined as:*

$$\zeta^{(i)}[\ell] \coloneqq \sum_{j=1}^{i} \log\bigl[1 + \exp\{-(1 - 2\hat{u}_j[\ell]) \cdot \lambda_m^{(j)}[\ell]\}\bigr], \tag{4.9}$$

where

$$\lambda_m^{(i)}[\ell] = \log\left[\frac{W_n^{(i)}(y^n, \hat{u}^{i-1}[\ell]|0)}{W_n^{(i)}(y^n, \hat{u}^{i-1}[\ell]|1)}\right] \tag{4.10}$$

is the log-likelihood ratio of bit u_i given the channel output y^n and the past trajectory of the path $\hat{u}_0^{i-1}[\ell]$. Then, if all the information bits are uniformly distributed in $\mathbb{F}_2$, $\zeta^{(i)}[\ell]$ is a strictly decreasing function of $W_n^{(i)}(y^n, \hat{u}^{i-1}[\ell]|\hat{u}_i[\ell])$.

In view of Theorem 4.1, we can implement the SCL decoder by using L parallel low-complexity *and stable* LLR-based SC decoders as the underlying building blocks. In addition, we need to keep track of L path-metrics. The metrics can be updated successively as the decoder proceeds by setting

$$\zeta^{(i)}[\ell] = \zeta^{(i-1)}[\ell] + \Delta\bigl(\lambda_m^{(i)}[\ell], \hat{u}_i[\ell]\bigr), \tag{4.11a}$$

where $\Delta\colon \mathbb{R} \times \mathbb{F}_2 \to \mathbb{R}$ is

$$\Delta(\lambda, u) \coloneqq \log\bigl[1 + \exp\{-(1 - 2u)\lambda\}\bigr]. \tag{4.11b}$$

As shown in Algorithm 4, using the values of the associated path metrics, the paths can be compared based on their likelihood.

Before proving Theorem 4.1 let us provide an intuitive interpretation of our metric. Since

$$\log[1 + \exp(\alpha)] \approx \begin{cases} 0 & \text{if } \alpha < 0, \\ \alpha & \text{if } \alpha \geq 0, \end{cases} \tag{4.12}$$

the update rule of (4.11) is well-approximated if we replace Δ with $\tilde{\Delta} : \mathbb{R} \times \mathbb{F}_2 \to \mathbb{R}$ defined as

$$\tilde{\Delta}(\lambda, u) \coloneqq \begin{cases} 0 & \text{if } u = \frac{1}{2}[1 - \operatorname{sign}(\lambda)], \\ |\lambda| & \text{otherwise.} \end{cases} \tag{4.13}$$

Algorithm 4: LLR-Based Formulation of SC List Decoding

```
$\mathcal{L} \leftarrow \{1\}$ ;                                   // start with a single active thread
for $i = 1$ to $n$ do
    Compute $\lambda_m^{(i)}[\ell]$, $\forall \ell \in \mathcal{L}$ ;                // parallel SC decoders
    if $i \notin \mathcal{I}$ then                                             // frozen bits
        $\forall \ell \in \mathcal{L}$, $\hat{u}_i[\ell] \leftarrow u_i$;
        $\forall \ell \in \mathcal{L}$, $\zeta^{(i)}[\ell] \leftarrow \zeta^{(i-1)}[\ell] + \Delta(\lambda_m^{(i)}[\ell], u_i)$ ;          // see (4.11b)
    else                                                        // information bits
        $\forall \ell \in \mathcal{L}$ and $\forall u \in \mathbb{F}_2$, set $P_{\ell,u} := \zeta^{(i)}[\ell] + \Delta(\lambda_n^{(i-1)}[\ell], u)$ ;
        // see (4.11b)
        if $|\mathcal{L}| < L$ then                             // duplicate all the threads
            foreach $\ell \in \mathcal{L}$ do
                forkPath($\ell$);
        else                                              // choose $L$ best extensions
            $\tau \leftarrow$ the median of $2L$ likelihoods $P_{\ell,u}$ ;
            foreach $\ell \in \mathcal{L}$ such that $P_{\ell,0} > \tau$ and $P_{\ell,1} > \tau$ do
                Kill the thread $\ell$ and set $\mathcal{L} \leftarrow \mathcal{L} \setminus \{\ell\}$;
            for $\ell \in \mathcal{L}$ do
                if $P_{\ell,u} < \tau$ while $P_{\ell,u\oplus 1} > \tau$ then
                    $\hat{u}_i[\ell] \leftarrow u$;
                    $\zeta^{(i)}[\ell] \leftarrow P_{\ell,u}$;
                else                                       // $P_{\ell,0} \leq \tau$ and $P_{\ell,1} \leq \tau$
                    forkPath($\ell$);
$\ell^* \leftarrow \arg\min_{\ell \in \mathcal{L}} \zeta^{(n)}[\ell]$;
return $\hat{u}^n[\ell^*]$;
subroutine forkPath($\ell$)
    Copy the thread $\ell$ into a new thread $\ell' \notin \mathcal{L}$;
    $\mathcal{L} \leftarrow \mathcal{L} \cup \{\ell'\}$;
    $\hat{u}_i[\ell] \leftarrow 0$;
    $\zeta_\ell^{(i)} \leftarrow P_{\ell,0}$;
    $\hat{u}_i[\ell'] \leftarrow 1$;
    $\zeta_{\ell'}^{(i)} \leftarrow P_{\ell,1}$;
```

We also note that $\frac{1}{2}[1 - \text{sign}(\lambda_m^{(i)}[\ell])]$ is the direction that the LLR (given the past trajectory $\hat{u}^{i-1}[\ell]$) suggests. This is the same decision that a SC decoder would take if it estimates the value of u_i at step i given the past set of decisions $\hat{u}^{i-1}[\ell]$ (cf. line 5 in Algorithm 1). Equation (4.13) shows that if at step i the ℓ^{th} path does not follow the direction suggested by $\lambda_m^{(i)}[\ell]$, it will receive a *penalty* roughly equal to $|\lambda_m^{(i)}[\ell]|$, which is the *reliability* of that LLR. It could immediately be concluded, based on such an interpretation, that the path that SC decoder follows will always have the lowest penalty thus is always declared as the output of the SCL decoder. So why should the SCL decoder exhibit a better performance compared to the SC decoder? The answer is that such a reasoning is correct only if *all* the elements of u^n are information bits. As soon as the decoder encounters a frozen bit, the path metric is updated based on the likelihood of the frozen bit, given the past trajectory of the path and the a-priori known value of that bit (cf. line 6 in Algorithm 4). If the value of that frozen bit does not agree with the LLR, given the past trajectory (which is an indication of a preceding erroneous decision), this can penalize the SC path by a considerable amount while keeping some other paths unpenalized. This is exactly where the constraints put by the subsequent frozen bits show their effect and correct the decisions (whereas in the SC decoder such an incompatibility would have been ignored).

Let us conclude this section by proving Theorem 4.1:

Lemma 4.2. *If U_i is uniformly distributed on $\mathbb{F}_2$, then,*

$$\frac{W_n^{(i)}(y^n, u^{i-1}|u_i)}{\Pr\{U_0^i = u^i|Y^n = y^n\}} = 2\Pr\{Y^n = y^n\}. \tag{4.14}$$

Proof. Since $\Pr\{U_i = u_i\} = 1/2$ for $\forall u_i \in \{0,1\}$,

$$\begin{aligned}\frac{W_n^{(i)}(y^n, u^{i-1}|u_i)}{\Pr\{U^i = u^i|Y^n = y^n\}} &= \frac{\Pr\{Y^n = y^n, U^i = u^i\}}{\Pr\{U_i = u_i\}\Pr\{U^i = u^i|Y^n = y^n\}} && (4.15)\\ &= \frac{\Pr\{Y^n = y^n\}\Pr\{U^i = u^i|Y^n = y^n\}}{\Pr\{U_i = u_i\}\Pr\{U^i = u^i|Y^n = y^n\}} && (4.16)\\ &= 2\Pr\{Y^n = y^n\}. && \square\end{aligned}$$

Proof of Theorem 4.1. It is sufficient to show

$$\zeta^{(i)}[\ell] = -\log\left[\Pr\{U^i = \hat{u}^i[\ell]|Y^n = y^n\}\right]. \tag{4.17}$$

Having shown (4.17), Theorem 4.1 will follow as an immediate corollary to Lemma 4.2 (because the channel output y^n is fixed for all decoding paths). Since the path index ℓ is fixed on both sides of (4.9) we will drop it in the proof. Let

$$\Lambda_m^{(i)} := \frac{W_n^{(i)}(y^n, \hat{u}^{i-1}|0)}{W_n^{(i)}(y^n, u^{i-1}|1)} = \frac{\Pr\{Y^n = y^n, U^{i-1} = \hat{u}^{i-1}, U_i = 0\}}{\Pr\{Y^n = y^n, U^{i-1} = \hat{u}^{i-1}, U_i = 1\}} \tag{4.18}$$

(the last equality follows since $\Pr\{U_i = 0\} = \Pr\{U_i = 1\}$). Observe that showing (4.17) is equivalent to proving

$$\Pr\{U^i = \hat{u}^i|Y^n = y^n\} = \prod_{j=1}^{i} \frac{1}{1 + (\Lambda_m^{(j)})^{-(1-2\hat{u}_j)}}. \tag{4.19}$$

Since

$$\Pr\{Y^n = y^n, U^{i-1} = \hat{u}^{i-1}\} = \sum_{\hat{u}_i \in \{0,1\}} \Pr\{Y^n = y^n, U^i = \hat{u}^i\} \tag{4.20}$$

$$= \Pr\{Y^n = y^n, U^i = \hat{u}^i\}\big[1 + (\Lambda_m^{(i)})^{-(1-2\hat{u}_i)}\big], \tag{4.21}$$

we have,

$$\Pr\{Y^n = y^n, U^i = \hat{u}^i\} = \frac{1}{1 + (\Lambda_m^{(i)})^{-(1-2\hat{u}_i)}} \Pr\{Y^n = y^n, U^{i-1} = \hat{u}^{i-1}\} \tag{4.22}$$

Repeated application of (4.22) (for $i-1, i-2, \ldots, 1$) yields

$$\Pr\{Y^n = y^n, U^i = \hat{u}^i\} = \prod_{j=0}^{i} \frac{1}{1 + (\Lambda_m^{(j)})^{-(1-2\hat{u}_i)}} \Pr\{Y^n = y^n\}. \tag{4.23}$$

Dividing both sides by $\Pr\{Y^n = y^n\}$ proves (4.19). □

4.2.2 Advantages of LLR-Based Formulation

As we have discussed earlier in this section, an important drawback of the log-likelihood-based implementations of the SC list decoder [10, 71, 72, 122, 126] is their irregular memory structure and the need for the processing elements that support large bit-widths. The LLR-based formulation of list decoder solves both of this problems.

A successive-cancellation list decoder can be implemented using L LLR-based SC decoders as core components. As LLRs are signed numbers and the update rules for computing decision LLRs from channel LLRs (cf. § 2.3.2) involve both additions and subtractions, the dynamic range of intermediate LLRs is smaller than that of log-likelihoods; hence all of them can be quantized using the same number of bits. (For instance, quantizing all LLRs as 6-bit signed integers is shown to be sufficient to minimize the performance degradation due to fixed-point quantization for length-1024 decoder in [8].)

In addition, L recursively updated path metrics need to be stored, based on them, the decoding paths will be ordered in each step. The path-metric update rule of (4.11b) can safely be approximated by (4.13) to avoid expensive arithmetics in hardware. The path metrics are unsigned numbers. If the LLRs are represented as Q-bit signed integers (i.e., with $Q-1$ bits for magnitude plus 1 bit for the sign), in theory, we have to consider $Q - 1 + \log_2(n)$ bits for an overflow-free storage of path metrics (hence a sorter[7] that supports this bit-width for ranking the paths in line 13 of Algorithm 4). However, in practice, any path that gets continually harshly penalized will be eliminated soon.

[7] Although only the median of $2L$ metrics needs to be found in line 11 of Algorithm 3 (respectively, line 13 of Algorithm 4), in practice, as L is small, it is faster to just sort the $2L$ numbers to rank the paths.

Hence, fewer bits turn out to be sufficient for storing the path metrics without performance degradation. (E.g., only 8 bits as opposed to the theoretical value of 15 bits for the situation considered in [8].)

Moreover, the recursive updates of the path metric (4.11) impose a particular structure on the input sequence of the sorter when the decoder proceeds through estimating a chunk of information bits indexed by a contiguous subset of information indices $\mathcal{I}$: L (out of of the $2L$) metrics to be sorted are the previously sorted metrics (as a result of decoding the previous information bits) and the remaining L are obtained by adding a positive number to the previous L metrics. Therefore, some of $\binom{L}{2}$ relations between the metrics are known. This structure can be exploited to decrease the sorting complexity [8,9,20,62].

We refer the reader to [8] for the details of the hardware architecture for an LLR-based implementation of the successive-cancellation list decoder. To showcase the significant gains in the performance of an LLR-based implementation of the decoder (compared to previously existing log-likelihood based decoders), in Table 4.1, we compare the throughput and area of log-likelihood-based and LLR-based implementations of list decoders with list sizes $L = 2$, $L = 4$, $L = 8$ for a polar code of length $n = 1024$ from [8]. We see that an LLR-based implementation of the decoder has up to 53% higher throughput and occupies roughly 33% less silicon area.

	LL-Based [10]	LLR-Based [8]	Speedup
$L = 2$	314	335	7%
$L = 4$	228	307	35%
$L = 8$	161	246	53%

(a) Throughput (Mbps)

	LL-Based [10]	LLR-Based [8]	Reduction
$L = 2$	1.38	0.88	36%
$L = 4$	2.62	1.78	32%
$L = 8$	5.38	3.58	33%

(b) Area (mm^2)

Table 4.1: The LLR-based formulation of list decoding leads to faster and smaller implementations of the decoder.

4.3 Other Design Considerations

4.3.1 Choice of CRC Length

As we briefly discussed in § 4.1.2, in the design of a modified polar code (to be decoded with a CRC-aided list decoder) the length of CRC has to be carefully chosen: By increasing the length of CRC, the decoder will be able to identify incorrect decisions more often. Meanwhile, the performance of the inner polar code will be degraded (as its rate needs to be increased to keep the effective information rate fixed). In this section, we provide numerical examples that exhibit this phenomena.

For our experiments, we picked four different CRCs of lengths $r = 4$, $r = 8$, $r = 16$, and $r = 32$ from [114] with the following generator polynomials:

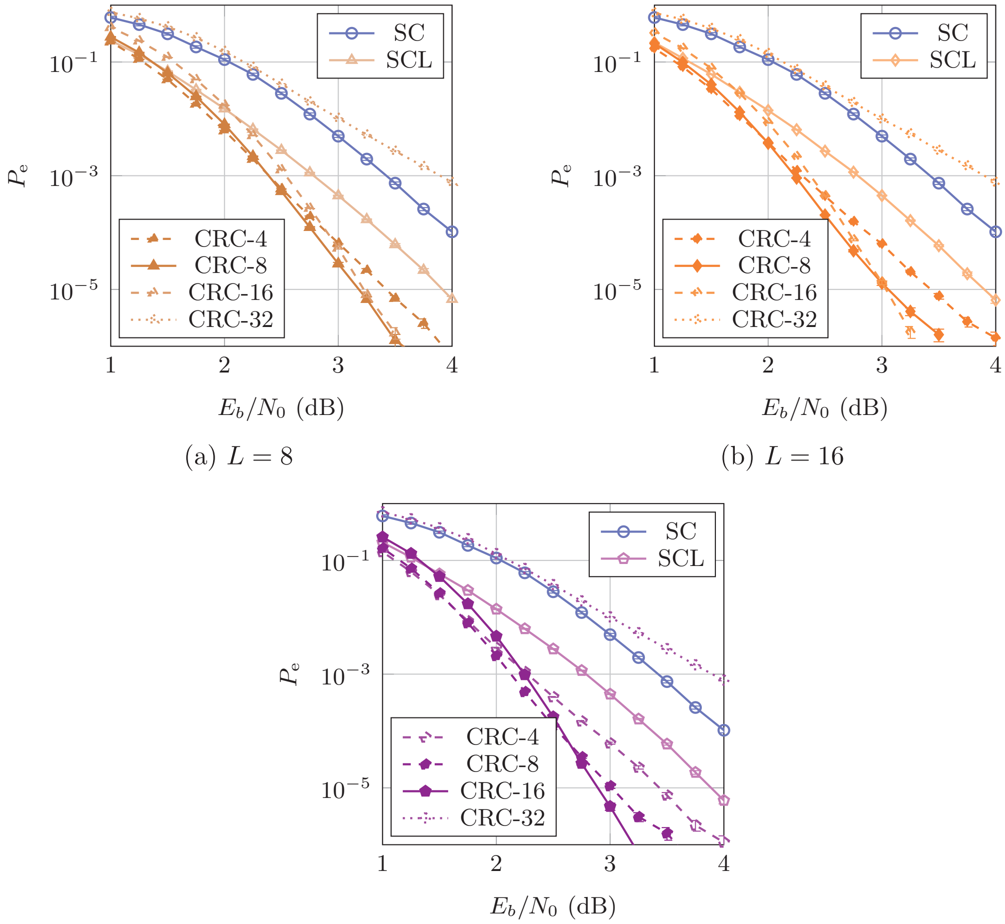
$$\begin{aligned} \text{CRC-4}: &\quad x^4 + x + 1, && (4.24)\\ \text{CRC-8}: &\quad x^8 + x^7 + x^6 + x^4 + x^2 + 1, && (4.25)\\ \text{CRC-16}: &\quad x^{16} + x^{15} + x^2 + 1, && (4.26)\\ \text{CRC-32}: &\quad x^{32} + x^{26} + x^{23} + x^{22} + x^{16} + x^{12} + x^{11} \\ &\qquad + x^{10} + x^8 + x^7 + x^5 + x^5 + x^2 + x + 1. && (4.27) \end{aligned}$$

In Figure 4.8 we plot the empirical block-error probability of different modified polar codes of block-length $n = 512$ and rate half, obtained by setting last r information bits of $(512, 256 + r)$ polar codes to CRC-rs (defined with the above-mentioned polynomials) of the preceding 256 information bits, decoded with list decoders of list sizes $L = 8$, $L = 16$ and $L = 32$. The plots in Figures 4.8a and 4.8b show that, for the list sizes of $L = 8$ and $L = 16$, using CRCs of length 4 or 8 improves the performance. Increasing the CRC length to 16 does *not* improve the performance further and even causes the CRC-aided list decoder to perform worse than the plain list decoder. More importantly, increasing the length of CRC to 32 will degraded the performance of the polar code so much that even with the help of CRC the decoder has a block-error probability *higher* than the conventional SC decoder. On the contrary, as shown in Figure 4.8c, if we increase the list size to $L = 32$, due to the lower block-error probability of the inner polar code under list decoding, we can increase the length of CRC to 16 and still improve the overall performance. But even with such a relatively large list size, increasing the CRC length to 32 degrades the performance drastically.

Repeating the same experiment for block-lengths of $n = 1024$ and $n = 2048$ results in the same phenomena (cf. Figures 4.9 and 4.10): Increasing the length of CRC up to 16 improves the performance, increasing it to 32 bits leads to a block-error probability worse than that of the plain list decoder.

4.3.2 SC or SCL?

Modern communication standards sometimes permit very long codes to be used. As a result, a natural question in the design of a polar coding scheme, to achieve a target block-error probability, would be to choose between long polar codes decoded under less complex SC decoding algorithm, or short codes decoded with SC list decoding. In this section, to study this trade-off, we compare the performance of the rate-1/2 polar codes of length $n = 512$, $n = 1024$, and $n = 2048$ under list decoding to that of longer polar codes under conventional SC decoding.



(a) $L = 8$

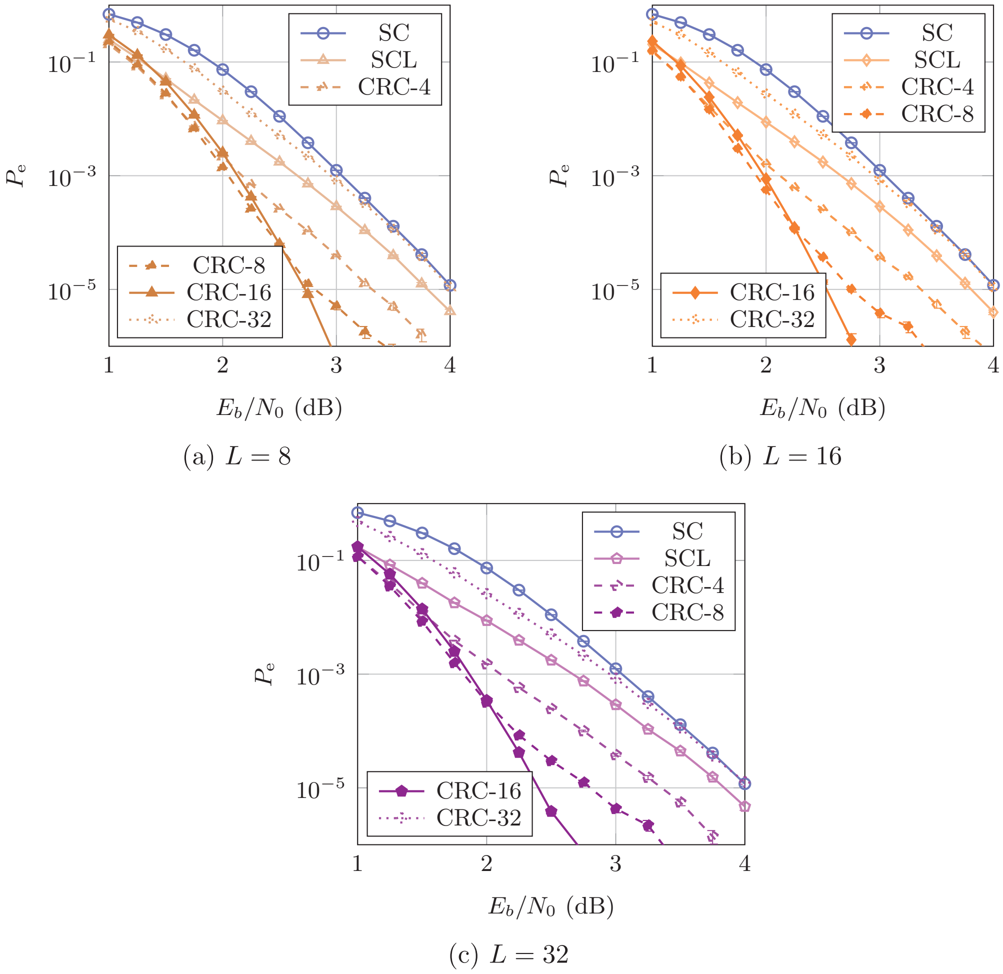
(b) $L = 16$

(c) $L = 32$

Figure 4.8: Performance of Modified Polar Codes of Block-length $n = 512$ and Rate $1/2$ under Various List Sizes

From a high-level perspective, the memory requirement of a successive-cancellation list decoder of block-length n and list size L is the same as that of a successive-cancellation decoder of block-length $L \cdot n$ (provided that the space-efficient implementation of the decoder is employed). Moreover, to simplify the problem, we would err on the side of conventional successive-cancellation decoder and pretend that a successive-cancellation decoder at block-length $L \cdot n$ has the same computational complexity as a successive-cancellation list decoder of block-length n and list size L.[8] Therefore, to make an equi-complexity comparison, we compare the performance of polar codes of block-length n under list decoding with list size L to that of polar codes of block-length $L \cdot n$ under conventional SC decoding. Moreover, as we discussed in § 4.1.2, modifying polar codes by concatenating them with a CRC and then running the CRC-

[8] In reality, the computation power required for the list decoder scales as $O(L \cdot n \log n)$ but that of the SC decoder scales like $O(L \cdot n \log n + L \cdot n \log L)$.



(a) $L = 8$

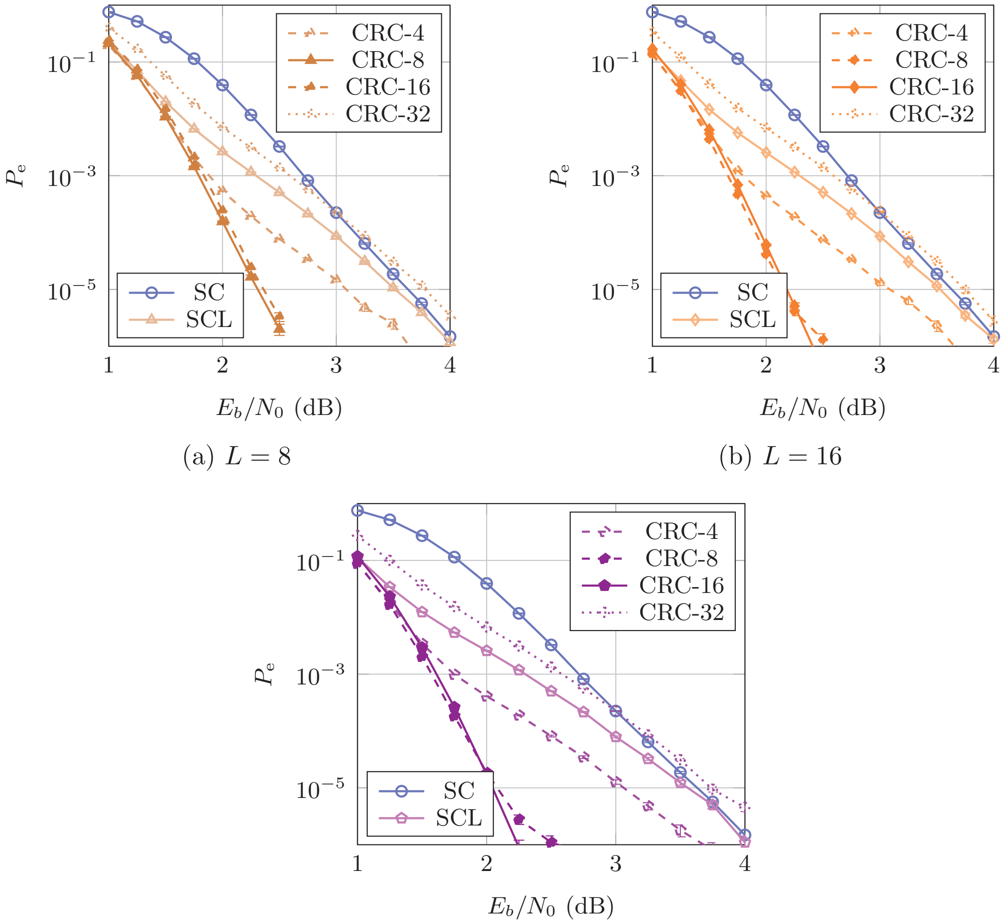
(b) $L = 16$

(c) $L = 32$

Figure 4.9: Performance of Modified Polar Codes of Block-length $n = 1024$ and Rate $1/2$ under Various List Sizes

aided list decoder has negligible additional costs. Hence, we choose the optimal CRC among those introduced in § 4.3.1 for each list size and block-length and consider the performance of the CRC-aided version of the list decoder.

Figure 4.11 shows the empirical performance of long polar codes under SC decoding, versus short modified polar codes under SC list decoding. We see in Figure 4.11a that modified polar codes of block-lengths $n = 512$, $n = 1024$, and $n = 2048$, under successive-cancellation list decoding with list sizes $L = 8$, $L = 4$, and $L = 2$, respectively, have block-error probabilities somewhat lower than that of a polar code of block-length $n = 4096$ under conventional SC decoding. If we can afford for a higher complexity of decoding a polar code of block-length $n = 8192$ or, equivalently, increasing the list sizes to $L = 16$, $L = 8$, and $L = 4$ for decoding the modified polar codes of block-lengths $n = 512$, $n = 1024$ and $n = 2048$, respectively, we see that except at low SNRs regime the length-512 modified polar code falls short whereas length-1024 and



(a) $L = 8$

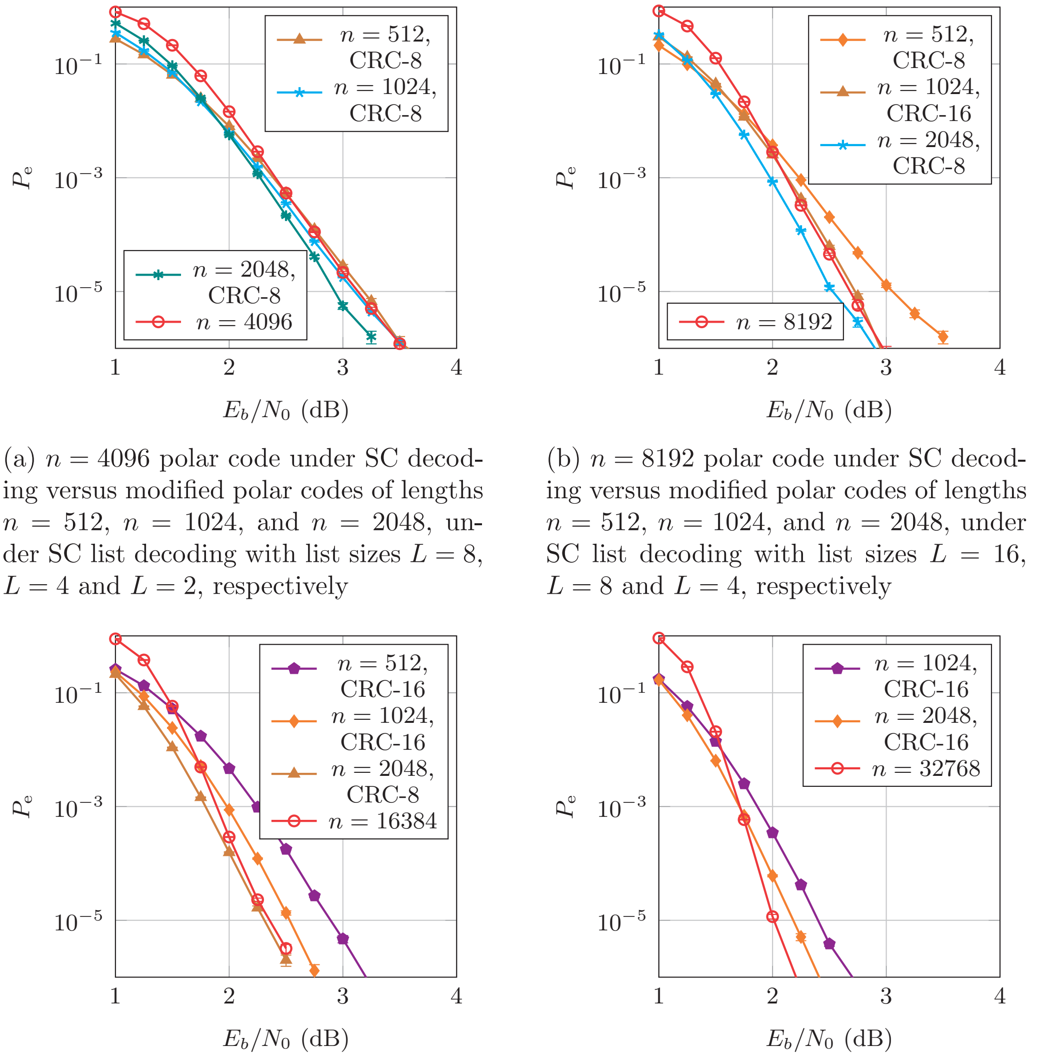
(b) $L = 16$

(c) $L = 32$

Figure 4.10: Performance of Modified Polar Codes of Block-length $n = 2048$ and Rate $1/2$ under Various List Sizes

length-2048 polar codes have the same block-error probability as the $n = 8192$ polar code under SC decoding. This phenomenon becomes more visible when we compare $n = 16384$ polar code under SC decoding with modified short polar codes in Figure 4.11c: here we see that only the length-2048 modified polar code under list decoding with list size of $L = 8$ performs better than the length-16384 polar code under SC decoding. Finally, we observe in Figure 4.11d that when we allow the block-length to increase to $n = 32768$, the performance of the long polar code under SC decoding improves significantly and neither our modified polar code of block-length $n = 1024$ nor the one with block-length $n = 2048$ perform better than the long polar code under SC decoding (except for the low SNR regime).

Our examples show that by using a CRC-aided successive-cancellation list decoder, we can achieve the same performance as that of conventional SC decoding with eight times smaller block-lengths and almost the same com-



(a) $n = 4096$ polar code under SC decoding versus modified polar codes of lengths $n = 512$, $n = 1024$, and $n = 2048$, under SC list decoding with list sizes $L = 8$, $L = 4$ and $L = 2$, respectively

(b) $n = 8192$ polar code under SC decoding versus modified polar codes of lengths $n = 512$, $n = 1024$, and $n = 2048$, under SC list decoding with list sizes $L = 16$, $L = 8$ and $L = 4$, respectively

(c) $n = 16384$ polar code under SC decoding versus modified polar codes of lengths $n = 512$, $n = 1024$, and $n = 2048$, under SC list decoding with list sizes $L = 32$, $L = 16$ and $L = 8$, respectively

(d) $n = 32768$ polar code under SC decoding versus modified polar codes of lengths $n = 1024$, and $n = 2048$, under SC list decoding with list sizes $L = 32$ and $L = 16$, respectively

Figure 4.11: Performance of Long Polar Codes under SC Decoding versus Short Codes under CRC-aided SC List Decoding (All codes have rate $1/2$.)

putational complexity per codeword. Hence, to achieve a particular block-error probability, successive-cancellation list decoding provides a solution for decreasing the decoding latency (by decreasing the block-length) without increasing the computational complexity, rendering polar codes suitable for low-latency applications.

Remark. We measured the *complexity* of decoding by a rough count of the number of memory elements and arithmetic operations required for decoding a codeword. Arguably, this might not be a precise measure — a finer measure depending on the implementation platform is more desirable. For example, in [8] we compare the hardware synthesis results for $n = 1024$ list decoders with list sizes $L = 2$ and $L = 4$ based on the hardware architecture proposed (in the paper) to the synthesis results for SC decoders of block-length $n = 2048$ and $n = 4096$ based on the architecture of [67]. We observe that even though we expect the SC list decoders to be at least as fast and as small as the corresponding SC decoders, due to the presence of extra processing elements (like the sorter and crossbars implementing the copy operations), they are slightly slower and about 15% larger.

4.4 Summary and Outlook

In this chapter, we have reviewed the successive-cancellation list-decoding algorithm, a low-complexity method to boost the performance of short polar codes. We have seen that combining successive-cancellation list decoding with CRC-concatenated polar codes, leads to an efficient error-correction scheme that is competitive with existing schemes.

We have proven that the successive-cancellation list-decoding algorithm can be formulated exclusively in terms of log-likelihood ratios, which leads to a numerically stable implementation of the decoder. A hardware implementation of the decoder by using the LLR-based formulation of the algorithm is up to 50% faster and around 30% smaller than the previous implementations of the decoder, due to the smaller dynamic range of LLRs [8].

We have also looked at the trade-off between the length of CRC and the performance of CRC-concatenated polar codes (when decoded with a CRC-aided list decoder) in the design of modified polar codes through numerical experiments in § 4.3.1. We observe that, given the block-length of the code, the list size of the decoder, and the channel quality, there is a limit on how much the length of CRC can be increased to improve the block-error probability of the code: increasing the length of CRC beyond this limit degrades the code performance.

Moreover, we have compared the performance of short modified polar codes under CRC-aided successive-cancellation list decoding and that of long polar codes under conventional successive-cancellation decoding in § 4.3.2. Our examples show that, to achieve a target block-error probability, CRC-aided successive-cancellation list decoding provides a means to reduce the block-length by a factor of eight and keeping the per-codeword decoding complexity roughly the same, compared to conventional successive-cancellation decoding. This fits polar codes for low-latency applications.

In conclusion to this chapter, we remark that list decoding combined with an outer code (to declare the final estimated codeword among candidates present in the output of the list decoder) is a general idea that, in principle, is applicable to any code. As we have seen, the structure of polar codes enables their efficient list decoding, that, together with the outer cyclic code, results in a competitive error-correction performance compared to existing error-correction codes. To have a more comprehensive comparison of polar codes with other error-correction schemes, it would be interesting to investigate the performance and complexity of decoding other codes (specifically, those compared to polar codes in this chapter and other works cited) with a list decoder (assisted with an outer code). To our knowledge, such a study has not yet been done.

Part II

Secrecy

The Wiretap Channel and Its Secrecy Exponents 5

Wyner [118] studied the problem of secure communications in the presence of an eavesdropper. In his model, called the *wiretap channel*, Alice wants to communicate with Bob through a noisy channel $W_{\mathrm{M}}\colon \mathcal{X} \to \mathcal{Y}$, while her transmitted signals are being wiretapped by Eve, through another noisy channel $W_{\mathrm{E}} : \mathcal{X} \to \mathcal{Z}$ (see Figure 5.1). The goal is to design a communication scheme that enables Alice to reliably communicate secret messages to Bob, while concealing them from Eve.

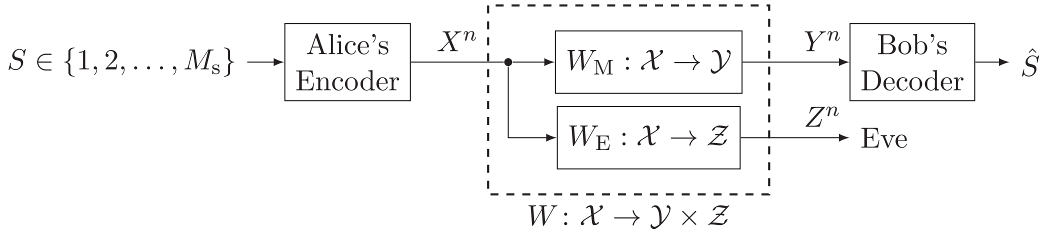

$S \in \{1, 2, \ldots, M_{\mathrm{s}}\}$
Alice's Encoder
X^n
$W_{\mathrm{M}} : \mathcal{X} \to \mathcal{Y}$
$W_{\mathrm{E}} : \mathcal{X} \to \mathcal{Z}$
Y^n
Z^n
Bob's Decoder
$\hat{S}$
Eve
$W\colon \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$

Figure 5.1: The Wiretap Channel Model

Assuming the eavesdropper's channel is *degraded* (see Appendix A) with respect to the legitimate receiver's channel, Wyner [118] characterized the *secrecy capacity* of the wiretap channel. The secrecy capacity is the highest rate R_{s}, for which there exist communication schemes in which Alice can communicate $M_{\mathrm{s}} \geq 2^{nR_{\mathrm{s}}}$ different secret messages via n channel uses to Bob with arbitrarily small error probability, and guarantee that the normalized amount of information that *leaks* to Eve about the secret message, $\frac{1}{n}I(S; Z^n)$, is arbitrarily small, provided that the block-length n is large enough. Wyner's work was subsequently extended by Csiszár and Körner [29] where, in particular, they relaxed the assumption on the degradedness of the eavesdroppers' channel and derived the secrecy capacity of an arbitrary wiretap channel.[1]

The information leakage rate is, arguably, a *weak* notion of secrecy: In a *weakly secure* system with large block-length n, upon observing Z^n, Eve could learn substantial amount of information about the secret message $I(S;Z^n)$ albeit small compared to n (and, hence, nR_s). Hence, it would be natural to design communication schemes that guarantee the *total* amount of information leaked to Eve, namely $I(S;Z^n)$, is small. Namely, to ensure *strong secrecy.* Rather surprisingly, the secrecy capacity does not change under this strong secrecy requirement. One way to prove this is via the method, proposed by Maurer and Wolf [77], for converting any weakly secure system to a strongly secure one with a negligible rate loss. Alternatively, as first proposed by Csiszár [27], a *resolvability-based* approach to secrecy leads to the construction of communication schemes for the wiretap channel that directly guarantee strong secrecy.

More strongly, a resolvability-based approach to the wiretap channel yields communication schemes in which the information leaked to Eve decays exponentially fast in block-length n. The rate of this exponential decay, called the *secrecy exponent* of the model, is a measure of the strength of the communication scheme.

In this chapter, we review the problem of secure communications in the presence of an eavesdropper, examine different achievability proofs, and show how a resolvability-based proof can establish strong secrecy. To prepare for our discussions of the following two chapters, we then formally define the notion of the *secrecy exponent*, discuss its relation to *resolvability exponent*, and give an overview of the existing works on the secrecy exponents.

An even more stringent notion of secrecy is *semantic secrecy* for the wiretap channel: it requires the information leaked to the eavesdropper $I(S;Z^n)$ to be small for *any* distribution of the secret messages picked by Alice [17]. We will see, in the proof of Theorem 5.4, that a resolvability-based approach to secrecy enables us to construct communication schemes that guarantee semantic secrecy at any rate below the secrecy capacity of the system. Thus, in particular, the secrecy capacity does not decrease, even if we ask for such a strict secrecy guarantee.

5.1 The Secrecy Capacity

A wiretap channel is described by a *joint* conditional probability $W\colon \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$, where $W(y,z|x)$ is the probability that the legitimate receiver (Bob) receives $y \in \mathcal{Y}$ and the eavesdropper (Eve) receives $z \in \mathcal{Z}$ when the input of

[1] In fact, [29] characterizes the achievable rate-region in a more general setting where Alice has two messages to communicate: a public message that has to be reliably transmitted to both Bob and Eve and a private message that needs to be decoded reliably by Bob while being hidden from Eve.

the channel is $x \in \mathcal{X}$. The joint conditional probability gives rise to a pair of marginals

$$W_{\mathrm{M}}\colon \mathcal{X} \to \mathcal{Y}, \qquad W_{\mathrm{M}}(y|x) = \sum_{z \in \mathcal{Z}} W(y, z|x), \tag{5.1}$$

and

$$W_{\mathrm{E}}\colon \mathcal{X} \to \mathcal{Z}, \qquad W_{\mathrm{E}}(z|x) = \sum_{y \in \mathcal{Y}} W(y, z|x). \tag{5.2}$$

These are the effective single-input single-output noisy channels through which the legitimate receiver and the eavesdropper perceive the signals transmitted by the sender, respectively. As we will see, as there is no interaction between the legitimate receiver and the eavesdropper, the system is characterized exclusively in terms of the two marginals W_{M} and W_{E}. In other words, two models $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ and $W' : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ that have the same marginals $W_{\mathrm{M}} = W'_{\mathrm{E}}$ and $W_{\mathrm{E}} = W'_{\mathrm{E}}$ are completely equivalent.

A code of block-length n and rate R_{s} for the wiretap channel is a collection of $M_{\mathrm{s}} \geq \exp(nR_{\mathrm{s}})$ distributions on the set of length-n channel input sequences, $\mathsf{E}\colon \{1, 2, \ldots, M_{\mathrm{s}}\} \to \mathcal{X}^n$. Note the encoding is assumed to be possibly randomized (and we will see that this indeed has to be the case). To communicate a particular message $s \in \{1, 2, \ldots, M_{\mathrm{s}}\}$, the encoder draws $X^n \sim \mathsf{E}(\cdot|s)$ and transmits it through the channel.

Definition 5.1. A rate R_{s} is an achievable *secrecy rate* over the wiretap channel $W\colon \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$, under the *weak secrecy* criteria, if for every pair $\epsilon_1 > 0$ and $\epsilon_2 > 0$, there exists n_0 such that for all $n \geq n_0$ there exists a code of block length n and secret-message size $M_{\mathrm{s}} \geq \exp(nR_{\mathrm{s}})$, $\mathsf{E}\colon \{1, 2, \ldots, M_{\mathrm{s}}\} \to \mathcal{X}^n$, using which

$$\Pr\{\hat{S}_{\mathrm{MAP}}(Y^n) \neq S\} \leq \epsilon_1 \tag{5.3}$$

$$\frac{1}{n} I(S; Z^n) \leq \epsilon_2. \tag{5.4}$$

In the above $\hat{S}_{\mathrm{MAP}}$ is the optimal (MAP) decision on S upon observing Y^n, the collection $S \mathrel{-\!\!\circ\!\!-} X^n \mathrel{-\!\!\circ\!\!-} (Y^n, Z^n)$ has distribution $P_S(s)\mathsf{E}(x^n|s)W^n(y^n, z^n|x^n)$ (as the diagram of Figure 5.1 shows), and P_S is the uniform distribution on $\{1, 2, \ldots, M_{\mathrm{s}}\}$.

The supremum of all achievable secrecy rates is the *secrecy capacity* of the wiretap channel $W\colon \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$.

Definition 5.2. A rate R_{s} is an achievable *secrecy rate* over the wiretap channel $W\colon \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ under the *strong secrecy* criteria, if for every pair $\epsilon_1 > 0$ and $\epsilon_2 > 0$ there exists a code with the properties described in Definition 5.1 that guarantees

$$I(S; Z^n) \leq \epsilon_2. \tag{5.5}$$

(Obviously, (5.5) is stronger than (5.4) hence replaces it.)

The *strong secrecy capacity* of the system is defined as the supremum of all achievable secrecy rates under the strong secrecy criteria.

Remark. Both (5.3) and (5.4) (or (5.5)) are *average-case* measures. We will show in Theorem 5.4 that by *expurgating* a good average-case code we can obtain a code that performs essentially as well as the original code for *any* distribution of the secret messages — i.e., a code that guarantees *semantic secrecy* [17]. Having such a *worst-case* guarantee is crucial from a practical point of view — it is the end-user that picks the secret messages to be transmitted and the performance guarantees for the system must *not* rely on a specific choice of hers.

Theorem 5.1 ([29, 118])**.** *Given any distribution on the input alphabet $\mathcal{X}$, $P_X \in \mathcal{P}(\mathcal{X})$, any rate*

$$R_{\mathrm{s}} < I(X;Y) - I(X;Z) \tag{5.6}$$

is an achievable secrecy rate in the sense of Definition 5.1. In the above, $(X,Y,Z) \sim P_X(x)W(y,z|x)$.

Equation (5.6) implies the secrecy capacity is lower-bounded as

$$C_{\mathrm{s}}(W) \geq \max_{P_X \in \mathcal{P}(\mathcal{X})} \{I(X;Y) - I(X;Z)\}. \tag{5.7}$$

Can the above lower bound be improved? In some cases, yes: Suppose Alice prefixes the channel with an *artificial* channel $P_{X|U} : \mathcal{U} \to \mathcal{X}$. That is to say, she uses a code for the input alphabet $\mathcal{U}$ instead of $\mathcal{X}$ and, to transmit each codeword $u^n \in \mathcal{U}^n$, she generates $X^n \sim P^n_{X|U}(x^n|u^n)$ and transmits it through the physical channel. This is equivalent to a wiretap channel with transition probability $P_{YZ|U}(y,z|u) = \sum_{x\in\mathcal{X}} P_{X|U}(x|u)W(y,z|x)$. Consequently, in view of Theorem 5.1, given any $P_{UX} = P_U \times P_{X|U}$, any rate up to $I(U;Y) - I(U;Z)$ will be achievable.

Theorem 5.2 ([29] also [37, Theorem 22.1])**.** *The secrecy capacity of the wiretap channel $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ is given by*

$$C_{\mathrm{s}}(W) = \max_{P_{UX} \in \mathcal{P}(\mathcal{U}\times\mathcal{X})} \{I(U;Y) - I(U;Z)\} \tag{5.8}$$

where $U \multimap X \multimap (Y,Z) \sim P_{UX}(u,x)W(y,z|x)$ and $|\mathcal{U}| \leq |\mathcal{X}|$.

The secrecy capacity equals

$$\max_{P_X \in \mathcal{P}(\mathcal{X})} \{I(X;Y) - I(X;Z)\} \tag{5.9}$$

if $W_{\mathrm{E}} \preceq_{\mathrm{c}} W_{\mathrm{M}}$, that is, the legitimate receiver's channel is more capable *than the eavesdropper's. Moreover, the secrecy capacity is positive unless $W_{\mathrm{M}} \preceq_{\mathrm{n}} W_{\mathrm{E}}$, i.e., the eavesdropper's channel is* less noisy *than the legitimate receiver's. (See Appendix A for the definitions of more capable and less noisy channels.)*

Remark 1. Channel prefixing is also useful for treating the channels with cost constraints [47]. (The auxiliary channel $P_{X|U} : \mathcal{U} \to \mathcal{X}$ can be chosen such that its output sequence satisfies the cost constraints of the physical channel $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$.)

Remark 2. It is obvious that any coding scheme for 'un-prefixed' system is immediately extensible to the prefixed system by replacing the channel $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ with the prefixed channel $P_{YZ|U} : \mathcal{U} \to \mathcal{Y} \times \mathcal{Z}$ (and setting the input of the prefixing channel $P_{X|U} : \mathcal{U} \to \mathcal{X}$ as the interface between the encoder and the channel). Consequently, in the remainder of this thesis we will not state the results explicitly for the prefixed channel, as we know that all presented results are applicable to the prefixed setting.

Here, we review the achievability proof (i.e., the proof of Theorem 5.1) and refer the reader interested in the converse proof (which, in turn, establishes Theorem 5.2) to [29].

Proof of Theorem 5.1 [76]. The result is established via a method called *random binning*: Fix $R > 0$ (to be tuned later). Given P_X, R_{s}, R, and n, let $M_{\text{s}} := \lceil \exp(nR_{\text{s}}) \rceil$ and $M := \lfloor \exp(nR) \rfloor$ and generate, a (large) codebook containing $M \times M_{\text{s}}$ codewords of block-length n, by sampling each codeword independently from the product distribution P_X^n (i.e., an i.i.d. random code of rate $R + R_{\text{s}} + o(1)$). This code can be partitioned into M_{s} sub-codes (or bins) $\mathscr{C}_s$, $s \in \{1, 2, \ldots, M_{\text{s}}\}$ of size M:

$$\begin{array}{rcl} \mathscr{C}_1 & = & \begin{pmatrix} x_{1,1}^n & x_{1,2}^n & \ldots & x_{1,M} \end{pmatrix} \\ \mathscr{C}_2 & = & \begin{pmatrix} x_{2,1}^n & x_{1,2}^n & \ldots & x_{2,M} \end{pmatrix} \\ & \vdots & \\ \mathscr{C}_{M_{\text{s}}} & = & \begin{pmatrix} x_{M_{\text{s}},1}^n & x_{M_{\text{s}},2}^n & \ldots & x_{M_{\text{s}},M}^n \end{pmatrix}. \end{array} \tag{5.10}$$

To communicate a message $s \in \{1, 2, \ldots, M_{\text{s}}\}$, the encoder transmits a uniformly chosen codeword from the sub-code $\mathscr{C}_s$ associated with the message s (via n independent uses of the channel). That is to say,

$$\mathsf{E}(x^n|s) = \begin{cases} \frac{1}{M} & \text{if } x^n \in \mathscr{C}_s \\ 0 & \text{otherwise.} \end{cases} \tag{5.11}$$

Stating differently, to communicate a message s, the encoder picks J uniformly from $\{1, 2, \ldots, M\}$ and transmits $x_{s,J}^n$ through the channel. In other words, it maps a *pair* of messages (S, J), where S is the secret message and J is a 'junk' message (drawn uniformly at random from $\{1, 2, \ldots, M\}$) to a codeword from the codebook $\mathscr{C}$.

The code and its partitioning, as well as the encoding scheme, are revealed to all parties (Alice, Bob, and Eve).

We now show that such a randomly constructed code, with high probability over the choice of the code, leads to a reliable and secure communication scheme (provided that n is large enough). More precisely, we will prove that $\forall \epsilon_1 \in (0, 1/2), \forall \epsilon_2 \in (0, 1/2)$, and large enough n, with probability $1 - (\epsilon_1 + \epsilon_2)$ over the choice of the code both (5.3) and (5.4) are satisfied.

Reliability Shannon's noisy channel coding theorem implies that, as long as,

$$R_{\rm s} + R < I(X;Y), \tag{5.12}$$

Bob can decode *both* messages (S, J) (thus, in particular S) with an arbitrarily small error probability, provided that the block-length of the code is large enough. More precisely, if $\hat{S}_{\rm MAP}(y^n)$ denotes the optimal (MAP) estimation of S, given Bob's channel output sequence y^n, there exists n_1 such that for $n \geq n_1$,

$$\mathbb{E}_{\mathscr{C}}\big[\Pr\{\hat{S}_{\rm MAP}(Y^n) \neq S\}\big] \leq \epsilon_1^2 \tag{5.13}$$

Therefore, Markov inequality implies, with probability at least $1 - \epsilon_1$ over the choice of the code,

$$\Pr\{\hat{S}_{\rm MAP}(Y^n) \neq S\} \leq \epsilon_1. \tag{5.14}$$

Secrecy Suppose a (bad) genie discloses the secret message S to Eve. Then, as long as

$$R < I(X;Z), \tag{5.15}$$

Eve can decode J with an arbitrarily small error probability. This is because knowing S, Eve knows which bin the sent codeword X^n is chosen from and, as the sub-codes are random codes of rate R themselves, Shannon's noisy channel coding theorem implies, for any $\delta > 0$ (to be tuned later), there exists n_2 such that for $n \geq n_2$,

$$\mathbb{E}_{\mathscr{C}_s}\big[\Pr\{\hat{J}_{\rm MAP}(S, Z^n) \neq J | S = s\}\big] \leq \delta\frac{\epsilon_2}{2}. \tag{5.16}$$

In the above $\hat{J}_{\rm MAP}(s, z^n)$ stands for the MAP estimation of J upon observing z^n and s. Consequently,

$$\mathbb{E}_{\mathscr{C}}\big[\Pr\{\hat{J}_{\rm MAP}(S, Z^n) \neq J\}\big] \leq \delta\frac{\epsilon_2}{2} \tag{5.17}$$

Therefore, with probability at least $1 - \epsilon_2/2$ over the choice of codewords

$$\Pr\{\hat{J}_{\rm MAP}(S, Z^n) \neq J\} < \delta. \tag{5.18}$$

This, together with Fano's inequality [30, Lemma 3.8] implies that with probability at least $1 - \epsilon_2/2$,

$$\frac{1}{n}H(J|S; Z^n) \leq \frac{1}{n} + R \cdot \delta \tag{5.19}$$

which, for $n \geq 4/\epsilon_2 =: n_3$, can be further upper bounded as

$$\frac{1}{n}H(J|S, Z^n) \leq \frac{\epsilon_2}{4} + R \cdot \delta. \tag{5.20}$$

Since J is uniformly distributed on $\{1, 2, \ldots, M\}$ and independent of S,

$$\frac{1}{n}I(S, J; Z^n) = \frac{1}{n}[I(S; Z^n) + I(J; Z^n|S)] \tag{5.21}$$

$$= \frac{1}{n}[I(S; Z^n) + I(J; S, Z^n)] \tag{5.22}$$

$$= \frac{1}{n}I(S; Z^n) + R - \frac{1}{n}H(J|S, Z^n). \tag{5.23}$$

Using (5.20) in (5.23) we conclude that, for $n \geq \max\{n_2, n_3\}$ with probability at least $1 - \epsilon_2/2$,

$$\frac{1}{n}I(S, J; Z^n) \geq \frac{1}{n}I(S; Z^n) + R(1 - \delta) - \frac{\epsilon_2}{4}. \tag{5.24}$$

Moreover, since $(S, J) \multimap X^n \multimap (Y^n, Z^n)$,

$$\frac{1}{n}I(S, J; Z^n) \leq \frac{1}{n}I(X^n; Z^n) \leq \frac{1}{n}\sum_{i=1}^{n} I(X_i; Z_i). \tag{5.25}$$

The right-hand side of the above is the average of i.i.d. random variables. As n grows, it concentrates around its mean, that is

$$\mathbb{E}_{\mathscr{C}}\left[I(X_i, Z_i)\right] \leq I(X; Z) \tag{5.26}$$

The last inequality follows since $I(X; Z)$ is a concave function of P_X and $\mathbb{E}_{\mathscr{C}}[P_{X_i}] = P_X$. Consequently, there exists n_4 such that for $n \geq n_4$, with probability at least $1 - \epsilon_2/2$,

$$\frac{1}{n}\sum_{i=1}^{n} I(X_i; Z_i) < I(X; Z) + \frac{\epsilon_2}{4} \tag{5.27}$$

which, together with (5.25), implies with probability at least $1 - \epsilon_2/2$,

$$\frac{1}{n}I(S, J; Z^n) \leq I(X; Z) + \frac{\epsilon_2}{4}. \tag{5.28}$$

Combining (5.24) and (5.28) we conclude that for $n \geq \max\{n_2, n_3, n_4\}$, with probability at least $1 - \epsilon_2$,

$$\frac{1}{n}I(S; Z^n) \leq I(X; Z) - R(1 - \delta) + \frac{\epsilon_2}{2} \tag{5.29}$$

Set $R = I(X; Z) - \epsilon_2/4$ and observe that this ensures (5.15) is satisfied. Let also $\delta = \epsilon_2/(4 \log |\mathcal{X}|)$. Plugging these values in (5.29) shows that, with probability at least $1 - \epsilon_2$,

$$\begin{aligned}\frac{1}{n}I(S; Z^n) &\leq I(X; Z) - \Big(I(X; Z) - \frac{\epsilon_2}{4}\Big)\Big(1 - \frac{\epsilon_2}{4 \log |\mathcal{X}|}\Big) + \frac{\epsilon_2}{2} \\ &= \frac{\epsilon_2}{4 \log |\mathcal{X}|} \cdot I(X; Z) + \frac{\epsilon_2}{4}\Big(1 - \frac{\epsilon_2}{4 \log |\mathcal{X}|}\Big) + \frac{\epsilon_2}{2} \leq \epsilon_2.\end{aligned} \tag{5.30}$$

(Note that $I(X;Z) \leq \log |\mathcal{X}|$.)

Finally, $R_{\rm s} < I(X;Y) - I(X;Z)$ implies $R + R_{\rm s} < I(X;Y) - \epsilon_2/4$, therefore (5.12) holds. Thus, in particular, for $n \geq n_1$, with probability $1 - \epsilon_1$ over the choice of codes (5.3) is satisfied. Therefore, for $n \geq \max\{n_1, n_2, n_3, n_4\} =: n_0$, with probability at least $1 - (\epsilon_1 + \epsilon_2) > 0$, both (5.3) and (5.4) are satisfied simultaneously. □

Note that the encoder defined in the achievability proof needs access to a random-number generator with entropy rate R. The entropy rate required by the stochastic encoder is called the *random-binning rate*. As can be seen through the proof, the role of the external randomness is to pack the eavesdropper's channel with junk in order to not leave any more room for useful information about the secret message S to be carried through the channel. We can also check that by varying the random-binning rate, we can trade the information leakage rate for the secret-message rate [29, 118]. We are surprised to observe that (5.15) puts an *upper bound* on the random-binning rate. This is counter-intuitive: putting more 'junk' in Eve's channel should improve the secrecy. What happens if we increase the random-binning rate above $I(X;Z)$? We will answer this in § 5.3.

5.2 Channel Resolvability

Let us put aside the problem of secure communications for a moment and formally define the notion of *channel resolvability* [48, 117].[2]

Consider the setting represented in Figure 5.2: If the input to the n-fold use of a discrete memoryless channel (DMC) $P_{V|U} : \mathcal{U} \to \mathcal{V}$ is drawn from distribution P_{U^n}, it is obvious that its output sequence V^n will have distribution $P_{V^n} = P_{U^n} \circ P^n_{V|U}$ (see Figure 5.2a). One particular way to feed the channel is to have its input sequence generated by a *deterministic* encoder, Enc : $\{1, 2, \ldots, M\} \to \mathcal{U}^n$, which maps a uniformly distributed word $J \in \{1, 2, \ldots, M\}$ to codewords of length n. This corresponds to taking P_{U^n} to be the uniform distribution over the codebook

$$\mathscr{C} := \big(u^n_{(j)} = \mathrm{Enc}(j) \colon j \in \{1, 2, \ldots, M\}\big) \tag{5.31}$$

and results in the output distribution

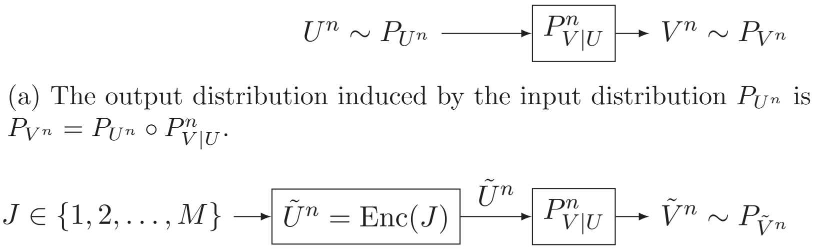
$$P_{\tilde{V}^n}(v^n) = P_{\mathscr{C}}(v^n) := \frac{1}{M} \sum_{j=1}^{M} P^n_{V|U}\big(v^n | u^n_{(j)}\big). \tag{5.32}$$

In this case (cf. Figure 5.2b), the encoder requires access to a random-number generator of rate $\frac{1}{n}\log(M)$ bits per channel use, whereas, to feed the channel

[2] To our knowledge, the term "resolvability" is coined by Han and Verdú in [48]. In some works, including Wyner's [117], the problem is studied under the name "soft covering."

with sequences drawn from an arbitrary P_{U^n}, an entropy rate of $\frac{1}{n}H(U^n)$ is required.

Given any P_{U^n} that induces $P_{V^n} = P_{U^n} \circ P^n_{V|U}$ at the output of the channel, is it possible to design a code $\mathscr{C}$ such that the output of the channel, $P_{\tilde{V}^n}$ is close to P_{V^n} — i.e., to make it hard for an observer who perceives V^n or $\tilde{V}^n$ to distinguish between the cases (a) and (b) in Figure 5.2? If yes, how much entropy rate R is required for the *resolvability* encoder to simulate the desired distribution P_{V^n} arbitrarily accurately?


$U^n \sim P_{U^n}$
$P^n_{V|U}$
$V^n \sim P_{V^n}$
$J \in \{1, 2, \dots, M\}$
$\tilde{U}^n = \text{Enc}(J)$
$\tilde{U}^n$
$P^n_{V|U}$
$\tilde{V}^n \sim P_{\tilde{V}^n}$

(a) The output distribution induced by the input distribution P_{U^n} is $P_{V^n} = P_{U^n} \circ P^n_{V|U}$.

(b) The output distribution when uniformly chosen $J \in \{1, 2, \dots, M\}$ is encoded into a codeword and transmitted is as (5.32).

Figure 5.2: Channel Resolvability

Definition 5.3. A rate R is said to be an *achievable resolvability rate* over the channel $P_{V|U} : \mathcal{U} \to \mathcal{V}$ and with respect to the sequence of reference measures $(P_{V^n} \in \mathcal{P}(\mathcal{V}^n), n \in \mathbb{N})$, if for every $\epsilon > 0$, there exists n_0 such that for every $n \geq n_0$, we can find a resolvability encoder of block-length n and rate at most R, Enc: $\{1, 2, \dots, M\} \to \mathcal{U}^n$, with $M \leq \exp(nR)$, such that if $P_{\tilde{V}^n}$ is the output distribution of $P^n_{V|U}$ (the n-fold use of the channel $P_{V|U}$) when its input is $\tilde{U}^n = \text{Enc}(J)$ with J uniformly distributed in $\{1, 2, \dots, M\}$ (see Figure 5.2b), we have

$$\frac{1}{n}D(P_{\tilde{V}^n} \| P_{V^n}) \leq \epsilon. \tag{5.33}$$

The infimum of all achievable resolvability rates over the channel $P_{V|U} : \mathcal{U} \to \mathcal{V}$ with respect to the sequence of reference measures $(P_{V^n}, n \in \mathbb{N})$ is called the *resolution* of the channel $P_{V|U}$ with respect to the sequence of reference measures $(P_{V^n}, n \in \mathbb{N})$.

Theorem 5.3. *The resolution of the channel $P_{V|U} : \mathcal{U} \to \mathcal{V}$ with respect to the sequence of i.i.d. reference measures $(P^n_V, n \in \mathbb{N})$ equals [117]:*

$$\min_{\substack{(\tilde{U},\tilde{V}): \\ P_{\tilde{V}|\tilde{U}} = P_{V|U} \\ P_{\tilde{V}} = P_V}} I(\tilde{U}; \tilde{V}). \tag{5.34}$$

The above holds also if the distance measure in (5.33) *is replaced by ℓ_1 norm [48], i.e., we ask for*

$$|P_{\tilde{V}^n} - P_{V^n}| \leq \epsilon, \tag{5.35}$$

or unnormalized divergence [51, 57], i.e., if we require that

$$D(P_{\tilde{V}^n} \| P_{V^n}) \leq \epsilon. \tag{5.36}$$

Theorem 5.3 has a natural interpretation if we regard the entropy rate as a valuable resource: Generating an i.i.d. sequence, distributed according to P_V, requires an entropy rate of $H(V)$ bits per symbol. The channel provides us with $H(V|U)$ bits per symbol. Hence, if we compensate for the difference, which is $I(U;V)$, we should be able to achieve our goal.

Our results in Chapter 6 imply the achievability part of Theorem 5.3 and its converse part will follow from the more general converse proof of Theorem 7.1 in Chapter 7 (cf. § 7.2.1). Therefore we will not prove Theorem 5.3 here. We just mention that finding a good resolvability code is easy based on the probabilistic method: Let $\mathscr{C}_n$ be a random code of block length n and rate $R > I(U;V)$ obtained by sampling the codewords independently form the product measure P_U^n (for some P_U that induces P_V at the output of the channel $P_{V|U}$). For sufficiently large n, the *ensemble average* of the divergence between the output distribution induced by the code (5.32) and the reference product measure $\mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n} \| P_V^n)]$ will be arbitrarily small. Therefore, such a random code is, with high probability, a good resolvability code. Note the analogy between the error-correction and resolvability problem: A random code of rate $R < I(U;V)$ is (with high probability) a good error-correction code, whereas if the rate is increased above $I(U;V)$, it will (with high probability) be good resolvability code.

5.3 Strong Secrecy from Channel Resolvability

Common sense turns out to be right! The fact that we have an upper bound on the random-binning rate is only an artifact of the proof technique: The achievability proof we reviewed at the end of § 5.1, treats $I(X;Z)$ as the capacity of Eve's channel $W_{\mathrm{E}} : \mathcal{X} \to \mathcal{Z}$ and builds upon squeezing random junk (independent of the secret message) into her channel as much as possible so that nothing further can pass through that channel. With a slight abuse of terminology, such a proof is called a *capacity-based* achievability proof [21].

However (again with a slight abuse of terminology) $I(X;Z)$ also equals resolution of the channel $W_{\mathrm{E}} : \mathcal{X} \to \mathcal{Z}$ (see Theorem 5.3). Suppose the random-binning rate R is *above* $I(X;Z)$. To communicate each secret message $s \in \{1, 2, \ldots, M_{\mathrm{s}}\}$, Alice transmits a uniformly chosen codeword from its corresponding bin $\mathscr{C}_s$ that is a code of rate $R > I(X;Z)$. Therefore, $P_{Z^n|S=s}$, the conditional distribution of Eve's observation, given the secret message, will be close to P_Z^n, no matter which secret message is transmitted. Hence, Z^n must contain almost no information about S.

Let us formalize this argument: Obviously, if $R > I(X;Z)$, the 'secrecy' part of the proof does not work because (5.15) is violated to begin with. (But the scheme is still reliable provided that $R_{\mathrm{s}} + R < I(X;Y)$.)

However, as the achievability proof for Theorem 5.3 will show (see Chapter 6), each bin $\mathscr{C}_s = (X^n_{s,1}, X^n_{s,2}, \dots, X^n_{s,M})$ is (with high probability) a good resolvability code for the channel $W_\mathrm{E} : \mathcal{X} \to \mathcal{Z}$ with respect to the product measure P^n_Z. In fact, given that a particular secret message $S = s$ is sent, the distribution of Alice's observable, $P_{Z^n|S=s}$, is the distribution induced at the output of W^n_E when it carries a uniformly chosen codeword from $\mathscr{C}_s$, that is, $P_{Z^n|S=s} = P_{\mathscr{C}_s}$. Therefore, for any $\epsilon_2 > 0$, there exists n_2 such that for $n \geq n_2$,

$$\mathbb{E}_{\mathscr{C}_s}\big[D(P_{Z^n|S=s}\|P^n_Z)\big] = \mathbb{E}_{\mathscr{C}_s}\big[D(P_{\mathscr{C}_s}\|P^n_Z)\big] \leq \epsilon_2^2, \tag{5.37}$$

which, together with the linearity of expectation and the fact that the subcodes have the same distribution, yields

$$\mathbb{E}_{\mathscr{C}}\big[D(P_{Z^n|S}\|P^n_Z|P_S)\big] \leq \epsilon_2^2. \tag{5.38}$$

Consequently, with probability at least $1 - \epsilon_2$ over the choice of the code,

$$D(P_{Z^n|S}\|P^n_Z|P_S) \leq \epsilon_2. \tag{5.39}$$

Moreover, for any distribution $Q_{Z^n} \in \mathcal{P}(\mathcal{Z}^n)$,

$$I(S; Z^n) = D(P_{Z^n|S}\|Q_{Z^n}|P_S) - D(P_{Z^n}\|Q_{Z^n}) \leq D(P_{Z^n|S}\|Q_{Z^n}|P_S). \tag{5.40}$$

Therefore, taking $Q_{Z^n} = P^n_Z$ in (5.40), we have from (5.39) that, for $n \geq n_2$, with probability at least $1 - \epsilon_2$ over the choice of codes,

$$I(S; Z^n) \leq \epsilon_2, \tag{5.41}$$

which not only implies (5.4) but also establishes *strong secrecy*: Setting the random-binning rate *just above* $I(X; Z)$ guarantees that the *unnormalized* amount of information leaked to Eve will be arbitrarily small (provided that the block-length n is sufficiently large).

This actually shows that Theorem 5.1, and as a consequence Theorem 5.2, hold under the strong secrecy criteria. That is, they characterize the achievable secrecy rates and secrecy capacity of the system, respectively, in the sense of Definition 5.2

Remark 1. In addition to simplifying the secrecy part of the achievability proof for the wiretap channel model, a resolvability-based approach to secrecy is useful for establishing secrecy for many other variants of the wiretap channel model (e.g., non-stationary channels, channels with cost constraint, ...) as discussed in [21].

Remark 2. If the distance metric for resolvability is the ℓ_1 distance (which is, due to Pinsker's inequality [30, Exercise 3.18], weaker than the KL divergence), it is still possible to establish strong secrecy from channel resolvability by using the absolute continuity of entropy [30, Lemma 2.7] and the exponential decay of the ℓ_1 distance between the code-induced output distribution (5.32) and the reference (product) measure in n (cf. [27] for example).

Remark 3. A (good) resolvability-based coding scheme for the wiretap channel guarantees that $D(P_{Z^n|S}\|P_Z^n|P_S)$ is as small as desired (see (5.39)). Such a code can, in turn, be expurgated to a code of essentially the same rate for which $\max_s D(P_{Z^n|S=s}\|P_Z^n)$ is small (see the proof of Theorem 5.4), which is a worst-case guarantee (it establishes semantic secrecy). This, also implies *stealth*: Eve cannot even detect if Alice communicates messages to Bob by encoding them to codewords from the codebook and transmitting them through the channel or if Alice is just feeding the channel with random symbols drawn from the input distribution — let aside estimating the content of those messages [58].

5.4 The Secrecy Exponent

So far we have learned that, in the presence of an eavesdropper, secure and reliable communication is feasible, provided that the communication rate is below the secrecy capacity of the wiretap channel (5.8).

A natural question is to wonder how large n needs to be in order to achieve a desired reliability–secrecy level of (ϵ_1, ϵ_2) (as discussed in Definition 5.2, see equations (5.3) and (5.5)) and if we can trade the secret-message rate R_{s} for better secrecy or reliability.

The (first order) answer to the question on the reliability should be easy: We know that the error-probability of a randomly constructed code (as we described in the achievability proof of Theorem 5.1) decays (on average) exponentially fast in the block-length, and we also know how to compute the rate of this decay [42, Theorem 5.6.2]. Hence, it follows straightforwardly that it is possible to construct coding schemes, of secret-message rate R_{s} and random-binning rate R, for the wiretap channel whose probability of error decays (at least) as fast as $\exp\{-nE_{\mathrm{r}}(R+R_{\mathrm{s}})\}$ where E_{r} is the random-coding error exponent (evaluated for the legitimate receiver's channel W_{M}).

It turns out that, with the same coding scheme, information leakage $I(S;Z^n)$ also decays exponentially fast in the block-length [27, 47, 51–54, 58]. Therefore, it makes sense to measure the *secrecy exponent* of the system.

Definition 5.4. Given a rate pair (R_{s}, R) and a wiretap channel $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$, a number E is an achievable *secrecy exponent* if for every $\epsilon > 0$ and $\delta > 0$, there exists n_0 such that for every $n \geq n_0$, there exists a coding scheme of block-length n, $\mathsf{E}\colon \{1, 2, \ldots, M_{\mathrm{s}}\} \to \mathcal{X}^n$, with secret-message size $M_{\mathrm{s}} \geq \exp(nR_{\mathrm{s}})$ and random-binning rate at most R,

$$\frac{1}{n}H(X^n|S) \leq R \tag{5.42}$$

which is reliable, i.e., yields,

$$\Pr\{\hat{S}_{\mathrm{MAP}}(Y^n) \neq S\} \leq \epsilon, \tag{5.43}$$

and guarantees

$$I(S;Z^n) \leq \exp\{-n[E-\delta]\}. \tag{5.44}$$

As before, in the above, $\hat{S}_{\text{MAP}}$ is the optimal (MAP) decision on S given Y^n, the collection $S \multimap X^n \multimap (Y^n, Z^n)$ has distribution $P_S(s)\mathsf{E}(x^n|s)W^n(z^n, y^n|x^n)$ (cf. Figure 5.1), and P_S is the uniform distribution on $\{1, 2, \ldots, M_{\text{s}}\}$.

Hayashi [51] was the first to derive a lower bound to the achievable secrecy exponents by using a resolvability-based construction of wiretap channel codes (as we described in § 5.3). Later on, he showed that this lower bound can be improved by *privacy amplification* — specifically, by using a random hash function, on top of a random code in the construction of the encoder–decoder pair [52]. More recently, it was shown (see special cases of [47, Theorem 3.1], [54, Theorem 2], or the proof given in [13]) that privacy amplification is unnecessary: when a randomly constructed wiretap channel code (as we described in § 5.3) is used for communication, the exponent derived in [52] lower-bounds the exponential decay rate of the ensemble average of the information leaked to Eve.

In fact, the exponential decay of the information leakage in the wiretap model follows by proving that, when $\mathscr{C}_n$ is a randomly constructed code, the divergence between the distribution $P_{\mathscr{C}_n}$ (as defined in (5.32)) and the reference product measure P_V^n will be exponentially small in block-length n (in expectation) [13, 21, 27, 31, 32, 47, 48, 51–54, 57, 58]. This, together with (5.40), implies the exponential decay of the information leakage. Let us define the *resolvability exponent* as well.

Definition 5.5. Given a stationary memoryless channel $P_{V|U} : \mathcal{U} \to \mathcal{V}$, a rate R, and a sequence of target distributions $(P_{V^n} \in \mathcal{P}(\mathcal{V}^n), n \in \mathbb{N})$, a number $E(R)$ is an achievable *resolvability exponent* over the channel $P_{V|U}$, at rate R, with respect to $(P_{V^n}, n \in \mathbb{N})$ if for every $\delta > 0$, there exists n_0 such that for every $n \geq n_0$ we can find a resolvability encoder of block-length n and rate at most R, that is, a deterministic mapping, $\text{Enc}\colon \{1, 2, \ldots, M\} \to \mathcal{U}^n$, with $M \leq \exp(nR)$, using which

$$D(P_{\tilde{V}^n} \| P_{V^n}) \leq \exp\{-n[E(R) - \delta]\}. \tag{5.45}$$

In the above, $P_{\tilde{V}^n}$ is the output distribution of $P_{V|U}^n$ when its input is $\tilde{U}^n = \text{Enc}(J)$, with J uniformly distributed on $\{1, 2, \ldots, M\}$ (cf. Figure 5.2b).

Instead of looking for sophisticated prescriptions on how to construct good resolvability codes and computing the resolvability exponents they achieve, we can rely on the probabilistic method and analyze the achievable resolvability exponents via an ensemble of random codes.

Definition 5.6. Given $\Pi = \big(P_{U^n} \in \mathcal{P}(\mathcal{U}^n), n \in \mathbb{N}\big)$, a sequence of probability distributions on $\mathcal{U}^n$, an *ensemble of random codes* of rate (at most) R is a sequence of random codes $\mathscr{C}_n$ of block-length n and size $M = \lfloor \exp(nR) \rfloor$ obtained by sampling the codewords independently from the distribution P_{U^n}.

In other words,

$$\Pr\bigl\{\mathscr{C}_n = \bigl(u^n_{(1)}, \ldots, u^n_{(M)}\bigr)\bigr\} = \prod_{i=1}^{M} P_{U^n}\bigl(u^n_{(i)}\bigr). \tag{5.46}$$

Two important classes of random codes are the ensemble of *i.i.d. random codes*, defined by the sequence of i.i.d. codeword-sampling distributions

$$P_{U^n}(u^n) = P_U^n(u^n) \tag{5.47}$$

for some $P_U \in \mathcal{P}(\mathcal{U})$, and the ensemble of *constant-composition* random codes, defined by the sequence of codeword-sampling distributions

$$P_{U^n}(u^n) = \frac{\mathbb{1}\bigl\{u^n \in \mathcal{T}^n_{P_U^{(n)}}\bigr\}}{\bigl|\mathcal{T}^n_{P_U^{(n)}}\bigr|} \tag{5.48}$$

where $(P_U^{(n)} \in \mathcal{P}_n(\mathcal{U}), n \in \mathbb{N})$ is a sequence of n-types that converge to some $P_U \in \mathcal{P}(\mathcal{U})$ and $\mathcal{T}^n_{P_U^{(n)}} \subseteq \mathcal{U}^n$ is the set of all sequences of type $P_U^{(n)}$ (see § 6.1 for formal definitions).

Definition 5.7. Given $\Pi = \bigl(P_{U^n} \in \mathcal{P}(\mathcal{U}^n), n \in \mathbb{N}\bigr)$, a stationary memoryless channel $P_{V|U} : \mathcal{U} \to \mathcal{V}$, and a rate R, a number $\underline{E}_{\mathrm{s}}(\Pi, P_{V|U}, R)$ is an achievable resolvability exponent for the ensemble of random codes of rate (at most) R defined by Π, over the channel $P_{V|U}$, if for every $\delta > 0$, there exists n_0 such that for all $n \geq n_0$,

$$\mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n} \| P_{V^n})] \leq \exp\bigl\{-n\bigl[\underline{E}_{\mathrm{s}}(\Pi, P_{V|U}, R) - \delta\bigr]\bigr\}. \tag{5.49}$$

In the above $\mathscr{C}_n$ is a random code of size $M = \lfloor \exp(nR) \rfloor$ distributed according to (5.46), $P_{\mathscr{C}_n}$ is the distribution of the output sequence of $P^n_{V|U}$ when a uniformly chosen codeword from $\mathscr{C}_n$ is transmitted through the channel (see (5.32)), and the target measure P_{V^n} is the distribution induced by the codeword-sampling distribution P_{U^n} at the output of the product channel $P^n_{V|U}$, i.e.,

$$P_{V^n}(v^n) := (P_{U^n} \circ P^n_{V|U})(v^n) = \sum_{u^n \in \mathcal{U}^n} P_{U^n}(u^n) P^n_{V|U}(v^n|u^n). \tag{5.50}$$

If $\underline{E}_{\mathrm{s}}$ is an achievable resolvability exponent for an ensemble of random codes, a substantial portion of the codes in the ensemble achieve the exponent $\underline{E}_{\mathrm{s}}$ with respect to the sequence of reference measures $(P_{V^n}, n \in \mathbb{N})$ defined in (5.50): Given any $\delta > 0$ and $\epsilon > 0$, choose $\delta' < \delta$ and n large enough such that

$$\mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n} \| P_{V^n})] \leq \exp\bigl\{-n\bigl[\underline{E}_{\mathrm{s}}(\Pi, P_{V|U}, R) - \delta'\bigr]\bigr\}$$

Markov inequality implies that for any $\epsilon > 0$, at any block-length n, a fraction $(1-\epsilon)$ of codes satisfy

$$D(P_{\mathscr{C}_n} \| P_{V^n}) \leq \frac{1}{\epsilon}\, \mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n} \| P_{V^n})]. \tag{5.51}$$

Using any such code in a resolvability encoder that simply maps the input word J, to the codeword it indexes $u^n_{(J)}$, yields $P_{\tilde{V}^n} = P_{\mathscr{C}_n}$ and

$$\begin{aligned} D(P_{\tilde{V}^n} \| P_{V^n}) &\leq \frac{1}{\epsilon}\, \mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n} \| P_{V^n})] \\ &\leq \exp\Big\{-n\Big[\underline{E}_{\mathrm{s}}(\Pi, P_{V|U}, R) - \Big(\delta' + \frac{\log(1/\epsilon)}{n}\Big)\Big]\Big\} \end{aligned} \tag{5.52}$$

Since $\delta' < \delta$, for large enough n, $\delta' + \frac{\log(1/\epsilon)}{n} \leq \delta$, which shows $\underline{E}_{\mathrm{s}}$ is achievable.

Also note that, in the passage to the probabilistic method, we restricted the sequence of target measures to those induced by the codeword-sampling distribution P_{U^n} at the output of the n-fold use of $P_{V|U}$, as formalized in (5.50). There is a simple reason for this: When $\mathscr{C}_n$ is a random code whose codewords are drawn independently from P_{U^n}, for any distribution $Q_{V^n} \in \mathcal{P}(\mathcal{V}^n)$,

$$\mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n} \| Q_{V^n})] = \mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n} \| P_{V^n})] + D(P_{V^n} \| Q_{V^n}). \tag{5.53}$$

Therefore, to show the existence of good resolvability codes for approximating a sequence of target distributions $(Q_{V^n}, n \in \mathbb{N})$ via random-coding arguments, we can consider exclusively the ensembles of random codes whose sampling distributions $(P_{U^n}, n \in \mathbb{N})$ induce $(Q_{V^n}, n \in \mathbb{N})$ at the output of the n-fold use of the channel — any other ensemble is *suboptimal* due to the residual divergence $D(P_{V^n} \| Q_{V^n})$.[3]

For the sake of completeness, let us also formally define the error exponent for an ensemble of random codes.

Definition 5.8. Given $\Pi = \big(P_{U^n} \in \mathcal{P}(\mathcal{U}^n), n \in \mathbb{N}\big)$, a stationary memoryless channel $P_{V|U} \colon \mathcal{U} \to \mathcal{V}$, and a rate R, a number $\underline{E}_{\mathrm{r}}(\Pi, P_{V|U}, R)$ is called an achievable *error exponent* of the ensemble Π at rate R on channel $P_{V|U}$, if for every $\delta > 0$, there exists n_0 such that for every $n \geq n_0$,

$$\mathbb{E}_{\mathscr{C}_n}[\Pr\{\hat{S}_{\mathrm{MAP}}(V^n) \neq S\}] \leq \exp\big\{-n\big[\underline{E}_{\mathrm{r}}(\Pi, P_{V|U}, R) - \delta\big]\big\} \tag{5.54}$$

when $\mathscr{C}_n$, a random code of size $M = \lceil \exp(nR) \rceil$ is used to communicate a uniformly chosen message $S \in \{1, 2, \ldots, M\}$ via n independent uses of $P_{V|U}$, V^n is the output sequence of $P^n_{V|U}$, and $\hat{S}_{\mathrm{MAP}}(v^n)$ is the optimal (MAP) estimation of S given the output sequence v^n.

[3] This actually explains why in most works on channel resolvability, e.g., [48, 51, 117], where the aim is to approximate an i.i.d. product distribution at the output of a DMC, the code for achievability proof is chosen randomly from the ensemble of i.i.d. random codes.

We can now characterize the error and secrecy exponents, of a randomly constructed sequence of coding schemes for the wiretap channel, in terms of error and resolvability exponents of the ensemble of random codes from which they are sampled. In fact, as we promised earlier, we also turn our average case guarantees to worst-case guarantees:

Theorem 5.4 ([13, 16])**.** *Let $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ be a wiretap channel and $W_{\mathrm{M}} : \mathcal{X} \to \mathcal{Y}$ and $W_{\mathrm{E}} : \mathcal{X} \to \mathcal{Z}$ be its corresponding legitimate receiver's and wiretapper's marginals, respectively (see Figure 5.1). Let also $\Pi = \big(P_{X^n} \in \mathcal{P}(\mathcal{X}^n), n \in \mathbb{N}\big)$ be a sequence of codeword-sampling distributions that define an ensemble of random codes (see Definition 5.6).*

If $\underline{E}_{\mathrm{r}}(\Pi, W_{\mathrm{M}}, R)$ is an achievable error exponent for the ensemble Π over the channel W_{M} at rate R that is continuous in R, and if $\underline{E}_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R)$ is an achievable resolvability exponent of the ensemble Π over the channel W_{E} (cf. Definitions 5.8 and 5.7, respectively), then, for every $\delta > 0$, there exists a coding scheme of (possibly large) block-length n for the wiretap channel, denoted as $\mathsf{E}\colon \{1, 2, \ldots, M_{\mathrm{s}}\} \to \mathcal{X}^n$, with secret-message size $M_{\mathrm{s}} \geq \exp(nR_{\mathrm{s}})$ and random-binning rate at most R (i.e., $\frac{1}{n}H(X^n|S) \leq R$), using which

$$\max_{P_S}\big\{\Pr\{\hat{S}_{\mathrm{MAP}}(Y^n) \neq S\}\big\} \leq \exp\big\{-n\big[\underline{E}_{\mathrm{r}}(\Pi, W_{\mathrm{M}}, R + R_{\mathrm{s}}) - \delta\big]\big\}, \tag{5.55}$$

$$\max_{P_S}\big\{I(S; Z^n)\big\} \leq \exp\big\{-n\big[\underline{E}_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R) - \delta\big]\big\} \tag{5.56}$$

when $S \multimap X^n \multimap (Y^n, Z^n)$ has distribution $P_S(s)\mathsf{E}(x^n|s)W^n(z^n, y^n|x^n)$ (and $\hat{S}_{\mathrm{MAP}}(Y^n)$ is the MAP estimation of S given Y^n).

Proof. Consider a randomly constructed code for the wiretap channel of block-length n, secret-message size $2M_{\mathrm{s}}$, with $M_{\mathrm{s}} = \lceil \exp(nR_{\mathrm{s}}) \rceil$, and random-binning rate R as described in the proof of Theorem 5.1: Specifically, those obtained by partitioning a random code of size $2M_{\mathrm{s}} \times M$ into $2M_{\mathrm{s}}$ sub-codes of size $M = \lfloor \exp(nR) \rfloor$, labeled as $\mathscr{C}_s$, $s = 1, 2, \ldots, 2M_{\mathrm{s}}$, with encoder $\mathsf{E} : \{1, 2, \ldots, 2M_{\mathrm{s}}\} \to \mathcal{X}^n$ defined as

$$\mathsf{E}(x^n|s) = \frac{1}{M}\mathbb{1}\{x^n \in \mathscr{C}_s\}. \tag{5.57}$$

Let

$$\begin{aligned}
\bar{P}_{\mathrm{e}} &:= \mathbb{E}_{\mathscr{C}}[\Pr\{\hat{S}_{\mathrm{MAP}}(Y^n) \neq S\}] && (5.58)\\
&= \mathbb{E}_{\mathscr{C}}\left[\frac{1}{2M_{\mathrm{s}}}\sum_{s=1}^{2M_{\mathrm{s}}} \Pr\{\hat{S}_{\mathrm{MAP}}(Y^n) \neq S | S = s\}\right], && (5.59)\\
\bar{D} &:= \mathbb{E}_{\mathscr{C}}\left[\frac{1}{2M_{\mathrm{s}}}\sum_{s=1}^{2M_{\mathrm{s}}} D(P_{\mathscr{C}_s} \| P_{Z^n})\right]. && (5.60)
\end{aligned}$$

with Y^n denoting the output sequence of the legitimate receiver's channel (as in Figure 5.1); $P_{\mathscr{C}_s}$ being the distribution of wiretapper's channel output sequence when a uniformly chosen codeword from the sub-code $\mathscr{C}_s$ is transmitted (see (5.32)); and $P_{Z^n} = P_{X^n} \circ W_{\mathrm{E}}^n$, the distribution induced by the codeword-sampling distribution at the output of the wiretapper's channel (see (5.50)). Recall that in (5.58) we have assumed that S is uniformly distributed on $\{1, 2, \dots, 2M_{\mathrm{s}}\}$.

Pick any $\delta' \in (0, \delta)$. As we have seen in the proof of Theorem 5.1, the probability of misdecoding the secret message s is upper bounded by the probability of decoding either the secret message and the junk message (i.e., the random index of the codeword in bin $\mathscr{C}_s$) incorrectly. As the error exponent is assumed to be continuous in rate, for sufficiently large n,

$$\bar{P}_{\mathrm{e}} \leq \exp\big\{-n\big[\underline{E}_{\mathrm{r}}(\Pi, W_{\mathrm{M}}, R_{\mathrm{s}} + R) - \delta'\big]\big\}. \tag{5.61}$$

Furthermore, using the linearity of expectation and the fact that the sub-codes $\mathscr{C}_s$ are identically distributed, for sufficiently large n,

$$\bar{D} \leq \exp\big\{-n\big[\underline{E}_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R) - \delta'\big]\big\}. \tag{5.62}$$

Markov's inequality implies, with probability of at least 2/3 over the choice of random codes,

$$\frac{1}{2M_{\mathrm{s}}} \sum_{s=1}^{2M_{\mathrm{s}}} \Pr\{\hat{S}_{\mathrm{MAP}}(Y^n) \neq S | S = s\} \leq 3\bar{P}_{\mathrm{e}}, \tag{5.63}$$

and, with probability of at least 2/3

$$\frac{1}{2M_{\mathrm{s}}} \sum_{s=1}^{2M_{\mathrm{s}}} D(P_{\mathscr{C}_s} \| P_{Z^n}) \leq 3\bar{D}. \tag{5.64}$$

Therefore, with probability of at least 1/3, the random code is chosen such that both bounds of (5.63) and (5.64) simultaneously hold. Let $(\mathscr{C}_s^\star, s \in \{1, 2, \dots, 2M_{\mathrm{s}}\})$ be the collection of sub-codes that define any such good code. Since the summands in the summation of (5.63) are all positive, there exists a subset $\mathcal{S}_{n,\mathrm{e}} \subseteq \{1, 2, \dots, 2M_{\mathrm{s}}\}$ of cardinality $|\mathcal{S}_{n,\mathrm{e}}| > \frac{3}{2}M_{\mathrm{s}}$ such that $\forall s \in \mathcal{S}_{n,\mathrm{e}}$,

$$\Pr\{\hat{S}_{\mathrm{ML}}(Y^n) \neq S | S = s\} \leq 12\bar{P}_{\mathrm{e}}. \tag{5.65}$$

Similarly, since the summands in (5.64) are positive, there exists a subset $\mathcal{S}_{n,\mathrm{s}} \subseteq \{1, 2, \dots, 2M_{\mathrm{s}}\}$ of cardinality $|\mathcal{S}_{n,\mathrm{s}}| > \frac{3}{2}M_{\mathrm{s}}$ such that $\forall s \in \mathcal{S}_{n,\mathrm{s}}$

$$D(P_{\mathscr{C}_s} \| P_{Z^n}) \leq 12\bar{D}. \tag{5.66}$$

Set $\mathcal{S}_n := \mathcal{S}_{n,\mathrm{e}} \cap \mathcal{S}_{n,\mathrm{s}}$ and note that $|\mathcal{S}_n| = |\mathcal{S}_{n,\mathrm{e}} \cap \mathcal{S}_{n,\mathrm{s}}| \geq M_{\mathrm{s}}$. Consider the encoder $\mathsf{E}^\star \colon \mathcal{S}_n \to \mathcal{X}^n$ defined as

$$\mathsf{E}^\star(x^n | s) = \frac{1}{M} \mathbb{1}\{x^n \in \mathscr{C}_s\}. \tag{5.67}$$

The secret-message rate of the code $\mathsf{E}^\star$ is at least $\log(M_\mathrm{s})/n = R_\mathrm{s}$ and its random-binning rate is $\log(M)/n \leq R$. Moreover, when it is employed with any prior P_S on secret messages, it satisfies

$$\Pr\{\hat{S}_\mathrm{MAP}(Y^n) \neq S\} \leq 12\bar{P}_\mathrm{e}, \tag{5.68}$$

due to (5.65), and

$$I(S; Z^n) \leq D(P_{Z^n|S} \| P_{Z^n} | P_S) \leq 12\bar{D}, \tag{5.69}$$

due to (5.66) and to the fact that $P_{Z^n|S=s} = P_{\mathscr{C}_s}$. Using this expurgated code,

$$\Pr\{\hat{S}_\mathrm{ML}(Y^n) \neq S\} \leq \exp\Big\{-n\Big[\underline{E}_\mathrm{r}(\Pi, W_\mathrm{M}, R_\mathrm{s}+R) - \Big(\delta' + \frac{\log(12)}{n}\Big)\Big]\Big\} \tag{5.70}$$

by combining (5.68) and (5.61), and

$$I(S; Z^n) \leq \exp\Big\{-n\Big[\underline{E}_\mathrm{s}(\Pi, W_\mathrm{E}, R_\mathrm{s}) - \Big(\delta' + \frac{\log(12)}{n}\Big)\Big]\Big\} \tag{5.71}$$

by combining (5.69) and (5.62), respectively. Taking n large enough so that $\delta' + \frac{\log(12)}{n} \leq \delta$ establishes (5.55) and (5.56). □

Remark 1. The secrecy part of the proof hinges on finding $\exp(nR_\mathrm{s})$ good resolvability codes via expurgation: we first generated twice as many resolvability codes as we needed and then threw away the 'bad' half. Recently, in [32], it was shown that the probability of choosing a bad resolvability code, namely a code $\mathscr{C}$ (of block-length n) for which the ℓ_1 distance between the output distribution $P_\mathscr{C}$ (5.32) and the reference measure P_{Z^n} is more than $\exp(-n\gamma)$ for some exponent γ, is *doubly exponentially small* in n. This suggests that even if we draw $\exp(nR_\mathrm{s})$ codes in a single shot from the ensemble, with very high probability they are *all* good resolvability codes. Nevertheless, we do not know if the results of [32] hold for any achievable exponent. (Also to establish the secrecy exponent, we need an exponentially small KL divergence between the code-induced distribution and the reference measure as opposed to ℓ_1 norm. But, at least for the i.i.d. random-coding ensemble, the KL divergence has the same exponential decay rate as the ℓ_1 distance [31, Equation (30)].)

Remark 2. Theorem 5.4 relates the error and resolvability exponents of an ensemble of random codes in the *ensemble-average sense* — i.e., the exponential decay rate of the ensemble average of the error probability and output divergence, respectively — to the error and secrecy exponents of a randomly constructed code for the wiretap channel from this ensemble. The symmetry that *random coding* puts into the problem is crucial in deriving the results: For instance, we might be able to find better resolvability codes (than an average random code) by carefully crafting the codewords. However, such codes might not be readily usable for constructing a code for the wiretap channel because of the poor error-correction performance of the union of M_s such codes.

Remark 3. Equations (5.55) and (5.56) suggest a trade-off in code design, in terms of the choice of input distributions, $\Pi = (P_{X^n} \in \mathcal{P}(\mathcal{X}^n), n \in \mathbb{N})$. The sequence of input distributions Π that maximizes $\underline{E}_\mathrm{s}$ might not coincide with the one that maximizes $\underline{E}_\mathrm{r}$.

Remark 4. To prove Theorem 5.4, we expurgated a good average-case code for the wiretap channel to obtain a good worst-case code of essentially the same rate. This, in particular, shows that the secrecy capacity remains the same under the *semantic secrecy* requirement.

5.5 Summary

In this chapter, to study data transmission in the presence of an eavesdropper, we have reviewed the wiretap channel model, introduced by Wyner [118] and generalized later by Csiszár and Körner [29]. We have seen that, as long as the rate of secret information is below the secrecy capacity of the channel (see Theorem 5.2), reliable *and secure* communication with the legitimate receiver is feasible: At a sufficiently large block-length, we can construct communication schemes in which the legitimate receiver can decode the secret messages with arbitrarily low error probability, while keeping the amount of information the eavesdropper learns about the secret message arbitrarily small. In fact, associating each secret message with a randomly constructed code, whose rate is just above the mutual information developed across the wiretapper's channel, and using a random-binning encoder, with high probability over the choice of codes, guarantees that the legitimate receiver can decode the secret message with *exponentially small* probability of error; and the amount of information leaked to the eavesdropper about the secret message is also exponentially small in block-length. Thus, by studying the error and secrecy exponents of the model, we can understand how the block-error probability and information leakage of the system scale in block-length. (We refer the reader interested in the trade-off between error and secrecy exponents to [24].)

Error exponents have been the subject of research for decades and, currently, we have a good understanding of achievable error exponents through random codes, of methods for achieving higher exponents compared to what an average code achieves, and of the best error exponents that can hoped to achieve (known as the *sphere-packing* exponent) [30, 42]. The state-of-the-art on secrecy exponents is, however, less mature. To our knowledge, to date, only lower bounds on achievable secrecy exponents via random-coding arguments are known — the best of which are those reported in [13, 47, 52, 54]. How tight are those lower bounds? In Chapter 6, we partially answer this question. We derive exponentially tight bounds on the ensemble average of the information leaked to the eavesdropper. Among other conclusions, our results will show that the lower bounds of [13, 47, 52, 54] on the achievable secrecy exponents are indeed tight for an average code.

6 Exact Random-Coding Secrecy Exponents for the Wiretap Channel

In Chapter 5, we have seen that when a randomly constructed code of large block length n for the wiretap channel is employed for communicating secret messages in presence of an eavesdropper (as illustrated in Figure 5.1), the information the eavesdropper learns about the secret message will be exponentially small in n. We denoted, as the *secrecy exponent* of the model, the rate of the exponential decay of the information leaked to the eavesdropper (cf. Definition 5.4). So far, the largest known lower bound on the achievable secrecy exponents is reported in [47, 52, 54], but the optimality of this bound is unclear. In fact, to our knowledge, no *upper bound* on the best achievable secrecy exponents for the wiretap channel is known. In this chapter, we derive the *ensemble-optimal* random-coding secrecy exponents for the wiretap channel, i.e., we compute

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{E}_{\mathscr{C}_n}[I(S;Z^n)]$$

(where S is the secret message, Z^n is the eavesdropper's observation, and a sequence of randomly constructed codes of block-length n, $\mathscr{C}_n$, are used for communication).

When the random code is sampled from the ensemble of i.i.d. random codes, the exponent derived from our method matches the one reported in [47,52,54]. This shows that the previously known lower bound on the achievable secrecy exponent is indeed *tight* for an *average* code. Stated differently, by looking at the ensemble average of the information leaked to Eve in the wiretap channel model, no better exponent can be shown to be achievable.[1]

[1] We remind the reader that in [52] the achievability of the exponent is shown using a randomly constructed code sampled from the ensemble of i.i.d. random codes *and* a randomly sampled hash function in the construction of the communication scheme (i.e., using

We also extend our analysis to the randomly constructed codes for the wiretap channel sampled from the ensemble of *constant-composition* random codes. Constructing codes for the wiretap channel from the ensemble of random constant-composition codes has been adopted in [53] in order to study the *universally attainable* (in the sense defined in [64]) secrecy exponents. In [53], a lower bound to the achievable secrecy exponent, when such wiretap codes are used in conjunction with privacy amplification, is derived. This lower bound is smaller than the lower bound of [52] on the achievable secrecy exponent using i.i.d. random codes. Our analysis shows that the exact secrecy exponent for the wiretap channel codes constructed from constant-composition random codes is larger than the lower bound derived in [53] and there are examples where this dominance is strict.

More importantly, these examples show that, in general, there is no ordering between the secrecy exponents of the ensembles of i.i.d. and constant-composition codes. In other words, for some channels the i.i.d. random-coding ensemble yields a better secrecy exponent, whereas for the other channels, the constant-composition ensemble prevails (see § 6.3.2).

Theorem 5.4 shows that if we know by using an ensemble of random codes we can achieve a resolvability exponent, then the same exponent is a lower bound to the achievable secrecy exponent by using that ensemble. The first step in our analysis is to complement Theorem 5.4 by showing that the *exact* resolvability exponent for an ensemble *equals* the *exact* secrecy exponent of the ensemble (see Theorem 6.1 in § 6.2); deriving the exact resolvability exponent for an ensemble is easier. We present our main result on the exact secrecy exponents in § 6.3 (see Theorem 6.3) and compare the exponents with previously known lower bounds; we defer the proof to § 6.4.

In information theory, random coding is a standard method for establishing achievability results. However, in practice, it is desirable to use more structured codes to reduce the complexity of transceivers. It is well known that random *linear* codes — which are much more structured compared to random codes — have the same error-correction capabilities as i.i.d. random codes [42, Chapter 6]. In § 6.5 we establish a similar result for their resolvability and, in view of Theorem 6.1, their secrecy performance. In fact, the achievability part of our proof in § 6.4.2 depends only on the first- and second-order statistics of the code distribution; these statistics are the same for i.i.d. random codes and random linear codes. Hence it follows that, for resolvability, random linear codes must perform as well as i.i.d. random codes. We also show that the exact resolvability exponent for the ensemble of random linear codes is the same as that of the ensemble of i.i.d. random codes.

random coding and *privacy amplification*). It was later shown that privacy amplification is unnecessary and the exponent reported in [52] lower-bounds the exponential decay rate of the expected amount of information leaked to Eve, when the code is randomly constructed (as we described in § 5.3) by sampling codewords from i.i.d. random-coding ensemble (see, for instance [47, Theorem 3.1] and [54, Theorem 2]).

Part of the results presented in this chapter are obtained in collaboration with N. Merhav and were published in [15, 16].

6.1 The Method of Types

The analysis method we present in this chapter is based on the method of types [28]. A distribution $P \in \mathcal{P}(\mathcal{X})$ is an n-type if for all $x \in \mathcal{X}$, $nP(x) \in \mathbb{Z}$. We denote the set of n-types on $\mathcal{X}$ as $\mathcal{P}_n(\mathcal{X})$ and use the fact that $|\mathcal{P}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|}$ [30, Lemma 2.2] repeatedly.

We denote the *type* of a sequence $x^n \in \mathcal{X}^n$ as $\hat{\mathsf{Q}}_{x^n}$,

$$\hat{\mathsf{Q}}_{x^n}(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{x_i = x\}. \tag{6.1}$$

If $P \in \mathcal{P}_n(\mathcal{X})$, we denote the set of all sequences of type P as $\mathcal{T}_P^n \subset \mathcal{X}^n$, i.e.,

$$\mathcal{T}_P^n := \{x^n \in \mathcal{X}^n \colon \hat{\mathsf{Q}}_{x^n} = P\}. \tag{6.2}$$

The size of the type class $\mathcal{T}_P^n$ (for $P \in \mathcal{P}_n(\mathcal{X})$) is bounded as (cf. [30, Lemma 2.3])

$$(n+1)^{-|\mathcal{X}|} \exp\{nH(P)\} \leq \left|\mathcal{T}_P^n\right| \leq \exp\{nH(P)\}. \tag{6.3}$$

We write $a(n) \dot{\leq} b(n)$ if there exists a function $p(n)$ such that

$$\limsup_{n\to\infty} \frac{\log[p(n)]}{n} \leq 0 \qquad \text{and} \qquad \forall n \in \mathbb{N}\colon a(n) \leq p(n)b(n). \tag{6.4}$$

As noted in [28, p. 2507], when $a(n)$ and $b(n)$ depend on variables other than n, it is understood that $p(n)$ can depend only on the *fixed parameters* of the problem such as channel transition probabilities, the cardinality of its input and output alphabet, and its input distribution — not the other parameters $a(n)$ and $b(n)$ might depend on. We write $a(n) \doteq b(n)$ if $a(n) \dot{\leq} b(n)$ and $b(n) \dot{\leq} a(n)$.

Remark. The reason for restricting the dependence of $p(n)$ on the fixed parameters of the problem is as follows: We often encounter expressions such as

$$a_\theta(n) \dot{\leq} b_\theta(n), \tag{6.5}$$

for some parameter θ that takes values in a parameter space Θ, from which we want to conclude

$$\sup_{\theta\in\Theta} a_\theta(n) \dot{\leq} \sup_{\theta\in\Theta} b_\theta(n). \tag{6.6}$$

It is obvious that (6.4) enables us to immediately draw such a conclusion because

$$\sup_{\theta\in\Theta} a_\theta(n) \leq p(n) \sup_{\theta\in\Theta} b_\theta(n). \tag{6.7}$$

Had we let p depend on θ as well, such a conclusion would not have always been true. For example, assume $\Theta = \mathbb{R}$ and let

$$b_\theta(n) \coloneqq 1 \qquad \text{and} \qquad a_\theta(n) \coloneqq \exp(n\mathbb{1}\{n \le \theta\}). \tag{6.8}$$

Then with $p_\theta(n) \coloneqq a_\theta(n)$, we have

$$a_\theta(n) \le p_\theta(n) b_\theta(n) \tag{6.9}$$

and for $\forall\theta$,

$$\limsup_{n\to\infty} \frac{\log[p_\theta(n)]}{n} = \limsup_{n\to\infty} \mathbb{1}\{n \le \theta\} = 0, \tag{6.10}$$

which would have shown $a_\theta(n) \mathrel{\dot{\le}} b_\theta(n)$. However,

$$\sup_{\theta\in\Theta} a_\theta(n) = \exp(n) \mathrel{\dot{\not\le}} \sup_{\theta\in\Theta} b_\theta(n) = 1. \tag{6.11}$$

In other words, with our definition $a_\theta(n) \mathrel{\dot{\le}} b_\theta(n)$ implies

$$\limsup_{n\to\infty} \Big\{ \sup_{\theta\in\Theta} \frac{1}{n} \log\Big[\frac{a_\theta(n)}{b_\theta(n)}\Big]\Big\} \le 0 \tag{6.12}$$

which is stronger than

$$\sup_{\theta\in\Theta} \Big\{ \limsup_{n\to\infty} \frac{1}{n} \log\Big[\frac{a_\theta(n)}{b_\theta(n)}\Big]\Big\} \le 0. \tag{6.13}$$

6.2 Exact Secrecy Exponent versus Exact Resolvability Exponent

We have seen in § 5.3 that channel resolvability is a convenient and powerful tool for establishing secrecy. Specifically, the exponential decay of the information leaked to the eavesdropper in the wiretap channel model is implied because the divergence between the distribution of the output of the n-fold use of a channel, when its input is a uniformly chosen codeword from a (randomly constructed) codebook, and the reference measure decays exponentially fast in n (see Theorem 5.4). We now strengthen Theorem 5.4 by proving that the ensemble-optimal resolvability exponent equals the ensemble-optimal secrecy exponent. Let us first formally define the notion of the *exact* resolvability exponent for an ensemble:

Definition 6.1. The *exact* resolvability exponent of the ensemble of random codes of rate (at most) R defined via the sequence of distributions $\Pi = \{P_{U^n} \in \mathcal{P}(\mathcal{U}^n)\}_{n\in\mathbb{N}}$ (see Definition 5.6), over the channel $P_{V|U}\colon \mathcal{U} \to \mathcal{V}$, is defined as

$$E_{\mathrm{s}}(\Pi, P_{V|U}, R) \coloneqq \lim_{n\to\infty} -\frac{1}{n} \log \mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n} \| P_{V^n})], \tag{6.14}$$

where $P_{V^n} = P_{U^n} \circ P^n_{V|U}$ (cf. (5.50)) and $P_{\mathscr{C}_n}$ is the output distribution of $P^n_{V|U}$ when its input is a uniformly chosen codeword from $\mathscr{C}_n$ (cf. (5.32)), *provided that the limit exists.*

Remark. The *exact* error exponents for an ensemble of random codes is, in the same way, defined as the exact exponential decay rate of the ensemble-average of the probability of MAP decoding error at the receiver when a random code from the ensemble is used for communication. For both ensembles of i.i.d. and constant-composition random codes, the exact error exponents are well known [30, 42, 43]. (The exactness of the random-coding exponent of [30, Theorem 10.2] follows from the exponential tightness of the truncated union bound [103, Appendix A].)

Theorem 6.1. *Let $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$ be a wiretap channel with $W_{\mathrm{M}} : \mathcal{X} \to \mathcal{Y}$ and $W_{\mathrm{E}} : \mathcal{X} \to \mathcal{Z}$ being its corresponding legitimate receiver's and wiretapper's stationary memoryless marginals, respectively (see Figure 5.1). Fix a sequence of codeword-sampling distributions $\Pi = (P_{X^n} \in \mathcal{P}(\mathcal{X}^n), n \in \mathbb{N})$ that define an ensemble of random codes as in Definition 5.6. Let $E_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R)$ be the* exact *resolvability exponent of the ensemble Π over the channel W_{E} at rate R (see Definition 6.1).*

Consider a sequence of randomly constructed codes $(\mathscr{C}_n,\ n \in \mathbb{N})$ of secret message rate R_{s} used for secure communications over the wiretap channel $W : \mathcal{X} \to \mathcal{Y} \times \mathcal{Z}$, by using a random-binning encoder that associates each secret message with a random code of rate (at most) R and block-length n in the ensemble (as we discussed in the proof of Theorem 5.1). Then, for any rate pair (R_{s}, R) such that $E_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R + R_{\mathrm{s}}) > E_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R)$, when the secret message S is uniformly distributed,

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{E}_{\mathscr{C}_n}[I(S; Z^n)] = E_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R). \tag{6.15}$$

In other words, E_{s} (evaluated at the random-binning rate R) is also the exact secrecy exponent *for the ensemble Π.*

Proof. For every $s \in \{1, 2, \ldots, M_{\mathrm{s}}\}$, let $\mathscr{C}_n^s$ denote the sub-code (bin) associated with the secret message s and $\mathscr{C}_n = (\mathscr{C}_n^1, \mathscr{C}_n^2, \ldots, \mathscr{C}_n^{M_{\mathrm{s}}})$ be the entire collection of codes used by the random-binning encoder. Since, to communicate a particular message $s \in \{1, 2, \ldots, M_{\mathrm{s}}\}$, the encoder transmits a codeword from $\mathscr{C}_n^s$, conditioned on $S = s$ the output of W_{E}^n has distribution $P_{\mathscr{C}_n^s}$ and, since S is uniformly distributed, the *unconditional* output distribution of W_{E}^n will be $P_{\mathscr{C}_n}$ (cf. (5.32)). Therefore, taking $Q_{Z^n} = P_{Z^n} := P_{X^n} \circ W_{\mathrm{E}}^n$ in the identity (5.40) yields

$$\mathbb{E}[I(S; Z^n)] = \mathbb{E}[D(P_{\mathscr{C}_n^S} \| P_{Z^n} | P_S)] - \mathbb{E}[D(P_{\mathscr{C}_n} \| P_{Z^n})]. \tag{6.16}$$

Using the linearity of expectation and the fact that the sub-codes $\mathscr{C}_n^s$ are identically distributed we get

$$\begin{aligned}\mathbb{E}[D(P_{\mathscr{C}_n^S} \| P_{Z^n} | P_S)] &= \sum_{s=1}^{M_{\mathrm{s}}} P_S(s)\, \mathbb{E}[D(P_{\mathscr{C}_n^s} \| P_{Z^n})] \\ &= \mathbb{E}[D(P_{\mathscr{C}_n^1} \| P_{Z^n})]. \end{aligned} \tag{6.17}$$

Thus, by (6.14), we have

$$\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{E}[D(P_{\mathscr{C}_n^s}\|P_{Z^n}|P_S)] = E_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R), \tag{6.18}$$

$$\begin{aligned}\lim_{n\to\infty} -\frac{1}{n}\log \mathbb{E}[D(P_{\mathscr{C}_n}\|P_{Z^n})] &= E_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R+R_{\mathrm{s}}) \\ &> E_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R).\end{aligned} \tag{6.19}$$

where the last inequality follows from the assumption that $E_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R+R_{\mathrm{s}}) > E_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R)$. Using (6.18) and (6.19) in (6.16) concludes the proof. □

Remark. The assumption on the uniform prior of the secret messages is essential for establishing Theorem 6.1. Without such an assumption $I(S; Z^n) = 0$, i.e., the secrecy exponent is infinity if P_S is positive only for a single secret message. However, as we have discussed in Chapter 5, in a practical setting, it is incorrect to assume that the user draws the secret messages from a particular distribution. It is obvious that the *exact* resolvability exponent is the best achievable exponent for the ensemble (according to Definition 5.4), and that Theorem 5.4 guarantees the existence of a sequence of coding schemes with error and secrecy exponents $\underline{E}_{\mathrm{r}}(\Pi, W_{\mathrm{M}}, R+R_{\mathrm{s}})$ and $E_{\mathrm{s}}(\Pi, W_{\mathrm{E}}, R)$, respectively, for *any* secret message distribution P_S.

Theorem 6.1 reduces the problem of deriving the exact secrecy exponent of the ensemble to that of deriving the exact resolvability exponent of the ensemble, which is easier; the former involves the divergence between two random distributions $P_{\mathscr{C}_n^s}$ and $P_{\mathscr{C}_n}$, whereas the latter depends only on $P_{\mathscr{C}_n^s}$.

6.3 Exact Resolvability Exponents

In light of Theorem 6.1, we focus on deriving the exact resolvability exponents for the ensembles of i.i.d. and constant-composition random codes.[2]

6.3.1 Main Result

Theorem 6.2. *Let $\mathscr{C}_n$ be a random code of block-length n and rate R constructed by sampling $M = \lfloor \exp(nR) \rfloor$ codewords independently from the distribution $P_{U^n} \in \mathcal{P}(\mathcal{U}^n)$ (see* (5.46)*). Let $P_{V|U} : \mathcal{U} \to \mathcal{V}$ be a discrete memoryless channel and $P_{\mathscr{C}_n}$ be the (random) output distribution of $P_{V|U}^n$ when a uniformly chosen codeword from $\mathscr{C}_n$ is transmitted via n independent uses of $P_{V|U}$ (see* (5.32)*). Then, the following hold:*

[2] Accordingly, $\mathscr{C}_n$ will denote the random resolvability code of block-length n, and not the entire wiretap channel code.

(i) If $P_{U^n} = P_U^n$ for some $P_U \in \mathcal{P}(\mathcal{U})$,

$$\mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n}\|P_{V^n})] \doteq \begin{cases} \exp\big\{-nE_n^{\text{i.i.d.}}(P_U, P_{V|U}, R)\big\} & \textit{if } I(U;V) > 0, \\ 0 & \textit{if } I(U;V) = 0, \end{cases} \tag{6.20}$$

where

$$E_n^{\text{i.i.d.}}(P_U, P_{V|U}, R) = \min_{Q_{UV}\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \big\{D(Q_{UV}\|P_{UV}) + [R - f(Q_{UV}\|P_{UV})]^+\big\}, \tag{6.21a}$$

with $P_{UV} = P_U \times P_{V|U}$ and

$$f(Q\|P) \coloneqq \sum_{(u,v)\in\mathcal{U}\times\mathcal{V}} Q(u,v)\log\Big[\frac{P(u,v)}{P_U(u)P_V(v)}\Big], \tag{6.21b}$$

for any pair of joint distributions $P, Q \in \mathcal{P}(\mathcal{U}\times\mathcal{V})$.

(ii) If P_{U^n} is the uniform distribution over the type-class $\mathcal{T}^n_{P_U^{(n)}}$ for a sequence of n-types ($P_U^{(n)} \in \mathcal{P}_n(\mathcal{U})$, $n \in \mathbb{N}$) that converge to P_U, i.e., $\lim_{n\to\infty}|P_U^{(n)} - P_U| = 0$,

$$\mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n}\|P_{V^n})] \doteq \begin{cases} \exp\big\{-nE_n^{\text{c.c.}}(P_U^{(n)}, P_{V|U}, R)\big\} & \textit{if } I(U;V) > 0, \\ 0 & \textit{if } I(U;V) = 0, \end{cases} \tag{6.22}$$

where

$$E_n^{\text{c.c.}}(P_U^{(n)}, P_{V|U}, R) = \min_{\substack{Q_{UV}\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V}):\\ Q_U = P_U^{(n)}}} \big\{D\big(Q_{UV}\|P_{UV}^{(n)}\big) + \big[R - g_n\big(Q_{UV}\|P_{UV}^{(n)}\big)\big]^+\big\}, \tag{6.23a}$$

with $P_{UV}^{(n)} \coloneqq P_U^{(n)} \times P_{V|U}$ and

$$g_n(Q\|P) \coloneqq \sum_{(u,v)\in\mathcal{U}\times\mathcal{V}} Q(u,v)\log P_{V|U}(v|u) + H(Q_V) + \min_{\substack{Q'\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V}):\\ Q'_U = Q_U, Q'_V = Q_V}} D(Q'\|P). \tag{6.23b}$$

for any pair of distributions $P, Q \in \mathcal{P}(\mathcal{U}\times\mathcal{V})$.

Recall that in the above $P_{V^n} = P_{U^n} \circ P_{V|U}^n$ (see (5.50)) and $I(U;V)$ is evaluated with $P_{UV} = P_U \times P_{V|U}$.

Theorem 6.2 gives exponentially tight bounds on the expected divergence between the output distribution of $P_{V|U}^n$ — when its input is a uniformly chosen codeword from a randomly chosen code — and the distribution induced by the codeword-sampling distribution at any finite (but possibly large) block-length n. As a consequence, the exact exponential decay rate of the aforementioned divergence, namely the exact resolvability exponent for the ensembles of interest, is the limit of the exponents of (6.20) and (6.22) as n goes to infinity. The exact resolvability exponents have the same forms as (6.21) and (6.23), except that the search space of the minimizations will change from the grid of empirical distributions to the set of all distributions.

Theorem 6.3.

(i) For the sequence of i.i.d. random codes of rate R, i.e., those defined via the sequence of sampling distributions $(P_{U^n} = P_U^n, n \in \mathbb{N})$ for some $P_U \in \mathcal{P}(\mathcal{U})$,

$$\lim_{n\to\infty} -\frac{1}{n}\log\bigl(\mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n}\|P_{V^n})]\bigr) = \begin{cases} E_{\mathrm{s}}^{\text{i.i.d.}}(P_U, P_{V|U}, R) & \textit{if } I(U;V) > 0,\\ +\infty & \textit{if } I(U;V) = 0,\end{cases} \tag{6.24}$$

where

$$E_{\mathrm{s}}^{\text{i.i.d.}}(P_U, P_{V|U}, R) = \min_{Q_{UV}\in\mathcal{P}(\mathcal{U}\times\mathcal{V})} \bigl\{D(Q_{UV}\|P_{UV}) + [R - f(Q_{UV}\|P_{UV})]^+\bigr\}, \tag{6.25}$$

with $P_{UV} = P_U \times P_{V|U}$ and f as defined in (6.21b)*.*

(ii) For the sequence of constant-composition random codes of rate R, specifically, those defined via the sequence of uniform distributions over the type class $\mathcal{T}^n_{P_U^{(n)}}$, for a sequence of n-types $(P_U^{(n)} \in \mathcal{P}_n(\mathcal{U}), n \in \mathbb{N})$ that converge to P_U, i.e., $\lim_{n\to\infty} |P_U^{(n)} - P_U| = 0$,

$$\lim_{n\to\infty} -\frac{1}{n}\log\bigl(\mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n}\|P_{V^n})]\bigr) = \begin{cases} E_{\mathrm{s}}^{\text{c.c.}}(P_U, P_{V|U}, R) & \textit{if } I(U;V) > 0,\\ +\infty & \textit{if } I(U;V) = 0,\end{cases} \tag{6.26}$$

where

$$E_{\mathrm{s}}^{\text{c.c.}}(P_U, P_{V|U}, R) = \min_{\substack{Q_{UV}\in\mathcal{P}(\mathcal{U}\times\mathcal{V}):\\ Q_U = P_U}} \bigl\{D(Q_{UV}\|P_{UV}) + [R - g(Q_{UV}\|P_{UV})]^+\bigr\}, \tag{6.27a}$$

with $P_{UV} = P_U \times P_{V|U}$ and

$$g(Q\|P) := \sum_{(u,v)\in\mathcal{U}\times\mathcal{V}} Q(u,v)\log P_{V|U}(v|u) + H(Q_V) + \min_{\substack{Q'\in\mathcal{P}(\mathcal{U}\times\mathcal{V}):\\ Q'_U = Q_U, Q'_V = Q_V}} D(Q'\|P), \tag{6.27b}$$

for any pair of distributions $P, Q \in \mathcal{P}(\mathcal{U}\times\mathcal{V})$.

Both exponents $E_{\mathrm{s}}^{\text{i.i.d.}}$ and $E_{\mathrm{s}}^{\text{c.c.}}$ are positive and strictly increasing in R for $R > I(U;V)$. Moreover, the value of $E_{\mathrm{s}}^{\text{i.i.d.}}$ can be computed through

$$E_{\mathrm{s}}^{\text{i.i.d.}}(P_U, P_{V|U}, R) = \max_{0\leq\lambda\leq 1}\{\lambda R - F_0(P_{UV},\lambda)\} \tag{6.28a}$$

with

$$F_0(P_{UV},\lambda) := \log\Bigg[\sum_{(u,v)\in\mathcal{U}\times\mathcal{V}} \frac{P_{UV}(u,v)^{1+\lambda}}{P_U(u)^{\lambda}P_V(v)^{\lambda}}\Bigg] \tag{6.28b}$$

Deducing Theorem 6.3 from Theorem 6.2 involves many technical details. The proof is hence relegated to Appendix 6.A.

Corollary 6.4. *The exponents $E_{\mathrm{s}}^{\text{i.i.d.}}(P_X, W_{\mathrm{E}}, R)$ and $E_{\mathrm{s}}^{\text{c.c.}}(P_X, W_{\mathrm{E}}, R)$ of* (6.25) *and* (6.27) *are the exact secrecy exponents for the ensembles of random wiretap channel codes of random-binning rate R and secret message rate R_{s} constructed from the ensembles of random i.i.d. and constant-composition codes, respectively, provided that $R_{\mathrm{s}} > 0$ and $R > I(X;Z)$.*

6.3.2 Comparison of Exponents

The exponent $E_{\mathrm{s}}^{\text{i.i.d.}}$ has already been shown to lower-bound the exponential decay rate of the information leaked to the eavesdropper when the code is sampled from the i.i.d. random-coding ensemble in [47,54] (and in conjugation with privacy amplification in [52]). Corollary 6.4 states that this exponent is indeed the exact secrecy exponent for the ensemble of i.i.d. random codes. (The exponent is expressed in the form of (6.28) in [47, 52, 54].) In contrast, it can be shown that $E_{\mathrm{s}}^{\text{c.c.}}$, the exact secrecy exponent for the ensemble of constant-composition random codes, is larger than the previously derived lower bound in [53]:

$$\underline{E}_{\mathrm{s}}(P_X, W_{\mathrm{E}}, R) = \max_{0\leq\lambda\leq 1}\{\lambda R - E_0(P_X, W_{\mathrm{E}},\lambda)\}, \tag{6.29a}$$

with

$$E_0(P_X, W_{\mathrm{E}},\lambda) := \log\Bigg\{\sum_{z\in\mathcal{Z}}\Bigg[\sum_{x\in\mathcal{X}} P_X(x)W_{\mathrm{E}}(z|x)^{\frac{1}{1-\lambda}}\Bigg]^{1-\lambda}\Bigg\}. \tag{6.29b}$$

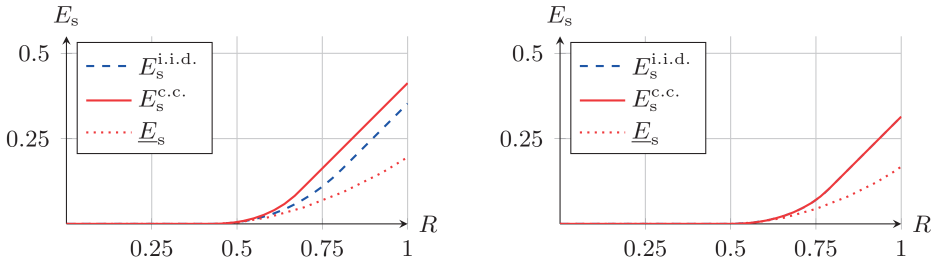
(Note that the function E_0 in (6.29b) is essentially Gallager's E_0 [42] up to a minus sign.)

Lemma 6.5. *For every discrete memoryless channel $P_{V|U} : \mathcal{U} \to \mathcal{V}$,*

$$E_{\mathrm{s}}^{\text{c.c.}}(P_U, P_{V|U}, R) \geq \underline{E}_{\mathrm{s}}(P_U, P_{V|U}, R). \tag{6.30}$$

Lemma 6.5 follows from the fact that for any pair of joint distributions P and Q with the same u-marginal, $g(Q\|P) \leq I(Q)$, by following similar steps as in [30, Problem 10.24] to derive Gallager-style expressions of random-coding error exponents. (See Appendix 6.B for a complete proof.)

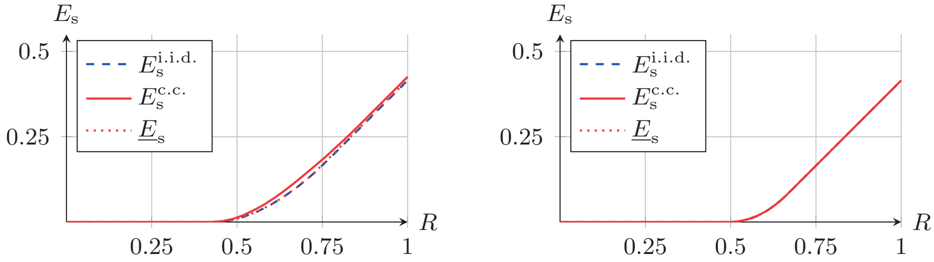
More importantly, as for the comparison of the secrecy exponents $E_{\mathrm{s}}^{\mathrm{i.i.d.}}$ and $E_{\mathrm{s}}^{\mathrm{c.c.}}$, numerical examples show that, in general, there is no ordering between them. In particular, as shown in Figures 6.1 and 6.2, for the binary symmetric channel and the binary erasure channel, the ensemble of constant-composition random codes leads to a larger exponent than the ensemble of i.i.d. random codes. The two exponents are equal when the input distribution is uniform. On the other side, in Figures 6.3 and 6.4, we see that for asymmetric channels (the Z-channel and the binary asymmetric channel) the ensemble of constant-composition random codes has a smaller secrecy exponent compared to the ensemble of i.i.d. random codes.



(a) $P_X(0) = 0.3, P_X(1) = 0.7$

(b) $P_X(0) = P_X(1) = 0.5$

Figure 6.1: Comparison of Secrecy Exponents when W_{E} is a Binary Symmetric Channel with Crossover Probability 0.11



(a) $P_X(0) = 0.28, P_X(1) = 0.72$

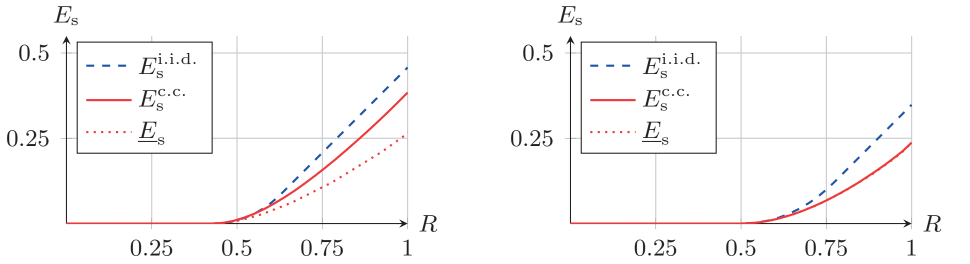
(b) $P_X(0) = P_X(1) = 0.5$

Figure 6.2: Comparison of Secrecy Exponents when W_{E} is a Binary Erasure Channel with Erasure Probability 0.5

The reader can find details on how the exponents are computed in Appendix 6.C.

6.4 Proof of Theorem 6.2

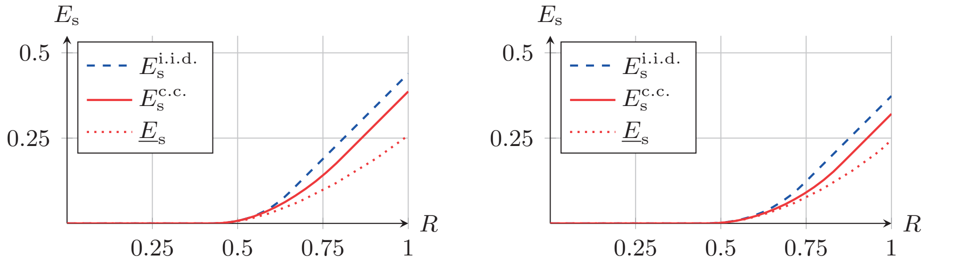
In this section, we fix P_U, accordingly, $P_{UV}(u, v) = P_U(u)P_{V|U}(v|u)$ where $P_{V|U}$ is the channel transition probability. Moreover, without essential loss of



(a) $P_X(0) = 0.36, P_X(1) = 0.64$

(b) $P_X(0) = 0.58, P_X(1) = 0.42$ (capacity-achieving)

Figure 6.3: Comparison of Secrecy Exponents when W_{E} is a Z-channel with $W_{\mathrm{E}}(0|1) = 0.303$



(a) $P_X(0) = 0.42, P_X(1) = 0.58$

(b) $P_X(0) = 0.57, P_X(1) = 0.43$ (capacity-achieving)

Figure 6.4: Comparison of Secrecy Exponents when W_{E} is a Binary Asymmetric Channel with $W_{\mathrm{E}}(1|0) = 0.01$, $W_{\mathrm{E}}(0|1) = 0.303$

generality, we assume that (i) $\mathrm{supp}(P_U) = \mathcal{U}$ (and for the constant-composition codes, $\forall n$, $\mathrm{supp}(P_U^{(n)}) = \mathcal{U}$), and (ii) for every $v \in \mathcal{V}$, there exists at least one $u \in \mathcal{U}$ such that $P_{V|U}(v|u) > 0$ (if this is not the case, we can shrink $\mathcal{V}$).

Recall that the setting we consider is as follows: A random code $\mathscr{C}_n = (U_{(1)}^n, U_{(2)}^n, \ldots, U_{(M)}^n)$ of block-length n and size $M = \lfloor \exp(nR) \rfloor$ is constructed by sampling each codeword independently from distribution P_{U^n}. A uniformly chosen codeword from this code is transmitted through the product channel $P_{V|U}^n$ and the (random) distribution of its output sequence is

$$P_{\mathscr{C}_n}(v^n) = \frac{1}{M} \sum_{j=1}^{M} P_{V|U}^n\big(v^n | U_{(j)}^n\big). \tag{6.31}$$

Trivial Case (zero-capacity channel) If P_U is such that $I(U;V) = 0$, then $\forall u \in \mathcal{U}$ and $\forall v \in \mathcal{V}$, $P_{V|U}(v|u) = P_V(v)$. This implies that *for any code of block-length n, $\mathscr{C}_n$, $P_{\mathscr{C}_n} = P_V^n$.* Moreover,

$$P_{V^n} = P_{U^n} \circ P_{V|U}^n = P_V^n \tag{6.32}$$

as well. Thus, $D(P_{\mathscr{C}_n} \| P_{V^n}) = 0$ (with probability 1 for a random code), which in turn implies $\mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n} \| P_{V^n})] = 0$.

Here, we begin the non-trivial part of the proof, specifically when the channel output sequence V^n is correlated with its input. For any fixed $v^n \in \mathcal{V}^n$, $P_{\mathscr{C}_n}(v^n)$ (as defined in (6.31)) is an average of M i.i.d. random variables $P^n_{V|U}(v^n|U^n_{(j)})$, $j = 1, 2, \dots, M$. Hence, it is expected to concentrate around its mean that is exactly $P_{V^n}(v^n)$. However, as the distribution of each of the summands depends on n, a plain application of the law of large numbers is not possible in this setting. Let

$$L(v^n) := \begin{cases} \frac{P_{\mathscr{C}_n}(v^n)}{P_{V^n}(v^n)} & \text{if } P_{V^n}(v^n) > 0, \\ 1 & \text{otherwise,} \end{cases} \tag{6.33}$$

denote the (random) likelihood ratio of each sequence $v^n \in \mathcal{V}^n$. By construction,

$$\mathbb{E}_{\mathscr{C}_n}[L(v^n)] = 1, \qquad \forall v^n \in \mathcal{V}^n. \tag{6.34}$$

Moreover, it follows that $P_{\mathscr{C}_n} \ll P_{V^n}$ with probability 1 (see Lemma 6.6 below). Hence, the linearity of expectation yields

$$\begin{aligned} \mathbb{E}_{\mathscr{C}_n}[D(P_{\mathscr{C}_n} \| P_{V^n})] &= \mathbb{E}_{\mathscr{C}_n}\left[\sum_{v^n \in \mathcal{V}^n} P_{\mathscr{C}_n}(v^n) \log\left(\frac{P_{\mathscr{C}_n}(v^n)}{P_{V^n}(v^n)}\right)\right] && (6.35) \\ &= \sum_{v^n \in \mathcal{V}^n} \mathbb{E}_{\mathscr{C}_n}\left[P_{\mathscr{C}_n}(v^n) \log\left(\frac{P_{\mathscr{C}_n}(v^n)}{P_{V^n}(v^n)}\right)\right] && (6.36) \\ &= \sum_{v^n \in \mathcal{V}^n} P_{V^n}(v^n)\, \mathbb{E}_{\mathscr{C}_n}[L(v^n) \log L(v^n)] && (6.37) \end{aligned}$$

To prove Theorem 6.2, we derive exponentially tight bounds on the value of $\mathbb{E}_{\mathscr{C}_n}[L(v^n) \log L(v^n)]$ (for each individual $v^n \in \mathcal{V}^n$). And, to derive the exponents of Theorem 6.2, we combine these bounds in (6.37).

6.4.1 Basics

Lemma 6.6. *Let P_{V^n} be as defined in* (5.50). *Then:*

(i) $P_{\mathscr{C}_n} \ll P_{V^n}$ with probability 1.

(ii) For any codeword-sampling distribution $P_{U^n} \in \mathcal{P}(\mathcal{U}^n)$ that depends on u^n only through its type, $P_{V^n}(v^n)$ will depend on v^n only through its type.

(iii) For both choices of P_{U^n} in Theorem 6.2 (i.e., i.i.d. and constant-composition codes),

$$\forall v^n \in \operatorname{supp}(P_{V^n}), \quad P_{V^n}(v^n) \geq \frac{1}{\alpha^n} \tag{6.38}$$

where

$$\alpha := \begin{cases} \frac{1}{P_U^{\min} P_{V|U}^{\min}} & \textit{if } P_{U^n} = P_U^n, \\ \frac{|\mathcal{U}|}{P_{V|U}^{\min}} & \textit{if } P_{U^n} \textit{ is the uniform distribution over } \mathcal{T}^n_{P_U^{(n)}}, \end{cases} \tag{6.39}$$

with

$$P_U^{\min} := \min_{u \in \mathcal{U}} P_U(u) \quad \textit{and} \quad P_{V|U}^{\min} := \min_{\substack{(u,v)\in\mathcal{U}\times\mathcal{V}: \\ P_{V|U}(v|u)>0}} P_{V|U}(v|u). \tag{6.40}$$

Remark. For the i.i.d. random-coding ensemble, i.e., when $P_{U^n} = P_U^n$, the reference measure P_{V^n} equals the product measure P_V^n hence $\operatorname{supp}(P_{V^n}) = \mathcal{V}^n$ (because we assumed $\operatorname{supp}(P_U) = \mathcal{U}$ and for every $v \in \mathcal{V}$ there exists at least one $u \in \mathcal{U}$ such that $P_{V|U}(v|u) > 0$). In contrast, when P_{U^n} is the uniform distribution over the sequences of type $P_U^{(n)}$, (i.e., for the constant-composition random-coding ensemble) the support of P_{V^n} need not necessarily be $\mathcal{V}^n$. For instance, consider a binary erasure channel and let $P_U^{(n)}$ be the uniform distribution on $\{0,1\}$, for even n, i.e. P_{U^n} is the uniform distribution on the set of $\binom{n}{n/2}$ sequences with equal zeros and ones. Then P_{V^n} puts mass neither on the all-zero output sequence and, by symmetry, nor on the all-one sequence.

Lemma 6.7. *Let A be an arbitrary non-negative random variable. Then, for any $\theta > 0$,*

$$c(\theta)\Big[\frac{\operatorname{var}(A)}{\mathbb{E}[A]} - \tau_\theta(A)\Big] \le \mathbb{E}\Big[A \ln\Big(\frac{A}{\mathbb{E}[A]}\Big)\Big] \le \frac{\operatorname{var}(A)}{\mathbb{E}[A]} \tag{6.41}$$

where

$$\tau_\theta(A) := \mathbb{E}[A]\left[\theta^2 \Pr\{A > (\theta+1)\,\mathbb{E}[A]\} + 2\int_\theta^{+\infty} t \Pr\{A > (t+1)\,\mathbb{E}[A]\}\mathrm{d}t\right], \tag{6.42}$$

and

$$c(\theta) := \frac{(1+\theta)\ln(1+\theta) - \theta}{\theta^2}. \tag{6.43}$$

Lemmas 6.6 and 6.7 are proven in Appendices 6.D and 6.E, respectively.

Remark. Jensen's inequality yields $\mathbb{E}[A\ln(A/\mathbb{E}[A])]) \ge 0$. Lemma 6.7 improves this lower bound for random variables with sufficiently small tails.

Unfortunately, $L(v^n)$ has heavy tails and a direct application of Lemma 6.7 to $L(v^n)$ will not result in exponentially tight bounds on $\mathbb{E}\big[L(v^n)\log L(v^n)\big]$.[3] However, it turns out that $L(v^n)$ can be split into light- and heavy-tail components. As we will see shortly, the heavy-tail component contributes to

[3] For simplicity, we drop the subscript $\mathscr{C}_n$ after the expectation operator. It is clear from the context that the expectations are taken over the choice of the code.

$\mathbb{E}\big[L(v^n)\log L(v^n)\big]$ only via its mean, and Lemma 6.7 can be applied to the light-tail component to obtain exponentially tight bounds on $\mathbb{E}\big[L(v^n)\log L(v^n)\big]$.

Since $P_{V^n}(v^n)$ depends on v^n only through its type, we can use type enumeration method [78, 79] and write

$$L(v^n) = \frac{1}{M}\sum_{i=1}^{M}\frac{P^n_{V|U}\big(v^n|U^n_{(i)}\big)}{P_{V^n}(v^n)} \tag{6.44}$$

$$= \frac{1}{M}\sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} N_Q(v^n)\ell(Q) \tag{6.45}$$

where

$$\ell(Q) \coloneqq \frac{P^n_{V|U}(v^n|u^n)}{P_{V^n}(v^n)} \tag{6.46}$$

is the common value of the right-hand side for $\forall(u^n, v^n) \in \mathcal{T}^n_Q$, and

$$N_Q(v^n) \coloneqq \big|\{u^n \in \mathscr{C}_n : (u^n, v^n) \in \mathcal{T}^n_Q\}\big| \tag{6.47}$$

is the number of codewords in $\mathscr{C}_n$ that have joint type Q with v^n. Therefore, $\{N_Q(v^n)\colon Q \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})\}$ is a multinomial collection with cluster size M and success probabilities

$$p_Q(v^n) = \frac{|\mathcal{T}^n_Q|}{|\mathcal{T}^n_{Q_V}||\mathcal{T}^n_{Q_U}|}P_{U^n}(\mathcal{T}^n_{Q_U})\mathbb{1}\{v^n \in \mathcal{T}^n_{Q_V}\}, \tag{6.48}$$

for any codeword-sampling distribution $P_{U^n}(u^n)$ that depends on u^n through its type — including our cases of interest. (The above equality is proven in Appendix 6.F.)

Partition the set of n-types on $\mathcal{U}\times\mathcal{V}$, $\mathcal{P}_n(\mathcal{U}\times\mathcal{V}) = \mathcal{Q}'_n \cup \mathcal{Q}''_n$ as

$$\mathcal{Q}'_n \coloneqq \{Q \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V}) : \ell(Q) \le e^2M\}, \tag{6.49}$$

$$\mathcal{Q}''_n \coloneqq \{Q \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V}) : \ell(Q) > e^2M\}, \tag{6.50}$$

and, accordingly, split $L(v^n) = L_1(v^n) + L_2(v^n)$ as

$$L_1(v^n) \coloneqq \frac{1}{M}\sum_{Q\in\mathcal{Q}'_n} N_Q(v^n)\ell(Q), \tag{6.51}$$

$$L_2(v^n) \coloneqq \frac{1}{M}\sum_{Q\in\mathcal{Q}''_n} N_Q(v^n)\ell(Q). \tag{6.52}$$

Indeed, L_1 turns out to be the light-tail component of L and L_2 its heavy-tail part. Let also,

$$\nu(v^n) \coloneqq \mathrm{var}\big(L_1(v^n)\big) + \frac{1}{M}\,\mathbb{E}[L_1(v^n)]^2, \qquad \text{and} \tag{6.53}$$

$$\mu(v^n) \coloneqq \mathbb{E}[L_2(v^n)]. \tag{6.54}$$

Using the elementary properties of the multinomial distribution, it can be verified[4] that

$$\nu(v^n) = \frac{1}{M} \sum_{Q \in \mathcal{Q}'_n} \ell(Q)^2 p_Q(v^n) \tag{6.55a}$$

$$\mu(v^n) = \sum_{Q \in \mathcal{Q}''_n} \ell(Q) p_Q(v^n). \tag{6.55b}$$

In the following two subsections, we prove that $\forall v^n \in \operatorname{supp}(P_{V^n})$,

$$\mathbb{E}\big[L(v^n) \ln L(v^n)\big] + \frac{1}{M} \doteq \nu(v^n) + \mu(v^n). \tag{6.56}$$

Since v^n is fixed in both sides of (6.56), we drop it in Subsections 6.4.2 and 6.4.3 to avoid cumbersome notation.

6.4.2 Achievability

For non-negative l_1 and l_2, and $l = l_1 + l_2$,

$$\begin{aligned} l \ln(l) &= l_1 \ln(l) + l_2 \ln(l) && (6.57)\\ &= l_1 \ln(l_1) + l_1 \ln(1 + l_2/l_1) + l_2 \ln(l) && (6.58)\\ &\le l_1 \ln(l_1) + l_2(1 + \ln(l)) && (6.59) \end{aligned}$$

(since $\ln(1 + l_2/l_1) \le l_2/l_1$). Thus,

$$\begin{aligned} \mathbb{E}[L \ln L] &\le \mathbb{E}[L_1 \ln L_1] + \mathbb{E}[L_2(1 + \ln L)] && (6.60)\\ &\overset{(*)}{\le} \mathbb{E}[L_1 \ln L_1] + (1 + n \ln \alpha)\, \mathbb{E}[L_2] && (6.61) \end{aligned}$$

where $(*)$ follows from (iii) in Lemma 6.6 (as $L = L(v^n) \le 1/P_{V^n}(v^n)$). The upper bound of (6.41) implies

$$\mathbb{E}[L_1 \ln L_1] \le \mathbb{E}[L_1] \ln\big(\mathbb{E}[L_1]\big) + \frac{\operatorname{var}(L_1)}{\mathbb{E}[L_1]} \overset{(*)}{\le} \frac{\operatorname{var}(L_1)}{\mathbb{E}[L_1]} \tag{6.62}$$

where $(*)$ follows since $\mathbb{E}[L_1] \le \mathbb{E}[L] = 1$. Moreover, using (6.53) and the fact that $\mathbb{E}[L_1] + \mathbb{E}[L_2] = 1$ we have

$$\begin{aligned} \frac{\operatorname{var}(L_1)}{\mathbb{E}[L_1]} &= \frac{\nu}{\mathbb{E}[L_1]} - \frac{\mathbb{E}[L_1]}{M} && (6.63)\\ &= \nu\Big(1 + \frac{\mathbb{E}[L_2]}{\mathbb{E}[L_1]}\Big) - \frac{1 - \mathbb{E}[L_2]}{M} && (6.64)\\ &= \nu + \mathbb{E}[L_2]\Big(\frac{\nu}{\mathbb{E}[L_1]} + \frac{1}{M}\Big) - \frac{1}{M}. && (6.65) \end{aligned}$$

[4] A proof of (6.55) is given in Appendix 6.G for completeness.

Since $\ell(Q) \leq Me^2$ for $Q \in \mathcal{Q}'_n$, using (6.55a) we have

$$\nu \leq \frac{1}{M} \sum_{Q \in \mathcal{Q}'_n} e^2 M \cdot \ell(Q) p_Q = e^2 \, \mathbb{E}[L_1]. \tag{6.66}$$

Using the above in (6.65) and replacing $\mathbb{E}[L_2] = \mu$, we get

$$\frac{\mathrm{var}(L_1)}{\mathbb{E}[L_1]} + \frac{1}{M} \leq \nu + \mathbb{E}[L_2]\Big(e^2 + \frac{1}{M}\Big) \leq \nu + (1+e^2)\mu. \tag{6.67}$$

Finally, using (6.67) in (6.62) yields,

$$\mathbb{E}[L_1 \ln L_1] + \frac{1}{M} \overset{\cdot}{\leq} \nu + \mu. \tag{6.68}$$

Using (6.68) in (6.61) (and noting that $\alpha \geq 1$ only depends on $|\mathcal{U}|$, P_U, and $P_{V|U}$), we conclude that

$$\mathbb{E}[L \ln L] + \frac{1}{M} \overset{\cdot}{\leq} \nu + \mu. \tag{6.69}$$

6.4.3 Ensemble Converse

The choice of $\mathcal{Q}''_n$ implies

$$\Pr\big\{L_2 \in (0, e^2)\big\} = 0. \tag{6.70}$$

This holds since either $\forall Q \in \mathcal{Q}''_n \colon N_Q = 0$ which implies $L_2 = 0$ or $\exists Q_0 \in \mathcal{Q}''_n$ such that $N_{Q_0} \geq 1$, in which case,

$$L_2 \geq \frac{1}{M}\ell(Q_0)N_{Q_0} \geq \frac{1}{M}\ell(Q_0) \geq e^2, \tag{6.71}$$

(because $\forall Q \in \mathcal{Q}''_n$, $\ell(Q) > e^2 M$). Consequently,

$$\begin{aligned} \mathbb{E}[L_2 \ln L_2] &= \sum_{l \geq e^2} l \, \ln(l) \Pr\{L_2 = l\} &(6.72)\\ &\geq \ln(e^2) \sum_{l \geq e^2} l \Pr\{L_2 = l\} = 2\, \mathbb{E}[L_2]. &(6.73) \end{aligned}$$

For positive l_1 and l_2, and $l = l_1 + l_2 \geq \max\{l_1, l_2\}$,

$$\begin{aligned} l \ln(l) &= l_1 \ln(l) + l_2 \ln(l) &(6.74)\\ &\geq l_1 \ln(l_1) + l_2 \ln(l_2). &(6.75) \end{aligned}$$

Therefore,

$$\mathbb{E}[L \ln L] \geq \mathbb{E}[L_1 \ln L_1] + \mathbb{E}[L_2 \ln L_2]. \tag{6.76}$$

Using the lower bound of (6.41) (with $\tau_\theta(L_1)$ and $c(\theta)$ defined as in (6.42) and (6.43), respectively), $\forall\theta > 0$:

$$\begin{aligned}
\mathbb{E}[L_1 \ln L_1] &\geq \mathbb{E}[L_1] \ln(\mathbb{E}[L_1]) + c(\theta)\Big[\frac{\mathrm{var}(L_1)}{\mathbb{E}[L_1]} - \tau_\theta(L_1)\Big] && (6.77)\\
&\overset{\text{(a)}}{=} (1 - \mathbb{E}[L_2]) \ln(1 - \mathbb{E}[L_2]) + c(\theta)\Big[\frac{\mathrm{var}(L_1)}{\mathbb{E}[L_1]} - \tau_\theta(L_1)\Big] && (6.78)\\
&\overset{\text{(b)}}{\geq} -\mathbb{E}[L_2] + c(\theta)\Big[\frac{\mathrm{var}(L_1)}{\mathbb{E}[L_1]} - \tau_\theta(L_1)\Big]. && (6.79)
\end{aligned}$$

In the above, (a) follows since $\mathbb{E}[L_1] = 1 - \mathbb{E}[L_2]$ and (b) follows since $(1 - t)\ln(1-t) \geq -t$. Using (6.73) and (6.79) in (6.75) shows that $\forall\theta > 0$:

$$\mathbb{E}[L \ln L] \geq c(\theta)\Big[\frac{\mathrm{var}(L_1)}{\mathbb{E}[L_1]} - \tau_\theta(L_1)\Big] + \mathbb{E}[L_2]. \qquad (6.80)$$

Let us now upper-bound $\tau_\theta(L_1)$. Starting by bounding the tail of L_1 we have

$$\begin{aligned}
\Pr\{L_1 \geq (t+1)\,\mathbb{E}[L_1]\} &= \Pr\left\{\sum_{Q\in\mathcal{Q}'_n} \ell(Q)(N_Q - Mp_Q) \geq Mt\,\mathbb{E}[L_1]\right\} && (6.81)\\
&\leq \Pr\left\{\bigcup_{Q\in\mathcal{Q}'_n} \left\{\ell(Q)(N_Q - Mp_Q) \geq \frac{Mt\,\mathbb{E}[L_1]}{|\mathcal{Q}'_n|}\right\}\right\} && (6.82)\\
&\overset{\text{(a)}}{\leq} \sum_{Q\in\mathcal{Q}'_n} \Pr\left\{\ell(Q)(N_Q - Mp_Q) \geq \frac{Mt\,\mathbb{E}[L_1]}{|\mathcal{Q}'_n|}\right\} && (6.83)\\
&\overset{\text{(b)}}{\leq} \sum_{Q\in\mathcal{Q}'_n} \frac{\mathbb{E}[\ell(Q)^4(N_Q - Mp_Q)^4]}{(Mt\,\mathbb{E}[L_1]/|\mathcal{Q}'_n|)^4} && (6.84)\\
&= \frac{|\mathcal{Q}'_n|^4}{t^4(\mathbb{E}[L_1])^4}\frac{1}{M^4} \sum_{Q\in\mathcal{Q}'_n} \ell(Q)^4\,\mathbb{E}[(N_Q - Mp_Q)^4], && (6.85)
\end{aligned}$$

where (a) is the union bound and (b) follows by Markov inequality. For $N \sim \mathrm{Binomial}(M, p)$,

$$\begin{aligned}
\mathbb{E}[(N - Mp)^4] &= Mp(1-p)[1 + 3(M-2)p(1-p)] && (6.86)\\
&\leq \mathrm{var}(N) + 3\,\mathrm{var}(N)^2. && (6.87)
\end{aligned}$$

Continuing (6.85) we have

$$\frac{1}{M^4}\sum_{Q\in\mathcal{Q}'_n} \ell(Q)^4\,\mathbb{E}[(N_Q - Mp_Q)^4] \leq \frac{1}{M^4}\sum_{Q\in\mathcal{Q}'_n} \ell(Q)^4\big(\mathrm{var}(N_Q) + 3\,\mathrm{var}(N_Q)^2\big)$$

$$\overset{(a)}{\dot{\le}} \frac{1}{M^2}\sum_{Q\in\mathcal{Q}_n'} \ell(Q)^2 \operatorname{var}(N_Q) + 3\frac{1}{M^4}\sum_{Q\in\mathcal{Q}_n'} \ell(Q)^4 \operatorname{var}(N_Q)^2 \tag{6.88}$$

$$\overset{(b)}{\le} \frac{1}{M^2}\sum_{Q\in\mathcal{Q}_n'} \ell(Q)^2 \operatorname{var}(N_Q) + 3\Big[\frac{1}{M^2}\sum_{Q\in\mathcal{Q}_n'} \ell(Q)^2 \operatorname{var}(N_Q)\Big]^2 \tag{6.89}$$

$$\overset{(c)}{\le} \nu + 3\nu^2 \overset{(d)}{\doteq} \nu, \tag{6.90}$$

where (a) follows since $\ell(Q) \le e^2 M \doteq M$ for $Q \in \mathcal{Q}_n'$, (b) follows since for positive summands, the sum of the squares is less than the square of the sums, (c) holds since $\operatorname{var}(N_Q) \le M p_Q$, and (d) follows since $\nu \le e^2\,\mathbb{E}[L_1] \le e^2$ (see (6.66)). Plugging (6.90) into (6.85) yields

$$\Pr\{L_1 \ge (t+1)\,\mathbb{E}[L_1]\} \,\dot{\le}\, \frac{|\mathcal{Q}_n'|^4\nu}{(\mathbb{E}[L_1])^4}\cdot\frac{1}{t^4}. \tag{6.91}$$

Using the above in (6.42) we get

$$\tau_\theta(L_1) = \mathbb{E}[L_1]\left[\theta^2 \Pr\{L_1 > (\theta+1)\,\mathbb{E}[L_1]\} + 2\int_\theta^{+\infty} t\Pr\{L_1 > (t+1)\,\mathbb{E}[L_1]\}\mathrm{d}t\right]$$

$$\dot{\le}\, \mathbb{E}[L_1]\left[\frac{\theta^2}{\theta^4} + 2\int_\theta^{+\infty}\frac{t}{t^4}\mathrm{d}t\right]\frac{|\mathcal{Q}_n'|^4}{\mathbb{E}[L_1]^4}\nu \tag{6.92}$$

$$\doteq \frac{\nu}{\mathbb{E}[L_1]^3}\cdot\frac{|\mathcal{Q}_n'|^4}{\theta^2}. \tag{6.93}$$

Equation (6.93) implies

$$\tau_\theta(L_1) \le d(n)\frac{|\mathcal{Q}_n'|^4\nu}{\big(\theta^2\,\mathbb{E}[L_1]^3\big)} \tag{6.94}$$

for some sub-exponentially increasing sequence $d(n)$ (which only depends on $|\mathcal{U}|$ and $|\mathcal{V}|$). Therefore, taking

$$\theta_n := 2\sqrt{d(n)}\frac{|\mathcal{Q}_n'|^2}{\mathbb{E}[L_1]}, \tag{6.95}$$

we will have

$$\tau_{\theta_n}(L_1) \le \frac{1}{4}\cdot\frac{\nu}{\mathbb{E}[L_1]}. \tag{6.96}$$

Using (6.53) and (6.96) in (6.80) yields

$$\mathbb{E}[L(v^n)\ln L(v^n)] \ge c(\theta_n)\Big[\frac{\operatorname{var}(L_1)}{\mathbb{E}[L_1]} - \tau_{\theta_n}(L_1)\Big] + \mathbb{E}[L_2] \tag{6.97}$$

$$\ge c(\theta_n)\Big[\frac{\nu}{\mathbb{E}[L_1]} - \frac{1}{M}\,\mathbb{E}[L_1] - \frac{1}{4}\cdot\frac{\nu}{\mathbb{E}[L_1]}\Big] + \mathbb{E}[L_2] \tag{6.98}$$

$$\overset{(*)}{\ge} c(\theta_n)\Big[\frac{3}{4}\cdot\frac{\nu}{\mathbb{E}[L_1]} - \frac{1}{M}\Big] + \mathbb{E}[L_2] \tag{6.99}$$

(where $(*)$ follows because $\mathbb{E}[L_1] \le 1$). Because for $\theta > 0$, $c(\theta) \le c(0) = 1/2 < 1$, we can further lower-bound (6.99) as

$$\mathbb{E}[L \ln L] \ge \frac{3}{4} c(\theta_n) \frac{\nu}{\mathbb{E}[L_1]} + \mathbb{E}[L_2] - \frac{1}{M} \tag{6.100}$$

Moreover,

$$\begin{aligned} c(\theta_n) &= \frac{1}{\theta_n} \cdot \frac{(1+\theta_n)\ln(1+\theta_n) - \theta_n}{\theta_n} && (6.101) \\ &\overset{\text{(a)}}{\ge} \frac{1}{\theta_n} \cdot \frac{(1+\mathbb{E}[L_1]\theta_n)\ln(1+\mathbb{E}[L_1]\theta_n) - \mathbb{E}[L_1]\theta_n}{\mathbb{E}[L_1]\theta_n} && (6.102) \\ &= \mathbb{E}[L_1] \frac{(1+\mathbb{E}[L_1]\theta_n)\ln(1+\mathbb{E}[L_1]\theta_n) - \mathbb{E}[L_1]\theta_n}{(\mathbb{E}[L_1]\theta_n)^2} && (6.103) \\ &\overset{\text{(b)}}{\dot{\ge}} \mathbb{E}[L_1], && (6.104) \end{aligned}$$

where (a) follows since $[(1+\theta)\ln(1+\theta) - \theta]/\theta$ is increasing in θ and $\mathbb{E}[L_1] \le 1$, and (b) holds since $[(1+\theta)\ln(1+\theta) - \theta]/\theta^2$ is decreasing in θ (see Lemma 6.14 in Appendix 6.E) and $\mathbb{E}[L_1]\theta_n = 2\sqrt{d(n)}|\mathcal{Q}_n'|^2 \le 2\sqrt{d(n)}(n+1)^{2|\mathcal{U}||\mathcal{V}|}$. Using this lower bound in (6.100), we get

$$\mathbb{E}[L \ln L] + \frac{1}{M} \dot{\ge} \nu + \mu. \tag{6.105}$$

6.4.4 Derivation of Exponents for Each Ensemble

Equations (6.69) and (6.105) prove (6.56). Plugging in the values of $\nu(v^n)$ and $\mu(v^n)$ from (6.55a) and (6.55b) and continuing (6.56), we get

$$\begin{aligned} \mathbb{E}\big[L(v^n) \ln L(v^n)\big] + \frac{1}{M} &\doteq \nu(v^n) + \mu(v^n) && (6.106) \\ &= \sum_{Q \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \ell(Q) p_Q(v^n) \kappa\big(\ell(Q)/M\big) && (6.107) \end{aligned}$$

where

$$\kappa(\lambda) = \begin{cases} 1 & \lambda > e^2, \\ \lambda & \lambda \le e^2. \end{cases} \tag{6.108}$$

It is easy to check that

$$\min\{1, \lambda\} \le \kappa(\lambda) \le e^2 \min\{1, \lambda\} \tag{6.109}$$

Therefore, (6.107) can be simplified as

$$\mathbb{E}\big[L(v^n) \ln L(v^n)\big] + \frac{1}{M} \doteq \sum_{Q \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \ell(Q) p_Q(v^n) \min\Big\{1, \frac{\ell(Q)}{M}\Big\}. \tag{6.110}$$

Using the above in (6.37) we get

$$\mathbb{E}[D(P_{\mathscr{C}_n}\|P_{V^n})] + \frac{\log(e)}{M} \doteq \sum_{v^n\in\mathcal{V}^n} P_{V^n}(v^n) \sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \ell(Q)p_Q(v^n)\min\Big\{1, \frac{\ell(Q)}{M}\Big\} \tag{6.111}$$

$$= \sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \ell(Q)\min\Big\{1, \frac{\ell(Q)}{M}\Big\} \sum_{v^n\in\mathcal{V}^n} p_Q(v^n)P_{V^n}(v^n). \tag{6.112}$$

Plugging in the value of $p_Q(v^n)$ from (6.48), we get

$$\sum_{v^n\in\mathcal{V}^n} p_Q(v^n)P_{V^n}(v^n) = \frac{|\mathcal{T}_Q^n|}{|\mathcal{T}_{Q_U}^n||\mathcal{T}_{Q_V}^n|} P_{U^n}\big(\mathcal{T}_{Q_U}^n\big)P_{V^n}\big(\mathcal{T}_{Q_V}^n\big). \tag{6.113}$$

Moreover, defining

$$\omega(Q) := \sum_{(u,v)\in\mathcal{U}\times\mathcal{V}} Q(u,v)\log P_{V|U}(v|u), \tag{6.114}$$

and recalling that $P_{V^n}(v^n)$ depends on v^n only through its type, we deduce that

$$\ell(Q) = \frac{\exp\big(n\omega(Q)\big)|\mathcal{T}_{Q_V}^n|}{P_{V^n}\big(\mathcal{T}_{Q_V}^n\big)}. \tag{6.115}$$

Combining (6.113) and (6.115) yields

$$\ell(Q)\sum_{v^n} p_Q(v^n)P_{V^n}(v^n) = \exp\big\{n\omega(Q)\big\}|\mathcal{T}_Q^n|\frac{P_{U^n}\big(\mathcal{T}_{Q_U}^n\big)}{\big|\mathcal{T}_{Q_U}^n\big|} \tag{6.116}$$

$$\doteq \exp\big\{-nD(Q\|Q_U\times P_{V|U})\big\}P_{U^n}\big(\mathcal{T}_{Q_U}^n\big), \tag{6.117}$$

where the last equality follows from (6.3). Thus, we have

$$\mathbb{E}[D(P_{\mathscr{C}_n}\|P_{V^n})] + \frac{\log(e)}{M} \doteq \sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \exp\big\{-nD(Q\|Q_U\times P_{V|U})\big\}P_{U^n}\big(\mathcal{T}_{Q_U}^n\big)\min\Big\{1, \frac{\ell(Q)}{M}\Big\}. \tag{6.118}$$

Since

$$\ell(P_{UV}) \geq \exp\{n\omega(P_{UV})\}\big|\mathcal{T}_{P_V}^n\big| \mathrel{\dot{\geq}} \exp\{nI(U;V)\}, \tag{6.119}$$

taking $Q = P_{UV}$ shows that the right-hand side of (6.118) decays at most as fast as $\exp\{-n[R-I(U;V)]^+\}$, which is strictly slower than $1/M = \exp(-nR)$

as $I(U;V) > 0$. Consequently, we can eliminate the term $\log(e)/M$ on the left-hand side of (6.118) and conclude[5] that

$$\mathbb{E}[D(P_{\mathscr{C}_n}\|P_{V^n})] \doteq \sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \exp\bigl\{-nD(Q\|Q_U\times P_{V|U})\bigr\}\cdot P_{U^n}\bigl(\mathcal{T}^n_{Q_U}\bigr)\min\Bigl\{1,\frac{\ell(Q)}{M}\Bigr\}. \quad (6.120)$$

Ensemble of i.i.d. Random Codes

When $P_{U^n} = P_U^n$,

$$P_{U^n}(\mathcal{T}^n_{Q_U}) \doteq \exp\{-nD(Q_U\|P_U)\} \quad (6.121)$$

Moreover, $P_{V^n} = P_V^n$. Therefore,

$$P_{V^n}(v^n) = \exp\Bigl\{n\sum_v Q_V(v)\log P_V(v)\Bigr\} \quad \text{if } v^n\in\mathcal{T}^n_{Q_V}. \quad (6.122)$$

Hence,

$$\ell(Q) = \frac{\exp\{n\omega(Q)\}}{P_V^n(v^n)} = \exp\Bigl\{n\sum_{u,v} Q(u,v)\log\frac{P_{V|U}(v|u)}{P_V(v)}\Bigr\}$$
$$= \exp\bigl\{nf(Q\|P_{UV})\bigr\}. \quad (6.123)$$

where f is defined in (6.21b). As a consequence,

$$\min\{1,\ell(Q)/M\} \doteq \exp\bigl\{-n[R - f(Q\|P_{UV})]^+\bigr\}. \quad (6.124)$$

Using (6.121) and (6.124) in (6.120) (together with the fact that $|\mathcal{P}_n(\mathcal{U}\times\mathcal{V})| \le (n+1)^{|\mathcal{U}||\mathcal{V}|}$), we conclude that

$$\mathbb{E}[D(P_{\mathscr{C}_n}\|P_{V^n})] \doteq \exp\Bigl\{-n\min_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})}\bigl\{D(Q\|Q_U\times P_{V|U}) + D(Q_U\|P_U) + [R - f(Q\|P_{UV})]^+\bigr\}\Bigr\}. \quad (6.125)$$

Simplifying the above exponent yields (6.21).

Ensemble of Constant-Composition Random Codes

When the codeword-sampling distribution, P_{U^n}, is the uniform distribution over the sequences of type $P_U^{(n)}$, $P_{U^n}\bigl(\mathcal{T}^n_{Q_U}\bigr) = 0$ unless $Q_U = P_U^{(n)}$. Therefore (6.120) reduces to

$$\mathbb{E}[D(P_{\mathscr{C}_n}\|P_{V^n})] \doteq \sum_{\substack{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V}):\\ Q_U=P_U^{(n)}}} \exp\bigl\{-nD(Q\|P_{UV}^{(n)})\bigr\}\min\{1,\ell(Q)/M\}. \quad (6.126)$$

[5] The careful reader could argue that P_{UV} might not be an n-type for all n hence find our reasoning inaccurate, in the passage from (6.118) to (6.120). Although this concern is valid, the claim is true regardless; because we can always find a sequence of n-types that converge to P_{UV}. We give a rigorous and more detailed proof of (6.120) in Appendix 6.H.

It remains to evaluate

$$\ell(Q) = \frac{P^n_{V|U}(v^n|u^n)}{P_{V^n}(v^n)}, \tag{6.127}$$

for $(u^n, v^n) \in \mathcal{T}^n_Q$ (when $Q_U = P^{(n)}_U$). The numerator of the above equals

$$P^n_{V|U}(v^n|u^n) = \exp\{n\omega(Q)\}, \tag{6.128}$$

where ω is defined in (6.114), for any $(u^n, v^n) \in \mathcal{T}^n_Q$. To compute the denominator, note that, since $Q_U = P^{(n)}_U$,

$$\begin{aligned}
P_{V^n}(v^n) &= \frac{1}{|\mathcal{T}^n_{Q_U}|} \sum_{\tilde{u}^n \in \mathcal{T}^n_{Q_U}} P^n_{V|U}(v^n|\tilde{u}^n) && (6.129)\\
&= \frac{1}{|\mathcal{T}^n_{Q_U}|} \sum_{\tilde{u}^n \in \mathcal{T}^n_{Q_U}} P^n_{V|U}(v^n|\tilde{u}^n) \sum_{Q' \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \mathbb{1}\{(\tilde{u}^n, v^n) \in \mathcal{T}^n_{Q'}\} && (6.130)\\
&= \frac{1}{|\mathcal{T}^n_{Q_U}|} \sum_{\tilde{u}^n \in \mathcal{T}^n_{Q_U}} \sum_{Q' \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \mathbb{1}\{(\tilde{u}^n, v^n) \in \mathcal{T}^n_{Q'}\} P^n_{V|U}(v^n|\tilde{u}^n) && (6.131)\\
&= \frac{1}{|\mathcal{T}^n_{Q_U}|} \sum_{\tilde{u}^n \in \mathcal{T}^n_{Q_U}} \sum_{Q' \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \mathbb{1}\{(\tilde{u}^n, v^n) \in \mathcal{T}^n_{Q'}\} \exp\{n\omega(Q')\} && (6.132)\\
&= \sum_{Q' \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \left[\frac{1}{|\mathcal{T}^n_{Q_U}|} \sum_{\tilde{u}^n \in \mathcal{T}^n_{Q_U}} \mathbb{1}\{(\tilde{u}^n, v^n) \in \mathcal{T}^n_{Q'}\}\right] \exp\{n\omega(Q')\} && (6.133)
\end{aligned}$$

As we have already shown in the proof of (6.48) (cf. Appendix 6.F),

$$\begin{aligned}
\sum_{\tilde{u}^n \in \mathcal{T}^n_{Q_U}} \mathbb{1}\{(\tilde{u}^n, v^n) \in \mathcal{T}^n_{Q'}\} &= \mathbb{1}\{Q'_U = Q_U\} \sum_{\tilde{u}^n \in \mathcal{U}^n} \mathbb{1}\{(\tilde{u}^n, v^n) \in \mathcal{T}^n_{Q'}\} && (6.134)\\
&= \mathbb{1}\{Q'_U = Q_U\} \mathbb{1}\big\{v^n \in \mathcal{T}^n_{Q'_V}\big\} \frac{|\mathcal{T}^n_{Q'}|}{|\mathcal{T}^n_{Q'_V}|}. && (6.135)
\end{aligned}$$

Consequently, for $v^n \in \mathcal{T}^n_{Q_V}$,

$$\begin{aligned}
\frac{1}{|\mathcal{T}^n_{Q_U}|} \sum_{\tilde{u}^n \in \mathcal{T}^n_{Q_U}} \mathbb{1}\{(\tilde{u}^n, v^n) \in \mathcal{T}^n_{Q'}\} \; &= \mathbb{1}\{Q'_U = Q_U\}\mathbb{1}\{Q'_V = Q_V\} \frac{|\mathcal{T}^n_{Q'}|}{|\mathcal{T}^n_{Q'_U}||\mathcal{T}^n_{Q'_V}|}\\
&\overset{(*)}{\doteq} \mathbb{1}\{Q'_U = Q_U\}\mathbb{1}\{Q'_V = Q_V\} \exp\{-nI(Q')\} && (6.136)
\end{aligned}$$

where $(*)$ follows from (6.3). Combining the exponents, we see that each of the summands in (6.133) equals

$$\begin{aligned}
&\left[\frac{1}{|\mathcal{T}^n_{Q_U}|} \sum_{\tilde{u}^n \in \mathcal{T}^n_{Q_U}} \mathbb{1}\{(\tilde{u}^n, v^n) \in \mathcal{T}^n_{Q'}\}\right] \exp\{n\omega(Q')\}\\
&\qquad \doteq \exp\big\{-n[D(Q'\|P^{(n)}_{UV}) - D(Q_U\|P^{(n)}_U) + H(Q_V)]\big\} && (6.137)
\end{aligned}$$

when $Q'_U = Q_U$ and $Q'_V = Q_V$ (and is zero otherwise). Recall that in the above $P^{(n)}_{UV} = P^{(n)}_U P_{V|U}$. Moreover, since $Q_U = P^{(n)}_U$, using (6.137) in (6.133) yields

$$P_{V^n}(v^n) \doteq \exp\biggl\{-n\biggl[H(Q_V) + \min_{\substack{Q' \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V}):\\ Q'_U = Q_U, Q'_V = Q_V}} D(Q' \| P^{(n)}_{UV})\biggr]\biggr\} \tag{6.138}$$

which, in turn, shows

$$\ell(Q) \doteq \mathbb{1}\{Q_U = P^{(n)}_U\} \exp\bigl[-ng_n\bigl(Q \| P^{(n)}_{UV}\bigr)\bigr] \tag{6.139}$$

with g_n defined as in (6.23b). Therefore,

$$\min\{1, \ell(Q)/M\} \doteq \exp\bigl\{-n\bigl[R - g_n\bigl(Q \| P^{(n)}_{UV}\bigr)\bigr]^+\bigr\}. \tag{6.140}$$

when Q is such that $Q_U = P^{(n)}_U$. Using (6.140) in (6.126) proves (6.23). □

6.5 Random Linear Codes

We derived the exact random-coding resolvability exponents for the ensembles of i.i.d. and of constant-composition random codes. In practice, a randomly sampled code from either of the aforementioned ensembles is too complex to use — to begin with, the encoder needs to store exponentially many codewords. Consequently, it is desirable to use more structured codes. For the error-correction problem, which is a counterpart of the resolvability problem, we know that on average *random linear codes* perform as well as i.i.d. random codes (for uniform input distribution) [42, Section 6.2]. In terms of their ensemble-average behavior, random linear codes are the same as i.i.d. random codes, simply because the proof of channel coding theorem [42, Chapter 5] depends only on the pairwise independence of codewords. This pairwise independence also holds for the codewords of a random linear (or more precisely, affine) code. How well do random linear codes perform as a resolvability code?

Inspecting the achievability part of our proof (§ 6.4.2), we see that our exponentially decaying upper bound on $\mathbb{E}[L(v^n)\ln L(v^n)]$ depends only on the first- and second-order statistics of the type enumerators $N_Q(v^n)$. The fact that they form a multinomial collection is of no importance in the proof. As long as the codewords are pairwise independent, (6.55) holds. Thus, as random linear codes have the same first- and second-order statistics as i.i.d. random codes [42], it straightforwardly follows that the exponent $E_{\mathrm{s}}^{\text{i.i.d.}}$ is also achievable by random linear codes (for uniform input distribution).

In this section, we will show that $E_{\mathrm{s}}^{\text{i.i.d.}}$ is indeed ensemble-optimal for the ensemble of random affine codes as well. Note that here we assume the channel input $\mathcal{U}$ equals $\mathbb{F}_q$ for some q, and that the input distribution P_U is the uniform distribution on $\mathbb{F}_q$. The goal is hence to approximate P^n_V, where

$$P_V(v) = \frac{1}{q} \sum_{u \in \mathbb{F}_q} P_{V|U}(v|u), \tag{6.141}$$

at the output of the channel by transmitting a uniformly chosen codeword from a random affine code.

Definition 6.2 (Random Affine Code [42])**.** An (n, k) random *affine* code is defined as the (random) collection of codewords

$$\mathscr{C}_n := \left(U^n_{w^k} = Gw^k + D^n \colon w^k \in \mathbb{F}_q^k\right) \tag{6.142}$$

where $G \in \mathbb{F}_q^{n\times k}$ is a random $n \times k$ generator matrix whose elements $G_{i,j}$ are drawn independently and uniformly from $\mathbb{F}_q$ and $D^n \in \mathbb{F}_q^n$ is a random dither vector independent of G uniformly distributed on $\mathbb{F}_q^n$. The codewords are indexed by q-ary vectors of length k, $w^k \in \mathbb{F}_q^k$.

As shown in [42, p. 207], the ensemble of random affine codes has the same first- and second-order statistics as the ensemble of i.i.d. random codes, namely, $\forall w^k \in \mathbb{F}_q^k$,

$$\Pr\left\{U^n_{w^k} = u^n\right\} = q^{-n}, \qquad \forall u^n \in \mathbb{F}_q^n, \tag{6.143}$$

and for all unequal pairs w^k and $\tilde{w}^k$ in $\mathbb{F}_q^k$,

$$\Pr\left\{U^n_{w^k} = u^n, U^n_{\tilde{w}^k} = \tilde{u}^n\right\} = q^{-2n}, \qquad \forall u^n, \tilde{u}^n \in \mathbb{F}_q^n. \tag{6.144}$$

Here we prove the following result.

Theorem 6.8. *Let $P_{V|U} \colon \mathcal{U} \to \mathcal{V}$ be a discrete memoryless channel with input $\mathcal{U} = \mathbb{F}_q$ and consider the sequence $(\mathscr{C}_n,\ n \in \mathbb{N})$ of (n, k_n) random affine codes of rate (at most) R, i.e., those satisfying*

$$\limsup_{n\to\infty} \frac{k_n \log(q)}{n} \le R. \tag{6.145}$$

(See Definition 6.2.) Let $P_{\mathscr{C}_n}$ be the (random) output distribution of $P^n_{V|U}$ when a uniformly chosen codeword from $\mathscr{C}_n$ is transmitted via n independent uses of $P_{V|U}$ (see (5.32)*). Then,*

$$\mathbb{E}[D(P_{\mathscr{C}_n} \| P_{V^n})] \doteq \begin{cases} \exp\left\{-nE_n^{\text{i.i.d.}}(P_U, P_{V|U}, R)\right\} & \text{if } I(U;V) > 0, \\ 0 & \text{if } I(U;V) = 0, \end{cases} \tag{6.146}$$

where P_U is the uniform distribution on $\mathbb{F}_q$ (and hence, $I(U;V)$ is evaluated with $P_{UV}(u,v) = \frac{1}{q}P_{V|U}(v|u)$), P_V is defined in (6.141)*, and $E_n^{\text{i.i.d.}}$ is defined in* (6.21) *(see Theorem 6.2).*

Having established Theorem 6.8, it follows that $E_{\text{s}}^{\text{i.i.d.}}$ (as defined in (6.25)) is the exact random-coding resolvability exponent (hence, the exact random-coding secrecy exponent) for the ensemble of random affine codes.

Proof of Theorem 6.8. We show that our argument in § 6.4 is valid if the code $\mathscr{C}_n$ is sampled from the ensemble of random affine codes, instead of the i.i.d. random-coding ensemble when P_U is the uniform distribution on $\mathcal{U} = \mathbb{F}_q$, (obviously, we only focus on the case $I(U;V) > 0$).

First, note that even though $P_{\mathscr{C}_n}(v^n)$ (see (6.31)) is only the average of pairwise independent random variables, by the virtue of (6.143), the distribution of each summand is the same as that when the code was sampled from i.i.d. random-coding ensemble. Thus, (6.34) in particular holds. By the same token, Lemma 6.6 is also valid for a random affine code (with $\alpha = 1/(P_U^{\min} P_{V|U}^{\min}) = q/P_{V|U}^{\min}$).[6] We can still rewrite the summation of $M = q^k$ (pairwise independent) random variables in (6.44) into the summation of at most $(n+1)^{|\mathcal{U}||\mathcal{V}|}$ weighted type enumerators as in (6.45) with $N_Q(v^n)$ exactly as defined in (6.47).

Although the collection of type-enumerators $\big(N_Q(v^n), Q \in \mathcal{Q}_n(\mathcal{U} \times \mathcal{V})\big)$ is *not* a multinomial collection anymore, $p_Q(v^n)$ as defined in (6.48) (with $P_{U^n}(u^n) = (1/q)^n$ for all $u^n \in \mathcal{U}^n$) is still the probability of having a codeword u^n in the code whose joint type with v^n is Q. Furthermore, in view of (6.143) and (6.144), type-enumerators for random affine codes have exactly the same first- and second-order statistics (i.e., expectation vector and covariance matrix) as they had for i.i.d. random codes. Hence (6.55a) and (6.55b) hold.

6.5.1 Achievability

The achievability part of our proof for random codes (in § 6.4.2) involves only elementary mathematical manipulations and uses only the first- and second-order statistics of the type-enumerators. Therefore, it follows that

$$\mathbb{E}[D(P_{\mathscr{C}_n} \| P_{V^n})] \dot{\leq} \exp\big\{-nE_n^{\text{i.i.d.}}(P_U, P_{V|U}, R)\big\} \tag{6.147}$$

(when $\mathscr{C}_n$ is a randomly chosen affine code of rate at most R).

6.5.2 Ensemble Converse

For the ensemble converse of the proof in § 6.4.3, to upper-bound the tails of $L_1(v^n)$, we used the fourth central moments of the type-enumerators, that depend on the third- and fourth-order statistics of the code. In a random affine code, the codewords are not necessarily triple-wise and quadruple-wise independent. For example, consider a binary affine code and take $w_{(1)}^k$, $w_{(2)}^k$, $w_{(3)}^k$, and $w_{(4)}^k$ such that

$$w_{(1)}^k \oplus w_{(2)}^k = w_{(3)}^k \oplus w_{(4)}^k$$

Then, it is obvious that

$$U^n_{w^k_{(1)}} \oplus U^n_{w^k_{(2)}} = U^n_{w^k_{(3)}} \oplus U^n_{w^k_{(4)}},$$

[6] Also note that since in this case $P_{V^n} = P_V^n$, with P_V as defined in (6.141), the support of P_{V^n} is $\mathcal{V}^n$.

with probability 1. Hence, every four codewords indexed by linearly dependent w^ks are also linearly dependent. (We will see shortly that for binary alphabet, the codewords are, in fact, triple-wise independent.)

It might be possible to derive other bounds on the tails of $L_1(v^n)$ and show that it still concentrates around its mean sufficiently fast. We, instead, show that even though the codewords are not triple- and quadruple-wise independent in general, a random linear code does not contain too many such dependent codewords. In particular, the dependencies are weak enough to enable us to prove that (6.90) is still an upper bound (up to sub-exponential factors) on the summation on the right-hand side of (6.85).

Lemma 6.9. *Consider a q-ary random affine code as in Definition 6.2. Fix a collection of $m \le k$ vectors in $\mathbb{F}_q^k$, $w_{(1)}^k, w_{(2)}^k, \ldots, w_{(m)}^k$, and $u_{(1)}^n, u_{(2)}^n, \ldots, u_{(m)}^n \in \mathbb{F}_q^n$. Let*

$$\Omega\big(w_{(1)}^k, w_{(2)}^k, \ldots, w_{(m)}^k\big) \coloneqq \begin{bmatrix} w_{(1)}^k & w_{(2)}^k & \cdots & w_{(m)}^k \\ 1 & 1 & \cdots & 1 \end{bmatrix}. \tag{6.148}$$

and

$$\Xi\big(u_{(1)}^n, u_{(2)}^n, \ldots, u_{(m)}^n\big) \coloneqq \begin{bmatrix} u_{(1)}^n & u_{(2)}^n & \cdots & u_{(m)}^n \end{bmatrix}. \tag{6.149}$$

Then

$$\Pr\left(\bigcap_{j=1}^{m} \{U_{w_{(j)}^k}^n = u_{(j)}^n\}\right) = q^{-n\,\mathrm{rank}(\Omega)} \mathbb{1}\{\mathrm{kern}(\Omega) \subseteq \mathrm{kern}(\Xi)\}, \tag{6.150}$$

(in the above $\mathrm{kern}(A) = \{y : Ay = 0\}$ *and all matrix operations are assumed to be done in $\mathbb{F}_q$).*

Let us defer the proof of Lemma 6.9 to Appendix 6.I and prove the ensemble converse here.

Recall that we need to show that

$$\frac{1}{M^4} \sum_{Q \in \mathcal{Q}_n'} \ell(Q)^4\, \mathbb{E}[(N_Q(v^n) - Mp_Q(v^n))^4] \mathrel{\dot{\le}} \nu(v^n) = \frac{1}{M} \sum_{Q \in \mathcal{Q}_n'} \ell(Q)^2 p_Q(v^n). \tag{6.151}$$

We also recall that

$$\sum_{Q \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \ell(Q) p_Q(v^n) = \mathbb{E}[L(v^n)] = 1, \tag{6.152}$$

thus, in particular, $\ell(Q)p_Q(v^n) \le 1$. As before, we shall drop v^n from the argument of the variables as it is fixed on both sides of (6.151).

Expand N_Q as

$$N_Q = \sum_{w^k \in \mathbb{F}_q^k} \underbrace{\mathbb{1}\{(U_{w^k}^n, v^n) \in \mathcal{T}_Q^n\}}_{=:\chi(w^k)} \tag{6.153}$$

(where $U^n_{w^k} = Gw^k + D^n$ as described in (6.142)) and recall that $\mathbb{E}[\chi(w^k)] = p_Q$. Let $\psi(w^k) := \chi(w^k) - p_Q$ for the sake of brevity. Therefore,

$$\mathbb{E}\big[(N_Q - Mp_Q)^4\big] = \mathbb{E}\Bigg[\bigg(\sum_{w^k \in \mathbb{F}_q^k} (\chi(w^k) - p_Q)\bigg)^4\Bigg] \tag{6.154}$$

$$= \sum_{w^k_{(1)},\dots,w^k_{(4)} \in \mathbb{F}_q^k} \mathbb{E}\big[\psi(w^k_{(1)})\psi(w^k_{(2)})\psi(w^k_{(3)})\psi(w^k_{(4)})\big] \tag{6.155}$$

$$= \sum_{w^k \in \mathbb{F}_q^k} \mathbb{E}[\psi(w^k)^4] \tag{6.156}$$

$$+ \sum_{w^k_{(1)} \neq w^k_{(2)}} \Big\{ 2\,\mathbb{E}[\psi(w^k_{(1)})\psi(w^k_{(2)})^3] + 3\,\mathbb{E}[\psi(w^k_{(1)})^2\psi(w^k_{(2)})^2] + 2\,\mathbb{E}[\psi(w^k_{(1)})^3\psi(w^k_{(2)})]\Big\} \tag{6.157}$$

$$+ 4 \sum_{\substack{\text{distinct}\\ w^k_{(1)},w^k_{(2)},w^k_{(3)}}} \Big\{ \mathbb{E}[\psi(w^k_{(1)})^2\psi(w^k_{(2)})\psi(w^k_{(3)})] + \mathbb{E}[\psi(w^k_{(1)})\psi(w^k_{(2)})^2\psi(w^k_{(3)})] + \mathbb{E}[\psi(w^k_{(1)})\psi(w^k_{(2)})\psi(w^k_{(3)})^2]\Big\} \tag{6.158}$$

$$+ \sum_{\substack{\text{distinct}\\ w^k_{(1)},\dots,w^k_{(4)}}} \mathbb{E}[\psi(w^k_{(1)})\psi(w^k_{(2)})\psi(w^k_{(3)})\psi(w^k_{(4)})]. \tag{6.159}$$

Had the codewords been triple- and quadruple-wise independent, the summations in (6.158) and (6.159) would have evaluated to 0. Hence, because the codewords are still pairwise independent, as shown in (6.86), the sum of (6.156) and (6.157) is upper-bounded as

$$\sum_{w^k \in \mathbb{F}_q^k} \mathbb{E}[\psi(w^k)^4] + \sum_{w^k_{(1)} \neq w^k_{(2)}} \Big\{ 2\,\mathbb{E}[\psi(w^k_{(1)})\psi(w^k_{(2)})^3] + 3\,\mathbb{E}[\psi(w^k_{(1)})^2\psi(w^k_{(2)})^2] + 2\,\mathbb{E}[\psi(w^k_{(1)})^3\psi(w^k_{(2)})]\Big\} \leq \operatorname{var}(N_Q) + 3\operatorname{var}(N_Q)^2. \tag{6.160}$$

Furthermore, in (6.90) it has already been shown that

$$\frac{1}{M^4} \sum_{Q \in \mathcal{Q}'_n} \ell(Q)^4 \big(\operatorname{var}(N_Q) + 3\operatorname{var}(N_Q)^2\big) \mathrel{\dot{\leq}} \nu. \tag{6.161}$$

Therefore, to establish (6.151) it remains to prove

$$\frac{1}{M^4} \sum_Q \ell(Q)^4 [\Sigma_3(Q) + \Sigma_4(Q)] \mathrel{\dot{\leq}} \frac{1}{M} \sum_Q \ell(Q)^2 p_Q, \tag{6.162}$$

where

$$\Sigma_3(Q) \coloneqq 4 \sum_{\substack{\text{distinct}\\ w^k_{(1)},\dots,w^k_{(3)}}} \Big\{\mathbb{E}[\psi(w^k_{(1)})\psi(w^k_{(2)})\psi(w^k_{(3)})^2] + \mathbb{E}[\psi(w^k_{(1)})\psi(w^k_{(2)})\psi(w^k_{(3)})^2] + \mathbb{E}[\psi(w^k_{(1)})\psi(w^k_{(2)})\psi(w^k_{(3)})^2]\Big\} \tag{6.163}$$

$$\Sigma_4(Q) \coloneqq \sum_{\substack{\text{distinct}\\ w^k_{(1)},\dots,w^k_{(4)}}} \mathbb{E}[\psi(w^k_{(1)})\psi(w^k_{(2)})\psi(w^k_{(3)})\psi(w^k_{(4)})]. \tag{6.164}$$

Bound on Σ_3 Let $w^k_{(1)}$, $w^k_{(2)}$, and $w^k_{(3)}$ be three distinct vectors in $\mathbb{F}_q^k$ and $\Omega(w^k_{(1)}, w^k_{(2)}, w^k_{(3)})$ as defined in (6.148). We observe that Ω is either full-rank or has rank 2. It cannot have rank 1 since, in that case, we must have

$$\begin{bmatrix} w^k_{(2)} \\ 1 \end{bmatrix} = a \begin{bmatrix} w^k_{(1)} \\ 1 \end{bmatrix} \tag{6.165}$$

for some $a \in \mathbb{F}_q$ which implies $a = 1$ hence $w^k_{(1)} = w^k_{(2)}$.

If $\text{rank}(\Omega) = 3$ then, (6.150) implies $\psi(w^k_{(1)})$, $\psi(w^k_{(2)})$ and $\psi(w^k_{(3)})$ are independent thus all the expectations in (6.163) will be zero. Hence, we need to consider only the case where $\text{rank}(\Omega) = 2$; that is, for distinct $w^k_{(1)}$ and $w^k_{(2)}$,

$$\begin{bmatrix} w^k_{(3)} \\ 1 \end{bmatrix} = a \begin{bmatrix} w^k_{(1)} \\ 1 \end{bmatrix} + b \begin{bmatrix} w^k_{(2)} \\ 1 \end{bmatrix} \tag{6.166}$$

for some $(a, b) \in \mathbb{F}_q^2$ such that $(a, b) \neq (0, 1)$ and $(a, b) \neq (1, 0)$. It can be verified that there are $q - 2$ such pairs (b is determined as a function of a as $a + b = 1$ and a cannot be 0 or 1).[7] Therefore, among $M(M-1)(M-2)$ summands in (6.163) only, $(q-2)M(M-1)$ are non-zero.

Moreover,

$$\psi(w^k_{(1)})\psi(w^k_{(2)})\psi(w^k_{(3)})^2 = [\chi(w^k_{(1)}) - p_Q][\chi(w^k_{(2)}) - p_Q]\psi(w^k_{(3)})^2 \tag{6.167}$$

$$= \big[\chi(w^k_{(1)})\chi(w^k_{(2)}) - p_Q[\chi(w^k_{(1)}) + \chi(w^k_{(2)})] + p_Q^2\big]\psi(w^k_{(3)})^2 \tag{6.168}$$

$$\le \big[\chi(w^k_{(1)})\chi(w^k_{(2)}) + p_Q^2\big]\psi(w^k_{(3)})^2 \overset{(*)}{\le} \chi(w^k_{(1)})\chi(w^k_{(2)}) + p_Q^2, \tag{6.169}$$

where $(*)$ follows by verifying that $|\psi(w^k_{(3)})^2| \le 1$. Consequently, for distinct $w^k_{(1)}$, $w^k_{(2)}$, and $w^k_{(3)}$,

$$\mathbb{E}\big[\psi(w^k_{(1)})\psi(w^k_{(2)})\psi(w^k_{(3)})^2\big] \overset{\text{(a)}}{\le} \mathbb{E}\big[\chi(w^k_{(1)})\chi(w^k_{(2)})\big] + p_Q^2 \tag{6.170}$$

$$\overset{\text{(b)}}{=} 2p_Q^2. \tag{6.171}$$

[7] Consequently, for binary alphabet this case never occurs and $\text{rank}(\Omega) = 3$ always. Hence the codewords of a binary affine code are triple-wise independent as well.

where (a) follows by (6.169) and (b) holds because of the pairwise independence of codewords. Therefore, each non-zero summand in (6.163) is upper-bounded by

$$\mathbb{E}[\psi(w^k_{(1)})\psi(w^k_{(2)})\psi(w^k_{(3)})^2] + \mathbb{E}[\psi(w^k_{(1)})\psi(w^k_{(2)})^2\psi(w^k_{(3)})] \\ + \mathbb{E}[\psi(w^k_{(1)})^2\psi(w^k_{(2)})\psi(w^k_{(3)})] \leq 6p_Q^2 \quad (6.172)$$

Using (6.172), together with the fact that there are only $(q-2)M(M-2)$ non-zero terms in the sum of (6.163), we can conclude that

$$\Sigma_3(Q) \leq 24(q-2)M(M-1)p_Q^2. \quad (6.173)$$

The above in turn implies

$$\begin{aligned} \frac{1}{M^4}\ell(Q)^4\Sigma_3(Q) &\dot{\leq} \frac{1}{M^2}\ell(Q)^4 p_Q^2 && (6.174)\\ &\overset{(a)}{\dot{\leq}} \frac{1}{M}p_Q^2\ell(Q)^3 && (6.175)\\ &\overset{(b)}{\leq} \frac{1}{M}p_Q\ell(Q)^2, && (6.176)\end{aligned}$$

where (a) follows since $\ell(Q) \dot{\leq} M$ for $Q \in \mathcal{Q}'_n$ and (b) follows since $\ell(Q)p_Q \leq 1$.

Bound on Σ_4 Following the same lines, for distinct $w^k_{(1)}$, $w^k_{(2)}$, $w^k_{(3)}$, and $w^k_{(4)}$, the matrix $\Omega(w^k_{(1)}, w^k_{(2)}, w^k_{(3)}, w^k_{(4)})$ as defined in (6.148) can have rank 2, 3, or 4. If it is full-rank, (6.150) implies that $\psi(w^k_{(1)})$, $\psi(w^k_{(2)})$, …, $\psi(w^k_{(4)})$ are independent, thus

$$\mathbb{E}\left[\prod_{j=1}^{4}\psi(w^k_{(j)})\right] = \prod_{j=1}^{4}\mathbb{E}\left[\psi(w^k_{(j)})\right] = 0 \quad (6.177)$$

Now, suppose rank$(\Omega) = 2$. This means,

$$\begin{bmatrix} w^k_{(3)} \\ 1 \end{bmatrix} = a\begin{bmatrix} w^k_{(1)} \\ 1 \end{bmatrix} + b\begin{bmatrix} w^k_{(2)} \\ 1 \end{bmatrix} \quad (6.178a)$$

$$\begin{bmatrix} w^k_{(4)} \\ 1 \end{bmatrix} = a'\begin{bmatrix} w^k_{(1)} \\ 1 \end{bmatrix} + b'\begin{bmatrix} w^k_{(2)} \\ 1 \end{bmatrix} \quad (6.178b)$$

for some $(a, b, a', b') \in \mathbb{F}_q^4$. For every pair $\left(w^k_{(1)}, w^k_{(2)}\right)$ there exists at most q^2 choices of (a, b, a', b') that satisfy (6.178).

Expanding the product

$$\prod_{j=1}^{4}\psi(w^k_{(j)}) = \prod_{j=1}^{4}(\chi(w^k_{(j)}) - p_Q) \quad (6.179)$$

and using the pairwise independence of codewords, we can check that for distinct $w^k_{(1)}, \ldots, w^k_{(4)}$,

$$\mathbb{E}\left[\prod_{j=1}^{4} \psi\big(w^k_{(j)}\big)\right] \leq \mathbb{E}\left[\prod_{j=1}^{4} \chi\big(w^k_{(j)}\big)\right] + 2p_Q^4 \tag{6.180}$$

$$\overset{\text{(a)}}{\leq} \mathbb{E}[\chi(w^k_{(1)})\chi(w^k_{(2)})] + 2p_Q^4 \tag{6.181}$$

$$\overset{\text{(b)}}{=} p_Q^2 + 2p_Q^4 \leq 3p_Q^2 \tag{6.182}$$

where (a) follows because $\chi(w^k) \in \{0,1\}$ and (b) follows from the pairwise independence of codewords.

If $\text{rank}(\Omega) = 3$, then every three columns of Ω are linearly independent, thus, the random variables $\psi(w^k_{(1)}), \ldots, \psi(w^k_{(4)})$ are *triple-wise* independent. Moreover, in this case,

$$\begin{bmatrix} w^k_{(4)} \\ 1 \end{bmatrix} = a \begin{bmatrix} w^k_{(1)} \\ 1 \end{bmatrix} + b \begin{bmatrix} w^k_{(2)} \\ 1 \end{bmatrix} + c \begin{bmatrix} w^k_{(3)} \\ 1 \end{bmatrix} \tag{6.183}$$

for some $(a,b,c) \in \mathbb{F}_q^3$. Given any triplet $(w^k_{(1)}, w^k_{(2)}, w^k_{(3)})$ there are at most q^2 triplets $(a,b,c) \in \mathbb{F}_q^3$ that satisfy (6.183). Starting from the bound of (6.180), for distinct $w^k_{(1)}, \ldots, w^k_{(4)}$ that index triple-wise independent codewords we have

$$\mathbb{E}\left[\prod_{j=1}^{4} \psi\big(w^k_{(j)}\big)\right] \leq \mathbb{E}\left[\prod_{j=1}^{4} \chi\big(w^k_{(j)}\big)\right] + 2p_Q^4 \tag{6.184}$$

$$\overset{\text{(a)}}{\leq} \mathbb{E}[\chi(w^k_{(1)})\chi(w^k_{(2)})\chi(w^k_{(3)})] + 2p_Q^4 \tag{6.185}$$

$$\overset{\text{(b)}}{=} p_Q^3 + 2p_Q^4 \leq 3p_Q^3 \tag{6.186}$$

where, again, (a) follows since $\chi(w^k) \in \{0,1\}$ and (b) follows from the triple-wise independent of codewords.

So, in total, among $\binom{M}{4}$ terms in the summation of (6.164), there are at most q^2M^2 that are upper-bounded as in (6.182), at most q^2M^3 that are upper-bounded as in (6.186), and the rest correspond to quadruple-wise independent codewords, hence they are zero. Therefore,

$$\Sigma_4(Q) \leq 3q^2M^2p_Q^2 + 3q^3M^3p_Q^3, \tag{6.187}$$

which yields,

$$\frac{1}{M^4}\ell(Q)^4\Sigma_4(Q) \;\dot{\leq}\; \frac{1}{M^2}\ell(Q)^4p_Q^2 + \frac{1}{M}\ell(Q)^4p_Q^3 \tag{6.188}$$

$$= \left[\frac{1}{M}\ell(Q)^2p_Q + \ell(Q)^2p_Q^2\right]\frac{1}{M}\ell(Q)^2p_Q \tag{6.189}$$

$$\overset{(*)}{\dot{\leq}} \frac{1}{M}\ell(Q)^2p_Q, \tag{6.190}$$

where $(*)$ follows since $\ell(Q) \dot{\leq} M$ for $Q \in \mathcal{Q}'_n$ and $\ell(Q)p_Q \leq 1$. Summing up both sides of (6.176) and (6.190) over $Q \in \mathcal{Q}'_n$, yields (6.162) and concludes the proof. □

6.6 Summary and Outlook

We have studied the *exact* exponential decay rate of the information leaked to the eavesdropper in Wyner's wiretap channel model when an average wiretap channel code in the ensemble of random codes is used for communication. Our analysis shows that the existing lower bound on the secrecy exponent of i.i.d. random codes in [47,52,54] is, indeed, tight. Moreover, our result for constant-composition random codes improves upon that of [53] (see (6.30) and examples in § 6.3.2).

A key step in our analysis is to observe that when a code for the wiretap channel is constructed by associating the secret messages with identically distributed sub-codes (randomly chosen from the ensemble), the exact secrecy exponent of the system equals the exact resolvability exponent of the ensemble. Consequently, we have reduced the problem to the derivation of exact random-coding resolvability exponents. The latter is easier, as the informational divergence of interest (whose exponential decay rate is being assessed) involves a single random distribution (the output distribution), whereas the former involves two random distributions (the conditional and unconditional output distributions). As we have seen in Chapter 5, channel resolvability was already known to be a convenient and powerful tool for establishing the secrecy [21,27,47,51–54,57,58]. Theorem 6.1 highlights the usefulness of this tool by showing that the resolvability exponent is not only a lower bound to the secrecy exponent but also equals the secrecy exponent.

The technique we use to derive the exact resolvability exponents is based on the elementary properties of the random-coding ensemble (namely, pairwise independence of codewords for the achievability part, and *relative* triple- and quadruple-wise independence among codewords for the converse part — as we discussed in § 6.4.3 and § 6.5.2). As we have seen, in addition to the ensemble of i.i.d. random codes, which was the *de facto* random-coding ensemble for proving achievability results in channel resolvability, our method is conveniently applicable to the ensemble of constant-composition random codes and the ensemble of random linear codes.

As we might have guessed, the ensemble of random linear codes has the same behavior as the ensemble of i.i.d. random codes for resolvability, hence for secrecy. It is remarkable that i.i.d. random codes sometimes perform better than constant-composition random codes, as our examples in § 6.3.2 show (see Figures 6.3 and 6.4). This is in contrast to channel coding, where constant-composition random codes turn out to be never worse than i.i.d. random codes in terms of the error exponent [30]. We emphasize that the ensembles of constant-composition and of i.i.d. random codes are *incomparable* in terms of their *resolvability* exponents, as they approximate different reference measures. But, in the context of constructing coding schemes for the wiretap channel, we can compare their secrecy power by looking at their secrecy exponents. The examples presented in § 6.3.2 suggest that the superior ensemble (in terms of the secrecy exponent) depends on the channel alone (i.e., for a given channel, one of the ensembles yields a better secrecy exponent for all input distributions). A subject for future research would be to characterize the set of channels for which the ensemble of i.i.d. random codes results in a better secrecy exponent (and vice versa).

As we have discussed in the remarks following Theorem 5.2, our results (as well as those of others cited) are immediately extensible to prefixed wiretap channels. More precisely, for a given auxiliary channel $P_{X|U}$, the exponents of (6.25) and (6.27), evaluated for the effective channel $P_{Z|U}(z|u) = \sum_x P_{X|U}(x|u)W_{\mathrm{E}}(z|x)$ (instead of W_{E}) and the input distribution P_U are the ensemble-optimal secrecy exponents of both random-coding ensembles. Observe that in this setting $P_{X|U}$ (in addition to the random-binning rate R) is also a design parameter that can be exploited to optimize the secrecy exponent. Moreover, it should also be noted that in the prefixed setting, in addition to the entropy rate of R bits per channel use (for random binning), the encoder requires an entropy rate of $H(X|U)$ bits per channel use to simulate the channel $P_{X|U}$ that has to be taken into account when comparing the secrecy exponents.

6.A Proof of Theorem 6.3

As the results when $I(U;V)=0$ are trivial, we only proceed with the proofs assuming $I(U;V)>0$.

6.A.1 Proof of (i) on Page 114

We need to show that

$$\lim_{n\to\infty} E_n^{\text{i.i.d.}}(P_U, P_{V|U}, R) = E_{\mathrm{s}}^{\text{i.i.d.}}(P_U, P_{V|U}, R). \tag{6.191}$$

Recall that $E_n^{\text{i.i.d.}}$ and $E_{\mathrm{s}}^{\text{i.i.d.}}$ are defined in (6.21) and (6.25), respectively. Since $\mathcal{P}_n(\mathcal{U}\times\mathcal{V}) \subset \mathcal{P}(\mathcal{U}\times\mathcal{V})$,

$$\begin{aligned}\min_{Q_{UV}\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} &\bigl\{D(Q_{UV}\|P_{UV}) + [R - f(Q_{UV}\|P_{UV})]^+\bigr\}\\ &\geq \min_{Q_{UV}\in\mathcal{P}(\mathcal{U}\times\mathcal{V})} \bigl\{D(Q_{UV}\|P_{UV}) + [R - f(Q_{UV}\|P_{UV})]^+\bigr\}. \end{aligned}\tag{6.192}$$

Therefore,

$$\liminf_{n\to\infty} E_n^{\text{i.i.d.}}(P_U, P_{V|U}, R) \geq E_{\mathrm{s}}^{\text{i.i.d.}}(P_U, P_{V|U}, R). \tag{6.193}$$

On the other side, let

$$Q^{\star}_{UV} \coloneqq \operatorname*{arg\,min}_{Q_{UV} \in \mathcal{P}(\mathcal{U}\times\mathcal{V})} \{D(Q_{UV}\|P_{UV}) + [R - f(Q_{UV}\|P_{UV})]^{+}\}. \tag{6.194}$$

Note that $Q^{\star}_{UV} \ll P_{UV}$; if not the value of the objective function would be $+\infty$ whereas at $Q_{UV} = P_{UV}$ it evaluates to $[R - I(U;V)]^{+} < +\infty$. Since $\bigcup_{n\in\mathbb{N}} \mathcal{P}_n(\mathcal{U}\times\mathcal{V})$ is dense in $\mathcal{P}(\mathcal{U}\times\mathcal{V})$, there exists a sequence of n-types $\big(Q^{(n)}_{UV} \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V}),\, n \in \mathbb{N}\big)$ that are absolutely continuous with respect to $Q^{\star}_{UV}$ and converge to $Q^{\star}_{UV}$, i.e.,

$$\lim_{n\to\infty} |Q^{(n)}_{UV} - Q^{\star}_{UV}| = 0.$$

(See Lemma C.1 in Appendix C.) We, also have,

$$\begin{aligned} D(Q^{(n)}_{UV}\|P_{UV}) &+ [R - f(Q^{(n)}_{UV}\|P_{UV})]^{+} \\ &\geq \min_{Q_{UV}\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} \big\{D(Q_{UV}\|P_{UV}) + [R - f(Q_{UV}\|P_{UV})]^{+}\big\} \\ &= E^{\text{i.i.d.}}_{n}(P_U, P_{V|U}, R). \end{aligned} \tag{6.195}$$

Since, $\forall n \in \mathbb{N}$, $Q^{(n)}_{UV} \ll Q^{\star}_{UV} \ll P_{UV}$ and both $D(Q\|P)$ and $f(Q\|P)$ are continuous in Q over the set of distributions Q that are absolutely continuous with respect to P,

$$\begin{aligned} &\lim_{n\to\infty} \big\{D(Q^{(n)}_{UV}\|P_{UV}) + [R - f(Q^{(n)}_{UV}\|P_{UV})]^{+}\big\} \\ &\qquad = D(Q^{\star}_{UV}\|P_{UV}) + [R - f(Q^{\star}_{UV}\|P_{UV})]^{+} = E^{\text{i.i.d.}}_{\text{s}}(P_U, P_{V|U}, R). \end{aligned} \tag{6.196}$$

Using (6.195) in (6.196) yields,

$$E^{\text{i.i.d.}}_{\text{s}}(P_U, P_{V|U}, R) \geq \limsup_{n\to\infty} E^{\text{i.i.d.}}_{n}(P_U, P_{V|U}, R) \tag{6.197}$$

which, together with (6.193) prove (6.191).

6.A.2 Proof of (ii) on Page 114

To prove (6.26), we need to show that

$$\lim_{n\to\infty} E^{\text{c.c.}}_{n}(P^{(n)}_{U}, P_{V|U}, R) = E^{\text{c.c.}}_{\text{s}}(P_U, P_{V|U}, R) \tag{6.198}$$

for any sequence of n-types, $P^{(n)}_{U} \in \mathcal{P}_n(\mathcal{U})$ that converge to P_U. The proof is divided into a sequence of lemmas. Recall that, without essential loss of generality, we assume $\operatorname{supp}(P^{(n)}_{U}) = \operatorname{supp}(P_U) = \mathcal{U}$.

Lemma 6.10. *$\forall Q \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})$ such that $Q \ll P$, $|g_n(Q\|P)| < +\infty$ and the minimizing n-type Q' on the right-hand side of* (6.23b) *is also absolutely continuous with respect to P. The same statement holds for g and the minimizing distribution Q' on the right-hand side of* (6.27b) *for any distribution (not necessarily n-type) $Q \ll P$.*

Proof. $Q \ll P$ implies the summation

$$\sum_{(u,v)} Q(u,v) \log P_{V|U}(v|u) \tag{6.199}$$

is bounded. The entropy term $H(Q_V)$ is also bounded. Moreover,

$$\min_{\substack{Q' \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V}):\\ Q'_U = Q_U, Q'_V = Q_V}} D(Q' \| P) \leq D(Q\|P) < +\infty, \tag{6.200}$$

because $Q \ll P$ and $Q' = Q$ is a feasible point in the above minimization. Therefore, $g_n(P\|Q)$ is bounded and the optimizing Q' is absolutely continuous with respect to P (if not $D(Q'\|P) = +\infty$). (The claim on g follows exactly in the same way.) □

Corollary 6.11. *The minimizing n-type Q_{UV} in* (6.23a) *is absolutely continuous with respect to $P_{UV}^{(n)}$. Similarly, the minimizing distribution Q_{UV} in* (6.27a) *is absolutely continuous with respect to P_{UV}.*

Proof. Consider the minimizing Q_{UV} in (6.27a). If it is not absolutely continuous with respect to P_{UV}, $D(Q_{UV}\|P_{UV}) = +\infty$, however, taking $Q_{UV} = P_{UV}$ in (6.23a)

$$E_{\mathrm{s}}^{\text{c.c.}}(P_U, P_{V|U}, R) \leq [R - g(P_{UV}\|P_{UV})]^+ \tag{6.201}$$

which is finite because $g(Q\|P)$ is bounded if $Q \ll P$. (In fact, it follows that $g(P\|P) = I(P)$ thus, $E_{\mathrm{s}}^{\text{c.c.}}(P_U, P_{V|U}, R) \leq [R - I(U;V)]^+$.) Therefore no $Q_{UV} \not\ll P_{UV}$ can be the minimizer. The same reasoning shows that the minimizing Q_{UV} in (6.23a) is absolutely continuous with respect to $P_{UV}^{(n)}$. □

Lemma 6.12. *Let $\big(Q^{(n)} \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V}), n \in \mathbb{N}\big)$ be a sequence of joint n-types and $\big(P^{(n)} \in \mathcal{P}(\mathcal{U}\times\mathcal{V}), n \in \mathbb{N}\big)$ a sequence of joint distributions. Let*

$$Q := \lim_{n\to\infty} Q^{(n)} \qquad \textit{and} \qquad P := \lim_{n\to\infty} P^{(n)}. \tag{6.202}$$

(Note that since the distributions live in a compact space, $\mathcal{P}(\mathcal{U}\times\mathcal{V})$, by passing to a subsequence if necessary, both above limits exist.) Assume, furthermore, that

(i) all $P^{(n)}$s have the same support as P; and

(ii) $\forall n \in \mathbb{N}$, $Q^{(n)} \ll P^{(n)}$.

Then,

$$\lim_{n\to\infty} g_n\big(Q^{(n)}\|P^{(n)}\big) = g(Q\|P), \tag{6.203}$$

(where g_n and g are defined in (6.23b) *and* (6.27b) *respectively).*

Accepting Lemma 6.12 momentarily, we are ready to establish (6.198). Note that $\lim_{n\to\infty} P_{UV}^{(n)} = P_{UV}$ because of the assumption that $\lim_{n\to\infty} P_U^{(n)} = P_U$. Also since, for all $n \in \mathbb{N}$, $\operatorname{supp}(P_U^{(n)}) = \operatorname{supp}(P_U)$ (by assumption), $\operatorname{supp}(P_{UV}^{(n)}) = \operatorname{supp}(P_{UV})$, for all $n \in \mathbb{N}$. Let

$$\tilde{Q}_{UV}^{(n)} := \operatorname*{arg\,min}_{\substack{Q_{UV}\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})\\ Q_U=P_U^{(n)}}} \left\{D(Q_{UV}\|P_{UV}^{(n)}) + [R - g_n(Q_{UV}\|P_{UV}^{(n)})]^+\right\} \tag{6.204}$$

and (by passing to a subsequence if necessary)

$$\tilde{Q}_{UV} := \lim_{n\to\infty} \tilde{Q}_{UV}^{(n)} \tag{6.205}$$

Note that $\tilde{Q}_{UV}^{(n)} \ll P_{UV}^{(n)}$ (by Corollary 6.11), thus, by the continuity of divergence and Lemma 6.12 we have

$$\begin{aligned}\liminf_{n\to\infty} E_n^{\text{c.c.}}(P_U^{(n)}, P_{V|U}, R) &= \liminf_{n\to\infty}\left\{D\big(\tilde{Q}_{UV}^{(n)}\|P_{UV}^{(n)}\big) + \big[R - g_n\big(\tilde{Q}_{UV}^{(n)}\|P_{UV}^{(n)}\big)\big]^+\right\}\\ &= D(\tilde{Q}_{UV}\|P_{UV}) + [R - g(\tilde{Q}_{UV}\|P_{UV})]^+ \end{aligned}\tag{6.206}$$

Moreover, since $\tilde{Q}_U^{(n)} = P_U^{(n)}$ and $\lim_{n\to\infty} P_U^{(n)} = P_U$, $\tilde{Q}_U = P_U$. Therefore,

$$\begin{aligned}&D(\tilde{Q}_{UV}\|P_{UV}) + [R - g(\tilde{Q}_{UV}\|P_{UV})]^+\\ &\quad\geq \min_{\substack{Q_{UV}:\\ Q_U=P_U}} \left\{D(Q_{UV}\|P_{UV}) + [R - g(Q_{UV}\|P_{UV})]^+\right\} = E_{\text{s}}^{\text{c.c.}}(P_U, P_{V|U}, R).\end{aligned}\tag{6.207}$$

Consequently,

$$\liminf_{n\to\infty} E_n^{\text{c.c.}}(P_U^{(n)}, P_{V|U}, R) \geq E_{\text{s}}^{\text{c.c.}}(P_U, P_{V|U}, R). \tag{6.208}$$

To prove the reverse inequality, let

$$Q_{UV}^\star = \operatorname*{arg\,min}_{\substack{Q_{UV}\in\mathcal{P}(\mathcal{U}\times\mathcal{V}):\\ Q_U=P_U}} \left\{D(Q_{UV}\|P_{UV}) + [R - g(Q_{UV}\|P_{UV})]^+\right\}. \tag{6.209}$$

Corollary 6.11 implies $Q_{UV}^\star \ll P_{UV}$. Moreover, since $\lim_{n\to\infty} P_U^{(n)} = P_U = Q_U^\star$, there exists a sequence of n-types $Q_{UV}^{(n)} \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})$ such that

(a) $Q_{UV}^{(n)} \ll Q_{UV}$ (thus, $Q_{UV}^{(n)} \ll P_{UV}^{(n)}$),

(b) $\lim_{n\to\infty} |Q_{UV}^{(n)} - Q_{UV}^\star| = 0$, and

(c) $\forall n$, $Q_U^{(n)} = P_U^{(n)}$

(see Lemma C.2 in Appendix C). Consequently, using the continuity of divergence and Lemma 6.12 once again,

$$\begin{aligned} E_{\mathrm{s}}^{\mathrm{c.c.}}(P_U, P_{V|U}, R) &= \left\{D(Q^{\star}_{UV}\|P_{UV}) + [R - g(Q^{\star}_{UV}\|P_{UV})]^+\right\} \\ &= \lim_{n\to\infty}\left\{D\big(Q^{(n)}_{UV}\|P^{(n)}_{UV}\big) + \big[R - g_n\big(Q^{(n)}_{UV}\|P^{(n)}_{UV}\big)\big]^+\right\} \end{aligned} \tag{6.210}$$

Moreover, since $\forall n \in \mathbb{N}$, $Q^{(n)}_U = P^{(n)}_U$,

$$\begin{aligned} &D\big(Q^{(n)}_{UV}\|P^{(n)}_{UV}\big) + \big[R - g_n\big(Q^{(n)}_{UV}\|P^{(n)}_{UV}\big)\big]^+ \\ &\qquad \geq \min_{\substack{Q_{UV}\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V}):\\ Q_U = P^{(n)}_U}} \left\{D(Q_{UV}\|P^{(n)}_{UV}) + [R - g_n(Q_{UV}\|P^{(n)}_{UV})]^+\right\}. \end{aligned} \tag{6.211}$$

Combining (6.210) and (6.211) yields,

$$E_{\mathrm{s}}^{\mathrm{c.c.}}(P_U, P_{V|U}, R) \geq \limsup_{n\to\infty} E_n^{\mathrm{c.c.}}(P^{(n)}_U, P_{V|U}, R) \tag{6.212}$$

Uniting (6.208) and (6.212) proves (6.198).

It remains to prove Lemma 6.12.

Proof of Lemma 6.12. Define, for any pair of joint-distributions P, Q,

$$\phi_n(Q\|P) \coloneqq \min_{\substack{Q'\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V}):\\ Q'_U = Q_U, Q'_V = Q_V}} D(Q'\|P) \tag{6.213}$$

and

$$\phi(Q\|P) \coloneqq \min_{\substack{Q'\in\mathcal{P}(\mathcal{U}\times\mathcal{V}):\\ Q'_U = Q_U, Q'_V = Q_V}} D(Q'\|P). \tag{6.214}$$

Entropy $H(Q_V)$ is continuous in Q and

$$\sum_{u,v} Q(u,v)\log P_{V|U}(v|u)$$

is also continuous when $Q \ll P$. Thus, it suffices to show

$$\lim_{n\to\infty} \phi_n(Q^{(n)}\|P^{(n)}) = \phi(Q\|P), \tag{6.215}$$

to establish the claim. As $\mathcal{P}_n(\mathcal{U}\times\mathcal{V}) \subset \mathcal{P}(\mathcal{U}\times\mathcal{V})$,

$$\phi_n(Q^{(n)}\|P^{(n)}) \geq \phi(Q^{(n)}\|P^{(n)}). \tag{6.216}$$

Since by assumption all $P^{(n)}$ have the same support as P, for all n, the minimizing Q' (in evaluation of $\phi(Q^{(n)}\|P^{(n)})$) lies in the compact set of distributions that are absolutely continuous with respect to P. Therefore, applying Lemma B.1 (in Appendix B), we conclude that ϕ is convex and continuous in (P, Q). Consequently,

$$\liminf_{n\to\infty} \phi_n(Q^{(n)}\|P^{(n)}) \geq \lim_{n\to\infty} \phi(Q^{(n)}\|P^{(n)}) = \phi(Q\|P). \tag{6.217}$$

We now prove the reverse equality. To simplify the argument, let us start with an additional assumption, which we relax later, on the sequence of n-types $Q^{(n)}$. Assume for all (but finitely many) $n \in \mathbb{N}$,

$$\operatorname{supp}(Q_U^{(n)}) \subseteq \operatorname{supp}(Q_U) \qquad \text{and} \qquad \operatorname{supp}(Q_V^{(n)}) \subseteq \operatorname{supp}(Q_V). \tag{6.218}$$

Let

$$\tilde{Q} \coloneqq \operatorname*{arg\,min}_{\substack{Q' \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V}): \\ Q'_U = Q_U, Q'_V = Q_V}} D(Q' \| P). \tag{6.219}$$

Note that $\tilde{Q} \ll P$ (as we discussed in Lemma 6.10). We claim that there exists a sequence of n-types $(\tilde{Q}^{(n)} \in \mathcal{P}_n(\mathcal{U} \times \mathcal{V}),\ n \in \mathbb{N})$ such that

(a) $\tilde{Q}^{(n)} \ll P^{(n)}$,

(b) $\lim_{n\to\infty} |\tilde{Q}^{(n)} - \tilde{Q}| = 0$,

(c) $\tilde{Q}_U^{(n)} = Q_U^{(n)}$ and $\tilde{Q}_V^{(n)} = Q_V^{(n)}$.

Properties (a) and (b) follow, rather easily, from the denseness of the union of n-types in the simplex. Also, since $\tilde{Q}^{(n)}$ converges to $\tilde{Q}$ whose u- and v-marginals are, respectively, Q_U and Q_V, and $Q_U^{(n)}$ and $Q_V^{(n)}$ also converge to Q_U and Q_V, the u- and v-marginals of $\tilde{Q}^{(n)}$ must be *close* to $Q_U^{(n)}$ and $Q_V^{(n)}$, respectively, i.e.,

$$\lim_{n\to\infty} |\tilde{Q}_U^{(n)} - Q_U^{(n)}| = 0 \qquad \text{and} \qquad \lim_{n\to\infty} |\tilde{Q}_V^{(n)} - Q_V^{(n)}| = 0. \tag{6.220}$$

However, it is not clear if the marginals of $\tilde{Q}^{(n)}$ exactly match those of $Q^{(n)}$. It can be shown (see Lemma C.3 in Appendix C) that, under mild technical conditions, we can actually construct a sequence of n-types $(\tilde{Q}^{(n)}, n \in \mathbb{N})$ that satisfies (c) (as well as (a) and (b)).[8]

Properties (a) and (b), together with the continuity of divergence, imply

$$\phi(Q\|P) = D(\tilde{Q}\|P) = \lim_{n\to\infty} D(\tilde{Q}^{(n)}\|P^{(n)}). \tag{6.221}$$

Property (c) yields, $\forall n \in \mathbb{N}$,

$$D(\tilde{Q}^{(n)}\|P^{(n)}) \geq \min_{\substack{Q' \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V}): \\ Q'_U = Q_U^{(n)}, Q'_V = Q_V^{(n)}}} D(Q'\|P^{(n)}) = \phi_n(Q^{(n)}\|P^{(n)}) \tag{6.222}$$

Uniting (6.221) and (6.222) yields

$$\limsup_{n\to\infty} \phi_n(Q^{(n)}\|P^{(n)}) \leq \phi(Q\|P) \tag{6.223}$$

which, together with (6.217) establishes (6.215).

[8] Lemma C.3 constructs an n-type $\tilde{Q}^{(n)}$ that is absolutely continuous with respect to $\tilde{Q}$. Since $\tilde{Q} \ll P$ and all $P^{(n)}$s are assumed to have the same support as P property (a) follows.

Let us now verify that the requirements of Lemma C.3 on the support of $\tilde{Q}$ are satisfied. Following the remark after Lemma C.3 it is sufficient to show that $Q^{(n)} \ll \tilde{Q}$. As it is shown in Lemma 6.13 (in Appendix 6.C), $\tilde{Q}(u_0, v_0) = 0$ implies $P(u_0, v_0) = 0$, or $\tilde{Q}_U(u_0) = 0$, or $\tilde{Q}_V(v_0) = 0$. If $P(u_0, v_0) = 0$, since all $P^{(n)}$s have the same support as P, $\forall n$, $P^{(n)}(u_0, v_0) = 0$, which, since $Q^{(n)} \ll P^{(n)}$, implies $\forall n$, $Q^{(n)}(u_0, v_0) = 0$. Otherwise, if $\tilde{Q}_U(u_0) = 0$, $Q_U(u_0) = 0$ (since $\tilde{Q}$ has the same marginals as Q) and the assumption (6.218) yields $\forall n$, $Q_U^{(n)}(u_0) = 0$ which in turn implies $Q^{(n)}(u_0, v_0) = 0$. Similarly $\tilde{Q}_V(v_0) = 0$ implies $\forall n$, $Q^{(n)}(u_0, v_0) = 0$. Therefore, $Q^{(n)} \ll \tilde{Q}$, hence the requirements of Lemma C.3 are satisfied if (6.218) holds.

It now remains to relax the constraints (6.218). Let

$$\mathcal{U}' := \operatorname{supp}(Q_U) \subseteq \mathcal{U} \qquad \text{and} \qquad \mathcal{V}' := \operatorname{supp}(Q_V) \subseteq \mathcal{V} \tag{6.224}$$

and, $\forall n \in \mathbb{N}$, define

$$\lambda_n := Q^{(n)}(\mathcal{U}' \times \mathcal{V}'). \tag{6.225}$$

Since $Q^{(n)}$ is an n-type, $n\lambda_n \in \mathbb{Z}$. Moreover, since $\operatorname{supp}(Q) \subseteq \mathcal{U}' \times \mathcal{V}'$ and $\lim_{n\to\infty} Q^{(n)} = Q$,

$$\lim_{n\to\infty} \lambda_n = 1. \tag{6.226}$$

Define

$$Q_1^{(n\lambda_n)}(u, v) := \frac{Q^{(n)}(u, v)}{\lambda_n} \mathbb{1}\{(u, v) \in \mathcal{U}' \times \mathcal{V}'\} \tag{6.227}$$

and

$$Q_2^{(n[1-\lambda_n])}(u, v) := \frac{Q^{(n)}(u, v)}{1 - \lambda_n} \mathbb{1}\{(u, v) \in (\mathcal{U} \times \mathcal{V}) \setminus (\mathcal{U}' \times \mathcal{V}')\}. \tag{6.228}$$

Note that $Q_1^{(n\lambda_n)}$ is an $n\lambda_n$-type (similarly $Q_2^{n[1-\lambda_n]}$ is an $n[1 - \lambda_n]$-type) and, by construction, $Q_1^{(n\lambda_n)}$ satisfies the constraints of (6.218). Consequently, our previous argument shows that we can quantize $\tilde{Q}$ to a sequence of $n\lambda_n$-types $\hat{Q}^{(n\lambda_n)}$ that are absolutely continuous with respect to $\tilde{Q}$, converge to $\tilde{Q}$ as $n \to \infty$, and have the same u- and v-marginals as $Q_1^{(n\lambda_n)}$. Therefore,

$$\phi(Q\|P) = D(\tilde{Q}\|P) = \lim_{n\to\infty} D\big(\hat{Q}^{(n\lambda_n)} \,\big\|\, P^{(n)}\big). \tag{6.229}$$

Moreover, since $P^{(n)} \to P$ as $n \to \infty$, for large enough n,

$$|P^{(n)} - P| \le \frac{1}{2} P_{\min}, \tag{6.230}$$

where we have defined

$$P_{\min} := \min_{(u,v)\in\operatorname{supp}(P)} P(u, v). \tag{6.231}$$

Since all $P^{(n)}$ have the same support as P (6.230) implies for large enough n,

$$P^{(n)}(u, v) \ge \frac{1}{2} P_{\min}, \qquad \forall (u, v) \in \operatorname{supp}(P). \tag{6.232}$$

The above, together with the fact that $Q_2^{(n[1-\lambda_n])} \ll P^{(n)}$ implies for large n,

$$D\big(Q_2^{(n[1-\lambda_n])} \parallel P^{(n)}\big) \leq \log\Big[\frac{2}{P_{\min}}\Big] \tag{6.233}$$

Consequently, since $\lambda_n \to 1$,

$$\lim_{n\to\infty} (1-\lambda_n) D\big(Q_2^{(n[1-\lambda_n])} \parallel P^{(n)}\big) = 0 \tag{6.234}$$

Therefore,

$$\begin{aligned}\phi(Q\|P) = D(\tilde{Q}\|P) &= \lim_{n\to\infty} D\big(\hat{Q}^{(n\lambda_n)} \parallel P^{(n)}\big)\\ &\overset{(*)}{=} \lim_{n\to\infty} \big\{\lambda_n D\big(\hat{Q}^{(n\lambda_n)} \parallel P^{(n)}\big) + (1-\lambda_n) D\big(Q_2^{(n[1-\lambda_n])} \parallel P^{(n)}\big)\big\}\end{aligned} \tag{6.235}$$

In the above, $(*)$ follows from (6.234). Since KL divergence is convex, defining the n-type $\tilde{Q}^{(n)} := \lambda_n \hat{Q}^{(n\lambda_n)} + (1-\lambda_n) Q_2^{(n[1-\lambda_n])}$, for $\forall n$, we have

$$\begin{aligned}&\lambda_n D\big(\tilde{Q}^{(n\lambda_n)} \parallel P^{(n)}\big) + (1-\lambda_n) D\big(Q_2^{(n[1-\lambda_n])} \parallel P^{(n)}\big)\\ &\qquad \geq D\big(\lambda_n \hat{Q}^{(n\lambda_n)} + (1-\lambda_n) Q_2^{(n[1-\lambda_n])} \parallel P^{(n)}\big) = D(\tilde{Q}^{(n)}\|P^{(n)})\end{aligned} \tag{6.236}$$

Using the above in (6.235) yields

$$\phi(Q\|P) \geq \limsup_{n\to\infty} D(\tilde{Q}^{(n)}\|P). \tag{6.237}$$

Moreover, since $\hat{Q}^{(n\lambda_n)}$ has the same marginals as $Q_1^{(n\lambda_n)}$, it can be verified that $\tilde{Q}^{(n)}$, by construction, has the same marginals as $Q^{(n)}$. Therefore (6.222) holds. Once again, uniting (6.222) and (6.237) yields (6.223) which, together with (6.217), establishes (6.215). □

6.A.3 Strict Monotonicity of The Exponents in Rate

That $E_{\mathrm{s}}^{\text{i.i.d.}}$ is strictly increasing in R for $R > I(U;V)$ can be easily seen through the form of (6.28): $E_{\mathrm{s}}^{\text{i.i.d.}}$ is the supremum of affine functions of R thus is convex in R. Moreover, since $F_0(P_{UV}, \lambda)$ is a convex function of λ passing through the origin with slope $I(U;V)$, $E_{\mathrm{s}}^{\text{i.i.d.}}$ starts to increase above 0 once R exceeds $I(U;V)$ which means it will be strictly increasing for $R > I(U;V)$.

Hence, we only prove the claim for $E_{\mathrm{s}}^{\text{c.c.}}$. (This proof can also be used to show $E_{\mathrm{s}}^{\text{i.i.d.}}$ is strictly increasing in R, replacing g with f.) Note that

$$E_{\mathrm{s}}^{\text{c.c.}}(P_U, P_{V|U}, R) = \min\left\{\begin{array}{l}\displaystyle\min_{\substack{Q_{UV}:\\ Q_U=P_U,\\ g(Q_{UV}\|P_{UV})\geq R}} D(Q_{UV}\|P_{UV})\,,\\[2em] \displaystyle\min_{\substack{Q_{UV}:\\ Q_U=P_U,\\ g(Q_{UV}\|P_{UV})\leq R}} \big\{D(Q_{UV}\|P_{UV}) + R - g(Q_{UV}\|P_{UV})\big\}\end{array}\right\}. \tag{6.238}$$

Consider the first minimization

$$\min_{\substack{Q_{UV}:\\ Q_U=P_U,\\ g(Q_{UV}\|P_{UV})\geq R}} D(Q_{UV}\|P_{UV}) = \min_{\substack{Q_{UV}\ll P_{UV}:\\ Q_U=P_U,\\ g(Q_{UV}\|P_{UV})\geq R}} D(Q_{UV}\|P_{UV}). \tag{6.239}$$

(We can safely restrict the search space to distributions Q_{UV} that are absolutely continuous with respect to P_{UV} since, otherwise, $D(Q_{UV}\|P_{UV}) = +\infty$.) The function g is continuous over the compact set of distributions $Q_{UV} \ll P_{UV}$ (see the proof of Lemma 6.12). Moreover, $D(Q_{UV}\|P_{UV})$ is a convex function attaining its unique global minimum at $Q_{UV} = P_{UV}$. We can also verify that $g(P_{UV}) = I(U;V)$. Therefore, if $R \geq I(U;V)$, Lemma B.2 (in Appendix B) yields

$$\min_{\substack{Q_{UV}:\\ Q_U=P_U,\\ g(Q_{UV}\|P_{UV})\geq R}} D(Q_{UV}\|P_{UV}) = \min_{\substack{Q_{UV}:\\ Q_U=P_U,\\ g(Q_{UV}\|P_{UV})=R}} D(Q_{UV}\|P_{UV}) \tag{6.240a}$$

and for every Q_{UV} such that $g(Q_{UV}\|P_{UV}) > R$,

$$D(Q_{UV}\|P_{UV}) > \min_{\substack{Q_{UV}:\\ Q_U=P_U,\\ g(Q_{UV}\|P_{UV})=R}} D(Q_{UV}\|P_{UV}). \tag{6.240b}$$

Consequently, using (6.240a) in (6.238), we get

$$E_{\mathrm{s}}^{\text{c.c.}}(P_U, P_{V|U}, R) = \min_{\substack{Q_{UV}:\\ Q_U=P_U,\\ g(Q_{UV}\|P_{UV})\leq R}} \big\{D(Q_{UV}\|P_{UV}) + R - g(Q_{UV}\|P_{UV})\big\} \tag{6.241}$$

$$= R + \min_{\substack{Q_{UV}:\\ Q_U=P_U,\\ g(Q_{UV}\|P_{UV})\leq R}} \big\{D(Q_{UV}\|P_{UV}) - g(Q_{UV}\|P_{UV})\big\} \tag{6.242}$$

(when $R \geq I(U;V)$).

Take $R' > R \geq I(U;V)$ and let

$$Q^\star_{UV} = \operatorname*{arg\,min}_{\substack{Q_{UV}:\\ Q_U=P_U,\\ g(Q_{UV}\|P_{UV})\leq R'}} \big\{D(Q_{UV}\|P_{UV}) - g(Q_{UV}\|P_{UV})\big\} \tag{6.243}$$

If $g(Q^\star_{UV}\|P_{UV}) \leq R$, then

$$E_{\mathrm{s}}^{\text{c.c.}}(P_U, P_{V|U}, R') = R' + D(Q^\star_{UV}\|P_{UV}) - g(Q^\star_{UV}\|P_{UV}) \tag{6.244}$$

$$\geq R' + \min_{\substack{Q_{UV}:\\ Q_U=P_U,\\ g(Q_{UV}\|P_{UV})\leq R}} \big\{D(Q_{UV}\|P_{UV}) - g(Q_{UV}\|P_{UV})\big\} \tag{6.245}$$

$$> R + \min_{\substack{Q_{UV}:\\ Q_U=P_U,\\ g(Q_{UV}\|P_{UV})\leq R}} \big\{D(Q_{UV}\|P_{UV}) - g(Q_{UV}\|P_{UV})\big\} \tag{6.246}$$

$$= E_{\mathrm{s}}^{\text{c.c.}}(P_U, P_{V|U}, R). \tag{6.247}$$

Otherwise, we have the following chain of inequalities

$$E_{\mathrm{s}}^{\mathrm{c.c.}}(P_U, P_{V|U}, R) = R + \min_{\substack{Q_{UV}:\\ Q_U = P_U,\\ g(Q_{UV}\|P_{UV}) \le R}} \big\{D(Q_{UV}\|P_{UV}) - g(Q_{UV}\|P_{UV})\big\}$$

$$\le R + \min_{\substack{Q_{UV}:\\ Q_U = P_U,\\ g(Q_{UV}\|P_{UV}) = R}} \big\{D(Q_{UV}\|P_{UV}) - g(Q_{UV}\|P_{UV})\big\} \tag{6.248}$$

$$= \min_{\substack{Q_{UV}:\\ Q_U = P_U,\\ g(Q_{UV}\|P_{UV}) = R}} \big\{D(Q_{UV}\|P_{UV})\big\} \tag{6.249}$$

$$\overset{\text{(a)}}{<} D(Q^{\star}_{UV}\|P_{UV}) \tag{6.250}$$

$$\overset{\text{(b)}}{\le} R' + D(Q^{\star}_{UV}\|P_{UV}) - g(Q^{\star}_{UV}\|P_{UV}) = E_{\mathrm{s}}^{\mathrm{c.c.}}(P_U, P_{V|U}, R'). \tag{6.251}$$

In the above, (a) follows from (6.240b) and (b) holds since $g(Q^{\star}_{UV}\|P_{UV}) \le R'$.

Consequently, for $R' > R \ge I(U;V)$,

$$E_{\mathrm{s}}^{\mathrm{c.c.}}(P_U, P_{V|U}, R') > E_{\mathrm{s}}^{\mathrm{c.c.}}(P_U, P_{V|U}, R). \tag{6.252}$$

6.A.4 Alternative form of $E_{\mathrm{s}}^{\mathrm{i.i.d.}}$

As $[a]^+ = \max_{0\le\lambda\le 1} \lambda a$, we have

$$\min_{Q_{UV}} \big\{D(Q_{UV}\|P_{UV}) + [R - f(Q_{UV}\|P_{UV})]^+\big\}$$

$$= \min_{Q_{UV}} \left\{ D(Q_{UV}\|P_{UV}) + \max_{0\le\lambda\le 1} \lambda[R - f(Q_{UV}\|P_{UV})] \right\} \tag{6.253}$$

$$= \min_{Q_{UV}} \max_{0\le\lambda\le 1} \{\lambda R + D(Q_{UV}\|P_{UV}) - \lambda f(Q_{UV}\|P_{UV})\} \tag{6.254}$$

$$\overset{\text{(a)}}{=} \max_{0\le\lambda\le 1} \min_{Q_{UV}} \{\lambda R + D(Q_{UV}\|P_{UV}) - \lambda f(Q_{UV}\|P_{UV})\} \tag{6.255}$$

$$= \max_{0\le\lambda\le 1} \left\{ \lambda R + \min_{Q_{UV}} \{D(Q_{UV}\|P_{UV}) - \lambda f(Q_{UV}\|P_{UV})\} \right\} \tag{6.256}$$

$$\overset{\text{(b)}}{=} \max_{0\le\lambda\le 1} \{\lambda R - F_0(P_{UV}, \lambda)\}. \tag{6.257}$$

In the above, (a) follows since $D(Q_{UV}\|P_{UV}) - \lambda f(Q_{UV}\|P_{UV})$ is convex in Q_{UV} (recall that f is linear in Q_{UV}) and (b) follows since the concavity of logarithm implies

$$D(Q_{UV}\|P_{UV}) - \lambda f(Q_{UV}\|P_{UV})$$

$$= \sum_{(u,v)\in\mathcal{U}\times\mathcal{V}} Q_{UV}(u,v) \log \frac{Q_{UV}(u,v)}{P_{UV}(u,v)^{1+\lambda} P_U(u)^{-\lambda} P_V(v)^{-\lambda}} \tag{6.258}$$

$$\overset{(*)}{\ge} -\log \sum_{(u,v)\in\mathcal{U}\times\mathcal{V}} P_{UV}(u,v)^{1+\lambda} P_U(u)^{-\lambda} P_V(v)^{-\lambda} \tag{6.259}$$

$$= F_0(P_{UV}, \lambda), \tag{6.260}$$

with equality in $(*)$ iff $Q_{UV}(u,v) \propto P_{UV}(u,v)^{1+\lambda} P_U(u)^{-\lambda} P_V(v)^{-\lambda}$. □

6.B Proof of Lemma 6.5

Taking $Q' = Q$ in (6.27b) shows

$$g(Q) \le I(Q) + D(Q_U \| P_U). \tag{6.261}$$

Therefore,

$$\begin{aligned} E_{\mathrm{s}}^{\mathrm{c.c.}}(P_U, P_{V|U}, R) &= \min_{\substack{Q_{UV} \in \mathcal{P}(\mathcal{U}\times\mathcal{V}):\\ Q_U = P_U}} \bigl\{ D(Q_{UV} \| P_{UV}) + [R - g(Q_{UV} \| P_{UV})]^+ \bigr\} \\ &\ge \min_{\substack{Q_{UV} \in \mathcal{P}(\mathcal{U}\times\mathcal{V}):\\ Q_U = P_U}} \bigl\{ D(Q_{UV} \| P_{UV}) + [R - I(Q_{UV})]^+ \bigr\} && (6.262) \\ &\overset{\text{(a)}}{=} \min_{\substack{Q_{UV} \in \mathcal{P}(\mathcal{U}\times\mathcal{V}):\\ Q_U = P_U}} \bigl\{ D(Q_{UV} \| P_{UV}) + \max_{0 \le \lambda \le 1} \{\lambda R - \lambda I(P_{UV})\} \bigr\} && (6.263) \\ &\overset{\text{(b)}}{=} \max_{0 \le \lambda \le 1} \bigl\{ \lambda R + \min_{\substack{Q_{UV} \in \mathcal{P}(\mathcal{U}\times\mathcal{V}):\\ Q_U = P_U}} \{ D(Q_{UV} \| P_{VU}) - \lambda I(Q_{UV}) \} \bigr\}, && (6.264) \end{aligned}$$

where (a) follows since $[a]^+ = \max_{0 \le \lambda \le 1} \lambda a$ and (b) holds since $D(Q\|P) - \lambda I(Q)$ is convex in Q for $\lambda \in [0,1]$ (and linear in λ). To verify the latter, note that

$$I(Q) = \min_{Q'_V \in \mathcal{P}(\mathcal{V})} D(Q \| Q_U \times Q'_V). \tag{6.265}$$

Therefore,

$$\begin{aligned} D(Q\|P) - \lambda I(Q) &= \max_{Q'_V \in \mathcal{P}(\mathcal{V})} \{ D(Q\|P) - \lambda D(Q \| Q_U \times Q'_V) \} && (6.266) \\ &= \max_{Q'_V \in \mathcal{P}(\mathcal{V})} \Bigl\{ \sum_{(u,v) \in \mathcal{U}\times\mathcal{V}} Q(u,v) \log \Bigl[\frac{Q(u,v)^{1-\lambda}}{P(u,v) Q_U(u)^{-\lambda} Q'_V(v)^{-\lambda}} \Bigr] \Bigr\} && (6.267) \\ &= \frac{1}{t} \max_{Q'_V \in \mathcal{P}(\mathcal{V})} \Bigl\{ \sum_{(u,v) \in \mathcal{U}\times\mathcal{V}} Q(u,v) \log \Bigl[\frac{Q(u,v)}{P(u,v)^t Q_U(u)^{1-t} Q'_V(v)^{1-t}} \Bigr] \Bigr\}, && (6.268) \end{aligned}$$

where we have defined $t \coloneqq 1/(1-\lambda)$ in the last step. As $t \ge 1$, the objective function inside the $\max\{\cdot\}$ in (6.268) is convex in Q and since the supremum of convex functions is still convex, and $t \ge 0$ (because $\lambda \le 1$) the convexity of $D(Q\|P) - \lambda I(Q)$ in Q follows. It can also be seen that the objective function in (6.268) is concave in Q'_V. Therefore

$$\begin{aligned}
&\min_{\substack{Q_{UV}\in\mathcal{P}(\mathcal{U}\times\mathcal{V}):\\ Q_U=P_U}} \{D(Q_{UV}\|P_{VU})-\lambda I(Q_{UV})\}\\
&= \frac{1}{t}\min_{\substack{Q_{UV}:\\ Q_U=P_U}}\max_{Q'_V}\Big\{\sum_{(u,v)\in\mathcal{U}\times\mathcal{V}} Q(u,v)\log\Big[\frac{Q(u,v)}{P(u,v)^t Q_U(u)^{1-t}Q'_V(v)^{1-t}}\Big]\Big\} \qquad (6.269)\\
&\overset{(a)}{=} \frac{1}{t}\max_{Q'_V}\min_{\substack{Q_{UV}:\\ Q_U=P_U}}\Big\{\sum_{(u,v)\in\mathcal{U}\times\mathcal{V}} Q(u,v)\log\Big[\frac{Q(u,v)}{P(u,v)^t Q_U(u)^{1-t}Q'_V(v)^{1-t}}\Big]\Big\} \qquad (6.270)\\
&\overset{(b)}{\geq} \frac{1}{t}\max_{Q'_V}\Big\{-\log\Big[\sum_{(u,v)\in\mathcal{U}\times\mathcal{V}} P(u,v)^t P_U(u)^{1-t}Q'_V(v)^{1-t}\Big]\Big\} \qquad (6.271)\\
&= -\min_{Q'_V}\Big\{\frac{1}{t}\log\Big[\sum_{v\in\mathcal{V}} Q'_V(v)^{1-t}\sum_{u\in\mathcal{U}} P_U(u)^{1-t}P_{UV}(u,v)^t\Big]\Big\}, \qquad (6.272)
\end{aligned}$$

where (a) follows from the concavity of the objective function of (6.268) in Q'_V and (b) follows from the concavity of logarithm, together with the fact that $Q_U = P_U$. KKT conditions imply the solution to the minimization of (6.272) is

$$Q'_V(v) = c\Big[\sum_{u\in\mathcal{U}} P_U(u)^{1-t}P_{UV}(u,v)^t\Big]^{1/t} \qquad (6.273)$$

with $c^{-1} = \sum_{v\in\mathcal{V}}\big[\sum_{u\in\mathcal{U}} P_U(u)^{1-t}P_{UV}(u,v)^t\big]^{1/t}$. Plugging this into the objective function of (6.272) and substituting $t = 1/(1-\lambda)$, we have

$$\begin{aligned}
&\min_{\substack{Q_{UV}\in\mathcal{P}(\mathcal{U}\times\mathcal{V}):\\ Q_U=P_U}} \{D(Q_{UV}\|P_{VU})-\lambda I(Q_{UV})\}\\
&\quad = -\log\sum_{v\in\mathcal{V}}\Big[\sum_{u\in\mathcal{U}} P_U(u)P_{V|U}(v|u)^{\frac{1}{1-\lambda}}\Big]^{1-\lambda} = -E_0(P_U,P_{V|U},\lambda). \qquad (6.274)
\end{aligned}$$

Plugging (6.274) into (6.264) proves the claim. □

6.C Numerical Evaluation of The Exponents

6.C.1 Computing $E_{\mathrm{s}}^{\text{i.i.d.}}$ and $\underline{E}_{\mathrm{s}}$

Both $E_{\mathrm{s}}^{\text{i.i.d.}}$ and $\underline{E}_{\mathrm{s}}$ can be evaluated via the expressions (6.28) and (6.29) by using the fact that both F_0 and E_0 (defined in (6.28b) and (6.29b), respectively) are convex in λ and pass through the origin with slope $I(U;V)$.

For instance, to evaluate $E_{\mathrm{s}}^{\text{i.i.d.}}$, we know that

- for $R \leq I(U;V)$, the exponent is zero, i.e., $E_{\mathrm{s}}^{\text{i.i.d.}}(P_U,P_{V|U},R) = 0$;

- for $I(U;V) < R < R_c$, where the critical rate R_c is

$$R_c \coloneqq \frac{\partial}{\partial \lambda} F_0(P_{UV}, \lambda)\Big|_{\lambda=1}, \tag{6.275}$$

the pairs $(R, E_{\mathrm{s}}^{\text{i.i.d.}})$ are related parametrically as

$$R(\lambda) = \frac{\partial}{\partial \lambda} F_0(P_{UV}, \lambda) \tag{6.276a}$$
$$E_{\mathrm{s}}^{\text{i.i.d.}}(\lambda) = \lambda R(\lambda) - F_0(P_{UV}, \lambda) \tag{6.276b}$$

for the range of $\lambda \in [0, 1]$;

- finally, if $R \geq R_c$

$$E_{\mathrm{s}}^{\text{i.i.d.}}(P_U, P_{V|U}, R) = R - F_0(P_{UV}, 1). \tag{6.277}$$

It is clear that to evaluate $\underline{E}_{\mathrm{s}}$, we have to replace F_0 with E_0 and follow precisely the same steps.

6.C.2 Computing $E_{\mathrm{s}}^{\text{c.c.}}$

To compute $E_{\mathrm{s}}^{\text{c.c.}}$ (defined in (6.27)), we have to solve two minimizations: that of (6.27a) and that of (6.27b). The second turns out to be efficiently solvable by using standard convex optimization tools.

$$\begin{aligned}\min_{\substack{Q' \in \mathcal{P}(\mathcal{U}\times\mathcal{V}):\\ Q'_U = Q_U, Q'_V = Q_V}} D(Q' \| P) &= \min_{Q' \in \mathcal{P}(\mathcal{U}\times\mathcal{V})} \Bigg\{ D(Q'\|P) + \max_{\substack{\lambda \in \mathbb{R}^{|\mathcal{U}|},\\ \rho \in \mathbb{R}^{|\mathcal{V}|}}} \Bigg\{ \sum_{u \in \mathcal{U}} \lambda_u [Q_U(u) - Q'_U(u)] \\ &\qquad + \sum_{v \in \mathcal{V}} \rho_v [Q_V(v) - Q'_V(v)] \Bigg\}\Bigg\} \qquad (6.278) \\ &= \max_{\substack{\lambda \in \mathbb{R}^{|\mathcal{U}|},\\ \rho \in \mathbb{R}^{|\mathcal{V}|}}} \Bigg\{ \sum_{u \in \mathcal{U}} \lambda_u Q_U(u) + \sum_{v \in \mathcal{V}} \rho_v Q_V(v) \\ &\qquad + \min_{Q'} \Bigg\{ D(Q'\|P) - \sum_{u \in \mathcal{U}} \lambda_u Q'_U(u) - \sum_{v \in \mathcal{V}} \rho_v Q'_V(v) \Bigg\}\Bigg\} \qquad (6.279)\end{aligned}$$

where $\lambda \coloneqq (\lambda_u : u \in \mathcal{U})$ and $\rho \coloneqq (\rho_v : v \in \mathcal{V})$ and the last equality follows since $D(Q\|P)$ is convex in Q and the second term in (6.278) is linear in Q

Moreover, the inner unconstrained minimization in (6.279) can be solved as

$$\min_{Q'}\Big\{D(Q'\|P)-\sum_{u\in\mathcal{U}}\lambda_u Q'_U(u)-\sum_{v\in\mathcal{V}}\rho_v Q'_V(v)\Big\}$$

$$=\min_{Q'}\left\{\sum_{(u,v)\in\mathcal{U}\times\mathcal{V}}Q'(u,v)\log\Big[\frac{Q'(u,v)}{P(u,v)\exp(\lambda_u)\exp(\rho_v)}\Big]\right\}\tag{6.280}$$

$$=-\log\left[\sum_{(u,v)\in\mathcal{U}\times\mathcal{V}}P(u,v)\exp(\lambda_u)\exp(\rho_v)\right],\tag{6.281}$$

by using the concavity of logarithm. The minimum is attained if

$$Q'(u,v)=c\,P(u,v)\exp(\lambda_u)\exp(\rho_v)\tag{6.282}$$

where $c^{-1}=\sum_{u,v}P(u,v)\exp(\lambda_u)\exp(\rho_v)$. Plugging this into (6.279), we get

$$\min_{\substack{Q'\in\mathcal{P}(\mathcal{U}\times\mathcal{V}):\\ Q'_U=Q_U,Q'_V=Q_V}}D(Q'\|P)=\max_{\substack{\lambda\in\mathbb{R}^{|\mathcal{U}|},\\ \rho\in\mathbb{R}^{|\mathcal{V}|}}}\left\{\sum_{u\in\mathcal{U}}\lambda_u Q_U(u)+\sum_{v\in\mathcal{V}}\rho_v Q_V(v)\right.$$
$$\left.-\log\left[\sum_{(u,v)\in\mathcal{U}\times\mathcal{V}}P(u,v)\exp(\lambda_u)\exp(\rho_v)\right]\right\}\tag{6.283}$$

Remark 1. Using Hölder's inequality, it can be verified that the objective function of (6.283) is concave in (λ,ρ), thus can be efficiently maximized using standard numerical methods.

Proof. Since the first two sums in the objective function of (6.283) are linear in (λ,ρ) it is sufficient to prove that the function

$$(\lambda,\rho)\mapsto\log\left[\sum_{(u,v)\in\mathcal{U}\times\mathcal{V}}P(u,v)\exp(\lambda_u)\exp(\rho_v)\right]\tag{6.284}$$

is convex in (λ,ρ). Fix $s\in[0,1]$, $\lambda,\lambda'\in\mathbb{R}^{|\mathcal{U}|}$, and $\rho,\rho'\in\mathbb{R}^{|\mathcal{V}|}$. Then, Hölder's inequality implies

$$\sum_{(u,v)\in\mathcal{U}\times\mathcal{V}}P(u,v)\exp[s\lambda_u+s\rho_v+(1-s)\lambda'_u+(1-s)\rho'_v]$$
$$=\sum_{(u,v)\in\mathcal{U}\times\mathcal{V}}P(u,v)^s\exp[s\lambda_u+s\rho_v]P(u,v)^{1-s}\exp[(1-s)\lambda'_u+(1-s)\rho'_v]\tag{6.285}$$
$$\le\left[\sum_{(u,v)\in\mathcal{U}\times\mathcal{V}}P(u,v)\exp[\lambda_u+\rho_v]\right]^s\left[\sum_{(u,v)\in\mathcal{U}\times\mathcal{V}}P(u,v)\exp[\lambda'_u+\rho'_v]\right]^{1-s}.\tag{6.286}$$

Taking the logarithm of both sides proves the claim. □

Remark 2. It is easy to check that if $Q_U = P_U$ (respectively $Q_V = P_V$), $\lambda = (1, 1, \dots, 1)$ (respectively $\rho = (1, 1, \dots, 1)$) will be a stationary point, and by concavity, the maximizer of (6.283). Therefore, for evaluating $E_{\mathrm{s}}^{\mathrm{c.c.}}$, since we only consider the distributions Q_{UV} with u-marginal $Q_U = P_U$, we already know the optimal λ and need to only optimize over the choices of ρ.

A by-product of the above calculations is a characterization of the support of the minimizing Q' in (6.27b):

Lemma 6.13. *Let*

$$Q^\star = \mathop{\arg\min}_{\substack{Q' \in \mathcal{P}(\mathcal{U}\times\mathcal{V}): \\ Q'_U = Q_U, Q'_V = Q_V}} D(Q' \| P). \tag{6.287}$$

Then, $Q^\star(u, v) = 0$ *if and only if* $P(u, v) = 0$ *or* $Q^\star_U(u) = 0$, *or* $Q^\star_V(v) = 0$.

Proof. The claim follows from (6.282). The optimizing $Q^\star$ is obtained by setting the optimal λ and ρ in (6.282). If $Q(u_0, v_0) = 0$ but $P(u_0, v_0) > 0$, then we must either have $\exp(\lambda_{u_0}) = 0$ or $\exp(\rho_{v_0}) = 0$. The former implies $Q(u_0, v) = 0$ for $\forall v \in \mathcal{V}$, which in turn, implies $Q_U(u_0) = 0$. The latter, similarly, yields $Q_V(v_0) = 0$. □

It remains to solve the minimization of (6.27a) to evaluate $E_{\mathrm{s}}^{\mathrm{c.c.}}$, which can be done by exhaustive search for the small alphabets we considered in § 6.3.2.

6.D Proof of Lemma 6.6

(i) The linearity of expectation shows that

$$P_{V^n}(v^n) = \sum_{u^n \in \mathcal{U}^n} P_{U^n}(u^n) P^n_{V|U}(v^n|u^n), \tag{6.288}$$

is the expectation of the non-negative random variable $P_{\mathscr{C}_n}(v^n)$ (defined in (6.31)). Therefore, $P_{V^n}(v^n) = 0$ implies $P_{\mathscr{C}_n}(v^n) = 0$ almost surely.

(ii) Pick v^n and $\tilde{v}^n$ that have the same type. Therefore, there exists a permutation, call it $\pi\colon \mathcal{V}^n \to \mathcal{V}^n$, such that $\tilde{v}^n = \pi(v^n)$ and $v^n = \pi^{-1}(\tilde{v}^n)$. Then,

$$P_{V^n}(\tilde{v}^n) = \sum_{\tilde{u}^n \in \mathcal{U}^n} P_{U^n}(\tilde{u}^n) P^n_{V|U}(\tilde{v}^n|\tilde{u}^n) \tag{6.289}$$

$$\overset{\text{(a)}}{=} \sum_{u^n \in \mathcal{U}^n} P_{U^n}\big(\pi(u^n)\big) P^n_{V|U}\big(\pi(v^n)|\pi(u^n)\big) \tag{6.290}$$

$$\overset{\text{(b)}}{=} \sum_{u^n \in \mathcal{U}^n} P_{U^n}(u^n) P^n_{V|U}(v^n|u^n) = P_{V^n}(v^n). \tag{6.291}$$

where in (a) we have taken $u^n = \pi^{-1}(\tilde{u}^n)$ and (b) follows since $P_{U^n}(u^n)$ only depends on the type of u^n (and by construction u^n and $\pi(u^n)$ have the same type) and similarly $P^n_{V|U}\big(\pi(v^n) \bigm| \pi(u^n)\big) = P^n_{V|U}(v^n|u^n)$.

(iii) We have

$$P_{V^n}(v^n) = \sum_{u^n \in \mathcal{U}^n} P_{U^n}(u^n) P^n_{V|U}(v^n|u^n). \tag{6.292}$$

$P_{V^n}(v^n) > 0$ implies there exists at least one sequence $u_0^n \in \operatorname{supp}(P_{U^n})$ for which $P^n_{V|U}(v^n|u_0^n) > 0$. Therefore, $P^n_{V|U}(v^n|u_0^n) > (P^{\min}_{V|U})^n$. Thus (6.292) yields

$$P_{V^n}(v^n) \geq P_{U^n}(u_0^n)(P^{\min}_{V|U})^n. \tag{6.293}$$

For i.i.d. random-coding ensemble, $P_{U^n}(u^n) = P^n_U(u^n) \geq (P^{\min}_U)^n$ and for the constant-composition random-coding ensemble, $P_{U^n}(u^n) \geq (1/|\mathcal{U}|)^n$ (since, for any n-type $P \in \mathcal{P}_n(\mathcal{U})$, $\mathcal{T}^n_P \subseteq \mathcal{U}^n$). □

6.E Proof of Lemma 6.7

We prove the claim for unit-mean A. Specifically, we show that

$$c(\theta)\,(\operatorname{var}(A) - \tau_\theta(A)) \leq \mathbb{E}[A \ln(A)] \leq \operatorname{var}(A), \tag{6.294}$$

for any random variable A with $\mathbb{E}[A] = 1$ (where $c(\theta)$ and $\tau_\theta(A)$ are defined in (6.43) and (6.42), respectively). The claim for general A then follows by setting $A' = A/\mathbb{E}[A]$ in the above and noting that

$$\mathbb{E}\Big[A \ln\Big(\frac{A}{\mathbb{E}[A]}\Big)\Big] = \mathbb{E}[A]\,\mathbb{E}[A' \ln(A')] \quad \text{and} \quad \operatorname{var}(A') = \frac{\operatorname{var}(A)}{(\mathbb{E}[A])^2}. \tag{6.295}$$

The upper bound of (6.294) follows as

$$\begin{aligned} \mathbb{E}[A \ln(A)] &= \mathbb{E}[A \ln(A) - (A-1)] && (6.296)\\ &\leq \mathbb{E}[(A-1)^2] = \operatorname{var}(A), && (6.297) \end{aligned}$$

because $a \ln(a) - (a-1) \leq (a-1)^2$ (see Figure 6.5).

To prove the lower bound of (6.294), we have

$$a \ln(a) - (a-1) \geq c(\theta)(a-1)^2 \mathbb{1}\{a \leq \theta + 1\}. \tag{6.298}$$

This follows by observing that $\frac{a\ln(a)-(a-1)}{(a-1)^2}$ is a decreasing function of a (see Lemma 6.14 below or Figure 6.5). Thus,

$$\mathbb{E}[A \ln(A)] \geq c(\theta) \int_0^{\theta+1} (a-1)^2 \mathrm{d}F_A(a). \tag{6.299}$$

where $F_A(a)$ is the cumulative distribution function of a.

Furthermore,

$$\int_0^{\theta+1} (a-1)^2 \mathrm{d}F_A(a) = \operatorname{var}(A) - \int_{\theta+1}^{+\infty} (a-1)^2 \mathrm{d}F_A(a) \tag{6.300}$$

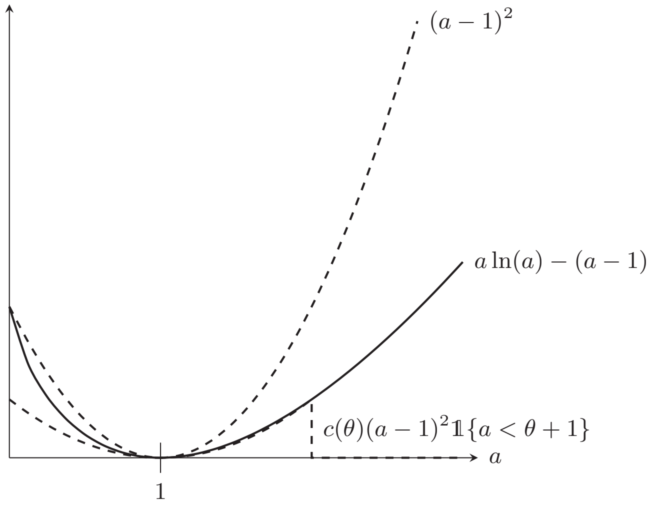


Figure 6.5: Bounds on $a\ln(a)-(a-1)$

Let $\bar{F}_A(a) = \Pr\{A > a\}$ be the complementary distribution function of A. Therefore,

$$\int_{\theta+1}^{+\infty} (a-1)^2 \mathrm{d}F_A(a) = -\int_{\theta+1}^{+\infty} (a-1)^2 \mathrm{d}\bar{F}_A(a) \tag{6.301}$$

$$= \left[-(a-1)^2\bar{F}_A(a)\right]_{\theta+1}^{+\infty} + 2\int_{\theta+1}^{+\infty} (a-1)\bar{F}_A(a)\mathrm{d}a \tag{6.302}$$

$$\overset{(*)}{=} \theta^2\bar{F}_A(\theta+1) + 2\int_{\theta}^{+\infty} t\bar{F}_A(t+1)\mathrm{d}t. \tag{6.303}$$

The equality in $(*)$ follows since we assumed the variance of A exists. Combining (6.303) and (6.300) in (6.299) proves the lower bound of (6.294). □

Lemma 6.14. *For $t \geq 0$,*

(i) the mapping $t \mapsto \dfrac{t\ln(t)-(t-1)}{t-1}$ is increasing in t;

(ii) the mapping $t \mapsto \dfrac{t\ln(t)-(t-1)}{(t-1)^2}$ is decreasing in t.

Proof.

(i)

$$\frac{\partial}{\partial t}\left\{\frac{t\ln(t)-(t-1)}{t-1}\right\} = \frac{(t-1)-\ln(t)}{(t-1)^2} \geq 0, \tag{6.304}$$

since $\ln(t) \leq t-1$.

(ii)

$$\frac{\partial}{\partial t}\Big\{\frac{t\ln(t)-(t-1)}{(t-1)^2}\Big\} = \frac{1}{(t-1)^2}\Big[2-\frac{t+1}{t-1}\ln(t)\Big]. \tag{6.305}$$

The curve $(t+1)\ln(t)$ is convex in t for $t \geq 1$ and concave in t for $t \leq 1$. (This is easy to check as its second derivative is $1/t - 1/t^2$.) Moreover, its tangent line at $t = 1$ is $2(t-1)$. Therefore,

$$(t+1)\ln(t) \leq 2(t-1) \qquad t \leq 1, \tag{6.306}$$

$$(t+1)\ln(t) \geq 2(t-1) \qquad t \geq 1. \tag{6.307}$$

Consequently,

$$\frac{t+1}{t-1}\ln(t) \geq 2, \tag{6.308}$$

hence, the term inside the square brackets in (6.305) is always negative. This proves (ii). □

6.F Proof of Equation (6.48)

Since $P_{U^n}(u^n)$ only depends on the type of u^n,

$$p_Q(v^n) = \sum_{u^n \in \mathcal{U}^n} \mathbb{1}\{(u^n,v^n) \in \mathcal{T}_Q^n\} P_{U^n}(u^n) \tag{6.309}$$

$$= \frac{P_{U^n}(\mathcal{T}_{Q_U}^n)}{|\mathcal{T}_{Q_U}^n|} \sum_{u^n \in \mathcal{U}^n} \mathbb{1}\{(u^n,v^n) \in \mathcal{T}_Q^n\}. \tag{6.310}$$

Furthermore, we have

$$|\mathcal{T}_Q^n| = \sum_{v^n \in \mathcal{V}^n} \sum_{u^n \in \mathcal{U}^n} \mathbb{1}\big\{(u^n,v^n) \in \mathcal{T}_Q^n\big\}. \tag{6.311}$$

The value of the inner sum in (6.311) depends only on the type of v^n (this can be easily checked using the same type of argument as we had in Appendix 6.D part (ii)) and, clearly, is zero if $v^n \notin \mathcal{T}_{Q_V}^n$. Thus

$$|\mathcal{T}_Q^n| = |\mathcal{T}_{Q_V}^n| \mathbb{1}\{v^n \in \mathcal{T}_{Q_V}^n\} \sum_{u^n \in \mathcal{U}^n} \mathbb{1}\big\{(u^n,v^n) \in \mathcal{T}_Q^n\big\}. \tag{6.312}$$

Plugging (6.312) into (6.310) yields (6.48). □

6.G Proof of Equations (6.55)

We prove only (6.55a) (as (6.55b) is trivial). We omit the dependence on v^n throughout the proof for notational brevity.

$$\operatorname{var}(L_1) = \sum_{Q \in \mathcal{Q}'_n} \frac{1}{M^2} \ell(Q)^2 \operatorname{var}(N_Q) + \sum_{\substack{(Q_1,Q_2) \in {\mathcal{Q}'_n}^2: \\ Q_1 \neq Q_2}} \frac{1}{M^2} \ell(Q_1)\ell(Q_2) \operatorname{cov}(N_{Q_1}, N_{Q_2}) \tag{6.313}$$

$$\overset{(*)}{=} \frac{1}{M} \sum_{Q \in \mathcal{Q}'_n} \ell(Q)^2 p_Q (1 - p_Q) - \frac{1}{M} \sum_{\substack{(Q_1,Q_2) \in {\mathcal{Q}'_n}^2 \\ Q_1 \neq Q_2}} \ell(Q_1)\ell(Q_2) p_{Q_1} p_{Q_2}, \tag{6.314}$$

where $(*)$ follows since for the multinomial collection $\{N_Q \colon Q \in \mathcal{P}_n(\mathcal{U} \times \mathcal{V})\}$, $\operatorname{var}(N_Q) = M p_Q (1 - p_Q)$ and $\operatorname{cov}(N_{Q_1}, N_{Q_2}) = -M p_{Q_1} p_{Q_2}$. Moreover,

$$\sum_{\substack{(Q_1,Q_2) \in {\mathcal{Q}'_n}^2: \\ Q_1 \neq Q_2}} \ell(Q_1)\ell(Q_2) p_{Q_1} p_{Q_2} = \sum_{Q_1 \in \mathcal{Q}'_n} \ell(Q_1) p_{Q_1} \sum_{Q_2 \in \mathcal{Q}'_n \setminus \{Q_1\}} \ell(Q_2) p_{Q_2} \tag{6.315}$$

$$= \sum_{Q_1 \in \mathcal{Q}'_n} \ell(Q_1) p_{Q_1} \Big(\mathbb{E}[L_1] - p_{Q_1} \ell(Q_1) \Big). \tag{6.316}$$

Using the above in (6.314), we get

$$\operatorname{var}(L_1) = \frac{1}{M} \sum_{Q \in \mathcal{Q}'_n} \ell(Q) p_Q \Big[(1 - p_Q)\ell(Q) - \big(\mathbb{E}[L_1] - p_Q \ell(Q) \big) \Big] \tag{6.317}$$

$$= \frac{1}{M} \sum_{Q \in \mathcal{Q}'_n} \ell(Q) p_Q \big[\ell(Q) - \mathbb{E}[L_1] \big] \tag{6.318}$$

$$= \frac{1}{M} \sum_{Q \in \mathcal{Q}'_n} \ell(Q)^2 p_Q - \frac{1}{M} \mathbb{E}[L_1]^2. \qquad \square$$

6.H Proof of Equation (6.120)

For $Q \in \mathcal{P}(\mathcal{U} \times \mathcal{V})$, let

$$a(Q) \coloneqq \exp\big\{ -nD(Q \| Q_U \times P_{V|U}) \big\} P_{U^n}\big(\mathcal{T}^n_{Q_U}\big) \min\Big\{ 1, \frac{\ell(Q)}{M} \Big\}, \tag{6.319}$$

for the sake of brevity.

Equation (6.118) immediately implies

$$\mathbb{E}[D(P_{\mathscr{C}_n} \| P_{V^n})] \stackrel{.}{\leq} \sum_{Q \in \mathcal{P}_n(\mathcal{U} \times \mathcal{V})} a(Q). \tag{6.320}$$

It remains to show

$$\mathbb{E}[D(P_{\mathscr{C}_n}\|P_{V^n})] \dot{\geq} \sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} a(Q), \tag{6.321}$$

to establish (6.120).

Recall that (6.118) means there exists a sub-exponentially increasing sequence $\beta(n)$ such that

$$\beta(n)\Big[\mathbb{E}[D(P_{\mathscr{C}_n}\|P_{V^n})] + \frac{\log(e)}{M}\Big] \geq \sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} a(Q) \tag{6.322}$$

Equation (6.115) implies $\forall Q \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})$,

$$\ell(Q) \geq \exp\big(n\omega(Q)\big)\big|\mathcal{T}^n_{Q_V}\big|. \tag{6.323}$$

Using the above and (6.3) we can verify that, $\forall Q \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V})$,

$$\ell(Q) \geq \exp\{n\omega(Q)\}\big|\mathcal{T}^n_{Q_V}\big| \tag{6.324}$$

$$\geq (n+1)^{-|\mathcal{V}|}\exp\{n[\omega(Q)+H(Q_V)]\} \tag{6.325}$$

$$= (n+1)^{-|\mathcal{V}|}\exp\{n[I(Q)-D(Q\|Q_U\times P_{V|U})]\}. \tag{6.326}$$

Let $(\hat{P}^{(n)}_{UV} \in \mathcal{P}_n(\mathcal{U}\times\mathcal{V}), n\in\mathbb{N})$ be a sequence of n-types that converges to P_{UV} and, for the constant composition ensemble, has the u-marginal $\hat{P}^{(n)}_U = P^{(n)}_U$. (The existence of such a sequence is guaranteed by Lemma C.1 or Lemma C.2 in Appendix C.) Thus, in particular, for every $\epsilon > 0$ there exists $n_0(\epsilon)$ such that $\forall n \geq n_0(\epsilon)$,

$$I\big(\hat{P}^{(n)}_{UV}\big) \geq I(U;V) - \epsilon/2, \tag{6.327a}$$

$$D\big(\hat{P}^{(n)}_{UV} \,\big\|\, \hat{P}^{(n)}_U \times P_{V|U}\big) \leq \epsilon/2, \qquad \text{and} \tag{6.327b}$$

$$P_{U^n}\Big(\mathcal{T}_{\hat{P}^{(n)}_U}\Big) \geq \exp\{-n\epsilon/2\}. \tag{6.327c}$$

Using (6.327a) and (6.327b) in (6.326) yields

$$\ell\big(\hat{P}^{(n)}_{UV}\big) \geq (n+1)^{-|\mathcal{V}|}\exp\{n(I(U;V)-\epsilon)\}, \tag{6.328}$$

which, in turn, implies

$$\min\Big\{1, \frac{\ell(\hat{P}^{(n)}_{UV})}{M}\Big\} \geq (n+1)^{-|\mathcal{V}|}\exp\{-n[R-I(U;V)+\epsilon]^+\}. \tag{6.329}$$

As a consequence, taking (6.327c) into account, $a(\hat{P}^{(n)}_{UV})$ is lower-bounded as

$$\begin{aligned} a\big(\hat{P}^{(n)}_{UV}\big) &= \exp\{-nD(\hat{P}^{(n)}_{UV}\|\hat{P}^{(n)}_U\times P_{V|U})\}P_{U^n}\big(\mathcal{T}^n_{\hat{P}^{(n)}_U}\big)\min\Big\{1, \frac{\ell(\hat{P}^{(n)}_{UV})}{M}\Big\} \\ &\geq (n+1)^{-|\mathcal{V}|}\exp\{-n(\epsilon+[R-I(U;V)+\epsilon]^+)\} \end{aligned} \tag{6.330}$$

We can verify that, setting

$$\epsilon := \min\{R/2, I(U;V)/3\} > 0, \tag{6.331}$$

yields

$$\epsilon + [R - I(U;V) + \epsilon]^+ = R - \epsilon. \tag{6.332}$$

Consequently, using this value of ϵ, (6.330) implies

$$a\big(\hat{P}_{UV}^{(n)}\big) \geq (n+1)^{-|\mathcal{V}|} \exp\{-n[R-\epsilon]\} \tag{6.333}$$

Obviously, $\exists n_1$ such that $\forall n \geq n_1$,

$$\begin{aligned}\beta(n)\frac{\log(e)}{M} &\leq 2\beta(n)\log(e)\exp(-nR)\\ &\leq \frac{1}{2}(n+1)^{-|\mathcal{V}|}\exp\{-n[R-\epsilon]\}.\end{aligned} \tag{6.334}$$

(The first inequality follows since $M = \lfloor \exp(nR) \rfloor \geq \exp(nR)/2$.) Hence, for $n \geq n_2 := \max\{n_0, n_1\}$,

$$\beta(n)\frac{\log(e)}{M} \leq \frac{1}{2}a\big(\hat{P}_{UV}^{(n)}\big) \tag{6.335}$$

Since $\forall Q \in \mathcal{P}(\mathcal{U}\times\mathcal{V})$, $a(Q) \geq 0$, for $n \geq n_2$, using (6.322) we have,

$$\begin{aligned}\beta(n)\Big[\mathbb{E}[D(P_{\mathscr{C}_n}\|P_{V^n})] + \frac{\log(e)}{M}\Big] &\geq \sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} a(Q) && (6.336)\\ &= \sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})\setminus\{\hat{P}_{UV}^{(n)}\}} a(Q) + a\big(\hat{P}_{UV}^{(n)}\big) && (6.337)\\ &\geq \frac{1}{2}\sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})\setminus\{\hat{P}_{UV}^{(n)}\}} a(Q) + a\big(\hat{P}_{UV}^{(n)}\big) && (6.338)\\ &= \frac{1}{2}\sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} a(Q) + \frac{1}{2}a\big(\hat{P}_{UV}^{(n)}\big) && (6.339)\\ &\overset{(*)}{\geq} \frac{1}{2}\sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} a(Q) + \beta(n)\frac{\log(e)}{M} && (6.340)\end{aligned}$$

where $(*)$ follows from (6.335).

Therefore, for $n \geq n_2$,

$$\sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} a(Q) \leq 2\beta(n)\,\mathbb{E}[D(P_{\mathscr{C}_n}\|P_{V^n})] \tag{6.341}$$

Taking

$$\beta'(n) := \begin{cases} +\infty & \text{if } n < n_2\\ 2\beta(n) & \text{otherwise,}\end{cases} \tag{6.342}$$

yields,

$$\sum_{Q\in\mathcal{P}_n(\mathcal{U}\times\mathcal{V})} a(Q) \leq \beta'(n)\,\mathbb{E}[D(P_{\mathscr{C}_n}\|P_{V^n})], \qquad \forall n\in\mathbb{N}. \tag{6.343}$$

Moreover, we have

$$\limsup_{n\to\infty}\frac{1}{n}\log\beta'(n) = \limsup_{n\to\infty}\frac{1}{n}\log\beta(n) = 0, \tag{6.344}$$

by assumption, and that β' depends only on the fixed parameters of the problem (because n_2 depends only on these parameters). Therefore, (6.343) establishes (6.321) and concludes the proof. □

6.I Proof of Lemma 6.9

We first prove that, with $r = \operatorname{rank}(\Omega) \leq m$,

$$\Pr\left(\bigcap_{j=1}^{m}\left\{U^n_{w^k_{(j)}} = u^n_{(j)}\right\}\right) = \Pr\left(\bigcap_{j=1}^{r}\left\{U^n_{w^k_{(j)}} = u^n_{(j)}\right\}\right)\mathbb{1}\{\operatorname{kern}(\Omega)\subseteq\operatorname{kern}(\Xi)\} \tag{6.345}$$

If $\operatorname{kern}(\Omega)\not\subseteq\operatorname{kern}(\Xi)$, the event of interest on the left-hand side of (6.345) has probability 0: If there exists a non-zero $(a_1,a_2,\ldots,a_m)\in\mathbb{F}_q^m$ such that $\Omega\cdot[a_1,a_2,\ldots,a_m]^T = 0$ but $\Xi\cdot[a_1,a_2,\ldots,a_m]^T \neq 0$,

$$\sum_{j=1}^{m} a_j U^n_{w^k_{(j)}} = \sum_{j=1}^{m} a_j \begin{bmatrix} G & D^n\end{bmatrix}\begin{bmatrix} w^n_{(j)} \\ 1\end{bmatrix} \tag{6.346}$$

$$= \begin{bmatrix} G & D^n\end{bmatrix}\sum_{j=1}^{m} a_j \begin{bmatrix} w^n_{(j)} \\ 1\end{bmatrix} = 0 \tag{6.347}$$

whereas $\bigcap_{j=1}^{m}\left\{U^n_{w^k_{(j)}} = u^n_{(j)}\right\}$ implies

$$\sum_{j=1}^{m} a_j U^n_{w^k_{(j)}} = \sum_{j=1}^{m} a_j u^n_{(j)} \neq 0. \tag{6.348}$$

When $\operatorname{kern}(\Omega)\subseteq\operatorname{kern}(\Xi)$,

$$\bigcap_{j=1}^{m}\left\{U^n_{w^k_{(j)}} = u^n_{(j)}\right\} = \bigcap_{j=1}^{r}\left\{U^n_{w^k_{(j)}} = u^n_{(j)}\right\}. \tag{6.349}$$

To prove (6.349), we show that $\bigcap_{j=1}^{r}\left\{U^n_{w^k_{(j)}} = u^n_{(j)}\right\}$, together with the fact that $\operatorname{kern}(\Omega)\subseteq\operatorname{kern}(\Xi)$, implies $\forall i = r+1,\ldots,m$, $\left\{U^n_{w^k_{(i)}} = u^n_{(i)}\right\}$.

Fix $i \in \{r+1, \dots, m\}$. Since $\operatorname{rank}(\Omega) = r$, there exists $(a_1, a_2, \dots, a_r) \in \mathbb{F}_q^r$ for which

$$\begin{bmatrix} w_{(i)}^n \\ 1 \end{bmatrix} = \sum_{j=1}^{r} a_j \begin{bmatrix} w_{(j)}^n \\ 1 \end{bmatrix} \quad \Longleftrightarrow \quad \sum_{j=1}^{r} a_j \begin{bmatrix} w_{(j)}^n \\ 1 \end{bmatrix} - \begin{bmatrix} w_{(i)}^n \\ 1 \end{bmatrix} = 0. \tag{6.350}$$

Therefore,

$$\begin{aligned}
U_{w_{(i)}^k}^n &= G w_{(i)}^k + D^n && (6.351)\\
&= \begin{bmatrix} G & D^n \end{bmatrix} \begin{bmatrix} w_{(i)}^k \\ 1 \end{bmatrix} && (6.352)\\
&= \begin{bmatrix} G & D^n \end{bmatrix} \sum_{j=1}^{r} a_j \begin{bmatrix} w_{(j)}^n \\ 1 \end{bmatrix} && (6.353)\\
&= \sum_{j=1}^{r} a_j \begin{bmatrix} G & D^n \end{bmatrix} \begin{bmatrix} w_{(j)}^n \\ 1 \end{bmatrix} && (6.354)\\
&= \sum_{j=1}^{r} a_j U_{w_{(j)}^k}^n && (6.355)\\
&\overset{(*)}{=} \sum_{j=1}^{r} a_j u_{(j)}^n && (6.356)
\end{aligned}$$

where $(*)$ follows by the assumption that $\bigcap_{j=1}^{r} \{U_{w_{(j)}^k}^n = u_{(j)}^n\}$. Moreover, since $\operatorname{kern}(\Omega) \subseteq \operatorname{kern}(\Xi)$, (6.350) implies

$$\sum_{j=1}^{r} a_j u_{(j)}^n - u_{(i)}^n = 0 \quad \Longleftrightarrow \quad \sum_{j=1}^{r} a_j u_{(j)}^n = u_{(i)}^n \tag{6.357}$$

Using the above in (6.356) shows

$$U_{w_{(i)}^k}^n = u_{(i)}^n. \tag{6.358}$$

Hence, it remains to compute

$$\begin{aligned}
&\Pr\left(\bigcap_{j=1}^{r} \{U_{w_{(j)}^k}^n = u_{(j)}^n\} \right)\\
&= \Pr\left\{ \begin{bmatrix} G & D^n \end{bmatrix} \begin{bmatrix} w_{(1)}^k & w_{(2)}^k & \cdots & w_{(r)}^k \\ 1 & 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} u_{(1)}^n & u_{(2)}^n & \cdots & u_{(r)}^n \end{bmatrix} \right\}. \qquad (6.359)
\end{aligned}$$

Since

$$A := \begin{bmatrix} w_{(1)}^k & w_{(2)}^k & \cdots & w_{(r)}^k \\ 1 & 1 & \dots & 1 \end{bmatrix} \in \mathbb{F}_q^{(k+1)\times r}$$

is a full rank matrix, given any $B \in \mathbb{F}_q^{n\times r}$, there are $q^{n\times(k-r+1)}$ choices of $X \in \mathbb{F}_q^{n\times(k+1)}$ that satisfy

$$X \cdot A = B. \tag{6.360}$$

This holds because we can split $A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$ where $A_1 \in \mathbb{F}_q^{r\times r}$ and $A_2 \in \mathbb{F}_q^{(k+1-r)\times r}$ and $X = \begin{bmatrix} X_1 & X_2 \end{bmatrix}$ where $X_1 \in \mathbb{F}_q^{n\times r}$ and $X_2 \in \mathbb{F}_q^{n\times(k+1-r)}$; and rewrite (6.360) as

$$X \cdot A = X_1 \cdot A_1 + X_2 \cdot A_2 = B \tag{6.361}$$

Since A_1 is a full-rank square matrix, given any $X_2 \in \mathbb{F}_q^{n\times(k+1-r)}$, we can uniquely determine X_1 that satisfies

$$X_1 \cdot A_1 = B - X_2 \cdot A_2.$$

As there are $q^{n\times(k-1+r)}$ choices for X_2, the linear system of (6.360) has $q^{n\times(k-1+r)}$ solutions.

The matrix $\begin{bmatrix} G & D^n \end{bmatrix}$ is uniformly distributed on $\mathbb{F}_q^{n\times(k+1)}$. Given the above considerations, out of all $q^{n\times(k+1)}$ choices of $\begin{bmatrix} G & D^n \end{bmatrix}$, $q^{n\times(k+1-r)}$ satisfy the equality on the right-hand side of (6.359). Therefore,

$$\Pr\left\{ \begin{bmatrix} G & D^n \end{bmatrix} \begin{bmatrix} w_{(1)}^k & w_{(2)}^k & \cdots & w_{(r)}^k \\ 1 & 1 & \dots & 1 \end{bmatrix} = \begin{bmatrix} u_{(1)}^n & u_{(2)}^n & \cdots & u_{(r)}^n \end{bmatrix} \right\} = q^{-r}. \tag{6.362}$$

which, together with (6.359) concludes the proof. □

7 Channel Resolvability in the Presence of Feedback

In Chapter 5, we have seen that channel resolvability is a powerful and convenient tool for establishing secrecy. Subsequently, in Chapter 6 we have derived ensemble-optimal resolvability exponents for the ensembles of i.i.d. and constant-composition random codes. The application of channel resolvability is not limited to information theoretic secrecy. In [117], Wyner developed his *soft covering lemma* — stating that the distribution induced at the output of a discrete memoryless channel when its input is a uniformly chosen codeword from a randomly constructed code is arbitrarily close to the i.i.d. measure — to prove achievability results on the *common information* of two random variables. Common information is defined as the minimum rate of common randomness that must be shared between two encoders (behaving, otherwise, independently) so that they can generate two correlated output sequences. Along the same lines, channel resolvability is also the building block for achievability results in *coordination* and *distributed channel synthesis*; the goal in such settings is to quantify the minimum communication rates required among different nodes of a network so that they can locally generate sequences that are jointly correlated (as if they were drawn from a prescribed joint distribution) [31, 33].

Channel resolvability and error correction are, in a sense, closely interconnected problems. In error correction, the aim is to *combat* the channel randomness so that, upon observing the output signal, the sent codeword is still distinguishable from the others, despite the noise added by the channel to the input signal. In channel resolvability, we *exploit* channel randomness so that, when it is combined with external randomness, the output sequence of the channel looks as if it were generated by a random-number generator.

In this chapter, we consider the problem of channel resolvability in the presence of causal feedback; specifically when the encoder of Figure 5.2b observes the previously received symbols $\tilde{V}^{i-1}$ before producing the i^{th} channel input symbol $\tilde{U}_i$. Hence, the encoder has the opportunity to decide about the value of $\tilde{U}_i$ based on the past behavior of the channel (see Figure 7.1). In this setting, the encoder can not be specified as a single deterministic function $\mathrm{Enc}\colon \{1, 2, \dots, M\} \to \mathcal{U}^n$. Instead, it is described by a *collection* of deterministic functions

$$\big(\mathrm{Enc}_i\colon \{1, 2, \dots, M\} \times \mathcal{V}^{i-1} \to \mathcal{U}, \quad i \in \{1, 2, \dots, n\}\big). \tag{7.1}$$

The function Enc_i determines the i^{th} channel input symbol $\tilde{U}_i$ as a function of the external randomness J and previously received symbols $\tilde{V}^{i-1}$.

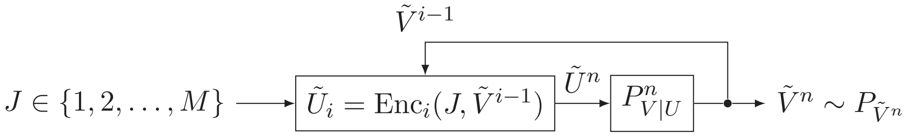

$\tilde{V}^{i-1}$
$J \in \{1, 2, \dots, M\}$
$\tilde{U}_i = \mathrm{Enc}_i(J, \tilde{V}^{i-1})$
$\tilde{U}^n$
$P^n_{V|U}$
$\tilde{V}^n \sim P_{\tilde{V}^n}$

Figure 7.1: Channel Resolvability in the Presence of Feedback

For its counterpart problem, i.e., channel coding, the utility of feedback is well studied in the literature. We know that feedback does not increase the capacity of the channel [42, Exercise 4.6], rather it enables the construction of higher-quality error-correction schemes. It was shown by Burnashev [23] that, in the presence of feedback (and using variable-length error-correction schemes), larger error exponents are achievable.

Given the analogy between resolvability and error correction, it is natural to wonder if similar advantages exist in the presence of feedback for channel resolvability? In this chapter, we study this problem and show that, although feedback does not reduce the resolution of the channel (see Theorem 7.1), it indeed enables us to improve the quality of distribution approximation at the output of the channel. We consider specifically two examples of binary erasure and binary symmetric channel and propose resolvability encoders that yield much larger resolvability exponents, compared to the achievable exponents of Chapter 6 (see Theorems 7.2 and 7.4 in § 7.3).

As we will see, the canonical benefit of feedback is to allow the encoder to *adapt*, based on the channel behavior, the rate of external randomness it consumes. This is done via variable-length resolvability schemes. Variable-length coding is necessary for achieving the exponents of Theorems 7.2 and 7.4: In Lemmas 7.3 and 7.5, we show that no block code can achieve the exponents of Theorems 7.2 and 7.4. We restrict our analysis to approximating an i.i.d. sequence at the output of channel $P_{V|U}$, i.e., resolvability with respect to the sequence of product measures ($P_{V^n} = P^n_V$, $n \in \mathbb{N}$). Apart from being the most widely used choice of the sequence of reference measures in prevalent applications of resolvability (particularly, in applications other than secrecy), such a restriction permits us to define fixed-to-variable-length resolvability encoders. Specifically, resolvability encoders that, instead of generating a fixed number of channel input symbols, given the external randomness and channel behavior, determine the length of channel input (and hence output) sequence, based on a stopping rule.

The results presented in this chapter were published in part in [14].

7.1 Variable-Length Resolvability Codes

The classic channel-resolvability problem is defined based on block codes: The goal is to make the distribution of a length-n block of the output $P_{\tilde{V}^n}$ close to i.i.d. P_V^n by mapping the input message $J \in \{1, 2, \ldots, M\}$ to a channel input sequence of length n, $\tilde{U}^n$ (in a deterministic manner) that will, in turn, be observed as $\tilde{V}^n$ at the output of the channel. It is useful to extend this notion to variable-length codes.

7.1.1 Variable-to-Fixed-Length Resolvability Codes

Instead of mapping a single input word J to n channel input symbols, the encoder can decide on how much randomness to consume, based on the channel behavior, by mapping a variable number of input words to n channel input symbols. This flexibility permits the encoder to begin the transmission upon reading a small number of random bits and then, based on the noise the channel injects to the system, to decide if more randomness is required for producing the subsequent symbols or if they can be computed as a deterministic function of the feedback signal and the randomness already available to the encoder. Such a variable-to-fixed-length resolvability encoder is formally defined as follows:

Definition 7.1. A *variable-to-fixed-length* resolvability encoder of block-length n for system with feedback, maps a variable number T of input words $J_1, J_2, \ldots, J_T$ into a sequence of length n of channel input symbols $\tilde{U}^n$. In its most generic form, such an encoder is defined via two collection of functions: A collection of *shift* functions,

$$\mathrm{Shift}_{i,t} \colon \{1, 2, \ldots, M\}^t \times \mathcal{V}^{i-1} \to \{0, 1\}, \qquad t \in \mathbb{N}, i = 1, 2, \ldots, n, \tag{7.2}$$

and a collection of encoding functions

$$\mathrm{Enc}_{i,t} \colon \{1, 2, \ldots, M\}^t \times \mathcal{V}^{i-1} \to \mathcal{U}, \qquad t \in \mathbb{N}, i = 1, 2, \ldots, n. \tag{7.3}$$

The encoder maps the input words $J_1, J_2, \ldots$ to $\tilde{U}^n$ using the following procedure:

Algorithm 5: Variable-to-Fixed-Length Resolvability Encoder

```
Input: $J_1, J_2, \ldots$, i.i.d. uniformly distributed on $\{1, 2, \ldots, M\}$
$t \leftarrow 1$;
for $i = 1$ to $n$ do
    while $\mathrm{Shift}_{i,t}(J^t, \tilde{V}^{i-1}) = 1$ do
        $t \leftarrow t + 1$;
    $\tilde{U}_i \leftarrow \mathrm{Enc}_{i,t}(J^t, \tilde{V}^{i-1})$;
    Transmit $\tilde{U}_i$ via the channel and record the feedback signal;
```

In other words, at any time instant $i \in \{1, 2, \ldots, n\}$, given the input words read so far $J^t = (J_1, J_2, \ldots, J_t)$ and the feedback signal $\tilde{V}^{i-1}$, the 'shift' function $\mathrm{Shift}_{i,t}$ decides whether to read a new word J_{t+1} from the input or not. Afterwards, the encoding function $\mathrm{Enc}_{i,t}$ produces the i^{th} channel input symbol $\tilde{U}_i$ as a function of the input words J^t and the feedback signal.

Remark. We can define a variable-to-fixed-length code for the system without feedback[1] by dropping the dependence of the encoding functions on the feedback signal. It is, however, not difficult to see that such a system can be reduced to a block resolvability code with an encoder that picks a codeword from the codebook (and transmits it through the channel) with a *non-uniform* distribution.

The performance of a variable-to-fixed-length resolvability code is measured by the output divergence $D(P_{\tilde{V}^n} \| P_V^n)$ and the average rate

$$R \coloneqq \frac{\mathbb{E}[T] \log(M)}{n}. \tag{7.4}$$

Indeed, by the law of large numbers, when the system is run many times, it consumes an average entropy rate of R bits per output symbol. Definitions 5.3 and 5.5 are extended straightforwardly (by replacing the rate R as defined in (7.4)) to variable-to-fixed-length codes.

7.1.2 Fixed-to-Variable-Length Resolvability Codes

An alternative strategy for adapting the entropy rate consumed by the encoder is to devise a fixed-to-variable-length resolvability encoder. Such an encoder takes a single message $J \in \{1, 2, \ldots, M\}$ as input but, based on a stopping rule, is allowed to produce as many channel input symbols as desired. The advantage of such a construction is that, when feedback is present, the encoder can adapt the code rate by amortizing the entropy of J over as many channel uses as possible; if the channel is too noisy, its output sequence has little correlation

[1] Variable-to-fixed-length resolvability codes for the system without feedback are also proposed independently and concurrently in [120], where their corresponding achievable rates over general (not necessary stationary nor memoryless) channels are studied using information spectrum methods.

with its input (which, in turn is a function of J), hence an observer sensing its output can detect that a sequence of deterministically constructed symbols are transmitted via the channel only after seeing a large number of output symbols. A fixed-to-variable-length resolvability code is, formally, defined as follows:

Definition 7.2. A *fixed-to-variable-length* resolvability code of size M for the system with feedback over the input and output alphabets $(\mathcal{U}, \mathcal{V})$ is defined via a collection of deterministic encoding functions

$$\mathrm{Enc}_n\colon \{1, 2, \ldots, M\} \times \mathcal{V}^{n-1} \to \mathcal{U} \cup \{\mathtt{STOP}\}, \qquad n \in \mathbb{N}, \tag{7.5}$$

where $\mathtt{STOP} \notin \mathcal{U}$ is a special symbol indicating the "end of transmission." Given the input word J and the past channel output symbols V^n, the encoding function Enc_{n+1} decides either to feed the channel with another input symbol in $\mathcal{U}$ (and increase the number of produced symbols to $n+1$) or to stop the encoding by outputting $\mathtt{STOP}$.

Using the above-mentioned collection of encoding functions, a resolvability encoder maps the input word J into a sequence of channel input symbols $\tilde{U}_1, \tilde{U}_2, \ldots$ using the following procedure:

Algorithm 6: Fixed-to-Variable-Length Resolvability Encoder

```
Input: $J$ uniformly distributed on $\{1, 2, \ldots, M\}$
$n \leftarrow 0$;
while $\mathrm{Enc}_{n+1}(J, \tilde{V}^n) \neq \mathtt{STOP}$ do
    $\tilde{U}_{n+1} \leftarrow \mathrm{Enc}_{n+1}(J, \tilde{V}^n)$;
    Transmit $\tilde{U}_{n+1}$ via the channel;
    $n \leftarrow n + 1$;
```

Remark. We can define a fixed-to-variable-length resolvability code for the system without feedback by removing the dependence of the encoding functions on the feedback signal. In this case, the system is equivalent to a deterministic encoder that maps the input words to codewords of variable lengths.

When a fixed-to-variable-length feedback resolvability encoder is employed, the *stopping time* of the encoder (hence the length of the channel output corresponding to a single run of the encoder) will be a random variable, which we denote by N. The stopping time depends both on the channel randomness and the randomness of the input word J:

$$N = \inf\{n \geq 0\colon\ \mathrm{Enc}_{n+1}(J, \tilde{V}^n) = \mathtt{STOP}\}. \tag{7.6}$$

Extending Definition 5.3, we say a rate R is an achievable resolvability rate with respect to the sequence of product measures $(P_V^n, n \in \mathbb{N})$, if for every $\epsilon > 0$ there exists a fixed-to-variable-length resolvability code as in Definition 7.1, with (possibly large) message size M, whose rate is at most R,

$$\frac{\log(M)}{\mathbb{E}[N]} \leq R, \tag{7.7}$$

and results in an expected output divergence of at most ϵ, i.e.,

$$\mathbb{D} := \sum_{n\geq 0} D(P_{\tilde{V}^n|N=n} \| P_V^n) \Pr\{N = n\} \leq \epsilon \tag{7.8}$$

Again, by the law of large numbers, when the scheme is run many times (each time producing a block of channel output symbols), the output sequence will have an average length of $\mathbb{E}[N]$ symbols per block. Hence, the system has consumed an average entropy rate of $\log(M)/\,\mathbb{E}[N]$ per output symbol. And the divergence between the distribution of the output string and the product distribution normalized by the number of blocks will be close to $\mathbb{D}$.

Likewise, following Definition 5.5, we call $E(R)$ an achievable resolvability exponent with respect to the sequence of product measures $(P_V^n, n \in \mathbb{N})$ if there exists a fixed-to-variable-length resolvability code of (possibly large) message size M, and rate at most R (cf. (7.7)) that yields

$$\mathbb{D} \leq \exp\bigl\{-\,\mathbb{E}[N](E(R) - \delta)\bigr\}. \tag{7.9}$$

Remark. Our choice of quality measure (7.8) is justified as follows. Consider successive independent runs of a classic (block) resolvability encoder. After B runs, the output sequence $\tilde{V}^{nB}$ is the concatenation of B sequences of length n,

$$\tilde{V}^{nB} = (\tilde{V}^n_{(1)}, \tilde{V}^n_{(2)}, \ldots, \tilde{V}^n_{(B)}) \tag{7.10}$$

where $\tilde{V}^n_{(t)}$ corresponds to the t^{th} run of the encoder. Since the successive executions are assumed to be independent, it is easy to see that

$$D(P_{\tilde{V}^{nB}} \| P_V^{nB}) = \sum_{t=1}^{B} D\bigl(P_{\tilde{V}^n_{(t)}} \parallel P_V^n\bigr). \tag{7.11}$$

Suppose an observer attempts to detect if an encoder is used at the input of the channel or a random-number generator drawing i.i.d. symbols from P_U. Any test that reliably identifies coded transmission will misidentify the transmission of i.i.d. P_U input symbols as coded transmission with probability $\approx \exp[-D(P_{\tilde{V}^{nB}} \| P_V^{nB})]$ [70, Chapter 3]. Consequently, unless the code leads to a perfect i.i.d. output sequence, the accumulated divergence between the distribution of the output sequence and the product measure increases with B and the observer will eventually be able to implement a decent classifier — no matter how powerful of a resolvability code we use. However, when the code guarantees that $D(P_{\tilde{V}^n} \| P_V^n) \leq \epsilon$ (over a single run), the per-block output divergence,

$$\frac{1}{B} D(P_{\tilde{V}^{nB}} \| P_V^{nB}) = \frac{1}{B} \sum_{t=1}^{B} D\bigl(P_{\tilde{V}^n_{(t)}} \parallel P_V^n\bigr) \tag{7.12}$$

will, indeed, be smaller than ϵ since each of the summands in the above sum are smaller than ϵ.

For a fixed-to-variable-length scheme, the fact that $\mathbb{D}$ is small (with $\mathbb{D}$ as defined in (7.8)) gives the same guarantee: when the encoder is run B times in a row (for large B) the accumulated divergence between the distribution of the output sequence and the product measure, normalized by the number of runs, B, will be close to $\mathbb{D}$ hence small.

7.2 Channel Resolution in the Presence of Feedback

Theorem 7.1. *In the presence of feedback, the resolution of the channel $P_{V|U} : \mathcal{U} \to \mathcal{V}$ with respect to the sequence of i.i.d. reference measures $(P_V^n, n \in \mathbb{N})$ is*

$$\min_{\substack{(\tilde{U},\tilde{V}):\\ P_{\tilde{V}|\tilde{U}}=P_{V|U}\\ P_{\tilde{V}}=P_V}} I(\tilde{U};\tilde{V}). \tag{7.13}$$

The above holds, even if a rate adaptation strategy (using variable-length codes) is employed.

In other words, feedback does not reduce the resolution of a channel. This can easily be explained by looking at the entropy rate as a resource: Neither feedback nor rate adaptation strategies add entropy rate to the system. Therefore, the difference between what we need, $H(V)$, and what the channel supplies, $H(V|U)$, must be offset in order to accurately approximate the product measure at the channel output.

We devote the rest of this section to proving Theorem 7.1. Clearly, we only need to give a converse proof, as the achievability proof follows from the achievability of rates below the channel's resolution without feedback. To this end, we first prove a converse result for block codes in the presence of feedback. Subsequently, to extend the proof to variable-length schemes, we show that based on every (good) variable-length code, we can construct a (good) block code of essentially the same rate. Thus, the converse for variable-length codes follows from the converse for block codes. We will also prove the converse under *weak* resolvability criteria. That is to say, we assume the code guarantees only a small *normalized* divergence between the output distribution and the product measure. For simplicity, and without loss of generality, throughout the proof we assume $\operatorname{supp}(P_V) = \mathcal{V}$.

7.2.1 Converse for Block Codes

Fix $\epsilon > 0$ and consider any resolvability encoder that maps uniformly distributed $J \in \{1, 2, \ldots, M\}$ to a channel input sequence $\tilde{U}^n \in \mathcal{U}^n$ and results in

$$\frac{1}{n} D(P_{\tilde{V}^n} \| P_V^n) \leq \epsilon, \tag{7.14}$$

where $\tilde{V}^n$ is the output sequence of n-fold use of $P_{V|U}$ to the input $\tilde{U}^n$.

We have

$$\frac{\log(M)}{n} = \frac{1}{n}H(J) \geq \frac{1}{n}I(J;\tilde{V}^n) \tag{7.15}$$

$$= \frac{1}{n}\Big[H(\tilde{V}^n) - \sum_{i=1}^{n} H(\tilde{V}_i|J,\tilde{V}^{i-1})\Big] \tag{7.16}$$

$$\overset{\text{(a)}}{=} \frac{1}{n}\Big[H(\tilde{V}^n) - \sum_{i=1}^{n} H(\tilde{V}_i|J,\tilde{V}^{i-1},\tilde{U}_i)\Big] \tag{7.17}$$

$$\overset{\text{(b)}}{=} \frac{1}{n}\Big[H(\tilde{V}^n) - \sum_{i=1}^{n} H(\tilde{V}_i|\tilde{U}_i)\Big]. \tag{7.18}$$

In the above, (a) follows since $\tilde{U}_i$ is a deterministic function of $\tilde{V}^{i-1}$ (the feedback signal) and J (the external randomness) and (b) because the channel is assumed to be memoryless, thus $(J,\tilde{V}^{i-1}) \multimap \tilde{U}_i \multimap \tilde{V}_i$. At this point the problem is reduced to proving the converse for the system without feedback, and we can proceed as in [117, Proof of Theorem 5.2].

It is easy to verify that

$$D(P_{\tilde{V}^n}\|P_V^n) = D\Big(P_{\tilde{V}^n}\,\Big\|\,\prod_{i=1}^{n} P_{\tilde{V}_i}\Big) + \sum_{i=1}^{n} D(P_{\tilde{V}_i}\|P_V) \tag{7.19}$$

Since both terms on the right-hand side of (7.19) are positive, (7.14) implies

$$\frac{1}{n}\sum_{i=1}^{n} D(P_{\tilde{V}_i}\|P_V) \leq \epsilon, \tag{7.20}$$

and

$$\frac{1}{n}D\Big(P_{\tilde{V}^n}\,\Big\|\,\prod_{i=1}^{n} P_{\tilde{V}_i}\Big) = \frac{1}{n}\Big[\sum_{i=1}^{n} H(\tilde{V}_i) - H(\tilde{V}^n)\Big] \leq \epsilon \tag{7.21}$$

Therefore, continuing (7.18), we have

$$\frac{\log(M)}{n} \geq \frac{1}{n}\Big[H(\tilde{V}^n) - \sum_{i=1}^{n} H(\tilde{V}_i|\tilde{U}_i)\Big] \tag{7.22}$$

$$\geq \frac{1}{n}\Big[\sum_{i=1}^{n} H(\tilde{V}_i) - \sum_{i=1}^{n} H(\tilde{V}_i|\tilde{U}_i)\Big] - \epsilon. \tag{7.23}$$

The uniform continuity of entropy [30, Lemma 2.7] implies

$$|H(\tilde{V}_i) - H(V)| \leq |P_{\tilde{V}_i} - P_V| \log\Big[\frac{|\mathcal{V}|}{|P_{\tilde{V}_i} - P_V|}\Big]. \tag{7.24}$$

Hence, as $\theta \mapsto \theta \log \frac{|\mathcal{V}|}{\theta}$ is concave in θ,

$$\frac{1}{n}\sum_{i=1}^{n} |H(\tilde{V}_i) - H(V)| \le \frac{1}{n}\sum_{i=1}^{n} |P_{\tilde{V}_i} - P_V| \log\left[\frac{|\mathcal{V}|}{|P_{\tilde{V}_i} - P_V|}\right] \tag{7.25}$$

$$\le \left[\frac{1}{n}\sum_{i=1}^{n} |P_{\tilde{V}_i} - P_V|\right] \log\left[\frac{|\mathcal{V}|}{\frac{1}{n}\sum_{i=1}^{n} |P_{\tilde{V}_i} - P_V|}\right]. \tag{7.26}$$

Moreover,

$$\frac{1}{n}\sum_{i=1}^{n} |P_{\tilde{V}_i} - P_V| \overset{\text{(a)}}{\le} \frac{1}{n}\sum_{i=1}^{n} \sqrt{2D(P_{\tilde{V}_i} \| P_V)} \tag{7.27}$$

$$\overset{\text{(b)}}{\le} \sqrt{2\frac{1}{n}\sum_{i=1}^{n} D(P_{\tilde{V}_i} \| P_V)} \tag{7.28}$$

$$\overset{\text{(c)}}{\le} \sqrt{2\epsilon} \tag{7.29}$$

where (a) follows by Pinsker's inequality [30, Exercise 3.18], (b) from the concavity of the square root, and (c) from (7.20). Using (7.29) in (7.26) we can further lower-bound (7.23) as

$$\frac{\log(M)}{n} \ge \frac{1}{n}\Big[\sum_{i=1}^{n} H(\tilde{V}_i) - \sum_{i=1}^{n} H(\tilde{V}_i|\tilde{U}_i)\Big] - \epsilon \tag{7.30}$$

$$\ge H(V) - \frac{1}{n}\sum_{i=1}^{n} H(\tilde{V}_i|\tilde{U}_i) - \Big[\epsilon + \sqrt{2\epsilon}\log\frac{|\mathcal{V}|}{\sqrt{2\epsilon}}\Big]. \tag{7.31}$$

Furthermore, since $\forall i = 1, 2, \ldots, n$, $P_{\tilde{V}_i|\tilde{U}_i} = P_{V|U}$,

$$H(\tilde{V}_i|\tilde{U}_i) \le \max_{\substack{(\tilde{U},\tilde{V}): \\ P_{\tilde{V}|\tilde{U}} = P_{V|U} \\ D(P_{\tilde{V}}\|P_V) \le D(P_{\tilde{V}_i}\|P_V)}} H(\tilde{V}|\tilde{U}). \tag{7.32}$$

Lemma B.3 (see Appendix B) shows that the right-hand side of the above is concave in $D(P_{\tilde{V}_i}\|P_V)$. Therefore

$$\frac{1}{n}\sum_{i=1}^{n} H(\tilde{V}_i|\tilde{U}_i) \le \frac{1}{n}\sum_{i=1}^{n} \max_{\substack{(\tilde{U},\tilde{V}): \\ P_{\tilde{V}|\tilde{U}} = P_{V|U} \\ D(P_{\tilde{V}}\|P_V) \le D(P_{\tilde{V}_i}\|P_V)}} H(\tilde{V}|\tilde{U}) \tag{7.33}$$

$$\le \max_{\substack{(\tilde{U},\tilde{V}): \\ P_{\tilde{V}|\tilde{U}} = P_{V|U} \\ D(P_{\tilde{V}}\|P_V) \le \frac{1}{n}\sum_{i=1}^{n} D(P_{\tilde{V}_i}\|P_V)}} H(\tilde{V}|\tilde{U}) \tag{7.34}$$

$$\overset{(*)}{\le} \max_{\substack{(\tilde{U},\tilde{V}): \\ P_{\tilde{V}|\tilde{U}} = P_{V|U} \\ D(P_{\tilde{V}}\|P_V) \le \epsilon}} H(\tilde{V}|\tilde{U}) \tag{7.35}$$

where, $(*)$ again follows from (7.20). Using the above in (7.31), we get

$$\frac{\log(M)}{n} \geq H(V) - \max_{\substack{(\tilde{U},\tilde{V}):\\ P_{\tilde{V}|\tilde{U}}=P_{V|U}\\ D(P_{\tilde{V}}\|P_V)\leq\epsilon}} H(\tilde{V}|\tilde{U}) - \left[\epsilon + \sqrt{2\epsilon}\log\frac{|\mathcal{V}|}{\sqrt{2\epsilon}}\right]. \tag{7.36}$$

Since (7.36) must hold for every $\epsilon > 0$ (but possibly large n and M), and $H(V|U)$ is continuous in P_{UV}, we get

$$\frac{\log(M)}{n} \geq \min_{\substack{(\tilde{U},\tilde{V}):\\ P_{\tilde{V}|\tilde{U}}=P_{V|U},\\ P_{\tilde{V}}=P_V}} I(\tilde{U};\tilde{V}). \tag{7.37}$$

7.2.2 Converse for Variable-to-Fixed-Length Codes

Now suppose we have a variable-to-fixed-length code, that maps T (a random number of) independently drawn input words $J_1, J_2, \ldots, J_T$, each uniformly distributed in $\{1, 2, \ldots, M\}$, into a sequence of length n of channel input symbols $\tilde{U}^n$ that, when transmitted through n-fold channel $P^n_{V|U}$, result in n approximately i.i.d. output symbols $\tilde{V}^n$, i.e.,

$$\frac{1}{n}D(P_{\tilde{V}^n}\|P^n_V) \leq \epsilon \tag{7.38}$$

for a given $\epsilon > 0$. We are set to prove that

$$\frac{\mathbb{E}[T]\log(M)}{n} \geq \min_{\substack{(\tilde{U},\tilde{V}):\\ P_{\tilde{V}|\tilde{U}}=P_{V|U},\\ P_{\tilde{V}}=P_V}} I(\tilde{U};\tilde{V}). \tag{7.39}$$

Let $B \in \mathbb{N}$ be a (large) integer and fix $\delta > 0$. The following procedure defines a block resolvability code of message size $M^{B(\mathbb{E}[T]+\delta)}$ and block-length nB (suppose $B(\mathbb{E}[T] + \delta)$ is integer for simplicity):

Algorithm 7: A Block Code Based on a Variable-to-Fixed-Length Code

```
Input: $J_1, J_2, \ldots, J_{B(\mathbb{E}[T]+\delta)}$ i.i.d. uniformly distributed on $\{1, 2, \ldots, M\}$
$S \leftarrow 0$;                                  // total number of words consumed
for $b = 1$ to $B$ do
    Run the variable-to-fixed-length scheme, feeding it with $T_b$ independent
    words $J_{S+1}, J_{S+2}, \ldots, J_{S+T_b}$ to produce $\tilde{U}^{nb}_{n(b-1)+1}$ ;      // $J_t = 1$ if
    $t > B(\mathbb{E}[T] + \delta)$
    $S \leftarrow S + T_b$;
```

In other words, our block encoder first buffers $B(\mathbb{E}[T]+\delta)$ input words and then runs the variable-to-fixed-length encoder B times successively, feeding the consecutive runs of the encoder with the input words from the buffer. In case it ends up with an empty buffer, it just uses dummy (deterministic) words to feed the variable-to-fixed-length encoders to finish B runs.

We claim that this code is a good block resolvability code, provided that B is sufficiently large. Define the event $\mathcal{H}$ as

$$\mathcal{H} := \{S > B(\mathbb{E}[T] + \delta) \text{ at the end of Algorithm 7}\}. \tag{7.40}$$

By the law of total probability, for any $v^{nB} \in \mathcal{V}^{nB}$,

$$P_{\tilde{V}^{nB}}(v^{nB}) = P_{\tilde{V}^{nB}|\mathcal{H}}(v^{nB}|\mathcal{H}) \Pr(\mathcal{H}) + P_{\tilde{V}^{nB}|\mathcal{H}^c}(v^{nB}|\mathcal{H}^c) \Pr(\mathcal{H}^c). \tag{7.41}$$

Since the KL divergence is convex,

$$D(P_{\tilde{V}^{nB}} \| P_V^{nB}) \leq D\big(P_{\tilde{V}^{nB}|\mathcal{H}} \,\big\|\, P_V^{nB}\big) \Pr(\mathcal{H}) + D\big(P_{\tilde{V}^{nB}|\mathcal{H}^c} \,\big\|\, P_V^{nB}\big) \Pr(\mathcal{H}^c). \tag{7.42}$$

The bound of (7.38) implies

$$\begin{aligned} D\big(P_{\tilde{V}^{nB}|\mathcal{H}^c} \,\big\|\, P_V^{nB}\big) \Pr(\mathcal{H}^c) &\leq D\big(P_{\tilde{V}^{nB}|\mathcal{H}^c} \,\big\|\, P_V^{nB}\big) && (7.43)\\ &= \sum_{b=1}^{B} D\Big(P_{\tilde{V}^{nb}_{n(b-1)+1}|\mathcal{H}^c} \,\Big\|\, P_V^n\Big) && (7.44)\\ &\leq nB\epsilon. && (7.45) \end{aligned}$$

Because, if $S \leq B(\mathbb{E}[T] + \delta)$, all the successive runs of the underlying variable-to-fixed-length encoder are fed with uniformly distributed words $J_1, J_2, \ldots, J_S$; thus, they all result in an approximately i.i.d. output sequence (as guaranteed by (7.38)).

To bound the second term in (7.42), we note that, for any distribution $Q_{V^{nB}} \in \mathcal{P}(\mathcal{V}^{nB})$,

$$D(Q_{V^{nB}} \| P_V^{nB}) \leq nB \log\Big[\frac{1}{P_{\min}}\Big] \tag{7.46}$$

where $P_{\min} := \min_{v \in \mathcal{V}} P_V(v)$. Moreover, since $S = \sum_{b=1}^{B} T_b$, we have

$$\begin{aligned} \Pr(\mathcal{H}) &= \Pr\left\{\sum_{b=1}^{B} T_b > B(\mathbb{E}[T] + \delta)\right\} && (7.47)\\ &= \Pr\left\{\frac{1}{B}\sum_{b=1}^{B} T_b \geq \mathbb{E}[T] + \delta\right\}. && (7.48) \end{aligned}$$

Since $T_1, T_2, \ldots, T_B$ are i.i.d. random variables with mean $\mathbb{E}[T]$, by the law of large numbers, we can guarantee that the above probability is smaller than any desired value if B is large enough. Choose B such that

$$\Pr(\mathcal{H}) \leq \frac{\epsilon}{\log[1/P_{\min}]}. \tag{7.49}$$

Then, using (7.46) and (7.49) we conclude that

$$D\big(P_{\tilde{V}^{nB}|\mathcal{H}} \,\big\|\, P_V^{nB}\big) \Pr(\mathcal{H}) \leq nB\epsilon \tag{7.50}$$

Using (7.45) and (7.50) in (7.42) yields

$$\frac{1}{nB} D(P_{\tilde{V}^{nB}} \| P_V^{nB}) \leq 2\epsilon \tag{7.51}$$

which shows the proposed scheme is a good block resolvability code. The rate of this block code is

$$\frac{\log(M) B(\mathbb{E}[T] + \delta)}{nB} = \frac{\mathbb{E}[T] \log(M)}{n} + \delta \frac{\log(M)}{n}, \tag{7.52}$$

and our converse result for block codes requires,

$$\frac{\mathbb{E}[T] \log(M)}{n} + \delta \frac{\log(M)}{n} \geq \min_{\substack{(\tilde{U}, \tilde{V}): \\ P_{\tilde{V}|\tilde{U}} = P_{V|U}, \\ P_{\tilde{V}} = P_V}} I(\tilde{U}; \tilde{V}). \tag{7.53}$$

As (7.53) holds for any infinitesimal $\delta > 0$, it implies (7.39).

7.2.3 Converse for Fixed-to-Variable Length Codes

We can prove the converse result for fixed-to-variable-length codes in a similar way. Specifically, given a fixed-to-variable-length resolvability code that maps $J \in \{1, 2, \ldots, M\}$ to a sequence of N of channel input symbols $\tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_N$ (where N is a variable-length), with average output length $\mathbb{E}[N]$, guaranteeing

$$\frac{1}{\mathbb{E}[N]} \sum_{n \geq 0} D(P_{\tilde{V}^n | N = n} \| P_V^n) \Pr\{N = n\} \leq \epsilon, \tag{7.54}$$

we construct a good block resolvability code whose rate is close to $\log(M)/\mathbb{E}[N]$. The converse, then, follows by the converse on block resolvability codes.

Fix $B \in \mathbb{N}$ (a large integer) and $\delta > 0$ and consider the following procedure:

Algorithm 8: A Semi-Block Resolvability Code Based on a Fixed-to-Variable-Length Code

```
Input: $J_1, J_2, \ldots, J_B$ i.i.d. uniformly distributed on $\{1, 2, \ldots, M\}$
$S \leftarrow 0$;                          // total number of symbols produced
for $t = 1$ to $B$ do
    Run the fixed-to-variable-length scheme, feeding it with $J_t$, to produce $\tilde{U}_{S+1}^{S+N_t}$;
    $S \leftarrow S + N_t$;                // $N_t$ symbols produced in the $t^{\text{th}}$ round
while $S < B(\mathbb{E}[N] - \delta)$ do
    $S \leftarrow S + 1$;
    $\tilde{U}_S \leftarrow u_0$;                // $u_0$ is any dummy symbol in $\mathcal{U}$
    Transmit $\tilde{U}_S$ via the channel;
```

Algorithm 8 runs the fixed-to-variable-length resolvability scheme B times in a row. After the last round, if the total number of symbols produced is less than $B(\mathbb{E}[N]-\delta)$, it produces extra dummy symbols to extend the length of its output sequence to $B(\mathbb{E}[N]-\delta)$. Note that the length of output sequence of the encoder defined in Algorithm 8 is still variable. (It terminates with $S \geq B(\mathbb{E}[N]-\delta)$ symbols.) By adding an extra 'early termination' condition, however, we can cut its output sequence short to contain only the first $b \coloneqq B(\mathbb{E}[N]-\delta)$ symbols and obtain a block resolvability code. (Again, suppose $B(\mathbb{E}[N]-\delta)$ is integer for simplicity.) We have intentionally kept the details of implementing the 'early termination' out of the algorithm, as we will see that analyzing the unterminated version is simpler and the performance of the terminated version is straightforwardly related to that of the unterminated version.

We now show that when $\tilde{V}^b$ is the first b symbols of the output sequence of the channel when it is fed by Algorithm 8,

$$\frac{1}{b} D(P_{\tilde{V}^b} \| P_V^b) \tag{7.55}$$

is small.

Define the event $\mathcal{H}$ as

$$\begin{aligned}\mathcal{H} \coloneqq \{&\text{the } \textbf{while} \text{ loop of line 5 of Algorithm 8 is executed}\} \\ &= \{N_1 + N_2 + \ldots + N_B < b\}\end{aligned} \tag{7.56}$$

where $N_1, N_2, \ldots, N_B$ are the stopping times of successive runs of the underlying fixed-to-variable-length resolvability encoder.

Same considerations as in (7.41) and (7.42) shows

$$D(P_{\tilde{V}^b} \| P_V^b) \leq D(P_{\tilde{V}^b|\mathcal{H}^c} \| P_V^b) \Pr(\mathcal{H}^c) + D(P_{\tilde{V}^b|\mathcal{H}} \| P_V^b) \Pr(\mathcal{H}). \tag{7.57}$$

We have

$$\Pr(\mathcal{H}^c) P_{\tilde{V}^b|\mathcal{H}^c}(v^b) = \Pr\Big\{\tilde{V}^b = v^b, \sum_{t=1}^{B} N_t \geq b\Big\} \tag{7.58}$$

$$\begin{aligned}= \sum_{\substack{n_1,n_2,\ldots,n_B:\\ n_1+n_2+\cdots+n_B \geq b}} &\Pr\{\tilde{V}^b = v^b | N_1 = n_1, \ldots, N_B = n_B\} \\ &\cdot \Pr\{N_1 = n_1, \ldots, N_B = n_b\}.\end{aligned} \tag{7.59}$$

Consequently, using the convexity of divergence we have

$$\begin{aligned}\Pr(\mathcal{H}^c) D(P_{\tilde{V}^b|\mathcal{H}^c} \| P_V^b) \leq \sum_{\substack{n_1,n_2,\ldots,n_B:\\ n_1+n_2+\cdots+n_B \geq b}} &D(P_{\tilde{V}^b|N_1=n_1,\ldots,N_B=n_B} \| P_V^b) \\ &\cdot \Pr\{N_1 = n_1, \ldots, N_B = n_B\}.\end{aligned} \tag{7.60}$$

Moreover, for every $s \geq b$, and any distribution $Q_{\tilde{V}^s} \in \mathcal{P}(\mathcal{V}^s)$, using the chain rule for the divergence we have

$$D(Q_{\tilde{V}^s} \| P_V^s) = D(Q_{\tilde{V}^b} \| P_V^b) + D(Q_{\tilde{V}^s_{b+1}|\tilde{V}^b} \| P_V^{b-s} | Q_{\tilde{V}^s_{b+1}}) \geq D(Q_{\tilde{V}^b} \| P_V^b). \tag{7.61}$$

In particular, taking $s = n_1 + n_2 + \cdots + n_B$ on the right-hand side of (7.60), we can continue our upper bound in (7.60) as

$$\begin{aligned}
\Pr(\mathcal{H}^c) D(P_{\tilde{V}^b|\mathcal{H}^c} \| P_V^b) &\leq \sum_{\substack{n_1, n_2, \ldots, n_B: \\ n_1 + n_2 + \cdots + n_B \geq b}} D(P_{\tilde{V}^b | N_1 = n_1, \ldots, N_B = n_B} \| P_V^b) \\
&\qquad \cdot \Pr\{N_1 = n_1, \ldots, N_B = n_B\} \qquad (7.62) \\
&\leq \sum_{\substack{n_1, n_2, \ldots, n_B: \\ n_1 + n_2 + \cdots + n_B \geq b}} D(P_{\tilde{V}^{n_1 + \cdots + n_B} | N_1 = n_1, \ldots, N_B = n_b} \| P_V^{n_1 + \cdots + n_B}) \\
&\qquad \cdot \Pr\{N_1 = n_1, \ldots, N_B = n_B\} \qquad (7.63) \\
&\overset{\text{(a)}}{=} \sum_{\substack{n_1, n_2, \ldots, n_B: \\ n_1 + n_2 + \cdots + n_B \geq b}} \sum_{t=1}^{B} D\Big(P_{\tilde{V}^{n_t}_{n_{t-1}+1} | N_t = n_t} \,\Big\|\, P_V^{n_t}\Big) \\
&\qquad \cdot \Pr\{N_1 = n_1, \ldots, N_B = n_B\} \qquad (7.64) \\
&\overset{\text{(b)}}{\leq} \sum_{\substack{n_1, n_2, \ldots, n_B: \\ n_1 \geq 0, n_2 \geq 0, \ldots, n_B \geq 0}} \sum_{t=1}^{B} D\Big(P_{\tilde{V}^{n_t}_{n_{t-1}+1} | N_t = n_t} \,\Big\|\, P_V^{n_t}\Big) \\
&\qquad \cdot \Pr\{N_1 = n_1, \ldots, N_B = n_B\} \qquad (7.65) \\
&= \sum_{t=1}^{B} \sum_{n \geq 0} D(P_{\tilde{V}^n | N_t = n} \| P_V^n) \Pr\{N_t = n\} \qquad (7.66) \\
&\overset{\text{(c)}}{\leq} \epsilon B \, \mathbb{E}[N] \qquad (7.67)
\end{aligned}$$

In the above, (a) follows since the successive runs of the original fixed-to-variable-length encoder are independent (in the inner summation we took $n_0 :=$ 0), (b) by extending the sum to all non-negative integers $n_1, n_2, \ldots, n_B$ (noting that the summands are non-negative), and (c) from (7.54).

Moreover, as we saw before

$$\Pr(\mathcal{H}) D(P_{\tilde{V}^b|\mathcal{H}} \| P_V^b) \leq b \Pr(\mathcal{H}) \log\Big[\frac{1}{P_{\min}}\Big] \tag{7.68}$$

with $P_{\min} := \min_{v \in \mathcal{V}} P_V(v)$.

Finally, since

$$\Pr(\mathcal{H}) = \Pr\left\{\frac{1}{B}\sum_{t=1}^{B} N_t \leq \mathbb{E}[N] - \delta\right\} \tag{7.69}$$

and $N_1, N_2, \ldots, N_B$ are i.i.d. random variables with mean $\mathbb{E}[N]$, due to the law of large numbers, by choosing B large enough, we can guarantee that

$$\Pr(\mathcal{H}) \leq \frac{\epsilon}{\log[1/P_{\min}]}. \tag{7.70}$$

Uniting (7.67)–(7.70) in (7.57), we conclude that, for large enough B,

$$\frac{1}{b} D(P_{\tilde{V}^b} \| P_V^b) \leq \frac{\mathbb{E}[N]}{\mathbb{E}[N] - \delta}\epsilon + \epsilon \leq 3\epsilon \tag{7.71}$$

where the last inequality follows by choosing $\delta \leq \mathbb{E}[N]/2$. Thus, we showed that the block code defined by a terminated version of Algorithm 8 is indeed a good resolvability code. Such a code produces $B(\mathbb{E}[N] - \delta)$ channel input symbols from B uniformly distributed input words in $\{1, 2, \ldots, M\}$, hence its rate is $\log(M)/(\mathbb{E}[N] - \delta)$. Therefore, it follows from the converse result we established in § 7.2.1 that

$$\frac{B\log(M)}{B(\mathbb{E}[N] - \delta)} = \frac{\log(M)}{\mathbb{E}[N] - \delta} \geq \min_{\substack{(\tilde{U},\tilde{V}): \\ P_{\tilde{V}|\tilde{U}} = P_{V|U}, \\ P_{\tilde{V}} = P_V}} I(\tilde{U}; \tilde{V}). \tag{7.72}$$

Since (7.72) holds for every $\delta > 0$, we deduce that

$$\frac{\log(M)}{\mathbb{E}[N]} \geq \min_{\substack{(\tilde{U},\tilde{V}): \\ P_{\tilde{V}|\tilde{U}} = P_{V|U}, \\ P_{\tilde{V}} = P_V}} I(\tilde{U}; \tilde{V}). \tag{7.73}$$

This concludes the proof. □

7.3 Improving Resolvability Exponents with Feedback

We have seen that feedback does not reduce the resolution of a channel; to approximate the product distribution P_V at the output of the channel $P_{V|U} : \mathcal{U} \times \mathcal{V}$ we need at least an entropy rate of $I(\tilde{U}; \tilde{V})$ at its input (where $\tilde{U}, \tilde{V} \sim P_U \times P_{V|U}$ for some P_U that induces P_V at the output of channel). Does this mean that feedback is totally useless?

Not at all! Similarly to the problem of error correction, feedback simplifies the design of encoder and improves the quality of approximation. We present two examples in this section: resolvability over a binary erasure channel and a binary symmetric channel. For the former, we construct a variable-to-fixed-length resolvability encoder that, by using the feedback signal and by consuming an entropy rate *approaching* the resolution of the channel (as the block-length increases), produces a *perfect i.i.d. string* at the output of the channel. That is to say, a resolvability exponent of *infinity* is achievable over the BEC in the presence of feedback. For the latter, we propose a fixed-to-variable-length resolvability encoder that, in the presence of feedback, achieves the straight-line resolvability exponent $[R-I(U;V)]^+$. Moreover, we will show that neither of these improved resolvability exponents are achievable by using block codes.

7.3.1 Resolvability over BEC in the Presence of Feedback

Theorem 7.2. *Assume the channel $P_{V|U}:\mathcal{U}\to\mathcal{V}$ is a $\mathsf{BEC}(p)$ with input alphabet $\mathcal{U}=\{0,1\}$ and output alphabet $\mathcal{V}=\{0,1,?\}$. Let $P_V\in\mathcal{P}(\{0,1,?\}$ be the distribution induced by the uniform distribution at the input of the channel at its output, i.e.,*

$$P_V(0)=P_V(1)=\frac{1-p}{2}, \qquad \textit{and} \qquad P_V(?)=p. \tag{7.74}$$

In the presence of feedback, there exists a variable-to-fixed-length resolvability encoder of block-length n that consumes, on average, an entropy rate of $(1-p)+p/n$ bits per channel use and, guarantees $P_{\tilde{V}^n}=P_V^n$. In other words 'perfect' resolvability encoders exist. Recall that $\tilde{V}^n$ denotes the output of the n-fold use of the channel when it is fed by the variable-to-fixed-length resolvability encoder. Therefore, the resolvability exponent

$$E_{\mathsf{BEC}}(p,R) := \begin{cases} 0 & \textit{if } R\le(1-p), \\ +\infty & \textit{if } R>(1-p), \end{cases} \tag{7.75}$$

is achievable over the channel with respect to the sequence of product reference measures P_V^n, using variable-to-fixed-length resolvability schemes.

Before proving Theorem 7.2 let us highlight the importance of variable-length coding through the following lemma.

Lemma 7.3. *Let P_V be the distribution induced at the output of a $\mathsf{BEC}(p)$ when its input has a uniform distribution (see (7.74)). Then, using any sequence of block resolvability encoders of block-length n (for the system with feedback) over a $\mathsf{BEC}(p)$,*

$$\limsup_{n\to\infty} -\frac{1}{n}\log D(P_{\tilde{V}^n}\|P_V^n) \le \begin{cases} 0 & \textit{if } R<(1-p), \\ 2d_2(R\|1-p) & \textit{if } R\ge(1-p), \end{cases} \tag{7.76}$$

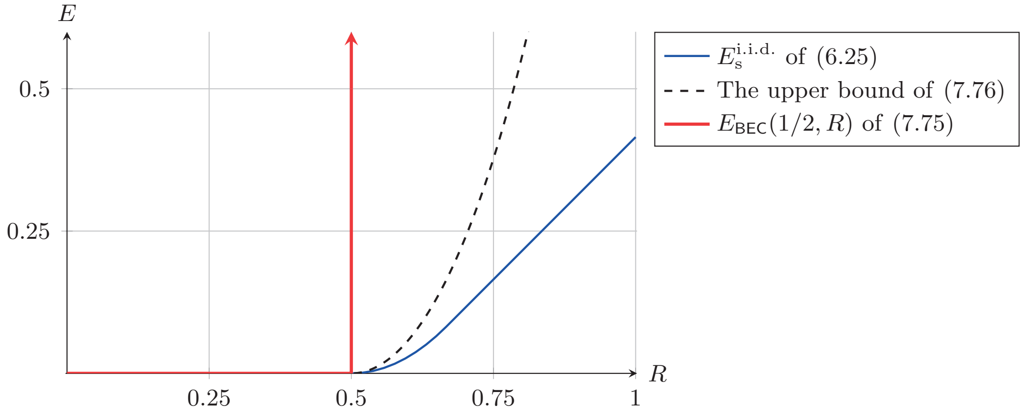


Figure 7.2: In the presence of feedback, perfect variable-to-fixed-length resolvability codes exist. Here, we evaluate the exponents for BEC$(1/2)$.

where

$$d_2(p\|q) \coloneqq p \log\left[\frac{p}{q}\right] + (1-p)\log\left[\frac{1-p}{1-q}\right] \tag{7.77}$$

is the binary divergence function. (Recall that $\tilde{V}^n$ denotes the output of the n-fold use of the channel when its input sequence is generated by the resolvability encoder.)

Lemma 7.3 follows along the same lines as [121, Corollary 6] by using Pinsker's inequality [30, Exercise 3.18]; it shows that 'perfect' block resolvability encoders of rates strictly less than 1 *cannot* exist. We defer the proof of Lemma 7.3 to Appendix 7.A. (In fact, [121, Corollary 6], combined with Pinsker's inequality, implies (7.76) for the system without feedback.) In Figure 7.2, we compare the upper bound of (7.76), the achievable exponent of (6.25) (evaluated for a BEC$(1/2)$ with uniform input distribution), and the exponent of (7.75).

Proof of Theorem 7.2. For any $n \in \mathbb{N}$, consider the variable-to-fixed-length resolvability encoder described as follows:

Algorithm 9: Variable-to-Fixed-Length Resolvability Encoder for BEC

```
Input: $U_1, U_2, \ldots$ i.i.d. uniformly distributed on $\{0,1\}$
$t \leftarrow 1$;
for $i = 1$ to $n$ do
    if $i > 1$ and $\tilde{V}_{i-1} \in \{0,1\}$ then
        $t \leftarrow t + 1$;
    $\tilde{U}_i \leftarrow U_t$;
    Transmit $\tilde{U}_i$ via the channel and record $\tilde{V}_i$;
```

Note that we have denoted the input words to the encoder as $U_1, U_2, \ldots$ that are i.i.d. uniformly distributed bits on $\{0,1\}$ (whereas, previously, we used to denote the input words to the resolvability encoder with symbol J).

If we wanted to formulate the encoding algorithm in terms of the collection of shifting and encoding functions (cf. Definition 7.1), the encoder would be defined via the collection of shifting functions $\mathrm{Shift}_{i,t}\colon \{0,1\}^t \times \mathcal{V}^{i-1} \to \{0,1\}$.

$$\mathrm{Shift}_{i,t}(u^t, v^{i-1}) = \mathbb{1}\{i > 1 \text{ and } v_{i-1} \in \{0,1\}\}, \qquad \forall t \in \mathbb{N}, \forall i \in \{1,2,\ldots,n\} \tag{7.78}$$

and encoding functions $\mathrm{Enc}_{i,t}\colon \{0,1\}^t \times \mathcal{V}^{i-1} \to \{0,1\}$,

$$\mathrm{Enc}_{i,t}(u^t, v^{i-1}) = u_t, \qquad \forall t \in \mathbb{N}, \forall i \in \{1,2,\ldots,n\}. \tag{7.79}$$

In other words, the encoder starts off by reading one uniformly distributed bit from the input and transmitting it through the channel. Then, at each subsequent time instant, $i \in \{2,3,\ldots,n\}$ checks if the previously transmitted bit is erased. If this is the case, it repeats the same bit. Otherwise, it reads a fresh random bit from the input and transmits it.

It is easy to check that $\tilde{V}^n$, the output of $P^n_{V|U}$ when the resolvability code of Algorithm 9 is used at its input, has distribution P^n_V. To verify this formally, we will show that

$$P_{\tilde{V}_i|\tilde{V}^{i-1}}(v_i|v^{i-1}) = \begin{cases} p, & \text{if } v_i = ?, \\ \frac{1-p}{2}, & \text{if } v_i \in \{0,1\}. \end{cases} \tag{7.80}$$

Using the law of total probability,

$$P_{\tilde{V}_i|\tilde{V}^{i-1}}(v_i|v^{i-1}) = \sum_{u_i \in \{0,1\}} \Pr\{\tilde{V}_i = v_i | \tilde{V}^{i-1} = v^{i-1}, \tilde{U}_i = u_i\} \cdot \Pr\{\tilde{U}_i = u_i | \tilde{V}^{i-1} = v^{i-1}\} \tag{7.81}$$

$$\overset{\text{(a)}}{=} \sum_{u_i \in \{0,1\}} P_{V|U}(v_i|u_i) \Pr\{\tilde{U}_i = u_i | \tilde{V}^{i-1} = v^{i-1}\} \tag{7.82}$$

$$\overset{\text{(b)}}{=} \begin{cases} p & \text{if } v_i = ? \\ (1-p)\Pr\{\tilde{U}_i = v_i | \tilde{V}^{i-1} = v^{i-1}\} & \text{if } v_i \in \{0,1\} \end{cases} \tag{7.83}$$

where (a) follows since the channel is memoryless and (b) using the transition probabilities of a $\mathsf{BEC}(p)$. Finally, we note that

$$\Pr\{\tilde{U}_i = v_i | \tilde{V}^{i-1} = v^{i-1}\} = \frac{1}{2} \quad \forall v_i \in \{0,1\}, \tag{7.84}$$

by construction. Using (7.84) in (7.83) shows $P_{\tilde{V}^n} = P^n_V$ (for every n).

Moreover, the average rate of the system is

$$\frac{\log(M)\,\mathbb{E}[T]}{n} = \frac{\mathbb{E}[T]}{n} = \frac{(n-1)(1-p)+1}{n} \tag{7.85}$$

$$= (1-p) + \frac{p}{n} \tag{7.86}$$

since T equals one plus the number of 'non-erasure' events in $n-1$ channel uses. Consequently, given any $R > (1-p)$ by taking n large enough, we will have a scheme whose rate is below R and guarantees

$$D(P_{\tilde{V}^n} \| P_V^n) = 0, \tag{7.87}$$

(since $\forall n$, $P_{\tilde{V}^n} = P_V^n$). □

Remark 1. The resolution of a $\mathsf{BEC}(p)$, with respect to the sequence of product measures P_V^n (with P_V as defined in (7.74)), is $(1-p)$ bits per channel use (because the uniform input distribution is the only distribution that induces P_V at the output of a BEC). Theorem 7.2 illustrates that by using an external entropy rate just above $I(U;V)$ we can perfectly simulate the i.i.d. measure P_V^n at the output of a BEC.

Remark 2. If we fed the encoder proposed in the proof of Theorem 7.2 with i.i.d. $U_1, U_2, \dots$ drawn from an arbitrary distribution $P_U \in \mathcal{P}(\mathcal{U})$, the scheme would lead to a perfect simulation of the product distribution P_V^n (where P_V is the distribution induced by P_U at the output of the BEC) consuming an average entropy rate of $H(U)(1-p) + o(1)$ at the encoder.

7.3.2 Resolvability over BSC in the Presence of Feedback

Theorem 7.4. *In the presence of feedback, the exponent*

$$E_{\mathsf{BSC}}(p, R) = [R - 1 + h_2(p)]^+ \tag{7.88}$$

is achievable over a $\mathsf{BSC}(p)$ *with respect to the sequence of uniform reference measures* $P_V^n \in \mathcal{P}(\mathcal{V}^n)$, $P_V(0) = P_V(1) = \frac{1}{2}$, *using fixed-to-variable-length resolvability schemes. In the above*

$$h_2(p) := p \log \frac{1}{p} + (1-p) \log \frac{1}{1-p} \tag{7.89}$$

is the binary entropy function.

The straight-line exponent of (7.88) is larger than the exponent of (6.25) as the objective function of (6.25) equals $[R - I(U;V)]^+$ at $Q_{UV} = P_{UV}$. Furthermore, we have the following upper bound on the best achievable resolvability exponent over a BSC using block codes in the presence of feedback.

Lemma 7.5. *Let* P_V *be the uniform distribution on* $\{0,1\}$. *Then using any sequence of block resolvability codes of length* n *and rate (at most)* R, *(for the system with feedback) over a* $\mathsf{BSC}(p)$,

$$\begin{aligned} &\limsup_{n\to\infty} -\frac{1}{n} \log[D(P_{\tilde{V}^n} \| P_V^n)] \\ &\qquad \le \begin{cases} 0 & \textit{if } R < 1 - h_2(p) \\ 2d_2\big(p - \frac{R-1+h_2(p)}{\log[(1-p)/p]} \,\big\|\, p\big) & \textit{if } 1 - h_2(p) \le R < \log[2(1-p)] \\ +\infty & \textit{if } R \ge \log[2(1-p)] \end{cases} \end{aligned} \tag{7.90}$$

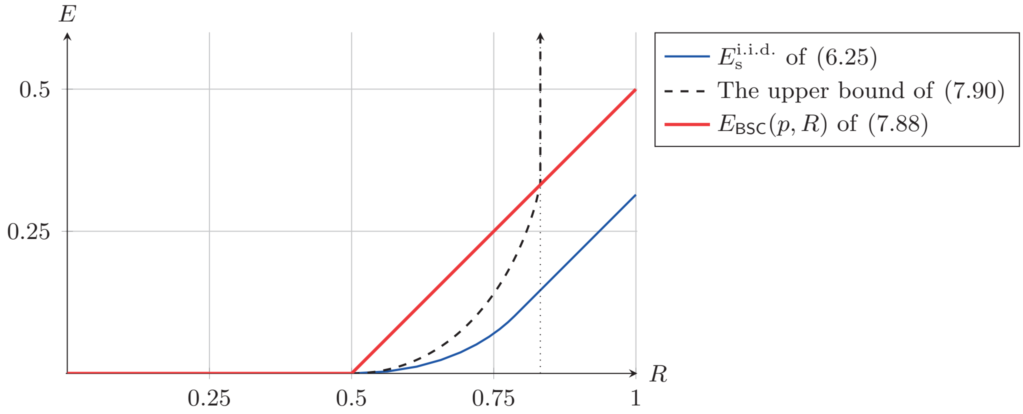


Figure 7.3: The straight-line exponent resolvability $[R - I(U;V)]^+$ is achievable over a BSC(p) with respect to the sequence of uniform product measures. The exponents are evaluated for a BSC$(1/2)$.

where $P_{\tilde{V}^n}$ is the output distribution of the n-fold channel when it is fed by the block resolvability encoder of length n. (In the above $d_2(\cdot\|\cdot)$ is the binary divergence function defined in (7.77)*.)*

We prove Lemma 7.5 in Appendix 7.A.[2]

In Figure 7.3, we compare the exponents of (7.88) and (6.25) and the upper bound of (7.90). It can easily be seen that, at least for rates close to the resolution of the channel, employing variable-length resolvability schemes is necessary; the best attainable resolvability exponent using block codes begins its increase above 0 with slope zero as R exceeds $1 - h_2(p)$.

Proof of Theorem 7.4. Consider the following fixed-to-variable-length resolvability scheme:

Fix $\alpha > 0$ and the message size $M \in \mathbb{N}$. The collection of encoding functions $(\mathrm{Enc}_n,\, n \in \mathbb{N})$ share a codebook of size M and infinite block-length,

$$\mathscr{C} := \big(u^\infty(j) \in \mathbb{F}_2^\infty : j \in \{1, 2, \dots, M\}\big) \tag{7.91}$$

(to be specified later) and are defined as

$$\mathrm{Enc}_1(j) = u_1(j), \quad \text{and} \tag{7.92a}$$

$$\mathrm{Enc}_{n+1}(j, v^n) = \begin{cases} \mathtt{STOP} & \text{if } \frac{\log(M)}{n} \le \alpha[1 - h_2(\hat{q}_n)], \\ u_{n+1}(j) & \text{otherwise,} \end{cases} \tag{7.92b}$$

where,

$$\hat{q}_n := \frac{\mathrm{d_H}(u^n(j), v^n)}{n} \tag{7.93}$$

[2] Similarly to our comment after Lemma 7.3, we remark that [121, Corollary 6], combined with Pinsker's inequality, implies (7.90) for the system without feedback.

is the fraction of flipped bits during the first n transmissions. (Recall that $\mathrm{d_H}$ denotes the Hamming distance between two sequences.)

In other words, given the input word j, the encoder successively transmits the letters of the corresponding codeword $u^\infty(j)$, until the transmission rate $\log(M)/n$ drops below a given multiple of the empirical capacity of the channel. This, in particular, implies that the stopping time N is larger than $\log(M)/\alpha$.

Lemma 7.6. *Given any $\delta > 0$, there exists $M_0(\delta)$ such that, for all $M \geq M_0$ using the proposed scheme,*

$$\alpha[1 - h_2(p)] - \delta \leq \frac{\log(M)}{\mathbb{E}[N]} \leq \alpha[1 - h_2(p)] + \delta \tag{7.94}$$

Proof. Let

$$\Phi_n \coloneqq \mathbb{1}\{\text{channel flips at time } n\}, \qquad \forall n \in \mathbb{N}. \tag{7.95}$$

Hence

$$n\hat{q}_n = \sum_{i=1}^{n} \Phi_i \tag{7.96}$$

where $(\Phi_n,\, n \in \mathbb{N})$ are i.i.d. Bernoulli(p) random variables. Let

$$S_n \coloneqq n\hat{q}_n - np. \tag{7.97}$$

The process $(S_n,\, n \in \mathbb{N})$ is a martingale with respect to the natural filtering $\big(\mathcal{F}_n = \sigma(\Phi_1, \ldots, \Phi_n),\, n \in \mathbb{N}\big)$. This follows simply because $\mathbb{E}[|S_n|] \leq 2n < +\infty$ and

$$\mathbb{E}[S_n|\mathcal{F}_{n-1}] = \mathbb{E}[\Phi_n] - p + S_{n-1} = S_{n-1} \tag{7.98}$$

The encoder stops at time

$$N = \inf\Big\{n \geq \frac{\log(M)}{\alpha} : 1 - h_2\big(\hat{q}_n\big) \geq \alpha^{-1}\frac{\log(M)}{n}\Big\}. \tag{7.99}$$

In terms of S_n, the stopping condition is

$$\log(M) \leq \alpha \cdot N\Big[1 - h_2\Big(p + \frac{S_N}{N}\Big)\Big]. \tag{7.100}$$

It easily can be verified (cf. Appendix 7.B) that $\forall p \in (0,1)$, $\forall \varepsilon \in (-p, 1-p)$,

$$h_2(p) + h_2'(p)\epsilon + h_2''(p)\epsilon^2 \leq h_2(p + \varepsilon) \leq h_2(p) + h_2'(p)\epsilon \tag{7.101}$$

Using the lower bound of (7.101) in (7.100), we get

$$\log(M) \leq \alpha N - \alpha N h_2(p) - \alpha N \frac{S_N}{N} h_2'(p) - \alpha N \frac{S_N^2}{N^2} h_2''(p) \tag{7.102}$$

$$= \alpha N[1 - h_2(p)] - \alpha S_N h_2'(p) - \alpha h_2''(p) \frac{S_N^2}{N}. \tag{7.103}$$

Taking the expectation of the right-hand side of (7.103), noting that $\mathbb{E}[S_N] = \mathbb{E}[S_{\lceil \log(M)/\alpha \rceil}] = 0$ (because a stopped martingale is also a martingale [44, Theorem 4, Chapter 7]), we get

$$\frac{\log(M)}{\mathbb{E}[N]} \leq \alpha[1 - h_2(p)] - \alpha h_2''(p) \frac{\mathbb{E}[S_N^2/N]}{\mathbb{E}[N]}. \tag{7.104}$$

(Note also that $h_2''(p) < 0$ since h_2 is a concave function.) The growth rate of the last term in (7.104) remains to be examined. Had we replaced the stopping time N with a fixed time n, the quantity of interest would have behaved like $1/n$ (since $\mathbb{E}[S_n^2/n]$ is a constant). It turns out that for a stopping time N, $\mathbb{E}[S_N^2/N]$ might not be a constant but will grow at most logarithmically in N: Lemma 7.7 (in Appendix 7.C) shows

$$\mathbb{E}\left[\frac{S_N^2}{N}\right] \leq p(1-p)\, \mathbb{E}[1 + \ln(N)]. \tag{7.105}$$

Consequently,

$$\begin{aligned}
\frac{\log(M)}{\mathbb{E}[N]} &\leq \alpha[1 - h_2(p)] - \alpha h_2''(p) p(1-p) \frac{\mathbb{E}[1 + \ln(N)]}{\mathbb{E}[N]} && (7.106)\\
&\overset{\text{(a)}}{\leq} \alpha[1 - h_2(p)] - \alpha h_2''(p) p(1-p) \frac{1 + \ln(\mathbb{E}[N])}{\mathbb{E}[N]} && (7.107)\\
&\overset{\text{(b)}}{\leq} \alpha[1 - h_2(p)] - \alpha h_2''(p) p(1-p) \frac{1 + \ln\bigl(\log(M)/\alpha\bigr)}{\log(M)/\alpha}, && (7.108)
\end{aligned}$$

where (a) follows from Jensen's inequality and (b) folows because $[1 + \ln(x)]/x$ is decreasing for $x \geq 1$ and $N \geq \log(M)/\alpha$. Therefore, by choosing $M \geq M_1(\delta)$ such that

$$-\alpha h_2''(p) p(1-p) \frac{1 + \ln\bigl(\log(M)/\alpha\bigr)}{\log(M)/\alpha} \leq \delta \tag{7.109}$$

the upper bound of (7.94) will be satisfied.

To lower-bound $\log(M)/\mathbb{E}[N]$, we note that $\forall n > 1$,

$$\hat{q}_n = \frac{n-1}{n} \hat{q}_{n-1} + \frac{1}{n} \Phi_n. \tag{7.110}$$

Since $h_2(\cdot)$ is concave, at the stopping time,

$$\begin{aligned}
h_2(\hat{q}_N) &\geq \frac{N-1}{N} h_2(\hat{q}_{N-1}) + \frac{1}{N} h_2(\Phi_N) && (7.111)\\
&\overset{(*)}{>} \frac{N-1}{N} \times \left[1 - \alpha^{-1} \frac{\log(M)}{N-1}\right] && (7.112)\\
&= \frac{N-1}{N} 1 - \alpha^{-1} \frac{\log(M)}{N}, && (7.113)
\end{aligned}$$

where $(*)$ follows from the stopping condition (7.99). Rearranging the above, we get,

$$\log(M) > \alpha(N-1)1 - \alpha N h_2(\hat{q}_N). \tag{7.114}$$

Substituting $\hat{q}_N = \frac{S_N}{N} + p$ yields

$$\begin{aligned}\log(M) &> \alpha(N-1) - \alpha N h_2\Big(\frac{S_N}{N} + p\Big) && (7.115)\\ &\overset{(*)}{\geq} \alpha(N-1)1 - \alpha N h_2(p) - \alpha N \frac{S_N}{N} h_2'(p) && (7.116)\\ &= \alpha N[1 - h_2(p)] - \alpha[1 + S_N h_2'(p)] && (7.117)\end{aligned}$$

where $(*)$ follows from the lower bound of (7.101). Taking the expectation of the right-hand side of (7.117) (and using the fact that $\mathbb{E}[S_N] = 0$ once again) we get,

$$\begin{aligned}\frac{\log(M)}{\mathbb{E}[N]} &\geq \alpha[1 - h_2(p)] - \frac{\alpha}{\mathbb{E}[N]}\\ &\overset{(*)}{\geq} \alpha[1 - h_2(p)] - \frac{\alpha^2}{\log(M)}. && (7.118)\end{aligned}$$

where $(*)$ follows because $N \geq \log(M)/\alpha$. Thus, choosing $M \geq M_2(\delta)$ such that

$$\frac{\alpha^2}{\log(M)} \leq \delta \tag{7.119}$$

ensures the lower bound of (7.94).

Taking $M \geq M_0(\delta) := \max\{M_1, M_2\}$ guarantees that both upper and lower bounds of (7.94) are satisfied. □

To complete the proof of Theorem 7.4, we need to bound the expected output divergence $\mathbb{D}$, as defined in (7.8), for an appropriate code. Let $h_2^{-1}(\cdot)$ denote the inverse of the binary entropy function $h_2(\cdot)$ (cf. (7.89)) when its domain is restricted to $[0, 1/2]$ and define $w\colon \{\lceil \log(M)/\alpha\rceil, \lceil \log(M)/\alpha\rceil + 1, \dots\} \to \mathbb{N}$ as

$$w(n) := \Big\lfloor n h_2^{-1}\Big(1 - \alpha^{-1}\frac{\log(M)}{n}\Big)\Big\rfloor. \tag{7.120}$$

Let $\Phi^n := (\Phi_1, \dots, \Phi_n)$ denote the flip pattern of n independent uses of the channel and

$$\mathcal{H}_n := \big\{\phi^n \in \{0,1\}^n : \{\Phi^n = \phi^n\} \subset \{N = n\}\big\} \tag{7.121}$$

denote the set of flip patterns that stop the encoder at time $N = n$. Using the fact that the process $n\hat{q}_n = \mathrm{w_H}(\Phi^n)$ is an integer-valued process and the stopping condition (7.99), we can conclude that (among other constraints) $\forall \phi^n \in \mathcal{H}_n$,

$$\mathrm{w_H}(\phi^n) = w(n) \quad \text{or} \quad n - \mathrm{w_H}(\phi^n) = w(n) \tag{7.122}$$

(see Figure 7.4), where $\mathrm{w_H}(\phi^n)$ denotes the Hamming weight of ϕ^n.

Note that $\mathcal{H}_n$ can be empty for some values of n, For example, if for some n, $\exists \ell \in \mathbb{N}$ such that $w(n-\ell) = w(n)$ then, $\mathcal{H}_n$ is empty because either the encoder stops at time $n - \ell$ or, if not, it will stop at some time $N > n$, because

$$\mathrm{w_H}(\Phi^n) \geq \mathrm{w_H}(\Phi^{n-\ell}) > w(n)$$

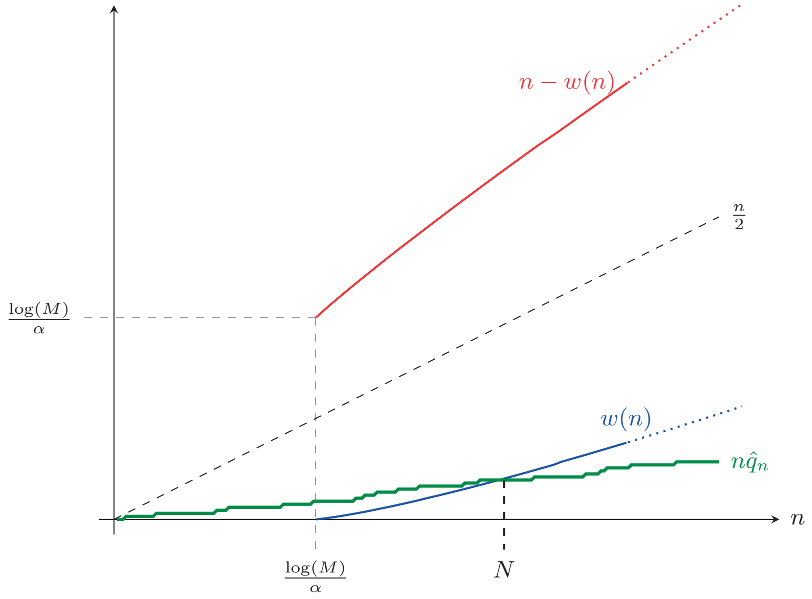


Figure 7.4: Stopping Time of the Proposed Encoder for BSC

and similarly $n - \mathrm{w_H}(\Phi^n) < n - w(n)$. Obviously $\Pr\{N = n\} = 0$ for such values of n, so we do not need to be concerned about them. Let

$$\mathcal{N} := \{n \geq \log(M)/\alpha : \Pr\{N = n\} > 0\} \tag{7.123}$$

be the support of N and assume $n \in \mathcal{N}$.

Partition $\mathcal{H}_n = \mathcal{H}_n^1 \cup \mathcal{H}_n^2$ where

$$\mathcal{H}_n^1 := \{\phi^n \in \mathcal{H}_n : \mathrm{w_H}(\phi^n) = w(n)\}, \tag{7.124}$$
$$\mathcal{H}_n^2 := \{\phi^n \in \mathcal{H}_n : \mathrm{w_H}(\phi^n) = n - w(n)\}. \tag{7.125}$$

It can easily be verified that $|\mathcal{H}_n^1| = |\mathcal{H}_n^2| = |\mathcal{H}_n|/2$. Indeed, the symmetry of stopping thresholds around $n/2$ (see Figure 7.4) implies $\phi^n \in \mathcal{H}_n^1$ if and only if its complement (i.e., the sequence obtained by flipping all elements of ϕ^n) is in $\mathcal{H}_n^2$. Consequently,

$$\Pr\{\Phi^n \in \mathcal{H}_n^1\} = \frac{1}{2}|\mathcal{H}_n| p^{w(n)}(1-p)^{n-w(n)}, \tag{7.126a}$$
$$\Pr\{\Phi^n \in \mathcal{H}_n^2\} = \frac{1}{2}|\mathcal{H}_n| p^{n-w(n)}(1-p)^{w(n)}. \tag{7.126b}$$

Assume, without essential loss of generality, that $p < \frac{1}{2}$. Then $\Pr\{\Phi^n \in \mathcal{H}_n^1\} \geq \Pr\{\Phi^n \in \mathcal{H}_n^2\}$. Hence,

$$\rho_n := \frac{\Pr\{\Phi^n \in \mathcal{H}_n^1\}}{\Pr\{\Phi^n \in \mathcal{H}_n^1\} + \Pr\{\Phi^n \in \mathcal{H}_n^2\}} \in [1/2 : 1]. \tag{7.127}$$

Moreover, since $\{N = n\} = \{\Phi^n \in \mathcal{H}_n\} = \{\Phi^n \in \mathcal{H}_n^1\} \cup \{\Phi^n \in \mathcal{H}_n^2\}$ and $\mathcal{H}_n^1$ and $\mathcal{H}_2^n$ are disjoint (by definition),

$$P_{\tilde{V}^n|N=n}(v^n) = \rho_n \Pr\{\tilde{V}^n = v^n|\Phi^n \in \mathcal{H}_n^1\} + (1-\rho_n)\Pr\{\tilde{V}^n = v^n|\Phi^n \in \mathcal{H}_n^2\}. \tag{7.128}$$

Given the specification of the encoder, we have,

$$\Pr\{\tilde{V}^n = v^n, \Phi^n \in \mathcal{H}_n^1\} = \frac{1}{M}\sum_{j=1}^{M} \Pr\{\tilde{V}^n = v^n, \Phi^n \in \mathcal{H}_n^1|\tilde{U}^n = u^n(j)\} \tag{7.129}$$

$$= \frac{1}{M}\sum_{j=1}^{M}\sum_{\phi^n \in \mathcal{H}_n^1} \Pr\{\tilde{V}^n = v^n, \Phi^n = \phi^n|\tilde{U}^n = u^n(j)\} \tag{7.130}$$

$$= \frac{1}{M}\sum_{j=1}^{M}\sum_{\phi^n \in \mathcal{H}_n^1} \Pr\{\tilde{V}^n = v^n|\Phi^n = \phi^n, \tilde{U}^n = u^n(j)\} \cdot \Pr\{\Phi^n = \phi^n|\tilde{U}^n = u^n(j)\} \tag{7.131}$$

$$\overset{\text{(a)}}{=} \frac{1}{M}\sum_{j=1}^{M}\sum_{\phi^n \in \mathcal{H}_n^1} \mathbb{1}\{v^n = \phi^n \oplus u^n(j)\}\Pr\{\Phi^n = \phi^n\} \tag{7.132}$$

$$\overset{\text{(b)}}{=} \frac{\Pr\{\Phi^n \in \mathcal{H}_n^1\}}{|\mathcal{H}_n^1|}\frac{1}{M}\sum_{j=1}^{M}\sum_{\phi^n \in \mathcal{H}_n^1} \mathbb{1}\{v^n = b^n \oplus u^n(j)\}, \tag{7.133}$$

where (a) follows since the channel behavior, Φ^n, is independent of its input $\tilde{U}^n$ and when its input is u^n and has flip pattern ϕ^n, its output is $u^n \oplus \phi^n$, and (b) follows since $\Pr\{\Phi^n = \phi^n\}$ only depends on $\mathrm{w_H}(\phi^n)$ and all $\phi^n \in \mathcal{H}_1^n$ have the same Hamming weight. As a consequence,

$$\Pr\{\tilde{V}^n = v^n|\Phi^n \in \mathcal{H}_n^1\} = \frac{1}{M|H_n^1|}\sum_{j=1}^{M} \mathbb{1}\{u^n(j) \oplus v^n \in \mathcal{H}_n^1\} \tag{7.134}$$

$$= \frac{1}{M|\mathcal{H}_n^1|}K_{\mathcal{H}_n^1}(v^n) \tag{7.135}$$

where for any $\mathcal{A}_n \subseteq \{0,1\}^n$, we have defined

$$K_{\mathcal{A}_n}(v^n) := \big|\{j \in \{1,2,\ldots,M\} : u^n(j) \oplus v^n \in \mathcal{A}_n\}\big|. \tag{7.136}$$

We similarly have

$$\Pr\{\tilde{V}^n = v^n|\Phi^n \in \mathcal{H}_n^2\} = \frac{1}{M|\mathcal{H}_n^2|}K_{\mathcal{H}_n^2}(v^n). \tag{7.137}$$

Since $P_V^n(v^n) = 2^{-n}$, combining (7.135) and (7.137), together with the fact that $|\mathcal{H}_n^1| = |\mathcal{H}_n^2| = \frac{1}{2}|\mathcal{H}_n|$ in (7.128), we get

$$L(v^n) := \frac{P_{\tilde{V}^n|N=n}(v^n)}{P_V^n(v^n)} = \frac{2^n}{M\frac{1}{2}|\mathcal{H}_n|}\big[\rho_n K_{\mathcal{H}_n^1}(v^n) + (1-\rho_n)K_{\mathcal{H}_n^2}(v^n)\big]. \tag{7.138}$$

Following the same considerations as in (6.37), we have

$$D(P_{\tilde{V}^n|N=n} \| P_V^n) = \sum_{v^n} P_V^n(v^n) L(v^n) \log L(v^n). \tag{7.139}$$

Now, assume the code shared by the encoding functions $(\mathrm{Enc}_n, n \in \mathbb{N})$ is sampled from the i.i.d. random-coding ensemble. Specifically, each codeword $U^\infty(j)$ is an infinite i.i.d. sequence of binary digits where each symbol is equally likely to take either value and the codewords are independent of each other. In this case, as we have seen in Chapter 6, $\{K_{\mathcal{H}_n^1}(v^n), K_{\mathcal{H}_n^2}(v^n)\}$ forms a multinomial collection with cluster size M and (equal) success probabilities $2^{-n}|\mathcal{H}_n|/2$. Thus, it can immediately be verified that $\mathbb{E}[L(v^n)] = 1$.

Since $L(v^n) \leq 2^n$,

$$\mathbb{E}[L(v^n) \log L(v^n)] \leq n\, \mathbb{E}[L(v^n)] = n. \tag{7.140}$$

Moreover, applying the upper bound of Lemma 6.7 to $L(v^n)$ yields

$$\mathbb{E}[L(v^n) \log L(v^n)] \leq \log(e) \operatorname{var}\bigl(L(v^n)\bigr). \tag{7.141}$$

Combining the previous two bounds we get,

$$\mathbb{E}[L(v^n) \log L(v^n)] \leq \min\bigl\{n, \log(e) \operatorname{var}\bigl(L(v^n)\bigr)\bigr\}. \tag{7.142}$$

Since

$$\begin{aligned} \operatorname{var}\bigl(K_{\mathcal{H}_n^1}(v^n)\bigr) &= \operatorname{var}\bigl(K_{\mathcal{H}_n^2}(v^n)\bigr) \\ &= M 2^{-n} \frac{1}{2} |\mathcal{H}| \Bigl(1 - 2^{-n} \frac{1}{2} |\mathcal{H}|\Bigr) \leq M 2^{-n} \frac{1}{2} |\mathcal{H}|, \end{aligned} \tag{7.143}$$

and $K_{\mathcal{H}_n^1}(v^n)$ and $K_{\mathcal{H}_n^2}(v^n)$ are negatively correlated, using (7.138) we get

$$\operatorname{var}\bigl(L(v^n)\bigr) \leq 2(\rho_n^2 + (1-\rho_n)^2) \frac{2^n}{M|\mathcal{H}_n|} \leq 2 \frac{2^n}{M|\mathcal{H}_n|}. \tag{7.144}$$

Using (7.144) in (7.142) and the linearity of the expectation together with (7.139), we conclude that

$$\mathbb{E}\bigl[D(P_{\tilde{V}^n|N=n} \| P_V^n)\bigr] \leq \min\Bigl\{n, 2\log_2(e) \frac{2^n}{M|\mathcal{H}_n|}\Bigr\}. \tag{7.145}$$

Since $\Pr\{N = n\} = \Pr\{\Phi^n \in \mathcal{H}_n^1\} + \Pr\{\Phi^n \in \mathcal{H}_n^2\}$ and $\Pr\{\Phi^n \in \mathcal{H}_n^1\} \geq \Pr\{\Phi^n \in \mathcal{H}_n^2\}$ (cf. (7.126)),

$$\Pr\{N = n\} \leq 2 \Pr\{\Phi^n \in \mathcal{H}_n^1\} = |\mathcal{H}_n| p^{w(n)} (1-p)^{n-w(n)}. \tag{7.146}$$

Multiplying the right-hand sides of (7.145) and (7.146), we get

$$\begin{aligned}
&\mathbb{E}[D(P_{\tilde{V}^n|N=n}\|P_V^n)]\Pr\{N=n\} \\
&\quad\le c_1 \min\Big\{n|\mathcal{H}_n|p^{w(n)}(1-p)^{n-w(n)}, \frac{2^n}{M}p^{w(n)}(1-p)^{n-w(n)}\Big\} && (7.147)\\
&\quad= c_1 p^{w(n)}(1-p)^{n-w(n)} \min\Big\{n|\mathcal{H}_n|, \frac{2^n}{M}\Big\} && (7.148)\\
&\quad\overset{(a)}{\le} c_1 p^{w(n)}(1-p)^{n-w(n)} \min\Big\{n2^{nh_2(w(n)/n)}, \frac{2^n}{M}\Big\} && (7.149)\\
&\quad\overset{(b)}{\le} c_1 p^{w(n)}(1-p)^{n-w(n)} \min\Big\{n2^{n[1-\alpha^{-1}\log(M)/n]}, \frac{2^n}{M}\Big\} && (7.150)\\
&\quad= c_1 2^n p^{w(n)}(1-p)^{n-w(n)} \min\Big\{n\frac{1}{M^{1/\alpha}}, \frac{1}{M}\Big\} && (7.151)\\
&\quad\le c_1 n 2^n p^{w(n)}(1-p)^{n-w(n)} \min\Big\{\frac{1}{M^{1/\alpha}}, \frac{1}{M}\Big\} && (7.152)\\
&\quad= c_1 n 2^n p^{w(n)}(1-p)^{n-w(n)} \frac{1}{M^{\max\{1,1/\alpha\}}}. && (7.153)
\end{aligned}$$

In the above $c_1 := 2\log(e)$, (a) holds since $\mathcal{H}_n$ is a subset of all binary sequences of length n and Hamming weight $w(n)$, and (b) follows by substituting the value of $w(n)$. (Recall that the binary entropy function is increasing when its argument is below 1/2.)

Since the stopping rule is independent of the choice of $\mathscr{C}$, using (7.153) in the definition of average output divergence $\mathbb{D}$ (see (7.8)) we get

$$\mathbb{E}[\mathbb{D}] \le c_1 \frac{1}{M^{\max\{1,1/\alpha\}}} \sum_{n\in\mathcal{N}} n2^n p^{w(n)}(1-p)^{n-w(n)}. \tag{7.154}$$

Define $\forall n \in \mathbb{N}$,

$$\gamma(n) := h_2^{-1}\Big(1-\alpha^{-1}\frac{\log(M)}{n}\Big) \in [0,1/2], \tag{7.155}$$

and note that $w(n) > n\gamma(n) - 1$. Therefore (since $p < \frac{1}{2}$),

$$2^n p^{w(n)}(1-p)^{n-w(n)} \le c_2 2^n p^{n\gamma(n)}(1-p)^{n[1-\gamma(n)]} \tag{7.156}$$

where $c_2 := \frac{1-p}{p}$. Thus, we can further upper-bound (7.154) as

$$\mathbb{E}[\mathbb{D}] \le c_1 c_2 \frac{1}{M^{\max\{1,1/\alpha\}}} \sum_{n\in\mathcal{N}} n2^n p^{\gamma(n)}(1-p)^{n-\gamma(n)}. \tag{7.157}$$

Moreover, we have

$$\begin{aligned}
2^n p^{n\gamma(n)}(1-p)^{n-n\gamma(n)} &= 2^{-n[d_2(\gamma(n)\|p)+h_2(\gamma(n))-]} && (7.158)\\
&\overset{(*)}{\le} 2^{-n[h_2(\gamma(n))-1]} && (7.159)\\
&= M^{1/\alpha} && (7.160)
\end{aligned}$$

(where $d_2(\cdot\|\cdot)$ is the binary KL divergence defined in (7.77)). In the above $(*)$ is equality iff $\gamma(n) = p$.

Let $\gamma^\star$ be the solution of

$$d_2(\gamma\|p) + h_2(\gamma) = 1. \tag{7.161}$$

(Note that $\gamma^\star \in (p, 1/2)$, because the right-hand side of the above is continuous and increasing in γ, at $\gamma = p$ it evaluates to $h_2(p) < 1$, and at $\gamma = 1/2$ evaluates to $d_2(1/2\|p) + 1 > 1$.)

Pick any $q \in (\gamma^\star, \frac{1}{2})$ and partition $\mathcal{N} = \mathcal{N}_1 \cup \mathcal{N}_2$ where

$$\mathcal{N}_1 \coloneqq \{n \in \mathcal{N} : \gamma(n) \leq q\} \tag{7.162}$$
$$\mathcal{N}_2 \coloneqq \{n \in \mathcal{N} : \gamma(n) > q\}. \tag{7.163}$$

Accordingly, split the summation in (7.157) as

$$\sum_{n\in\mathcal{N}} n2^n p^{\gamma(n)}(1-p)^{n-\gamma(n)} = \sum_{n\in\mathcal{N}_1} n2^n p^{\gamma(n)}(1-p)^{n-\gamma(n)} + \sum_{n\in\mathcal{N}_2} n2^n p^{\gamma(n)}(1-p)^{n-\gamma(n)}. \tag{7.164}$$

Since $\gamma(n)$ is increasing in n,

$$\sup\{n \in \mathcal{N}_1\} \leq \frac{\alpha^{-1}}{1-h_2(q)} \log(M). \tag{7.165}$$

Consequently, the first summation in (7.164) contains at most $\frac{\alpha^{-1}}{1-h_2(q)} \log(M) \eqqcolon c_3 \log(M)$ terms and, in view of (7.160), is upper-bounded[3] as

$$\sum_{n\in\mathcal{N}_1} n2^n p^{\gamma(n)}(1-p)^{n-\gamma(n)} \leq \big[c_3 \log(M)\big]^2 M^{1/\alpha}. \tag{7.166}$$

Defining $\eta \coloneqq d_2(q\|p) + h_2(q) - 1 > 0$, the second summation in (7.164) is upper bounded as

$$\sum_{n\in\mathcal{N}_2} n2^n p^{\gamma(n)}(1-p)^{n-\gamma(n)} = \sum_{n\in\mathcal{N}_2} n2^{-n[d_2(\gamma(n)\|p)+h_2(\gamma(n))-1]} \overset{\text{(a)}}{\leq} \sum_{n\in\mathcal{N}_2} n2^{-n\eta} \overset{\text{(b)}}{\leq} \sum_{n=0}^{\infty} n2^{-n\eta} = \frac{2^\eta}{[2^\eta - 1]^2}, \tag{7.167}$$

where (a) holds since $d_2(\gamma\|p) + h_2(\gamma)$ is increasing in γ and (b) follows by extending the sum to all integers. Since for large M, the bound of (7.166) is (much) larger than that of (7.167), in view of (7.164) we have

$$\sum_{n\in\mathcal{N}} n2^n p^{\gamma(n)}(1-p)^{n-\gamma(n)} \leq 2\big[c_3 \log(M)\big]^2 M^{1/\alpha}. \tag{7.168}$$

[3] Since $q > \gamma^\star > p$, for large M this summation contains n for which $\gamma(n) \approx p$. Therefore, the bound is tight up to logarithmic factors.

Using (7.168) in (7.157) we get

$$\begin{aligned}\mathbb{E}[\mathbb{D}] &\leq 2c_1c_2\big[c_3\log(M)\big]^2\frac{1}{M^{\max\{1,1/\alpha\}}}M^{1/\alpha}\\ &= 2c_1c_2\big[c_3\log(M)\big]^2M^{-[\alpha-1]^+/\alpha} \qquad (7.169)\end{aligned}$$

Thus, at least for half of the choices of $\mathscr{C}$,

$$\mathbb{D} \leq 4c_1c_2\big[c_3\log(M)\big]^2M^{-[\alpha-1]^+/\alpha} \qquad (7.170)$$

Now we are ready to conclude the proof of Theorem 7.4. We need to show that given and R and $\delta > 0$, we can tune the parameter α such that with large enough M,

$$\frac{\log(M)}{\mathbb{E}[N]} \leq R \qquad (7.171)$$

and

$$\mathbb{D} \leq \exp\{-\mathbb{E}[N]([R-1+h_2(p)]^+ - \delta)\}. \qquad (7.172)$$

Let $\delta' > 0$ be a small positive number. In view of Lemma 7.6, there exists $M_0(\delta')$ such that for $\forall M \geq M_0(\delta')$,

$$\frac{\log(M)}{\mathbb{E}[N]} \leq \alpha[1-h_2(p)] + \delta' \qquad (7.173)$$

Thus, taking $\alpha = [R-\delta']/[1-h_2(p)]$ in the above guarantees (7.171). Moreover, for any good code satisfying (7.170),

$$\begin{aligned}-\frac{\log(\mathbb{D})}{\mathbb{E}[N]} &= \frac{\log(\mathbb{D})}{\log(M)}\frac{\log(M)}{\mathbb{E}[N]} && (7.174)\\ &\overset{(a)}{\geq} \frac{\log(\mathbb{D})}{\log(M)}\big(\alpha[1-h_2(p)]-\delta'\big) && (7.175)\\ &\overset{(b)}{\geq} \Big(\frac{[\alpha-1]^+}{\alpha} - \epsilon_M\Big)\big(\alpha[1-h_2(p)]-\delta'\big) && (7.176)\\ &= [\alpha-1]^+[1-h_2(p)]\\ &\quad -\Big\{\big(\alpha[1-h_2(p)]-\delta'\big)\epsilon_M + \frac{[\alpha-1]^+}{\alpha}\delta'\Big\} && (7.177)\\ &\overset{(c)}{=} [R-1+h_2(p)-\delta']^+\\ &\quad -\Big\{(R-2\delta')\epsilon_M + \frac{[R-1+h_2(p)-\delta']^+}{R-\delta'}\delta'\Big\} && (7.178)\\ &\overset{(d)}{\geq} [R-1+h_2(p)]^+\\ &\quad -\Big\{\delta' + (R-2\delta')\epsilon_M + \frac{[R-1+h_2(p)-\delta']^+}{R-\delta'}\delta'\Big\} && (7.179)\end{aligned}$$

where (a) follows from the lower bound of (7.94) and (b) follows from (7.170) where ϵ_M is a quantity that approaches 0 as M grows large, (c) by substituting the value of α and (d) since $[a-\delta']^+ \geq [a]^+ - \delta'$.

Since ϵ_M approaches 0 as $M \to \infty$, by choosing δ' small enough and M sufficiently large, we can make sure that the term inside the curly brackets in (7.179) is smaller than δ to guarantee (7.172). □

7.4 Summary and Outlook

In this chapter, we have studied the problem of channel resolvability in the presence of feedback. We have shown that, although feedback does not decrease the channel resolution, in the presence of causal feedback, higher resolvability exponents are achievable, compared to what the block resolvability codes discussed in Chapter 6 attain.

Specifically, we have shown that we can perfectly simulate an i.i.d. string at the output of an erasure channel by using an external entropy rate *just above* the resolution of the erasure channel. We remark that for the counterpart problem, i.e., error correction, a similar method (retransmission of every bit until it is received unerased) permits zero-error data transmission over an erasure channel at rates arbitrarily close to its capacity, in the presence of feedback.

Moreover, we have also proposed a resolvability scheme that achieves the straight-line resolvability exponent $[R - I(U;V)]^+$, over a binary symmetric channel. This result is the analogue of establishing the achievability of the error exponent $[I(U;V) - R]^+$ in the presence of feedback (cf. [111, Section 2.1]) for communication. Burnashev's *optimal* exponent for error correction [18,23] is also a straight line but with a steeper slope. It would be interesting to investigate whether the results presented in this chapter can be extended to show the achievability of the exponent $[R - I(U;V)]^+$ in the presence of feedback over general channels and with respect to arbitrary target distributions.

Theorems 7.2 and 7.4 suggest that, by using variable-length resolvability schemes, much higher resolvability exponents compared to the exponents of Chapter 6 (which are optimal for an average code) are achievable in the presence of feedback. These improvements are, in essence, due to rate-adaptation gains that variable-length coding provides. In fact, for the settings considered in Theorems 7.2 and 7.4, Lemmas 7.3 and 7.5, respectively, upper-bound the best attainable resolvability exponent by using block-codes in the presence of feedback. Comparing those upper bounds to the exponents of Theorems 7.2 and 7.4 gives rise to two important conclusions: First, employing variable-length resolvability encoders is necessary; even the best block encoder cannot attain the exponents of Theorems 7.2 and 7.4, at least at rates close to channel's resolution. Second, the improvements in the exponent, demonstrated in this chapter, are exclusively due to the presence of feedback.

We remark that the bounds of Lemmas 7.3 and 7.5 — specifically Equations (7.76) and (7.90) — for the system without feedback were already implied by [121, Corollary 6]. We have shown (in Appendix 7.A) that, using a method similar to that of [121], we can establish the same bounds in the presence of feedback.

It is worth noting that, for the channel coding problem, Dobrushin [34] and Haroutunian [49] upper-bounded the best attainable error exponent in the presence of feedback by using block codes (This upper bound equals the sphere-packing exponent for symmetric channels [34] but is larger than that, for asymmetric ones [30, Exercise 10.36].) These results imply that employing variable-length error-correction schemes is necessary to achieve the higher exponents of [23]. Lemmas 7.3 and 7.5 give similar conclusions for channel resolvability; feedback does not increase the upper bound on the best attainable resolvability exponent by using block codes for the settings considered in theses lemmas.

7.A Upper Bounds on the Best Attainable Exponents

In this section we upper-bound the best achievable resolvability exponent by using block codes for the system with feedback to establish Lemmas 7.3 and 7.5. Our method is largely based on that of [121].

In view of Pinsker's inequality [30, Exercise 3.18], Lemma 7.3 follows if we prove that for any sequence of block resolvability codes (indexed by their block-length n), for the system with feedback,

$$\limsup_{n\to\infty} -\frac{1}{n}\log|P_{\tilde{V}^n} - P_V^n| \leq \begin{cases} 0 & \text{if } R < (1-p), \\ d_2(R\|1-p) & \text{if } R \geq (1-p). \end{cases} \tag{7.180}$$

Likewise, to prove Lemma 7.5 it suffices to show

$$\limsup_{n\to\infty} -\frac{1}{n}\log|P_{\tilde{V}^n} - P_V^n| \\ \leq \begin{cases} 0 & \text{if } R < 1-h_2(p), \\ d_2\big(p - \frac{R-1+h_2(p)}{\log[(1-p)/p]} \,\|\, p\big) & \text{if } 1-h_2(p) \leq R < \log[2(1-p)], \\ +\infty & \text{if } R \geq \log[2(1-p)]. \end{cases} \tag{7.181}$$

It is well known that for any two probability measures P and Q on $\mathcal{V}^n$

$$|P - Q| = 2\sup_{\mathcal{A}\subseteq\mathcal{V}^n}\{P(\mathcal{A}) - Q(\mathcal{A})\}. \tag{7.182}$$

Recall that a block resolvability code maps a uniformly distributed word $J \in \{1, 2, \ldots, M\}$ into a sequence of length-n of channel input symbols $\tilde{U}^n$ that are observed at the output of the channel as $\tilde{V}^n$. The collection $(J, \tilde{U}^n, \tilde{V}^n)$ has distribution

$$\Pr\{J = j, \tilde{U}^n = u^n, \tilde{V}^n = v^n\} = \frac{1}{M}\prod_{i=1}^{n} \mathbb{1}\{\mathrm{Enc}_i(j, \tilde{v}^{i-1}) = \tilde{u}_i\} P_{V|U}(\tilde{v}_i|\tilde{u}_i). \tag{7.183}$$

Therefore,

$$\Pr\{\tilde{V}^n = v^n | J = j\} = \prod_{i=1}^{n} P_{V|U}\big(v_i \mid \mathrm{Enc}_i(j, v^{i-1})\big). \tag{7.184}$$

For each $j = 1, 2, \ldots, M$, let

$$\mathcal{A}_j \coloneqq \Big\{ v^n \in \mathcal{V}^n \colon \frac{\Pr\{\tilde{V}^n = v^n | J = j\}}{P_V^n(v^n)} \geq 2M \Big\}, \tag{7.185}$$

Define also

$$\mathcal{A} \coloneqq \bigcup_{j=1}^{M} \mathcal{A}_j \tag{7.186}$$

In view of (7.182),

$$\frac{1}{2}|P_{\tilde{V}^n} - P_V^n| \geq P_{\tilde{V}^n}(\mathcal{A}) - P_V^n(\mathcal{A}). \tag{7.187}$$

We also have

$$P_{\tilde{V}^n}(\mathcal{A}) = \frac{1}{M}\sum_{j=1}^{M}\sum_{v^n \in \mathcal{V}^n} \Pr\{\tilde{V}^n = v^n | J = j\} \mathbb{1}\{v^n \in \mathcal{A}\} \tag{7.188}$$

$$\geq \frac{1}{M}\sum_{j=1}^{M}\sum_{v^n \in \mathcal{V}^n} \Pr\{\tilde{V}^n = v^n | J = j\} \mathbb{1}\{v^n \in \mathcal{A}_j\}, \tag{7.189}$$

where the inequality follows since $\mathcal{A}_j \subseteq \mathcal{A}$ (for every $j = 1, 2, \ldots, M$). Moreover, the union bound implies

$$P_V^n(\mathcal{A}) \leq \sum_{j=1}^{M} P_V^n(\mathcal{A}_j) = \sum_{j=1}^{M}\sum_{v^n \in \mathcal{V}^n} P_V^n(v^n)\mathbb{1}\{v^n \in \mathcal{A}_j\}. \tag{7.190}$$

Combining (7.189) and (7.190) in (7.187) yields

$$\frac{1}{2}|P_{\tilde{V}^n} - P_V^n| \geq \frac{1}{M}\sum_{j=1}^{M}\sum_{v^n \in \mathcal{V}^n} \big[\Pr\{\tilde{V}^n = v^n | J = j\} - M P_V^n(v^n)\big] \cdot \mathbb{1}\Big\{\frac{\Pr\{\tilde{V}^n = v^n | J = j\}}{P_V^n(v^n)} \geq 2M\Big\}. \tag{7.191}$$

7.A.1 Proof of Lemma 7.3

Given $j \in \{1, 2, \ldots, M\}$ and for a fixed block resolvability encoder, define

$$\mathcal{V}_n(j) \coloneqq \big\{ v^n \in \{0, 1, ?\}^n \colon \forall i = 1, 2, \ldots, n, v_i \in \{?, \mathrm{Enc}_i(v^{i-1}, j)\}\big\}. \tag{7.192}$$

Using the transition probabilities of $\mathsf{BEC}(p)$ and (7.184) we get

$$\Pr\{\tilde{V}^n = v^n | J = j\} = p^{n_e(v^n)}(1-p)^{n-n_e(v^n)}\mathbb{1}\{v^n \in \mathcal{V}_n(j)\}, \tag{7.193}$$

where for $v^n \in \{0, 1, ?\}^n$,

$$n_e(v^n) := \bigl|\{i \in \{1, 2, \ldots, n\} \colon v_i = ?\}\bigr| \tag{7.194}$$

is the number of erasures in v^n. Moreover

$$P_V^n(v^n) = p^{n_e(v^n)}\Bigl(\frac{1-p}{2}\Bigr)^{n-n_e(v^n)}. \tag{7.195}$$

Consequently (7.191) is simplified as

$$\begin{aligned}\frac{1}{2}|P_{\tilde{V}^n} - P_V^n| \geq \frac{1}{M}\sum_{j=1}^{M}\sum_{v^n \in \mathcal{V}^n}&\bigl[\Pr\{\tilde{V}^n = v^n | J = j\} - MP_V^n(v^n)\bigr] \\ &\cdot \mathbb{1}\Bigl\{\frac{\Pr\{\tilde{V}^n = v^n | J = j\}}{P_V^n(v^n)} \geq 2M\Bigr\} &(7.196)\\ = \frac{1}{M}\sum_{j=1}^{M}\sum_{v^n \in \mathcal{V}^n}&\Bigl[p^{n_e(v^n)}(1-p)^{n-n_e(v^n)}\mathbb{1}\{v^n \in \mathcal{V}_n(j)\} \\ &- Mp^{n_e(v^n)}\Bigl(\frac{1-p}{2}\Bigr)^{n-n_e(v^n)}\Bigr]\mathbb{1}\Bigl\{\frac{\mathbb{1}\{v^n \in \mathcal{V}_n(j)\}}{2^{-[n-n_e(v^n)]}} \geq 2M\Bigr\} &(7.197)\\ = \frac{1}{M}\sum_{j=1}^{M}\sum_{e=0}^{n}&K_j(e)p^e(1-p)^{n-e}\bigl[1 - M2^{-(n-e)}\bigr]\mathbb{1}\bigl\{2^{n-e} \geq 2M\bigr\}. &(7.198)\end{aligned}$$

where in the last step we have defined

$$K_j(e) := \bigl|\{v^n \in \mathcal{V}_n(j) \colon n_e(v^n) = e\}\bigr|. \tag{7.199}$$

We now note that there is a one-to-one correspondence between the set of sequences in $\mathcal{V}_n(j)$ with e erasures and the set of binary sequences of Hamming weight e: Given any sequence $v^n \in \mathcal{V}_n(j)$ with e erasures, we can construct a binary sequence of Hamming weight e by putting ones at coordinates where $v_i = ?$ and zeros elsewhere. Conversely, given a binary sequence of Hamming weight e, $b^n \in \{0, 1\}^n$, we can construct a sequence $v^n \in \mathcal{V}_n(j)$ by setting $v_i = \mathrm{Enc}_i(j, v^{i-1})$ if $b_i = 0$ and $v_i = ?$ otherwise, for $i = 1, 2, \ldots, n$. Therefore,

$$K_j(e) = \binom{n}{e}. \tag{7.200}$$

Using the above in (7.198) yields

$$\frac{1}{2}|P_{\tilde{V}^n} - P_V^n| \geq \sum_{e=0}^{n} \binom{n}{e} p^e(1-p)^{n-e}\big[1 - M2^{-(n-e)}\big]\mathbb{1}\big\{2^{n-e} \geq 2M\big\} \tag{7.201}$$

$$= \sum_{e=0}^{\lfloor n-\log(M)-1\rfloor} \binom{n}{e} p^e(1-p)^{n-e}\big[1 - M2^{-(n-e)}\big] \tag{7.202}$$

$$\overset{\text{(a)}}{\geq} \frac{1}{2} \sum_{e=0}^{\lfloor n-\log(M)-1\rfloor} \binom{n}{e} p^e(1-p)^{n-e} \tag{7.203}$$

$$\overset{\text{(b)}}{\geq} \frac{1}{2(n+1)} \sum_{e=0}^{\lfloor n-\log(M)-1\rfloor} 2^{-nd_2(e/n\|p)} \tag{7.204}$$

$$\geq \frac{1}{n+1} \max_{e\in\{0,1,\dots,\lfloor n-\log(M)-1\rfloor\}} 2^{-nd_2(e/n\|p)}. \tag{7.205}$$

In the above (a) follows since for $e \leq n - \log(M) - 1$, $M2^{-(n-e)} \leq \frac{1}{2}$ and (b) follows since

$$\binom{n}{k} \geq \frac{1}{(n+1)} 2^{nh_2(k/n)}. \tag{7.206}$$

Using the continuity of binary divergence, (7.205) yields

$$\limsup_{n\to\infty} -\frac{1}{n}\log|P_{\tilde{V}^n} - P_V^n| \leq \min_{0\leq q\leq 1-R} d_2(q\|p) \tag{7.207}$$

$$= \begin{cases} 0 & \text{if } R \leq 1-p, \\ d_2(1-R\|p) & \text{if } R > 1-p, \end{cases} \tag{7.208}$$

which establishes (7.180). □

7.A.2 Proof of Lemma 7.5

Let

$$\tilde{w}_j(v^n) := \big|\big\{i \in \{1,2,\dots,n\} \colon v_i \neq \mathrm{Enc}_i(j, v^{i-1})\big\}\big|. \tag{7.209}$$

As $P_{V|U}$ is the transition probability of a $\mathsf{BSC}(p)$, (7.184), yields

$$\Pr\{\tilde{V}^n = v^n | J = j\} = p^{\tilde{w}_j(v^n)}(1-p)^{n-\tilde{w}_j(v^n)}. \tag{7.210}$$

Furthermore, as P_V is assumed to be the uniform distribution on $\{0,1\}$, (7.191) is simplified as

$$\frac{1}{2}|P_{\tilde{V}^n} - P_V^n| \geq \frac{1}{M}\sum_{j=1}^{M}\sum_{v^n\in\mathcal{V}^n}\big[\Pr\{\tilde{V}^n = v^n|J=j\} - MP_V^n(v^n)\big]$$

$$\cdot\, \mathbb{1}\Big\{\frac{\Pr\{\tilde{V}^n = v^n|J=j\}}{P_V^n(v^n)} \geq 2M\Big\} \tag{7.211}$$

$$= \frac{1}{M} \sum_{j=1}^{M} \sum_{v^n \in \mathcal{V}^n} \left[p^{\tilde{w}_j(v^n)} (1-p)^{n-\tilde{w}_j(v^n)} - M2^{-n} \right] \cdot \mathbb{1}\left\{ \frac{p^{\tilde{w}_j(v^n)}(1-p)^{n-\tilde{w}_j(v^n)}}{2^{-n}} \geq 2M \right\} \quad (7.212)$$

$$= \frac{1}{M} \sum_{j=1}^{M} \sum_{w=0}^{n} K_j(w) \left[p^w (1-p)^{n-w} - M2^{-n} \right] \cdot \mathbb{1}\left\{ \frac{p^w(1-p)^{n-w}}{2^{-n}} \geq 2M \right\}, \quad (7.213)$$

where in the last step we have defined

$$K_j(w) \coloneqq \left| \left\{ v^n \in \{0,1\}^n \colon \tilde{w}_j(v^n) = w \right\} \right|. \quad (7.214)$$

Again, we observe a one-to-one correspondence between the elements of the set $\{v^n \in \{0,1\}^n \colon \tilde{w}_j(v^n) = w\}$ and the binary sequences of Hamming weight w: Any v^n with $\tilde{w}_j(v^n) = w$ can be mapped to a binary sequence of Hamming weight w by putting ones at coordinates for which $v_i \neq \mathrm{Enc}_i(j, v^{i-1})$ and zeros elsewhere. Conversely, given a binary sequence $b^n \in \{0,1\}^n$ of Hamming weight w we can construct v^n with $\tilde{w}_j(v^n) = w$ by setting $v_i = \mathrm{Enc}_i(j, v^{i-1}) \oplus b_i$. Consequently,

$$K_j(w) = \binom{n}{w}, \qquad \forall j = 1, 2, \ldots, M. \quad (7.215)$$

Using (7.215) in (7.213) yields

$$\frac{1}{2} |P_{\tilde{V}^n} - P_V^n| \geq \sum_{w=0}^{n} \binom{n}{w} \left[p^w (1-p)^{n-w} - M2^{-n} \right] \mathbb{1}\left\{ \frac{p^w(1-p)^{n-w}}{2^{-n}} \geq 2M \right\}. \quad (7.216)$$

As $p \leq 1/2$,

$$\sum_{w=0}^{n} \binom{n}{w} \left[p^w (1-p)^{n-w} - M2^{-n} \right] \mathbb{1}\left\{ \frac{p^w(1-p)^{n-w}}{2^{-n}} \geq 2M \right\} = \sum_{w=0}^{\lfloor \tau_n \rfloor} \binom{n}{w} \left[p^w (1-p)^{n-w} - M2^{-n} \right] \quad (7.217)$$

where, with $R \coloneqq \log(M)/n$,

$$\tau_n \coloneqq n \Big[p - \frac{1}{\log[(1-p)/p]} \Big(R + \frac{1}{n} - [1 - h_2(p)] \Big) \Big]. \quad (7.218)$$

Therefore, we can lower-bound $|P_{\tilde{V}^n} - P_V^n|$ as

$$\frac{1}{2}|P_{\tilde{V}^n} - P_V^n| \geq \sum_{w=0}^{\lfloor \tau_n \rfloor} \binom{n}{w} \left[p^w(1-p)^{n-w} - M2^{-n}\right] \tag{7.219}$$

$$\overset{\text{(a)}}{\geq} \sum_{w=0}^{\lfloor \tau_n \rfloor} \binom{n}{w} \frac{1}{2} p^w(1-p)^{n-w} \tag{7.220}$$

$$\overset{\text{(b)}}{\geq} \frac{1}{2(n+1)} \sum_{w=0}^{\lfloor \tau_n \rfloor} 2^{-nd_2(w/n\|p)} \tag{7.221}$$

$$\geq \frac{1}{2(n+1)} \max_{w\in\{0,1,\ldots,\lfloor \tau_n \rfloor\}} 2^{-nd_2(w/n\|p)}. \tag{7.222}$$

where (a) follows as for $w \leq \lfloor \tau_n \rfloor$, $M2^{-n} \leq \frac{1}{2}p^w(1-p)^{n-w}$ by definition, and (b) holds because of (7.206). Consequently, as

$$\lim_{n\to\infty} \frac{\lfloor \tau_n \rfloor}{n} = p - \frac{R - [1 - h_2(p)]}{\log[(1-p)/p]} =: q(R), \tag{7.223}$$

and the binary divergence function is continuous, we get

$$\limsup_{n\to\infty} -\frac{1}{n} \log |P_{\tilde{V}^n} - P_V^n| \leq \min_{q\in[0,q(R)]} d_2(q\|p). \tag{7.224}$$

Finally, we note that

$$\min_{q\in[0,q(R)]} d_2(q\|p) = \begin{cases} 0 & q(R) \geq p \\ d_2(q(R)\|p) & 0 < q(R) < p \\ +\infty & q(R) \leq 0. \end{cases} \tag{7.225}$$

This establishes (7.181) and concludes the proof. □

7.B Proof of Equation (7.101)

The upper bound follows from concavity of $h_2(\cdot)$. To establish the lower bound, recall that $h_2'(p) = \log\left[\frac{1-p}{p}\right]$ and $h_2''(p) = -\frac{\log(e)}{p(1-p)}$. Therefore,

$$h_2(p+\epsilon) - h_2(p) - \epsilon h_2'(p) = (p+\epsilon)\log\left[\frac{p}{p+\epsilon}\right] + (1-p-\epsilon)\log\left[\frac{1-p}{1-p-\epsilon}\right]$$

$$= -\left\{(p+\epsilon)\log\left[1+\frac{\epsilon}{p}\right] + (1-p-\epsilon)\log\left[1-\frac{\epsilon}{1-p}\right]\right\} \tag{7.226}$$

$$\overset{(*)}{\geq} -\log(e)\left\{(p+\epsilon)\frac{\epsilon}{p} - (1-p-\epsilon)\frac{\epsilon}{1-p}\right\} \tag{7.227}$$

$$= -\frac{\log(e)}{p(1-p)}\epsilon^2. \tag{7.228}$$

In the above $(*)$ follows since $\log(1+s) \leq \log(e)s$. □

7.C A Bound on the Empirical Second Moment of Stopped Martingales

Lemma 7.7. *Let $(\xi_n,\ n \in \mathbb{N})$ be i.i.d. zero-mean random variables and*

$$S_n \coloneqq \sum_{i=1}^{n} \xi_n, \qquad n \in \mathbb{N}.$$

Then the process $(S_n,\ n \in \mathbb{N})$ is a martingale with respect to the natural filtering $\big(\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n),\ n \in \mathbb{N}\big)$ and, if N is a stopping time,

$$\mathbb{E}\left[\frac{S_N^2}{N}\right] \leq \operatorname{var}(\xi_1)\,\mathbb{E}[1 + \ln(N)]. \tag{7.229}$$

Proof. That $(S_n,\ n \in \mathbb{N})$ is a martingale is trivial. We only prove (7.229). Let

$$N_m \coloneqq \min\{N, m\}, \qquad \forall m \in \mathbb{N}. \tag{7.230}$$

It is clear that $\forall m \in \mathbb{N}$, $N_m \in \{1, 2, \dots, m\}$, almost surely and N_m is a stopping time. The latter can be verified by noting that

$$\{N_m = n\} = \begin{cases} \{N = n\} & \text{if } n < m, \\ \{N \geq m\} & \text{if } n = m. \end{cases} \tag{7.231}$$

Thus for $n < m$, $\{N_m = n\} = \{N = n\} \in \mathcal{F}_n$ by the hypothesis that N is a stopping time, and for $n = m$,

$$\{N_m = m\} = \{N \geq m\} = \bigcap_{j=1}^{m-1} \{N \neq j\} \in \mathcal{F}_{m-1}, \tag{7.232}$$

and $\mathcal{F}_{m-1} \subseteq \mathcal{F}_m$ (hence $\{N_m = m\} \in \mathcal{F}_m$). Finally $N_1 = 1$ almost surely, hence,

$$\mathbb{E}\left[\frac{S_{N_1}^2}{N_1}\right] = \operatorname{var}(\xi_1). \tag{7.233}$$

We now have

$$\begin{aligned} \mathbb{E}\left[\frac{S_{N_m}^2}{N_m}\right] - \mathbb{E}\left[\frac{S_{N_{m-1}}^2}{N_{m-1}}\right] &= \mathbb{E}\left[\left(\frac{S_m^2}{m} - \frac{S_{m-1}^2}{m-1}\right)\mathbb{1}\{N \geq m\}\right] & (7.234) \\ &= \mathbb{E}\left[\frac{(m-1)\big(\xi_m^2 + 2\xi_m S_{m-1}\big) - S_{m-1}^2}{(m-1)m}\mathbb{1}\{N \geq m\}\right] \\ &\leq \frac{1}{m}\big(\mathbb{E}[\xi_m^2 \mathbb{1}\{N \geq m\}] + 2\,\mathbb{E}[\xi_m S_{m-1} \mathbb{1}\{N \geq m\}]\big) \\ &\overset{(*)}{=} \frac{1}{m}\operatorname{var}(\xi_m)\Pr\{N \geq m\}. & (7.235) \end{aligned}$$

In the above $(*)$ follows since, as shown in (7.232), $\{N \geq m\} \in \mathcal{F}_{m-1}$; thus $\mathbb{1}\{N \geq m\}$ is independent of ξ_m.

Using (7.235) repeatedly together with the fact that $\forall n \in \mathbb{N}$, $\mathrm{var}(\xi_n) = \mathrm{var}(\xi_1)$, we get

$$\mathbb{E}\left[\frac{S_{N_m}^2}{N_m}\right] \leq \mathbb{E}\left[\frac{S_{N_1}^2}{N_1}\right] + \mathrm{var}(\xi_1) \sum_{\ell=2}^{m} \frac{\Pr\{N \geq \ell\}}{\ell} \tag{7.236}$$

$$\overset{(*)}{=} \mathrm{var}(\xi_1) \sum_{\ell=1}^{m} \frac{\Pr\{N \geq \ell\}}{\ell}$$

$$\leq \mathrm{var}(\xi_1) \sum_{\ell \geq 1} \frac{\Pr\{N \geq \ell\}}{\ell}. \tag{7.237}$$

where $(*)$ follows from (7.233) and the fact that $N \geq 1$ almost surely. We finally have

$$\sum_{\ell \geq 1} \frac{\Pr\{N \geq \ell\}}{\ell} = \sum_{n \geq 1} \Pr\{N = n\} \sum_{\ell=1}^{n} \frac{1}{\ell} \tag{7.238}$$

$$\leq \sum_{n \geq 1} \Pr\{N = n\}(1 + \ln(n)) = \mathbb{E}[1 + \ln(N)]. \tag{7.239}$$

Using the above in (7.237) yields

$$\mathbb{E}\left[\frac{S_{N_m}^2}{N_m}\right] \leq \mathbb{E}[1 + \ln(N)], \qquad \forall m \in \mathbb{N}. \tag{7.240}$$

Now, since $\lim_{m\to\infty} N_n = N$ with probability 1

$$\mathbb{E}\left[\frac{S_N^2}{N}\right] = \mathbb{E}\left[\lim_{m\to\infty} \frac{S_{N_m}^2}{N_m}\right] = \mathbb{E}\left[\liminf_{m\to\infty} \frac{S_{N_m}^2}{N_m}\right] \overset{(a)}{\leq} \liminf_{m\to\infty} \mathbb{E}\left[\frac{S_{N_m}^2}{N_m}\right] \overset{(b)}{\leq} \mathbb{E}[1+\ln(N)], \tag{7.241}$$

where (a) follows from Fatou's lemma (applied to the sequence of non-negative random variables $S_{N_m}^2/N_m$, $m \in \mathbb{N}$) and (b) follows from (7.240). □

A Ordering the Noisy Channels

Let $W : \mathcal{X} \to \mathcal{Y}$ and $V : \mathcal{X} \to \mathcal{Z}$ be two discrete-input memoryless channels (DMCs) with the same input alphabet $\mathcal{X}$ and arbitrary output alphabets. We are interested in *comparing* these channels. For ease of presentation, we assume the channel has also a discrete output alphabet. The results presented here can straightforwardly be extended to continuous-output channels.

Definition A.1. The channel $V : \mathcal{X} \to \mathcal{Z}$ is *degraded* with respect to the channel $W : \mathcal{X} \to \mathcal{Y}$ (equivalently, W is *upgraded* with respect to V) iff there exist a stochastic matrix $Q : \mathcal{Y} \to \mathcal{Z}$ such that

$$V(z|x) = \sum_{y \in \mathcal{Y}} Q(z|y) W(y|x) \qquad \forall z \in \mathcal{Z}, \forall x \in \mathcal{X} \tag{A.1}$$

We write $V \preceq_{\mathrm{d}} W$ or $W \succeq_{\mathrm{d}} V$ to denote this.

Degradedness means by processing the output of W one can *simulate* V (see Figure A.1).

Definition A.2. The channels $W : \mathcal{X} \to \mathcal{Y}$ and $V : \mathcal{X} \to \mathcal{Z}$ are “equivalent” if $W \preceq_{\mathrm{d}} V$ and $W \succeq_{\mathrm{d}} V$.

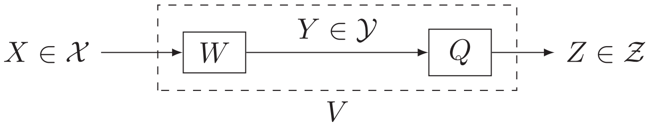


Figure A.1: $V \preceq_{\mathrm{d}} W$ or Equivalently $W \succeq_{\mathrm{d}} V$

Lemma A.1. *Suppose $W : \mathbb{F}_2 \to \mathcal{Y}$ is a binary-input DMC. Assume there exist two output symbols y' and y'' in its output alphabet $\mathcal{Y}$ with the same likelihood ratio,*

$$\frac{W(y'|0)}{W(y'|1)} = \frac{W(y''|0)}{W(y''|1)}. \tag{A.2}$$

Then, W is equivalent (in the sense of Definition A.2) to $W' : \mathbb{F}_2 \to \mathcal{Y} \setminus \{y', y''\} \cup \{(y', y'')\}$ with transition probabilities

$$W'(y|x) = \begin{cases} W(y|x) & \textit{if } y \neq (y', y'') \\ W(y'|x) + W(y''|x) & \textit{if } y = (y', y''). \end{cases} \tag{A.3}$$

Proof. Let $\mathcal{Y}' \coloneqq \mathcal{Y} \setminus \{y', y''\} \cup \{(y', y'')\}$ for the sake of brevity. Taking the degrading channel, $Q : \mathcal{Y} \to \mathcal{Y}'$, as the one that maps y' and y'' to (y', y'') (and acts as the identity channel on all other symbols), namely,

$$Q(y_1|y_2) = \mathbb{1}\big\{\big(y_2 \notin \{y', y''\} \wedge y_1 = y_2\big) \vee \big(y_2 \in \{y', y''\} \wedge y_1 = (y', y'')\big)\big\},$$

shows that $W' \preceq_{\mathrm{d}} W$. Whereas, because of the assumption (A.2), taking the channel $Q' : Q' \to \mathcal{Y}$ as

$$Q'(y_1|y_2) = \begin{cases} 1 & \text{if } y_2 \neq (y', y'') \text{ and } y_1 = y_2 \\ \frac{W(y'|0)}{W(y'|0)+W(y''|0)} & \text{if } y_2 = (y', y'') \text{ and } y_1 = y' \\ \frac{W(y''|0)}{W(y'|0)+W(y''|0)} & \text{if } y_2 = (y', y'') \text{ and } y_1 = y'' \end{cases} \tag{A.4}$$

shows that $W \preceq_{\mathrm{d}} W'$. □

Definition A.3. A channel $W : \mathcal{X} \to \mathcal{Y}$ is *less noisy* than channel $V : \mathcal{X} \to \mathcal{Z}$ if for every distribution P_{UX} such that $U \multimap X \multimap (Y, Z) \sim P_{UX}(u, x)W(y|x)V(z|x)$,

$$I(U; Y) \geq I(U; Z) \tag{A.5}$$

We write $V \preceq_{\mathrm{n}} W$ to show W is less noisy than V.

Definition A.4. A channel $W : \mathcal{X} \to \mathcal{Y}$ is *more capable* than channel $V : \mathcal{X} \to \mathcal{Z}$ if for every input distribution P_X, with $(X, Y, Z) \sim P_X(x)W(y|x)V(z|x)$,

$$I(X; Y) \geq I(X; Z) \tag{A.6}$$

We write $V \preceq_{\mathrm{c}} W$ if W is more capable than V.

Degradedness is a strictly stronger condition than being less noisy which is itself strictly stronger than being more capable [37, § 5.6]

Channel degradation and Arıkan's polar transform (cf. § 2.2) commute:

Theorem A.2 ([63, Lemma 4.7])**.** *Let $W : \mathcal{X} \to \mathcal{Y}$ and $V : \mathcal{X} \to \mathcal{Z}$ be two discrete memoryless channels. Then, if $V \preceq_{\mathrm{d}} W$, $\forall m \in \mathbb{N}$, $\forall \mathsf{s}^m \in \{-, +\}^m$,*

$$V^{\mathsf{s}^m} \preceq_{\mathrm{d}} W^{\mathsf{s}^m}.$$

We further know that if W is less noisy than V, the synthetic channels obtained from W are more capable (but not necessarily less noisy) than the corresponding synthetic channels obtained from V.

Theorem A.3 ([105, Theorem 10])**.** *Let $W : \mathcal{X} \to \mathcal{Y}$ and $V : \mathcal{X} \to \mathcal{Z}$ be two discrete memoryless channels. Then, if $V \preceq_{\mathrm{n}} W$, $\forall m \in \mathbb{N}$, and $\forall \mathsf{s}^m \in \{-,+\}^m$,*

$$V^{\mathsf{s}^m} \preceq_{\mathrm{c}} W^{\mathsf{s}^m}.$$

Useful Results from Convex Analysis

B

Lemma B.1. *Let $f\colon A \to \mathbb{R}$ be a convex and continuous function over a convex domain A and $L : A \to B$ be a continuous linear mapping from A onto B — i.e., $B = L(A)$ is the image of A under L. Define $g\colon B \to \mathbb{R}$ as*

$$g(y) \coloneqq \inf_{\substack{x\in A:\\ L(x)=y}} f(x). \tag{B.1}$$

Then g is convex in y hence continuous in the interior of B. Furthermore, if A is also compact, then g is continuous everywhere in B.

Proof. Pick y_1 and y_2 in B. By definition $\forall \epsilon > 0$, there exists $x_1^\star$ and $x_2^\star$ in A with

$$L(x_1^\star) = y_1 \qquad \text{and} \qquad L(x_2^\star) = y_2, \tag{B.2}$$

such that

$$g(y_1) + \epsilon > f(x_1^\star) \qquad \text{and} \qquad g(y_2) + \epsilon > f(x_2^\star). \tag{B.3}$$

Then for any $\lambda \in [0 : 1]$,

$$\begin{aligned}
\lambda g(y_1) + (1-\lambda) g(y_2) + \epsilon &> \lambda f(x_1^\star) + (1-\lambda) f(x_2^\star) && \text{(B.4)}\\
&\overset{\text{(a)}}{\geq} f\big(\lambda x_1^\star + (1-\lambda) x_2^\star\big) && \text{(B.5)}\\
&\overset{\text{(b)}}{\geq} \min_{\substack{x\in A:\\ L(x)=\lambda y_1+(1-\lambda)y_2}} f(x) = g\big(\lambda y_1 + (1-\lambda) y_2\big) && \text{(B.6)}
\end{aligned}$$

where (a) follows by the convexity of f and (b) because

$$L\big(\lambda x_1^\star + (1-\lambda) x_2^\star\big) = \lambda L(x_1^\star) + (1-\lambda) L(x_2^\star) \lambda y_1 + (1-\lambda) y_2, \tag{B.7}$$

as L is linear.

As (B.6) holds for every positive ϵ, it implies

$$\lambda g(y_1) + (1-\lambda)g(y_2) \geq g\big(\lambda y_1 + (1-\lambda)y_2\big) \tag{B.8}$$

which establishes the convexity of g.

Convexity implies continuity in the interior of B. The only discontinuity points of g could be at the boundaries of the set B where it might only be upper semi-continuous. In other words, for a sequence of points $(y_n, n \in \mathbb{N})$ in B with $y \coloneqq \lim_{n\to\infty} y_n$ on the boundaries of B,

$$\limsup_{n\to\infty} g(y_n) \leq g(y). \tag{B.9}$$

We now prove that if A is compact, then g is lower semi-continuous at any point, hence is continuous everywhere. Because L is a bounded linear mapping, the compactness of A implies B is also compact. Further, if A is compact the infimum in (B.1) is indeed a minimum.

Let $(y_n, n \in \mathbb{N})$ be *any* sequence of points in B and $y \coloneqq \lim_{n\to\infty} y_n$. Note that $y \in B$ since B is compact. Define again

$$x_n^\star \coloneqq \arg\min_{\substack{x\in A:\\ L(x)=y_n}} f(x), \tag{B.10}$$

and $x \coloneqq \lim_{n\to\infty} x_n^\star$ (by passing to a subsequence if necessary). The compactness of A implies $x \in A$. Since f is assumed to be continuous,

$$\lim_{n\to\infty} g(y_n) = \lim_{n\to\infty} f(x_n^\star) = f(x) \tag{B.11}$$

Because L is continuous

$$\lim_{n\to\infty} L(x_n^\star) = L(x). \tag{B.12}$$

Moreover,

$$\lim_{n\to\infty} L(x_n^\star) = \lim_{n\to\infty} y_n = y. \tag{B.13}$$

Hence, $L(x) = y$ which yields

$$f(x) \geq \min_{\substack{x'\in A:\\ L(x')=y}} f(x') = g(y). \tag{B.14}$$

Combining (B.11) and (B.14) yields

$$\liminf_{n\to\infty} g(y_n) \geq g(y). \tag{B.15}$$

Therefore, g is lower semi-continuous at any point hence is continuous over the entire set B. □

Lemma B.2. *Let $f : A \to \mathbb{R}^+$ be a non-negative, convex, and continuous function over a convex domain A and $g : A \to \mathbb{R}$ a continuous function. Let x_0 be a global minimizer of f in A. Then, if $g(x_0) \leq \alpha$*

$$\inf_{\substack{x \in A:\\ g(x) \geq \alpha}} f(x) = \inf_{\substack{x \in A:\\ g(x) = \alpha}} f(x) \tag{B.16}$$

Moreover, if x_0 is the unique *global minimizer, $\forall x_1 \in A$ with $g(x_1) > \alpha$,*

$$f(x_1) > \inf_{\substack{x \in A:\\ g(x) = \alpha}} f(x). \tag{B.17}$$

Proof. Obviously

$$\inf_{\substack{x \in A:\\ g(x) \geq \alpha}} f(x) \leq \inf_{\substack{x \in A:\\ g(x) = \alpha}} f(x). \tag{B.18}$$

To establish (B.16) we will show that

$$\inf_{\substack{x \in A:\\ g(x) \geq \alpha}} f(x) \geq \inf_{\substack{x \in A:\\ g(x) = \alpha}} f(x). \tag{B.19}$$

By definition, $\forall \epsilon > 0$, there exists $x^\star \in A$ with $g(x^\star) \geq \alpha$ such that

$$\inf_{\substack{x \in A:\\ g(x) \geq \alpha}} f(x) + \epsilon > f(x^\star). \tag{B.20}$$

Define $\gamma : [0,1] \to \mathbb{R}$ as

$$\gamma(\lambda) := g\big(\lambda x^\star + (1-\lambda)x_0\big). \tag{B.21}$$

Since g is continuous so is γ. Moreover $\gamma(0) = g(x_0) \leq \alpha$ and $\gamma(1) = g(x^\star) \geq \alpha$. Thus, there exists $0 \leq \lambda^\star \leq 1$ for which

$$\gamma(\lambda^\star) = \alpha. \tag{B.22}$$

Therefore, (B.20) yields

$$\begin{align}
\inf_{\substack{x \in A:\\ g(x) \geq \alpha}} f(x) + \epsilon &> f(x^\star) \tag{B.23}\\
&\overset{\text{(a)}}{\geq} \lambda^\star f(x^\star) + (1-\lambda^\star) f(x_0) \tag{B.24}\\
&\overset{\text{(b)}}{\geq} f\big(\lambda^\star x^\star + (1-\lambda^\star)x_0\big) \tag{B.25}\\
&\overset{\text{(c)}}{\geq} \inf_{\substack{x \in A:\\ g(x) = \alpha}} f(x). \tag{B.26}
\end{align}$$

where (a) follows as x_0 is the global minimizer of f and (b) from the convexity of f and (c) because

$$g\big(\lambda^\star x^\star + (1-\lambda^\star)x_0\big) = \alpha. \tag{B.27}$$

As (B.26) holds for every positive ϵ, it implies (B.19).

To prove (B.17), we use the same type of argument. Specifically, we define

$$\gamma'(\lambda) \coloneqq g\big(\lambda x_1 + (1-\lambda)x_0\big) \tag{B.28}$$

and note that there exists $0 \leq \lambda^\star < 1$ such that $\gamma'(\lambda^\star) = \alpha$. (Note that $\gamma'(1) > \alpha$ by assumption.) Therefore,

$$\begin{aligned} f(x_1) &\overset{\text{(a)}}{>} \lambda^\star f(x_1) + (1-\lambda^\star) f(x_0) && \text{(B.29)} \\ &\overset{\text{(b)}}{\geq} f\big(\lambda^\star x_1 + (1-\lambda^\star)x_0\big) && \text{(B.30)} \\ &\overset{\text{(c)}}{\geq} \inf_{\substack{x \in A: \\ g(x) = \alpha}} f(x) && \text{(B.31)} \end{aligned}$$

where (a) follows since $\lambda^\star < 1$ and x_0 is the unique global minimizer and (b) and (c) from the convexity of f and the fact that $g(\lambda^\star x_1 + (1-\lambda^\star)x_0) = \alpha$, respectively. □

Lemma B.3. *Let $f\colon A \to \mathbb{R}$ a convex function over some convex domain A and $L\colon A \to \mathbb{R}$ a linear function. Then, $g\colon \mathbb{R} \to \mathbb{R}$, defined as*

$$g(y) \coloneqq \sup_{\substack{x \in A: \\ f(x) \leq y}} L(x) \tag{B.32}$$

is concave in y.

Proof. Take $y_1, y_2 \in \mathbb{R}$. By definition, for every $\epsilon > 0$, there exists $x_1^\star \in A$ and $x_2^\star \in A$ with

$$f(x_1^\star) \leq y_1 \qquad \text{and} \qquad f(x_2^\star) \leq y_2, \tag{B.33}$$

such that

$$L(x_1^\star) > g(y_1) - \epsilon \qquad \text{and} \qquad L(x_2^\star) > g(y_2) - \epsilon. \tag{B.34}$$

Since L is linear, for any $\lambda \in [0:1]$,

$$\begin{aligned} L\big(\lambda x_1^\star + (1-\lambda)x_2^\star\big) &= \lambda L(x_1^\star) + (1-\lambda)L(x_2^\star) && \text{(B.35)} \\ &> \lambda g(y_1) + (1-\lambda)g(y_2) - \epsilon. && \text{(B.36)} \end{aligned}$$

Moreover, as f is convex,

$$f(\lambda x_1^\star + (1-\lambda)x_2^\star) \leq \lambda f(x_1^\star) + (1-\lambda) f(x_2^\star) \leq \lambda y_1 + (1-\lambda) y_2. \tag{B.37}$$

Consequently,

$$L\big(\lambda x_1^\star + (1-\lambda)x_2^\star\big) \leq \sup_{\substack{x \in A: \\ f(x) \leq \lambda y_1 + (1-\lambda)y_2}} L(x) = g\big(\lambda y_1 + (1-\lambda)y_2\big). \tag{B.38}$$

Combining (B.36) and (B.38) we have

$$g\big(\lambda y_1 + (1-\lambda)y_2\big) > \lambda g(y_1) + (1-\lambda)g(y_2) - \epsilon. \tag{B.39}$$

As (B.39) must hold for every positive ϵ, we conclude that

$$g\big(\lambda y_1 + (1-\lambda)y_2\big) \geq \lambda g(y_1) + (1-\lambda)g(y_2).$$

□

C Types versus Distributions

Recall that given a discrete alphabet $\mathcal{A}$, a distribution $P \in \mathcal{P}(\mathcal{A})$ is an n-type, iff $\forall a \in \mathcal{A}$, $nP(a) \in \mathbb{Z}$ and that we denote the set of n-types over alphabet $\mathcal{A}$ as $\mathcal{P}_n(\mathcal{A}) \subsetneq \mathcal{P}(\mathcal{A})$ (see § 6.1).

The union of n-types $\bigcup_{n\in\mathbb{N}} \mathcal{P}_n(\mathcal{A})$ is dense in the set of distributions on $\mathcal{A}$:

Lemma C.1. *Given any distribution $P \in \mathcal{P}(\mathcal{A})$ there exists a sequence of n-types ($P^{(n)} \in \mathcal{P}_n(\mathcal{A})$, $n \in \mathbb{N}$) such that*

$$\lim_{n\to\infty} |P - P^{(n)}| = 0. \tag{C.1}$$

Moreover the sequence of n-types can be chosen such that $P^{(n)} \ll P$ for all $n \in \mathbb{N}$ and, for $n \geq \lceil 1/P_{\min} \rceil$, $\operatorname{supp}(P^{(n)}) = \operatorname{supp}(P)$. Here, we have defined,

$$P_{\min} \coloneqq \min_{a\in\operatorname{supp}(P)} P(a). \tag{C.2}$$

Proof. Without loss of generality, suppose the support of P equals $\mathcal{A}$ (if not, just start with a smaller $\mathcal{A}$). Define, for every n, $\mathcal{A}_n \subseteq \mathcal{A}$ as

$$\mathcal{A}_n \coloneqq \{a \in \mathcal{A} : nP(a) \notin \mathbb{Z}\} \tag{C.3}$$

and let

$$k_n(a) \coloneqq \lfloor nP(a) \rfloor. \tag{C.4}$$

Obviously, if $a \in \mathcal{A}_n$

$$k_n(a) > nP(a) - 1 \tag{C.5}$$

otherwise,

$$k_n(a) = nP(a) \tag{C.6}$$

Therefore,

$$n - |\mathcal{A}_n| < \sum_{a \in \mathcal{A}} k_n(a) \leq n \tag{C.7}$$

Finally, let

$$r_n := n - \sum_{a \in \mathcal{A}} k_n(a). \tag{C.8}$$

Let $\mathcal{B}_n$ be any arbitrary subset of size $r_n < |\mathcal{A}_n|$ of $\mathcal{A}_n$ and set

$$\tilde{k}_n(a) := \begin{cases} k_n(a) & \text{if } a \notin \mathcal{B}_n \\ k_n(a) + 1 & \text{if } a \in \mathcal{B}_n \end{cases} \tag{C.9}$$

The integers $\tilde{k}_n(a)$, $a \in \mathcal{A}$ have the following properties:

(i) $\tilde{k}_n(a) \geq 0$ (by construction); and

(ii) $\sum_a \tilde{k}_n(a) = n$.

Consequently

$$P^{(n)}(a) := \frac{\tilde{k}(a)}{n} \tag{C.10}$$

is a valid n-type. Moreover, $\forall a \in \mathcal{A}$,

$$|\tilde{k}_n(a) - nP(a)| \leq 1 \tag{C.11}$$

hence,

$$|P^{(n)} - P| \leq \frac{|\mathcal{A}|}{n}, \tag{C.12}$$

which goes to 0 as $n \to \infty$.

Finally, note that if $n > 1/P_{\min}$, then $\forall a \in \mathcal{A}$

$$nP(a) > 1. \tag{C.13}$$

Therefore, $\forall a \in \mathcal{A}$, $k_n(a) > 0$. □

Constructing a sequence of n-types that approximate a given distribution was trivial: we just had to quantize the distribution to obtain an n-type. We sometimes need to quantize distributions to n-types with a prescribed marginal:

Lemma C.2. *Let $P \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a joint distribution on $\mathcal{X} \times \mathcal{Y}$ and P_X be its x-marginal. Let $Q_X^{(n)} \in \mathcal{P}_n(\mathcal{X})$ be an arbitrary sequence of n-types that converges to P_X. Then, there exists a sequence of n-types $P^{(n)} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ such that*

$$\lim_{n \to \infty} |P - P^{(n)}| = 0 \tag{C.14}$$

and

$$P_X^{(n)} = Q_X^{(n)}, \qquad \forall n \in \mathbb{N} \tag{C.15}$$

Moreover, $P^{(n)} \ll P$ for all $n \in \mathbb{N}$.

Proof. Once again, without loss of generality, assume $\mathrm{supp}(P_X) = \mathcal{X}$.

We know that every n, and every $x \in \mathcal{X}$, there exists an $nQ_X^{(n)}(x)$-type that is close to $P_{Y|X}(\cdot|x)$. More precisely, using Lemma C.1 we can construct an $nQ_X^{(n)}(x)$-type, call it $P_{Y|X}^{(n)}(\cdot|x)$, such that

$$\left|P_{Y|X}^{(n)}(\cdot|x) - P_{Y|X}(\cdot|x)\right| \leq \frac{|\mathcal{Y}|}{nQ_X^{(n)}(x)} \tag{C.16}$$

and $P_{Y|X}^{(n)}(\cdot|x) \ll P_{Y|X}(\cdot|x)$.

Now it is easy to check that $P^{(n)}(x,y) \coloneqq Q_X^{(n)}(x) \times P_{Y|X}^{(n)}(y|x)$ is an n-type with x-marginal $Q_X^{(n)}$. Moreover

$$\begin{aligned}\left|P^{(n)} - P\right| &= \sum_{x,y} \left|Q_X^{(n)}(x)P_{Y|X}^{(n)}(y|x) - P_X(x)P_{Y|X}(y|x)\right| && \text{(C.17)}\\ &\leq \sum_x Q_X^{(n)}(x) \sum_y \left|P_{Y|X}^{(n)}(y|x) - P_{Y|X}(y|x)\right| \\ &\quad + \sum_x \left|Q_X^{(n)}(x) - P_X(x)\right| \sum_y P_{Y|X}(y|x) && \text{(C.18)}\\ &\leq \frac{|\mathcal{X}||\mathcal{Y}|}{n} + \left|Q_X^{(n)} - P_X\right|. && \text{(C.19)}\end{aligned}$$

The above upper bound converges to 0 as $n \to \infty$ by the assumption that $\lim_{n\to\infty} |Q_X^{(n)} - P_X| = 0$. □

A more interesting scenario is to ask for quantizing a distribution such that *both* of the marginals of the quantized n-types match desired n-types. This is also possible but requires taking care of a few technical details.

Definition C.1. Given any joint distribution $P \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ the *support graph* of P is the bipartite graph $G = (V, E)$ with $V \coloneqq \mathcal{X} \cup \mathcal{Y}$ (here we assume, without loss of generality, that $\mathcal{X} \cap \mathcal{Y} = \emptyset$) and

$$E \coloneqq \{e = (x,y)\colon P(x,y) > 0\} \tag{C.20}$$

That is, $x \in \mathcal{X}$ is connected to $y \in \mathcal{Y}$ iff P puts a positive mass on (x,y).

Definition C.2. If We call $(\mathcal{X}_i \times \mathcal{Y}_i,\ i = 1,2,\ldots,k)$ the *block-partitioning* of the support of P if,

(i) $\mathrm{supp}(P) \subseteq \bigcup_{i=1}^{k} (\mathcal{X}_i \times \mathcal{Y}_i)$;

(ii) $(\mathcal{X}_1, \mathcal{X}_2, \ldots, \mathcal{X}_k)$ (respectively $(\mathcal{Y}_1, \mathcal{Y}_2, \ldots, \mathcal{Y}_k)$) partitions $\mathcal{X}$ (resp. $\mathcal{Y}$);

(iii) $\forall i = 1,2,\ldots,k$, the component corresponding to vertices in $\mathcal{X}_i \cup \mathcal{Y}_i$ of the connectivity graph of P (cf. Definition C.1) is connected.

(iv) $\forall i \neq j$, the component corresponding to vertices in $\mathcal{X}_i \cup \mathcal{Y}_i \cup \mathcal{X}_j \cup \mathcal{Y}_j$ in the connectivity graph of P is disconnected.

Obviously the block-partitioning of the support of a distribution is unique (up to reordering the sets) as it corresponds to partitioning the connectivity graph of the distribution to connected components.

Lemma C.3. *Let $P \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ be a joint distribution on $\mathcal{X} \times \mathcal{Y}$ and P_X and P_Y) its x- and y-marginals, respectively. Let $Q_X^{(n)} \in \mathcal{P}_n(\mathcal{X})$ and $Q_Y^{(n)} \in \mathcal{P}_n(\mathcal{Y})$ be two arbitrary sequences of n-types that converge to P_X and P_Y respectively. Let $(\mathcal{X}_i \times \mathcal{Y}_i,\ i = 1, 2, \ldots, k)$ be the block-partitioning of the support of P. Then, if*

$$Q_X^{(n)}(\mathcal{X}_i) = Q_Y^{(n)}(\mathcal{Y}_i), \qquad \forall i = 1, 2, \ldots, k, \tag{C.21}$$

there exists a sequence of n-types $P^{(n)} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y})$ such that

$$\lim_{n \to \infty} |P - P^{(n)}| = 0 \tag{C.22}$$

and

$$P_X^{(n)} = Q_X^{(n)}, \qquad \textit{and} \qquad P_Y^{(n)} = Q_Y^{(n)} \qquad \forall n \in \mathbb{N}. \tag{C.23}$$

Moreover, $P^{(n)} \ll P$ for all $n \in \mathbb{N}$.

Remark. Before proving Lemma C.3, it is worthwhile to mention that the conditions on the support of P are rather mild and, in most applications, we do not need to worry about them. For example, if the support graph of P (cf. Definition C.1) is connected (C.21) obviously holds. Another very common situation is when $Q_X^{(n)}$ and $Q_Y^{(n)}$ are marginals of a joint n-type $Q^{(n)}$, that is absolutely continuous with respect to P. Then, we can easily verify that even if the support graph of P is disconnected, (C.21) is automatically satisfied.

Proof of Lemma C.3. By the virtue of Lemma C.1, there exist a sequence of n-types $(\tilde{P}^{(n)} \in \mathcal{P}_n(\mathcal{X} \times \mathcal{Y}), n \in \mathbb{N})$ that converge to P as n grows large. Obviously the marginals of $\tilde{P}^{(n)}$s can be different than $Q_X^{(n)}$ and $Q_Y^{(n)}$ but since the latter converge to P_X and P_Y, respectively, the marginals of $\tilde{P}^{(n)}$ are *close* to $Q_X^{(n)}$ and $Q_Y^{(n)}$. It turns out that we can slightly *perturb* $\tilde{P}^{(n)}$ so that its marginals match $Q_X^{(n)}$ and $Q_Y^{(n)}$ *without changing its support.*

The constraint on the support turn out to be quite important. In fact, without such a constraint, much easier solutions exist.

To prove the claim, we shall show that $\forall \epsilon > 0$, $\exists n_0(\epsilon)$ such that $\forall n \geq n_0$,

$$|\tilde{P}^{(n)} - P| \leq \epsilon/2 \tag{C.24}$$

and we can find an integer-value mapping $d \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{Z}$ with the following properties:

1. With

$$r_X(x) \coloneqq n[Q_X^{(n)}(x) - \tilde{P}_X^{(n)}(x)], \qquad \text{and} \tag{C.25}$$

$$r_Y(y) \coloneqq n[Q_Y^{(n)}(y) - \tilde{P}_Y^{(n)}(y)], \tag{C.26}$$

we have

$$\forall x \in \mathcal{X}, \quad \sum_{y \in \mathcal{Y}} d(x,y) = r_X(x), \quad \text{and} \tag{C.27}$$

$$\forall y \in \mathcal{Y}, \quad \sum_{x \in \mathcal{X}} d(x,y) = r_Y(y). \tag{C.28}$$

2. $\forall (x,y) \in \mathcal{X} \times \mathcal{Y}$,

$$d(x,y) + n\tilde{P}^{(n)}(x,y) \geq 0 \tag{C.29}$$

with equality if $\tilde{P}^{(n)}(x,y) = 0$.

3.

$$|d| \coloneqq \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} |d(x,y)| \leq n\epsilon/2 \tag{C.30}$$

(Note that $d(x,y)$ also depends on n, but we do not show this dependence explicitly to keep the notation simple.) If such d can be found,

$$P^{(n)}(x,y) \coloneqq \tilde{P}^{(n)}(x,y) + \frac{1}{n} d(x,y) \tag{C.31}$$

will be an n-type (because d is integer-valued) whose x- and y-marginals are $Q_X^{(n)}$ and $Q_Y^{(n)}$, respectively, (due to the first property) and is absolutely continuous with respect to $\tilde{P}^{(n)}$ (due to the second property). Moreover, due to the third property, using the triangle inequality and (C.24) it easily follows that

$$|P^{(n)} - P| \leq |P^{(n)} - \tilde{P}^{(n)}| + |\tilde{P}^{(n)} - P| \leq \epsilon. \tag{C.32}$$

Hence, we have constructed the sequence of n-types with desired properties.

Let us assume $n \geq \lceil 1/P_{\min} \rceil$ (see (C.2)) so that $\tilde{P}^{(n)}$ has the same support as P and pick

$$\gamma \coloneqq \min\{(2/5)P_{\min}, \epsilon/(4|\mathcal{X}||\mathcal{Y}|)\}. \tag{C.33}$$

By definition, $\exists n_1(\gamma)$ such that for $n \geq n_1$,

$$|\tilde{P}^{(n)} - P| \leq \gamma/2. \tag{C.34}$$

(Hence, for such n (C.24) holds.) Moreover, (C.34) implies

$$|\tilde{P}_X^{(n)} - P_X| \leq \gamma/2 \qquad \text{and} \qquad |\tilde{P}_Y^{(n)} - P_Y| \leq \gamma/2. \tag{C.35}$$

On the other side, since $Q_X^{(n)}$ and $Q_Y^{(n)}$ converge to P_X and P_Y, respectively, $\exists n_2(\gamma)$ such that for $n \geq n_2$

$$|Q_X^{(n)} - P_X| \leq \gamma/2 \qquad \text{and} \qquad |Q_Y^{(n)} - P_Y| \leq \gamma/2 \tag{C.36}$$

Therefore, triangle inequality implies, for $n \geq \max\{n_1, n_2\}$,

$$|Q_X^{(n)} - \tilde{P}_X^{(n)}| \leq \gamma \qquad \text{and} \qquad |Q_Y^{(n)} - \tilde{P}_Y^{(n)}| \leq \gamma \tag{C.37}$$

Let G be the support graph of $\tilde{P}^{(n)}$ (see Definition C.1). Since $n \geq \lceil 1/P_{\min} \rceil$, this would be the same as the support graph of P.

The graph G has k connected components by definition. Consider the i^{th} connected component of G, and denote it as $G_i = (V_i, E_i)$. We know that $V_i = \mathcal{X}_i \cup \mathcal{Y}_i$. Let $T_i = (V_i, E_i')$, $E_i' \subseteq E_i$ be a spanning tree of that component. Pick any vertex, say $x_0 \in \mathcal{X}_i$, as the root of the tree.[1] Suppose this tree has height H_i. Let $V_i(h)$ denote the set of vertices at height h in the tree. (Thus $V_i(0) = \{x_0\}$.) For every vertex $v \in V_i(h)$, $h \geq 1$, let $p(v) \in V_i(h-1)$ be the parent of v and $K_i(v) := \{u \in V_i(h+1) : (v, u) \in E_i'\}$ be the children of v (with $K(v) = \emptyset$ for the leaves). Consider the following algorithm to populate the edges of the tree T_i with integer values $(d_e, e \in E_i')$:

Algorithm 10:

```
1 for $h = H$ to 1 do
2     foreach $v \in V_i(h)$ do
3         $d_{(v,p(v))} \leftarrow r(v) - \sum_{u \in K_i(v)} d_{(v,u)}$ ;        // $r(v)$ is defined below
```

In line 3 we have used the generic notation

$$r(v) := \begin{cases} r_X(x), & \text{if } v \in \mathcal{X}, \\ r_Y(y), & \text{if } v \in \mathcal{Y}. \end{cases} \tag{C.38}$$

Now, we claim that running Algorithm 10 on each connected component and setting

$$d(x, y) := \begin{cases} d_e & \text{if } (x, y) \in E' \\ 0 & \text{otherwise.} \end{cases} \tag{C.39}$$

results in the desired mapping $d \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{Z}$ with all desired properties.

Let us verify them step by step: Equation (C.27) and (C.28) hold by construction except for the roots of the trees. To prove that (C.27) holds for the roots to too we first note that for every i,

$$\sum_{(x,y) \in \mathcal{X}_i \times \mathcal{Y}_i} d(x, y) = \sum_{y \in \mathcal{Y}_i} \sum_{x \in \mathcal{X}_i} d(x, y) \overset{\text{(a)}}{=} \sum_{y \in \mathcal{Y}_i} r_Y(y) \overset{\text{(b)}}{=} n[Q_Y^{(n)}(\mathcal{Y}_i) - \tilde{P}_Y^{(n)}(\mathcal{Y}_i)] \tag{C.40}$$

where (a) follows since we know (C.28) holds for all $y \in \mathcal{Y}_i$ (and that $d(x, y) = 0$ if $y \in \mathcal{Y}_i$ but $x \notin \mathcal{X}_i$, by construction) and (b) from the definition of r_Y, (C.26).

[1] It is obvious that the root can also be picked from $\mathcal{Y}_i$ too.

Therefore,

$$n[Q_Y^{(n)}(\mathcal{Y}_i) - \tilde{P}_Y^{(n)}(\mathcal{Y}_i)] = \sum_{y\in\mathcal{Y}_i} d(x_0,y) + \sum_{x\in\mathcal{X}_i\setminus\{x_0\}} \sum_{y\in\mathcal{Y}_i} d(x,y) \tag{C.41}$$

$$\overset{(*)}{=} \sum_{y\in\mathcal{Y}_i} d(x_0,y) + \sum_{x\in\mathcal{X}_i\setminus\{x_0\}} r_X(x) \tag{C.42}$$

$$= \sum_{y\in\mathcal{Y}_i} d(x_0,y) - r_X(x_0) + \sum_{x\in\mathcal{X}_i} r_X(x) \tag{C.43}$$

where again $(*)$ follows since (C.27) holds for $x \in \mathcal{X}_i \setminus \{x_0\}$. Moreover, using the definition (C.25),

$$\sum_{x\in\mathcal{X}_i} r_X(x) = n[Q_X^{(n)}(\mathcal{X}_i) - \tilde{P}_X^{(n)}(\mathcal{X}_i)] \tag{C.44}$$

But since $\mathrm{supp}(\tilde{P}^{(n)}) = \mathrm{supp}(P) \subseteq \bigcup_{i=1}^k (\mathcal{X}_i \times \mathcal{Y}_i))$, we have

$$\tilde{P}_X^{(n)}(\mathcal{X}_i) = \tilde{P}_Y^{(n)}(\mathcal{Y}_i). \tag{C.45}$$

Using the above and the assumption (C.21) yields,

$$\sum_{x\in\mathcal{X}_i} r_X(x) = n[Q_X^{(n)}(\mathcal{X}_i) - \tilde{P}_X^{(n)}(\mathcal{X}_i)] = n[Q_Y^{(n)}(\mathcal{Y}_i) - \tilde{P}_Y^{(n)}(\mathcal{Y}_i)]. \tag{C.46}$$

Plugging (C.46) into (C.43) proves

$$\sum_{y\in\mathcal{Y}} d(x_0,y) = \sum_{y\in\mathcal{Y}_i} d(x_0,y) = r_X(x_0). \tag{C.47}$$

To verify (C.29) and (C.30) we use the following bound

$$|d(x,y)| \leq n\big[|Q_X^{(n)} - \tilde{P}_X^{(n)}| + |Q_Y^{(n)} - \tilde{P}_Y^{(n)}|\big] \leq n2\gamma. \tag{C.48}$$

The bound of (C.48) follows easily by nothing that the value associated with each edge of the tree in the algorithm is (in absolute value) at most as large as the sum of the absolute values of the values associated with the vertices in the sub-tree rooted at the end of that edge. This sum can be extended to the whole tree to derive the bound of (C.48).

Now (C.29) can be verified as follows: First, if $\tilde{P}^{(n)}(x,y) = 0$, then $(x,y) \notin E \supset E'$ and, as a consequence, $d(x,y) = 0$. Otherwise,

$$d(x,y) + n\tilde{P}^{(n)}(x,y) \overset{\text{(a)}}{\geq} -2n\gamma + \tilde{P}^{(n)}(x,y) \tag{C.49}$$

$$\overset{\text{(b)}}{\geq} -2n\gamma + nP(x,y) - n\gamma/2 \tag{C.50}$$

$$\geq n[P_{\min} - 5/2\gamma] \tag{C.51}$$

where (a) follows by (C.48) and (b) from (C.34). Since $\gamma \leq \frac{2}{5}P_{\min}$ (see (C.33)) the above is non-negative.

Finally since $\gamma \leq \epsilon/(4|\mathcal{X}||\mathcal{Y}|)$ (again, see (C.33)) the bound of (C.30) follows by summing the left-hand side of (C.48) over all $(x,y) \in \mathcal{X} \times \mathcal{Y}$. □

Bibliography

[1] 3$^{\text{rd}}$ Generation Partnership Project (3GPP). (2017, May) Final report of 3GPP TSG RAN WG1 #87 v1.0.0 (Reno, USA, 14th – 18th November 2016). [Online]. Available: http://www.3gpp.org/ftp/tsg_ran/wg1_rl1/TSGR1_88/Docs/R1-1701552.zip

[2] A. Alamdar-Yazdi and F. R. Kschischang, "A simplified successive-cancellation decoder for polar codes," *IEEE Communications Letters*, vol. 15, no. 12, pp. 1378–1380, Oct. 2011.

[3] M. Alsan and E. Telatar, "A simple proof of polarization and polarization for non-stationary memoryless channels," *IEEE Transactions on Information Theory*, vol. 62, no. 9, pp. 4873–4878, Sep. 2016.

[4] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jul. 2008, pp. 1173–1177.

[5] ——, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3051–3073, Jul. 2009.

[6] E. Arıkan and E. Telatar, "On the rate of channel polarization," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jul. 2009, pp. 1493 –1495.

[7] A. Balatsoukas-Stimming, M. Bastani Parizi, and A. P. Burg, "LLR-based successive cancellation list decoding of polar codes," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 3903–3907.

[8] ——, "LLR-based successive cancellation list decoding of polar codes," *IEEE Transactions on Signal Processing*, vol. 63, no. 19, pp. 5165–5179, Oct. 2015.

[9] ——, "On metric sorting for successive cancellation list decoding of polar codes," in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2015, pp. 1993–1996.

[10] A. Balatsoukas-Stimming, A. J. Raymond, W. J. Gross, and A. Burg, "Hardware architecture for list successive cancellation decoding of polar codes," *IEEE Transactions on Circuits and Systems—Part II: Express Briefs*, vol. 61, no. 8, pp. 609–613, May 2014.

[11] M. Bastani Parizi, "Polar codes: Finite length implementation, error correlations and multilevel modulation," Master's thesis, EPFL, Lausanne, Switzerland, 2012. [Online]. Available: http://infoscience.epfl.ch/record/196935

[12] M. Bastani Parizi and E. Telatar, "On the correlation between polarized BECs," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jul. 2013, pp. 784–788.

[13] ——, "On the secrecy exponent of the wire-tap channel," in *Proceedings of IEEE Information Theory Workshop (ITW)*, Oct. 2015, pp. 287–291.

[14] ——, "On channel resolvability in presence of feedback," in *Proceedings of Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2016, pp. 78–85.

[15] M. Bastani Parizi, E. Telatar, and N. Merhav, "Exact random coding secrecy exponents for the wiretap channel," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jul. 2016, pp. 1521–1525.

[16] ——, "Exact random coding secrecy exponents for the wiretap channel," *IEEE Transactions on Information Theory*, vol. 63, no. 1, pp. 509–531, Jan. 2017.

[17] M. Bellare, S. Tessaro, and A. Vardy, "Semantic security for the wiretap channel," in *Proceedings of the 32nd Annual Cryptology Conference on Advances in Cryptology — CRYPTO 2012*, vol. 7417. New York, NY, USA: Springer-Verlag New York, Inc., 2012, pp. 294–311.

[18] P. Berlin, B. Nakiboglu, B. Rimoldi, and E. Telatar, "A simple converse of burnashev's reliability function," *IEEE Transactions on Information Theory*, vol. 55, no. 7, pp. 3074–3080, Jul. 2009.

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

[20] V. Bioglio, F. Gabry, L. Godard, and I. Land, "Two-step metric sorting for parallel successive cancellation list decoding of polar codes," *IEEE Communications Letters*, vol. 21, no. 3, pp. 456–459, 2017.

[21] M. R. Bloch and J. N. Laneman, "Strong secrecy from channel resolvability," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8077–8098, Dec. 2013.

[22] D. K. Bose, Raj C.and Ray-Chaudhuri, "On a class of error correcting binary group codes," *Information and Control*, vol. 3, no. 1, pp. 68–79, 1960. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0019995860902874

[23] M. V. Burnashev, "Data transmission over a discrete channel with feedback: Random transmission time," *Problemy peredachi informatsii*, vol. 12, no. 4, pp. 250–265, 1976.

[24] T.-H. Chou, V. Y. F. Tan, and S. C. Draper, "The sender-excited secret key agreement model: Capacity, reliability, and secrecy exponents," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 609–627, Jan. 2015.

[25] A. Collins, G. Durisi, T. Erseghe, V. Kostina, J. Östman, Y. Polyanskiy, I. Tal, and W. Yang, "SPECTRE Short Packet Communication Toolbox," http://github.com/yp-mit/spectre, 2017.

[26] D. J. Costello and G. D. Forney, "Channel coding: The road to channel capacity," *Proceedings of the IEEE*, vol. 95, no. 6, pp. 1150–1177, Jun. 2007.

[27] I. Csiszár, "Almost independence and secrecy capacity," *Problems of Information Transmission*, vol. 32, no. 1, pp. 40–47, 1996.

[28] ——, "The method of types," *IEEE Transactions on Information Theory*, vol. 44, no. 6, pp. 2505–2523, Oct. 1998.

[29] I. Csiszár and J. Körner, "Broadcast channels with confidential messages," *IEEE Transactions on Information Theory*, vol. 24, no. 3, pp. 339–348, 1978.

[30] ——, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge University Press, 2011.

[31] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[32] ——, "Soft covering with high probability," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jul. 2016, pp. 2963–2967.

[33] P. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, Sep. 2010.

[34] R. L. Dobrushin, "Asymptotic bounds on the probability of error for the transmission of messages over a memoryless channel using feedback," *Probl. Kibern*, vol. 8, pp. 161–168, 1963.

[35] I. Dumer, "Soft-decision decoding of Reed–Muller codes: a simplified algorithm," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 954–963, Mar. 2006.

[36] I. Dumer and K. Shabunov, "Soft-decision decoding of Reed–Muller codes: recursive lists," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1260–1266, Mar. 2006.

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

[38] P. Elias, "Coding for noisy channels," *Proceedings of the Institute of Radio Engineers*, vol. 43, no. 3, pp. 356–356, 1955.

[39] ——, "List decoding for noisy channels," Massachusetts Institute of Technology, Research Laboratory of Electronics, Tech. Rep., 1957.

[40] Y. Fan and C.-Y. Tsui, "An efficient partial-sum network architecture for semi-parallel polar codes decoder implementation," *IEEE Transactions on Signal Processing*, vol. 62, no. 12, pp. 3165–3179, Jun. 2014.

[41] R. G. Gallager, "Low-density parity-check codes," Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, USA, 1963.

[42] ——, *Information Theory and Reliable Communication.* New York, NY, USA: John Wiley & Sons, Inc., 1968.

[43] ——, "The random coding bound is tight for the average code," *IEEE Transactions on Information Theory*, vol. 19, no. 2, pp. 244–246, Mar. 1973.

[44] ——, *Discrete Stochastic Processes.* Boston, MA, USA: Kluwer, 1996.

[45] M. J. E. Golay, "Notes on digital coding," *Proceedings of The Institute of Radio Engineers*, vol. 37, no. 6, p. 657, Jun. 1949.

[46] R. W. Hamming, "Error detecting and error correcting codes," *The Bell System Technical Journal*, vol. 29, no. 2, pp. 147–160, Apr. 1950.

[47] T. S. Han, H. Endo, and M. Sasaki, "Reliability and secrecy functions of the wiretap channel under cost constraint," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6819–6843, Nov. 2014.

[48] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Transactions on Information Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[49] E. A. Haroutunian, "Lower bound for error probability in channels with feedback," *Problemy peredachi informatsii*, vol. 13, no. 2, pp. 36–44, 1977.

[50] S. A. Hashemi, A. Balatsoukas-Stimming, P. Giard, C. Thibeault, and W. J. Gross, "Partitioned successive-cancellation list decoding of polar codes," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Mar. 2016, pp. 957–960.

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

[52] ——, "Exponential decreasing rate of leaked information in universal random privacy amplification," *IEEE Transactions on Information Theory*, vol. 57, no. 6, pp. 3989–4001, Jun. 2011.

[53] M. Hayashi and R. Matsumoto, "Universally attainable error and information exponents, and equivocation rate for the broadcast channels with confidential messages," in *Proceedings of Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2011, pp. 439–444.

[54] ——, "Secure multiplex coding with dependent and non-uniform multiple messages," *IEEE Transactions on Information Theory*, vol. 62, no. 5, pp. 2355–2409, May 2016.

[55] A. Hocquenghem, "Codes correcteurs d'erreurs," *Chiffres*, vol. 2, no. 2, pp. 147–56, 1959.

[56] J. Honda and H. Yamamoto, "Polar coding without alphabet extension for asymmetric models," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7829–7838, Dec. 2013.

[57] J. Hou and G. Kramer, "Informational divergence approximations to product distributions," in *Proceedings of Canadian Workshop on Information Theory (CWIT)*, Jun. 2013, pp. 76–81.

[58] ——, "Effective secrecy: Reliability, confusion and stealth," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jun. 2014, pp. 601–605.

[59] IEEE Standards Association, "IEEE standard for information technology—telecommunications and information exchange between systems local and metropolitan area networks—specific requirements," *IEEE Std 802.11n^{TM}-2009*, Oct. 2009.

[60] ——, “IEEE standard for air interface for broadband wireless access systems,” *IEEE Std 802.16^{TM}-2012*, Aug. 2012.

[61] ——, “IEEE standard for information technology—telecommunications and information exchange between systems local and metropolitan area networks—specific requirements,” *IEEE Std 802.11ad^{TM}-2012*, Dec. 2012.

[62] B. Y. Kong, H. Yoo, and I.-C. Park, “Efficient sorting architecture for successive-cancellation-list decoding of polar codes,” *IEEE Transactions on Circuits and Systems—Part II: Express Briefs*, vol. 63, no. 7, pp. 673–677, Jul. 2016.

[63] S. B. Korada, “Polar codes for channel and source coding,” Ph.D. dissertation, EPFL, Lausanne, Switzerland, 2009.

[64] J. Körner and A. Sgarro, “Universally attainable error exponents for broadcast channels with degraded message sets,” *IEEE Transactions on Information Theory*, vol. 26, no. 6, pp. 670–679, Nov. 1980.

[65] S. Kudekar, S. Kumar, M. Mondelli, H. D. Pfister, E. Sasoglu, and R. L. Urbanke, “Reed–muller codes achieve capacity on erasure channels,” *IEEE Transactions on Information Theory (to appear)*, 2017. [Online]. Available: http://doi.org/10.1109/TIT.2017.2673829

[66] S. Kudekar, T. Richardson, and R. L. Urbanke, “Spatially coupled ensembles universally achieve capacity under belief propagation,” *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7761–7813, Dec. 2013.

[67] C. Leroux, A. J. Raymond, G. Sarkis, and W. J. Gross, “A semi-parallel successive-cancellation decoder for polar codes,” *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 289–299, Jan. 2013.

[68] C. Leroux, I. Tal, A. Vardy, and W. J. Gross, “Hardware architectures for successive cancellation decoding of polar codes,” in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 1665–1668.

[69] S. K. Leung-Yan-Cheong and M. E. Hellman, “The gaussian wire-tap channel,” *IEEE Transactions on Information Theory*, vol. 24, no. 4, pp. 451–456, Jul. 1978.

[70] B. C. Levy, *Principles of Signal Detection and Parameter Estimation.* New York, NY, USA: Springer, 2008.

[71] J. Lin and Z. Yan, “Efficient list decoder architecture for polar codes,” in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, Jun. 2014, pp. 1022–1025.

[72] ——, “An efficient list decoder architecture for polar codes,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 11, pp. 2508–2518, Nov. 2015.

[73] F. J. MacWilliams and N. J. A. Sloane, *The Theory of Error Correcting Codes*, ser. North-Holland Mathematical Library. North-Holland, 1978.

[74] H. Mahdavifar and A. Vardy, “Achieving the secrecy capacity of wiretap channels using polar codes,” *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6428–6443, Oct. 2011.

[75] J. L. Massey, “Threshold decoding,” Massachusetts Institute of Technology, Research Laboratory of Electronics Cambridge, MA, USA, Tech. Rep., Apr. 1963.

[76] ——, “A simplified treatment of Wyner’s wire-tap channel.” in *Proceedings of Annual Allerton Conference on Communication, Control, and Computing*, Oct. 1983, pp. 268–276.

[77] U. Maurer and S. Wolf, “Information-theoretic key agreement: From weak to strong secrecy for free,” in *Advances in Cyptology — EUROCRYPT 2000*, ser. Lecture Notes in Computer Science, B. Preneel, Ed., vol. 1807. Springer-Verlag, May 2000, pp. 351–368.

[78] N. Merhav, “Statistical physics and information theory,” *Foundations and Trends in Communications and Information Theory*, vol. 6, no. 1–2, pp. 1–212, 2009. [Online]. Available: http://dx.doi.org/10.1561/0100000052

[79] ——, “Exact random coding error exponents of optimal bin index decoding,” *IEEE Transactions on Information Theory*, vol. 60, no. 10, pp. 6024–6031, Oct. 2014.

[80] A. Mishra, A. J. Raymond, L. Amaru, G. Sarkis, C. Leroux, P. Meinerzhagen, A. Burg, and W. J. Gross, “A successive cancellation decoder ASIC for a 1024-bit polar code in 180nm CMOS,” in *Proceedings of IEEE Asian Solid State Circuits Conference*, Nov. 2012, pp. 205–208.

[81] M. Mondelli, R. Urbanke, and S. H. Hassani, “How to achieve the capacity of asymmetric channels,” in *Proceedings of Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2014, pp. 789–796.

[82] R. Mori and T. Tanaka, “Channel polarization on q-ary discrete memoryless channels by arbitrary kernels,” in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jun. 2010, pp. 894–898.

[83] D. E. Muller, "Application of boolean algebra to switching circuit design and to error detection," *Transactions of the Institute of Radio Engineers Professional Group on Electronic Computers*, vol. EC-3, no. 3, pp. 6–12, Sep. 1954.

[84] R. Nasser and E. Telatar, "Polar codes for arbitrary DMCs and arbitrary MACs," *IEEE Transactions on Information Theory*, vol. 62, no. 6, pp. 2917–2936, Jun. 2016.

[85] K. Niu and K. Chen, "CRC-aided decoding of polar codes," *IEEE Communications Letters*, vol. 16, no. 10, pp. 1668–1671, Oct. 2012.

[86] A. Pamuk and E. Arıkan, "A two phase successive cancellation decoder architecture for polar codes," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jul. 2013, pp. 957–961.

[87] W. Park and A. Barg, "Polar codes for q-ary channels, $q = 2^r$," *IEEE Transactions on Information Theory*, vol. 59, no. 2, pp. 955–969, Feb. 2013.

[88] R. Pedarsani, S. H. Hassani, I. Tal, and E. Telatar, "On the construction of polar codes," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jul. 2011, pp. 11–15.

[89] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Transactions on Information Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[90] A. J. Raymond and W. J. Gross, "Scalable successive-cancellation hardware decoder for polar codes," in *Proceedings of IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, Dec. 2013, pp. 1282–1285.

[91] I. S. Reed, "A class of multiple-error-correcting codes and the decoding scheme," *Transactions of the Institute of Radio Engineers Professional Group on Information Theory*, vol. 4, no. 4, pp. 38–49, Sep. 1954.

[92] I. S. Reed and G. Solomon, "Polynomial codes over certain finite fields," *Journal of the society for industrial and applied mathematics*, vol. 8, no. 2, pp. 300–304, 1960.

[93] T. Richardson and R. Urbanke, *Modern Coding Theory.* New York, NY, USA: Cambridge University Press, 2008.

[94] A. G. Sahebi and S. S. Pradhan, "Multilevel channel polarization for arbitrary discrete memoryless channels," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 7839–7857, Dec. 2013.

[95] G. Sarkis, P. Giard, A. Vardy, C. Thibeault, and W. J. Gross, "Fast polar decoders: Algorithm and implementation," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 5, pp. 946–957, May 2014.

[96] G. Sarkis and W. J. Gross, "Increasing the throughput of polar decoders," *IEEE Communications Letters*, vol. 17, no. 4, pp. 725–728, Apr. 2013.

[97] E. Şaşoğlu, "Polar coding theorems for discrete systems," Ph.D. dissertation, EPFL, Lausanne, Switzerland, 2011.

[98] ——, "Polar codes for discrete alphabets," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jul. 2012, pp. 2137–2141.

[99] E. Şaşoğlu, E. Telatar, and E. Arıkan, "Polarization for arbitrary discrete memoryless channels," in *Proceedings of IEEE Information Theory Workshop (ITW)*, Oct. 2009, pp. 144–148.

[100] E. Şaşoğlu and A. Vardy, "A new polar coding scheme for strong security on wiretap channels," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jul. 2013, pp. 1117–1121.

[101] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, pp. 379–423, 623–656, 1948.

[102] ——, "Communication theory of secrecy systems," *The Bell System Technical Journal*, vol. 28, no. 4, pp. 656–715, Oct. 1949.

[103] N. Shulman, "Communication over an unknown channel via common broadcasting," Ph.D. dissertation, Department of Electrical Engineering Systems, Tel Aviv University, Tel Aviv, Israel, 2003.

[104] B. Shuval and I. Tal, "A lower bound on the probability of error of polar codes over BMS channels," `arXiv:1701.01628v1`, Jan. 2017. [Online]. Available: http://arxiv.org/abs/1701.01628v1

[105] D. Sutter and J. M. Renes, "Universal polar codes for more capable and less noisy channels and sources," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jun. 2014, pp. 1461–1465.

[106] D. Sutter, J. M. Renes, F. Dupuis, and R. Renner, "Achieving the capacity of any DMC using only polar codes," in *Proceedings of IEEE Information Theory Workshop (ITW)*, Sep. 2012, pp. 114–118.

[107] I. Tal, "Private communication," Aug. 2014.

[108] I. Tal and A. Vardy, "List decoding of polar codes," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jul. 2011, pp. 1–5.

[109] ——, “How to construct polar codes,” *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6562–6582, Oct. 2013.

[110] ——, “List decoding of polar codes,” *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.

[111] A. Tchamkerten, “Feedback communication over unknown channels,” Ph.D. dissertation, School of Computer and Communication Sciences, EPFL, Lausanne, 2005.

[112] A. J. Viterbi, “Error bounds for convolutional codes and an asymptotically optimum decoding algorithm,” *IEEE Transactions on Information Theory*, vol. 13, no. 2, pp. 260–269, Apr. 1967.

[113] T. Wang, D. Qu, and T. Jiang, “Parity-check-concatenated polar codes,” *IEEE Communications Letters*, vol. 20, no. 12, pp. 2342–2345, Dec. 2016.

[114] Wikipedia, “Polynomial representations of cyclic redundancy checks — wikipedia, the free encyclopedia,” Sep. 2014. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Polynomial_representations_of_cyclic_redundancy_checks&oldid=620254949

[115] J. M. Wozencraft, “List decoding,” Massachusetts Institute of Technology, Research Laboratory of Electronics, Tech. Rep., 1958.

[116] ——, “Sequential decoding for reilable communication,” Massachusetts Institute of Technology, Research Laboratory of Electronics Cambridge, MA, USA, Tech. Rep., Aug. 1957.

[117] A. D. Wyner, “The common information of two dependent random variables,” *IEEE Transactions on Information Theory*, vol. 21, no. 2, pp. 163–179, Mar. 1975.

[118] ——, “The wire-tap channel,” *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355–1387, Oct. 1975.

[119] A. D. Wyner and J. Ziv, “A theorem on the entropy of certain binary sequences and applications: Part I,” *IEEE Transactions on Information Theory*, vol. 19, no. 6, pp. 769–772, Nov. 1973.

[120] H. Yagi and T. S. Han, “Variable-length resolvability for general sources and channels,” `arXiv:1701.08712v1`, Jan. 2017. [Online]. Available: http://arxiv.org/abs/1701.08712

[121] W. Yang, R. F. Schaefer, and H. V. Poor, “Finite-blocklength bounds for wiretap channels,” in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, Jul. 2016, pp. 3087–3091.

[122] B. Yuan and K. K. Parhi, “Low-latency successive-cancellation list decoders for polar codes with multibit decision,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 23, no. 10, pp. 2268–2280, Oct. 2015.

[123] C. Zhang, B. Yuan, and K. K. Parhi, “Reduced-latency SC polar decoder architectures,” in *Proceedings of IEEE International Conference on Communications (ICC)*, Jun. 2012, pp. 3471–3475.

[124] C. Zhang and K. K. Parhi, “Low-latency sequential and overlapped architectures for successive cancellation polar decoder,” *IEEE Transactions on Signal Processing*, vol. 61, no. 10, pp. 2429–2441, Mar. 2013.

[125] ——, “Latency analysis and architecture design of simplified SC polar decoders,” *IEEE Transactions on Circuits and Systems—Part II: Express Briefs*, vol. 61, no. 2, pp. 115–119, Feb. 2014.

[126] C. Zhang, X. You, and J. Sha, “Hardware architecture for list successive cancellation polar decoder,” in *Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS)*, Jun. 2014, pp. 209–212.

[127] H. Zhou, C. Zhang, W. Song, S. Xu, and X. You, “Segmented CRC-aided SC list polar decoding,” in *Proceedings of IEEE Vehicular Technology Conference*, May 2016, pp. 1–5.

Curriculum Vitæ

Mani Bastani Parizi

EPFL IC IINFOCOM LTHI
INR 032 (Bâtiment INR)
Station 14
CH-1015 Lausanne

☎ +41 21 69 31359
✉ mani.bastaniparizi@epfl.ch
🖱 people.epfl.ch/mani.bastaniparizi

Research Interests

- Information and Coding Theory
- Statistical Signal Processing
- Machine Learning
- Wireless Communications

Education

2012–2017 **Ph.D. in Information Theory**
EPFL, Lausanne, Switzerland

2009–2012 **M.Sc. in Communication Systems**,
EPFL, Lausanne, Switzerland

2005–2009 **B.Sc. in Electrical Engineering**,
University of Tehran, Tehran, Iran

Professional Experience

2012–2017 **Doctoral Research Assistant**,
Information Theory Laboratory, EPFL, Switzerland

2012 **Research Assistant** (Internship),
Information Theory Laboratory, EPFL, Switzerland

2011 **M.Sc. Thesis** *"Polar Codes: Finite Length Implementation, Error Correlations, and Multilevel Modulation"*,
Information Theory Laboratory, EPFL, Switzerland

2011 **DSP Engineer** (Internship),
Marvell SÀRL, Etoy, Switzerland

2010 **Research Assistant**,
Information Theory Laboratory, EPFL, Switzerland

2009 **B.Sc. Thesis** *"Analysis of Synchronization and Time-offset Estimation Methods in OFDM Systems"*,
School of ECE, University of Tehran, Iran

2008 **Telecommunications Engineer** (Internship),
Iran Telecommunications Research Center, Tehran, Iran

Teaching Experience

EPFL (2010–2016)

- Advanced Digital Communications
- Applied Probability and Stochastic Processes
- Information Theory and Coding
- Principles of Digital Communications
- Analysis II
- Circuits and Systems II
- Discrete Structures
- Probability and Statistics

University of Tehran (2006–2008)

- Engineering Mathematics
- Probability and Statistics
- C++ Programming

Honors and Awards

2016 EPFL I&C Outstanding Teaching Assistant Award
2015 Finalist for the Best Student Paper Award at IEEE ISCAS 2015
2013 EPFL I&C Outstanding Teaching Assistant Award
2012 EPFL I&C Doctoral School Fellowship
2011 Ranked 2nd (GPA 5.9/6) among Communication System M.Sc. Students of 2012 Class
2010 EPFL Research Assistant Scholarship
2009 Ranked 3rd (GPA 18.27/20) among EE Students of 2009 Class with Telecommunication Major
2005 Ranked 220th among 500,000 participants in the nationwide university entrance exam in Iran

Publications

(1) M. Bastani Parizi and E. Telatar, "On the correlation between polarized BECs," in *Proceedings of 2013 IEEE International Symposium on Information Theory (ISIT)*, Jul. 2013, pp. 784–788.

(2) A. Balatsoukas-Stimming, M. Bastani Parizi, and A. P. Burg, “LLR-based successive cancellation list decoding of polar codes,” in *Proceedings of 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 3903–3907.

(3) A. Balatsoukas-Stimming, M. Bastani Parizi, and A. P. Burg, “LLR-based successive cancellation list decoding of polar codes,” *IEEE Transactions on Signal Processing*, vol. 63, no. 19, pp. 5165–5179, Oct. 2015.

(4) A. Balatsoukas-Stimming, M. Bastani Parizi, and A. P. Burg, “On metric sorting for successive cancellation list decoding of polar codes,” in *Proceedings of 2015 IEEE International Symposium on Cirtcuits and Systems (ISCAS)*, May 2015, pp. 1993–1996.

(5) M. Bastani Parizi and E. Telatar, “On the secrecy exponent of the wiretap channel,” in *Proceedings of 2015 IEEE Information Theory Workshop*, Oct. 2015, pp. 287–291.

(6) M. Bastani Parizi, E. Telatar, and N. Merhav, “Exact random coding secrecy exponents for the wiretap channel,” in *Proceedings of 2016 IEEE International Symposium on Information Theory (ISIT)*, Jul. 2016, pp. 1521–1525.

(7) M. Bastani Parizi and E. Telatar, “On channel resolvability in presence of feedback,” in *Proceedings of 54th Annual Allerton Conference on Communication, Control, and Computing*, Sep. 2016, pp. 78–85.

(8) M. Bastani Parizi, E. Telatar, and N. Merhav, “Exact random coding secrecy exponents for the wiretap channel,” *IEEE Transactions on Information Theory*, vol. 63, no. 1, pp. 509–531, Jan. 2017.

Invited Talks

- “Error Analysis and List Decoder Implementation for Polar Codes,” *IBM Zürich Research Labs,* Nov. 2016.
- “Channel Resolvability: Exact Random Coding Exponents and the Utility of Feedback,“ *Information Theory and Applications Workshop (ITA) — Graduation Day,* San Diego, Feb. 2017.