FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Adaptive Codes for Physical-Layer Security

João Paulo Patriarca de Almeida

Programa Doutoral em Telecomunicações

Orientador: Doutor João Francisco Cordeiro de Oliveira Barros, Professor Associado com Agregação do Departamento de Engenharia Eletrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto

24 de Julho de 2014

Adaptive Codes for Physical-Layer Security

João Paulo Patriarca de Almeida

Programa Doutoral em Telecomunicações

Aprovado em provas públicas pelo Júri:

Presidente: Doutor José Alfredo Ribeiro da Silva Matos, Professor Catedrático da Faculdade de Engenharia da Universidade do Porto

Arguente: Doutor Matthieu Bloch, Assistant Professor, School of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, USA

Vogal: Doutor Mikael Skoglund, Associate Professor, School of Electrical Engineering, KTH Royal Institute of Technology, Stockholm, Sweden;

Vogal: Doutor Adriano Jorge Cardoso Moreira, Professor Associado do Departamento de Sistemas de Informação da Universidade do Minho;

Vogal: Doutor Jaime dos Santos Cardoso, Professor Associado com Agregação do Departamento de Engenharia Eletrotécnica e de Computadores da Faculdade de Engenharia da Universidade do Porto

28 de Maio de 2014

Dedicated to Inês, Aurora and to what the future holds.

In memory of Kika.

Acknowledgments

The work presented in this thesis is a testament to how much I owe to my family, friends and professors. The following lines will certainly follow short on acknowledging how deeply grateful I am of being inspired by so many wonderful people.

First and foremost, a word to Prof. João Barros, the first responsible for having me started on this journey. From the moment we met, João put a high expectation on me and gave me the confidence to pursue any goals I was set out to reach. Among the many things I could thank him for, I choose to thank him for his faith on my work and skills, for the freedom he gave me to work on any research topic I would fall in love with and for always stimulating me with his curiosity and enthusiasm. It was indeed a pleasure to share these last years with him.

Second, I would like to thank the committee members, Professors Matthieu Bloch, Mikael Skoglund, Adriano Moreira, Jaime Cardoso and José Matos, for their availability to be part of the defense jury, but mostly for allowing me to be part of a great discussion on physical-layer security.

I would also like to thank the opportunity given by Professor Matthieu Bloch and Professor Muriel Médard for allowing me to spend some time with their research groups at Georgia Tech Lorraine and MIT, respectively. The experiences have been both rewarding and fulfilling. Additional thanks to Professors Cristiano Torezzan and Willie Harrison who visited us in our lab and from whom I learnt so much!

While the journey to the Ph.D. was long and weary, all my colleagues from the Networking and Information Security group at Instituto de Telecomunicações (IT-Porto) made it a lot more bearable. A salute to the groups former students João Vilela, Lu isa Lima, Mate Boban and Sérgio Crisóstomo and best wishes for all of you who are waiting in line: Hana, João Rodrigues, Mari, Pedro, Rui Meireles, Saurabh and Susana.

I was fortunate enough to meet some of my best friends while working at NIS. They are role models in every aspect I can think about and will forever stay in my heart. Thank you Diogo for your spirit, your enthusiasm, for letting me train my parenting skills with you and for being always a great friend. Thank you *minino* Lato for having the patience to teach me how to do research, for setting the bar so high and specially for sharing so many memorable and crazy moments, even those that we do not remember. Thank you Paulo, for always putting doubts in my mind with respect to anything possibly imaginable. While it was always a source of constant laughing, it made me revisit many things which I would otherwise miss. To Rui, for always taking me back to the roots of greatness and reminding me never to settle for less. For sharing his passion for science and discovery and the constant seek for elegance. To Tiago, for teaching me so many things and showing me that everything can be built from even the smallest example. For being a constant reference in principles and values that should always be a part of any scientist. To all of you, I owe this thesis. Not only for the help you provided me with when I was stuck in technical details, but also for your faith in the problems I tackled and for the constant reminders of what we were set out to get when we all started this journey!

Of course that I am also in debt to many of my friends outside work. To all of you, my sincere thanks. A special thank you to Eliana, André, their daughter Mafalda and their son Benjamim, for always reminding the values for which we should guide our lives with and for keeping in my mind that there is nothing more important than living fully, even among times where the hardest sacrifices are needed.

To my parents Vitorino and Belarmina, and my brother Zé , whose sacrifice, endurance, guidance and example led me to finish this tough path. Without you support, it would be impossible to be who I am today.

Lastly, to my wife Inês and my daughter Aurora. We have taken huge steps these last years, suffered great losses, overcame many obstacles and built so many beautiful things. Last time I wrote down we were going to write many stories. Eventually we were able to write the most beautiful fairy tale. Thank you for your infinite love and support. Thank you for the endless joy I felt from the moment we met. For being my future.

With the greatest gratitude and love,

João Almeida

Resumo

Os sistemas de comunicação vêem tomando um papel cada vez mais importante no nosso dia-a-dia. O uso difundido da Internet e sistemas de comunicação sem fios não só alteraram a forma como comunicamos, mas também o tipo de informação que comunicamos. Dado que a maior parte do canais de comunicação são susceptíveis a escutas, e frequentemente é pretendida a transmissão de dados sensíveis, existe uma clara necessidade de mecanismos que garantam confidencialidade de comunicação. Tradicionalmente, a confidencialidade de dados é gerida na camada de aplicação, usando primitivas criptográficas. No entanto, nos últimos anos, outros métodos de segurança foram desenvolvidos. Em particular, os métodos de *segurança na camada física* podem actuar como um complemento (ou alternativa) a soluções baseadas em criptografia. De uma forma geral, a ideia subjacente a estas técnicas é a utilização do ruído inerente aos canais de comunicação como fonte de aleatoriedade, um elemento essencial no desenho de sistemas de comunicação segura.

O tópico fundamental desta tese é precisamente o desenvolvimento de esquemas de segurança para a camada física. Neste contexto, são exploradas várias alternativas ao estado-da-arte na construção de códigos seguros. Em particular, são fornecidas construções explícitas para fontes contínuas e discretas, com enfoque em códigos de tamanho finito.

Primeiro, é desenvolvido um quantizador escalar com restrições de segurança. A ideia principal nesta construção é desenhar um código conjunto de fonte-canal que garanta que um atacante tenha uma distorção acima de um determinado limite. Tal objectivo é atingido usando um desenho cuidado dos parâmetros do quantizador escalar, nomeadamente as fonteiras de quantização e o número de níveis de quantização.

Em seguida, é proposto o uso de mapeamentos de expansão de largura de banda para canais de wiretap com ruído Gaussiano aditivo. Tais mapeamentos são caracterizados pela existência de erros anómalos, quando o ruído de canal se situa acima de um dado nível. Estes erros tipicamente levam a estimativas afectadas por uma grande distorção. A ideia aplicada na construção é desenhar um código que garanta que um atacante é afectado por erros anómalos com elevada probabilidade. Para isso, é usada uma construção denominada Torus Layer Spherical Codes, que permite um controle intuitivo do nível a partir do qual erros anómalos surgem.

Finalmente, é proposto o uso de puncionamento aleatório como meio de obter segurança em canais com apagamentos. O prícipio subjacente a esta construção é a interpretação da técnica de puncionamento aleatório como uma técnica de introdução de ruído artificial. Neste sentido, é possível utilizar o puncionamento aleatório para saturar o canal de um atacante com apagamentos, forąndo-o a operar com elevada equivocação. Duas instâncias do sistema são consideradas. Primeiro é assumido que o padrão de puncionamento é público. Neste caso, a equivocação do atacante está directamente relacionada com a sua capacidade de recuperar mensagens através de um canal de apagamentos, em que a probabilidade de apagamento é elevada. Num segundo caso, é assumido que o padrão de puncionamento é um segredo partilhado entre as entidades legítimas. Em particular, mostramos que o puncionamento aleatório introduz perdas de sincronização ao nível do *bit*, contribuindo para um grande acréscimo à equivocação do atacante.

Abstract

Communication systems have taken an increasingly important role for most aspects of our daily lives. The widespread use of the Internet and wireless communications not only have changed the way in which we communicate, but also what type of information we communicate. Owing to the fact that most of the communication channels are open to eavesdropping, and often the data we wish to transmit is of a sensitive nature, it is clear that mechanisms for ensuring confidential communications are required. Traditionally, confidentiality is managed at the application layer using cryptographic primitives. However, in recent years, other means of achieving confidential data transmission have emerged. Physical-layer security is one of such techniques, which can act as a complement (or sometimes alternative), to standard cryptographic solutions. Broadly, the idea of physical layer security is to use the noise inherent to communication channels as a source of randomness, an essential element to design secure communication systems.

This thesis is fundamentally concerned with the development of security schemes for the physical-layer. In this context, we explore several alternatives to current state-of-the-art secrecy codes. In particular, we provide explicit code constructions for both continuous and discrete sources, focusing on codes with finite block-length.

First, we develop channel-optimized scalar quantizers with secrecy constraints. The main idea is to design a joint source-channel code which guarantees that the eavesdropper's distortion lies above a prescribed threshold. This is achieved by a careful design of the parameters of the scalar quantizer, most notably the quantization boundaries and the number of quantizations levels.

Second, we propose the use of bandwidth expansion mappings over wiretap Gaussian channels. Bandwidth expansion mappings are characterized by the existence of anomalous errors, when the channel noise is above a given threshold. These errors typically lead to estimates of the transmitted messages with high distortion. The main idea of the proposed code construction is to design codes such that the eavesdropper is generally affected by these anomalous errors. To this purpose we employ an instance of spherical codes construction known as Torus Layer Spherical Codes, which allows for an intuitive control over the threshold above which anomalous errors appear.

Finally, we propose the use of random puncturing as means to obtain secrecy for the binary erasure wiretap channel. The underlying principle of the coding scheme is to look at random puncturing as a technique for introducing artificial noise. Hence, we can use random puncturing to saturate the eavesdroppers channel with erasures, forcing the eavesdropper to operate with high equivocation. Two instances of the system are considered. We first assume that the puncturing pattern is public. In this case the eavesdroppers equivocation is directly related to the ability of recovering messages from a binary erasure channel with a large erasure probability. We then move to the case

where the puncturing pattern is a shared secret between the legitimate parties. We show that random puncturing introduces loss of bit-level synchronization, which contributes to a greater increase in the eavesdroppers equivocation.

Contents

List of Figures

List of Tables

Notation

$\mathcal{X}$	Alphabet or set
p_X	Probability distribution of X
$X \sim p_X$	Random variable X follows distribution p_X
$p_{X\|Y}$	Conditional probability distribution of X given Y
E_X	Expected value over X
$H(X)$	Entropy of X
$H(X\|Y)$	Conditional entropy of X given Y
$I(X;Y)$	Mutual information between X and Y
$0,1^n$	Binary vector of length n
$\mathbb{R}$	Field of real numbers
$\mathbb{F}$	Galois Field
λ	Lagrangian multiplier
∇	Gradient function
$\mathcal{N}(\mu,\sigma^2)$	Gaussian distribution with mean μ and variance σ^2
$\|\|n\|\|$	Norm of vector n
$\mathbb{S}^1$	Sphere in the Euclidean space
$\delta_S B$	Distance between folds in a curve
δ_T	Distance between two tori

Abbreviations

AWGN	Additive White Gaussian Noise
BCC	Broadcast Channel with Confidential Messages
BER	Bit error rate
BDC	Binary Deletion Channels
BEC	Binary Erasure Channel
BEWC	Binary Erasure Wiretap Channel
BP	Belief Propagation
BSC	Binary Symmetric Channel
CO-SC	Channel-Optimized Scalar Quantizer
CSI	Channel State Information
CSNR	Channel Signal-to-noise Ratio
KKT	Karush-Kuhn-Tucker
LDPC	Low-Density Parity Check
LM-SC	Lloyd-Max Scalar Quantizer
MAP	Maximum a Posteriori
ML	Maximum Likelihood
MSE	Mean Square Error
MMSE	Minimum Mean Square Error
OPTA	Optimum Performance Theoretically Achievable
PDF	Probability Density Function
SC	Spherical codes
SK	Shannon-Kotel'nikov
SNR	Signal-to-noise Ratio
TLSC	Torus Layer Spherical Codes
WTC	Wiretap Channel Model
WTC-SK	Wiretap Channel Model with a Shared Key

Chapter 1

Introduction

Devising schemes for secret communications has been an object of study almost since the invention of the first alphabets. In ancient civilizations, where the first forms of encryption appeared, they were mostly used to create an aura of mystery around messages written in tombstones [1]. However, they soon became an essential tool for military purposes across the ages. The ability to encrypt (or hide) the contents of messages was a crucial advantage in preparing field operations, as well as for secret diplomatic communication [1]. On the other hand, the competence on performing cryptanalysis, i.e., the ability to break an encryption scheme and obtaining the respective contents of a hidden message, became an even greater advantage, as it revealed plans and information from adversaries, allowing to properly adopt any necessary counter-measures. While in the past the arts of cryptography and cryptanalysis where mostly restricted to the domain of military and diplomatic communications, the evolution of computer networks has changed this paradigm, establishing security as a ubiquitous concern.

In modern communication systems, entities interact using devices as proxies and communication takes place through channels that may be remotely eavesdropped/tampered. Such outline suggests a broad scope of security concerns [2]. For instance, communicating entities should be able to corroborate the identity of each other, which implies some sort of mechanism should provide for identity authentication. Users should be able to corroborate the source of a given received message, as well as being able to verify if that message have not been subject to changes. Therefore, mechanisms that guarantee message authentication and data integrity are imperative. It could also be the case that the origin or reception of a given message needs to be proven, for which schemes that ensure non-repudiation are required. Another relevant question is how to prevent unauthorized users from accessing some resource, which could be tackled with appropriate access control mechanisms. These examples illustrate some of the security objectives that should be met, if required by the communicating entities. Notwithstanding the emergence of these new security concerns, data confidentiality is still one of the most crucial security problems. In general, it is not possible to guarantee that the messages transmitted from some device will not be observed by an unintended third party. This fact is even more evident in wireless networks since wireless transmissions are by nature susceptible to eavesdropping [3, Chapter 5], where intercepting messages (which may contain sensitive data) can be done with any device equipped with a wireless interface. Such eavesdropping attacks are categorized as a passive attacks [4, Chapter 1.3][1]. Clearly, solutions are required to ensure that unauthorized entities (or eavesdroppers) are not able to decipher the contents of a captured message.

Within the context of modern communication networks, cryptography is extensively used as the standard technique to achieve data confidentiality. In particular, for networks organized in a layered architecture, it is common to find suites of security protocols (based on principles of modern cryptography) that provide confidentiality services for many network layers. One such example is the Internet protocol suite, which proposes the use of confidentiality mechanisms at the application, transport and network layers for the encryption of raw data, the payload of TCP/UDP packets and the payload of IP packets, respectively. Surprisingly, the physical layer is often neglected as a layer where security mechanisms should be implemented, even though it is ultimately the layer responsible for transmitting data (e.g. modern radio-based communication systems do not implement data confidentiality at the physical-layer, but at layers above). Moreover, it provides a natural source of randomness (the communication channel), an essential ingredient of any secrecy system. A possible justification lies on the fact that cryptographic-based solutions have reached a seemingly mature state. This contrasts with existing solutions that use the physical-layer to achieve security, which are still in an early stage of development. In addition, cryptographic solutions are not constrained by the type of source message and have only a small set of requirements (e.g. the existence of an encryption key). Schemes that exploit the benefits provided by the physical-layer tend to make several assumptions with respect to the source and the underlying communication channel, which may be hard to guarantee in practice. Consequently developing practical schemes with sufficient abstraction and generality for security purposes at the physical layer is not a simple process. However, in recent years, remarkable progress has been made both in the theory and practice of physical-layer security. Consequently, it is anticipated that these schemes will eventually achieve a state of maturity that allows their adoption on standard communication systems. The main topic of this thesis is precisely this: the design of practical security schemes for data confidentiality based on physical-layer security principles.

[1] While active attacks such as jamming might have severe consequences, they are in general easier to detect and appropriate counter-attack measures can be taken. On the other hand, passive attacks such as eavesdropping are almost impossible to detect. Consequently, one should pro-actively implement mechanisms that prevent eavesdroppers from acquiring any meaningful information, rather than to react to eavesdropping events.

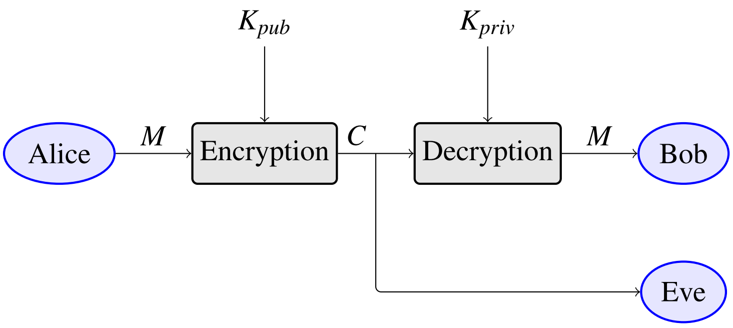


Figure 1.1: Example of asymmetric encryption.

1.1 Two security models: computational and information-theoretic security

The motivation behind this thesis will be more evident once we contrast the virtues and limitations of the computational security model and the information-theoretic security model (under which physical-layer security is based[2]).

Algorithms based on the computational security model rely on problems that are computationally hard to solve unless some side information is available to the user. These algorithms are based on the assumption of the existence of one-way functions [5, Chapter 2], i.e. functions that are easy to compute (given a function f and an input x, there exists a polynomial-time algorithm that computes $f(x)$) but hard to invert (for a possible input x, the average probability of successfully finding an inverse of x under f for any probabilistic polynomial-time algorithm is negligible) [5, Chapter 2]. Thus, the essence of the computational security model lies on considering an adversarial model where the malicious user has limited computational resources.

A broad class of primitives that fall in this domain are based on the concept of asymmetric cryptography (also known as public-key cryptography). A typical setup for public-key encryption/decryption is illustrated in Fig. 1.1. In this class of algorithms, users are equipped with a private/public key pair (K_{priv}, K_{pub}). They distribute their public keys, which are then used to encrypt messages that are destined to them. Once they receive an encrypted message, they use the private key to decrypt the transmitted cipher-text. Essentially, the key pair should be constructed in such a way that trying to obtain the private key from the public key requires to solve a hard problem (e.g. trying to invert a one-way function). The RSA [6] and ElGamal [7] cryptosystems are natural examples of public key cryptography. One possible attack on the RSA public key encryption scheme can

[2] The difference between information-theoretic security and physical-layer security is subtle but exists. The former is generally concerned with the characterization of the secrecy of a system, based on information theoretic quantities, and does not necessarily assume the existence of a communication channel. On the other hand, physical-layer security bases itself on the information-theoretic security model to provide secrecy at the physical-layer.

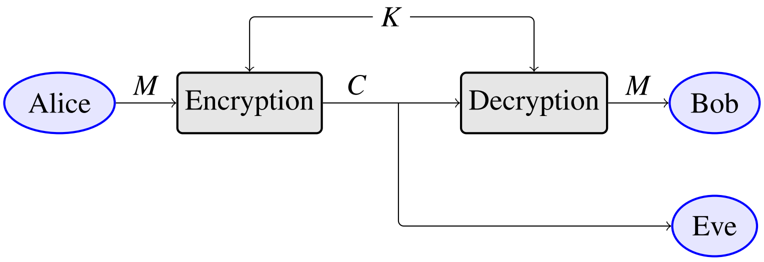


Figure 1.2: Example of symmetric encryption.

be performed by solving the problem of prime factorization while an attack on the El-Gamal public key encryption scheme can be performed by solving the discrete logarithm problem [8]. Both of these problems are believed to be one-way functions, and therefore hard to solve. Generally, the implementation of these primitives requires computationally expensive operations such as computing exponentiations and moduli. Therefore, these strategies are often used to share a single secret which will act as a key to a more computationally efficient (albeit generally less secure) symmetric-key encryption algorithm. In symmetric encryption algorithms, the same key is used both for encryption and decryption operations (see Fig. 1.2). Therefore, such key should be shared a priori (e.g. through the public-key cryptography framework described above). Symmetric ciphers make use of the principles of confusion and diffusion proposed by Shannon [9]. Confusion refers to creating a complex relationship between the cipher-text C and the secret key K while diffusion refers to creating a complex relationship between the message M and the cipher-text C[3]. Ultimately, these ciphers create an avalanche effect where changing one bit of either the message or the secret key leads to a cipher-text that is independent of the original cipher-text, which makes the system hard to attack. On the other hand, this also makes it vulnerable to channel noise, since the decryption of a cipher-text contaminated with errors will lead to decoding the wrong message.

Given the properties of the systems based on public-key/symmetric-key cryptography, it is not surprising that they have been largely adopted in current communication systems: they are reasonably efficient, data agnostic and apparently secure under the premise of limited computational resources. Indeed, the existence of one-way functions is not yet mathematically proven [5, Chapter 2] (essentially proving the existence of such functions would prove that the complexity classes P and NP are not the same, a long-standing problem in theoretical computer science). Even if this is the case, such functions may only be applied under the correct computational model (e.g. the problem of factoring large integers can be solved in polynomial time in a quantum computer [11]). Thus, the long term security of these schemes cannot be ensured if, for instance, quantum computers

[3] The most common way to implement these principles is to use substitution-permutation networks of Feistel networks [10].

become ubiquitous. Aside from the security notions, modern ciphers also present some practical concerns. It is not uncommon that these ciphers suffer from broken implementations which, in practice, means that these schemes can be attacked through alternative means. For instance, [12] lists several attacks that can be performed on the RSA cryptosystem that range from the wrong choice of parameters for key generation and partial key exposures to timing/power attacks or attacks based on faulty computations. These concerns are sufficient for at least considering the possibility of using a different security paradigm.

Information-theoretic security provides an alternative formulation of the secrecy problem. The field was born in Shannon's landmark paper in 1949 [9] as a natural extension of information / communication theory. The basic problem in information theory is to recover a message that is transmitted over a noisy channel with an arbitrarily small error. In information-theoretic security one needs not only to ensure the aforementioned condition for the legitimate party, but also that a malicious third party (eavesdropper), with access to the transmitted messages, possibly corrupted by channel noise, cannot reduce its uncertainty about the transmitted message. More precisely, if a user wishes to transmit a secret message M, encoded as C, the average uncertainty of the information obtained by the eavesdropper can be measured using the *conditional entropy* $H(M|C)$. This quantity is also commonly denoted as the *eavesdropper's equivocation*. If $H(M|C) = 0$, there is no uncertainty left in the information obtained by the eavesdropper, which means that his observation provides him with sufficient information to obtain M without any errors. On the other hand, if $H(M|C) = H(M)$, the eavesdropper's average uncertainty about M is the same with or without C. Hence, the captured message does not provide any information to the eavesdropper. A system that ensures $H(M|C) = H(M)$ is called *unconditionally secure* or *perfectly secure*. Such systems are immune to cryptanalysis [9].

The first aspect to be retained is that an information-theoretic formulation of secrecy provides a precise definition/measure of security. The second aspect is that assumptions regarding the resources available to the eavesdropper do not need to be made.

Shannon originally worked on an information-theoretic formulation of symmetric encryption, where the eavesdropper observes an error-free cryptogram (as in Fig. 1.2). He showed that communicating under unconditional security constraints is only possible if the entropy of the key is greater or equal than the entropy of the message [9]. This result implies that the key of a symmetric encryption scheme should be at least as large as the original message. Eventually, if a key with such characteristics is not available to the legitimate party, they will have to share it over a possibly unsecured channel. Hence, the problem of how to communicate the key over an insecure channel remains, which may be even harder to solve than the original problem. On the other hand, it also shows that symmetric ciphers based on small secret keys (used in many current systems) cannot ensure unconditionally secure communications, suggesting that such strategies may not be strong

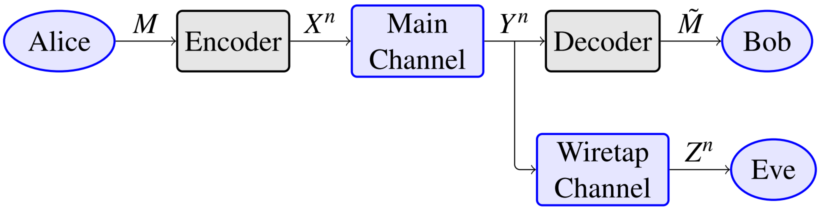


Figure 1.3: Physically degraded wiretap channel (DWTC) model.

enough to provide for secrecy. However, these pessimistic results are a consequence of the strict assumptions of the communication model, since the only available source of randomness is the secret key.

Inspired by Shannon's formulation, Wyner [13] contemplated the use of a different source of randomness, the *communication channel*. He proposed a new model for secrecy, which is now commonly known as the *wiretap model* [13]. Rather than giving the eavesdropper a noiseless copy of the cipher-text, Wyner assumed that the eavesdropper's observations were obtained through a channel that is corrupted by noise. The first wiretap model considered that the eavesdropper's observations were (physically) degraded versions of the messages obtained by the legitimate receiver (see Fig. 1.3). Additionally, Wyner relaxed Shannon's conditions for unconditionally secure communication. Instead of requiring that the eavesdropper's equivocation is equal to the source's entropy, Wyner proposed the use of equivocation rate as a secrecy metric, in which case it is required that the eavesdropper's equivocation rate is arbitrarily close to the entropy rate of the source, i.e. $\frac{1}{n}H(M|Z^n) \approx \frac{1}{n}H(M)$, for sufficiently large n. Then, he defined the *secrecy capacity* as the maximum rate that satisfies such condition, while ensuring that the error probability of the legitimate receiver is arbitrarily small.

Wyner's model implicitly assumed that the eavesdropper has access to degraded versions of the messages received by the legitimate receiver, and therefore enforces a very strict assumption with respect to the eavesdropper's observations. Csizár and Körner [14] generalized the degraded wiretap model to account for a broadcast channel from the source to the legitimate receiver and the eavesdropper. This model, illustrated in Fig. 1.4, can be thought of as a system with two parallel channels: one between the sender and legitimate receiver (main channel) and another between the sender and eavesdropper (wiretap channel). Furthermore, [14] extends the wiretap model in the following way: source messages have a private and public component. The private message M_1 is to be decoded only by the legitimate receiver and the public message M_0 is to be decoded by both the legitimate receiver and the eavesdropper. While both wiretap models contrast in a few aspects, the imposed reliability and secrecy constraints are very similar. However, the rate-equivocation region and secrecy capacity characterizations are more complex in the latter case.

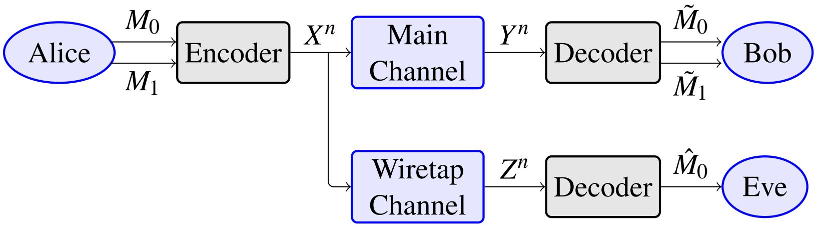


Figure 1.4: Broadcast channel with confidential messages (BCC).

One of the main achievements of information-theoretic security is precisely the characterization of the fundamental limits of secure communication under wiretap-like models, that is, of their secrecy capacity. Once these fundamental limits are established, practical code constructions can replace cryptographic algorithms to ensure secure communications. Unlike their cryptographic counterpart, these coding schemes do not require a secret key *a priori*. Instead, they encode a source message M onto a sequence X^n to be transmitted, making use of channel randomness (and possibly some local randomness known only by the sender). Each of the receiving users observes the transmitted messages through their respective channels and tries to recover the original message. The question to be answered is how can we design a code that guarantees a negligible error for the legitimate receiver and satisfies some secrecy requirement, using the randomness provided by the underlying communication channels. These requirements will be formalized later.

Clearly, physical-layer security solutions (inspired by the information-theoretic security model) have many benefits: secrecy can be measured through the average uncertainty of the information possessed by the eavesdropper, one can use the randomness provided by the channel (rather than using pseudo-random generators) and secure communication is possible without requiring a secret key a priori. However, the problem of designing practical codes that achieve the fundamental limits of secure communication is far from being trivial. In fact, the problem of coding for secrecy is unsolved in general. First, codes should be able to meet reliability and secrecy constraints, which are at odds with each other. Second, the construction of secrecy codes generally requires some *channel state information* (CSI) with respect to the main and wiretap channel. This constitutes one of the main difficulties in implementing physical layer security solutions, since in general it is hard to obtain CSI for the wiretap channel[4]. Third, these schemes generally require a proper environment for operation [15]. For instance, it is generally required that the wiretap channel is in some sense worse than the main channel.

It should be noted that the information-theoretic security model also has several drawbacks, other than the difficulty in code design. The secrecy constraints used in these models are, in general, probabilistic, i.e. they do not guarantee secrecy with probability

[4] It is always possible to make some assumptions regarding this channel, but a conservative estimate may cause us to operate far from the fundamental limits of secure communication.

one but rather with arbitrarily high probability. While this does not constitute a problem *per se*, care should be taken when specific codes are employed since they may not provide the level of secrecy one was expecting. Furthermore, the security notions are asymptotic by definition. Since any implementation of a physical-layer security system requires the use of finite block-lengths, one should proceed with caution when moving from code constructions based on asymptotic analysis to the finite block-length regime.

In summary, security schemes based on the information-theoretic model may provide several benefits over schemes based on the computational model, either in terms of measurable secrecy and computational efficiency (secrecy is obtained via coding, which is already implemented in any communication system for reliability purposes). However, it also has some disadvantages which may not be neglected such as strict assumptions on channel models or the need to restrict the transmission rate to account for secrecy. That being said, it is certainly true that such schemes could be used to enhance the security at the higher layers of the protocol stack. For instance, physical-layer security schemes can be coupled with cryptographic schemes and guarantee that, with high probability, an adversary will have access to a cipher-text that contains errors [16]. Clearly, the task of a cryptanalyst is made harder since cryptographic attacks are generally designed under the assumption of a correct cipher-text. They can also simplify the task of key distribution, since they do not require a secure channel *a priori*. One can use the principles of physical-layer security to develop key agreement schemes based on the fact that an eavesdropper receives a signal that is different from the legitimate receiver [17]. Both these aspects suggest that a cross-layer approach to secrecy may be desirable in many cases. Along these lines, physical-layer security can be useful to enhance the security levels of current systems or to simplify the design of secrecy systems. This represents a departure from the complex security architectures that are currently employed, which make use of third parties for key distribution. It also provides the means to effectively assess the security of a system, by filling the lack of secrecy metrics that currently exists.

1.2 Motivation

While the problem of coding for secrecy under the information-theoretic model is still unsolved in general, code designs that achieve secrecy capacity are in fact known. Many practical code constructions have been developed under the notion of *weak secrecy* proposed by Wyner [13]. Most of these code constructions share the same guideline, which is to map every message to possible multiple codewords and randomly choose a message within this set for transmission [15]. This principle can be put into practice using codes with a *nested structure*, where a codebook is partitioned onto several sub-codebooks, each one associated with a message to be transmitted. When the transmitter wishes to send a given message m, he randomly selects a message m' from the sub-codebook that is associated with m and transmits it. A sufficient condition to guarantee weak secrecy is to design the sub-codebooks to be capacity-achieving over the wiretap channel [3, Chapter 6]. Due to this seemingly simple constraint, nested codes became the prevailing practical code construction for physical-layer security. Consequently, most of the research efforts on coding for physical-layer security focus on finding codes, based on nested structures, that satisfy this condition.

While useful from a theoretic and practical perspective, the application of nested codes can be limited by the operational environment [15]. These constructions have several requirements, some of which we list next. First, codes must have an arbitrarily large block-length. Second, channel state information (CSI) for the main and wiretap channel is required to properly dimension the codebook. Third, the code is dependent on such channel state information. The following observations, connected to these requirements, motivate the need for alternative code designs for secrecy:

- In any communication system the employed codes must have a finite block-length. This remark has several ramifications: a) source-channel separation theorems may not hold, meaning that the optimal coding scheme for a communication system could involve solving a joint source-channel coding problem; b) the secrecy performance of a code may be far from the performance predicted by its asymptotic analysis; and c) the objective of achieving secrecy capacity becomes unreachable which may justify using alternative secrecy metrics.
- While legitimate users may cooperate in order to characterize their communication channel, it may be very hard to obtain the CSI for the wiretap channel in practice. Therefore, code constructions should provide a good secrecy performance either for a large range of channel parameters or, if some estimate of the quality of the wiretap channel is available, under the circumstances of channel mismatch.
- Depending on the communication environment, it may be the case that channel statistics vary over time. Since codes with nested structures vary the sizes of their sub-codebooks according the main and wiretap channel statistics, a change in channel condition may imply the design of a new code. Consequently, nested codes may be unfit for time-varying channels.

The code designs proposed in this thesis attempt to circumvent the aforementioned issues. More precisely, the proposed code constructions are of finite block-length. Consequently, the secrecy analysis associated with these codes will reflect this fact. A key point is that the proposed codes do not strive to achieve the secrecy capacity, but rather ensure that the eavesdropper's ability to estimate the sent messages is greatly impaired. We do require that this impact can be quantified through information theoretic quantities.

However, the secrecy criteria employed on the eavesdropper's side may not necessarily be the eavesdropper's equivocation, but could be, for instance, distortion. A second aspect is that the proposed codes are deterministic. This contrasts with the common approach used to design secrecy codes, which considers the use of stochastic encoders through instances of local randomness. The reason is that stochastic encoding is useful to cancel out the information leaked to the eavesdropper, but this requires CSI for the wiretap channel. Therefore, if such information is not available, it is not clear how to use the local randomness at the encoder to satisfy the secrecy constraints. Thus, our general approach to the problem of code design for secrecy focuses on meeting a certain reliability constraint while providing a *best-effort* approach with respect to secrecy. While deterministic constructions generally have a worse performance (in terms of secrecy) when compared to stochastic codes, they allow for a simplified design which is sufficient for the purposes we intend (design codes that provide a prescribed level of security for a large range of channel parameters). We do note that the proposed schemes can also be extended to include nested-like structures. Finally, we distinguish code constructions according to the type of source. We consider two types of sources: a) sources that are discrete in time and continuous in amplitude (herein referred as continuous sources) and b) sources that are discrete in time and amplitude (herein referred as discrete sources). While there exists a large body of research that addresses discrete sources, secrecy codes for continuous sources are almost non-existent[5]. The reason lies in the fact that, if source-channel separation theorems hold, the secrecy capacity may be achieved by using an optimal source encoder followed by an optimal wiretap code, and hence secrecy is achieved on the discrete part of the problem [18]. As mentioned before, these arguments may not hold and even if they do, both components may be extremely hard to design, thus motivating a different approach to the design of secrecy codes for continuous sources.

1.3 Outline and Main Contributions

In this thesis we propose three coding schemes for the problem of confidential data transmission. The first two schemes are directed towards continuous sources, while the third focuses on discrete sources. Within the domain of continuous sources we propose a joint-source channel coding scheme based on scalar quantizers and a coding scheme based on bandwidth expansion mappings. The objectives in each of these constructions are distinct. The former construction forces eavesdroppers to operate bellow a desired performance

[5] Continuous sources arise in many situations. Audio and video signals can be represented by continuous variables that are subject to digitalization prior to transmission. The coefficients of Fourier and other related transforms are also generally represented by continuous variables. For instance, the discrete cosine transform (DCT), that is widely used in image coding standards, outputs real-valued coefficients. Signal processing techniques make ample use of continuous random variables (filtering, signal acquisition, ...). Additionally, natural sources (e.g. the quantities measured by a sensor) or artificially induced sources (e.g. sources induced from channel gains) can be represented by continuous variables. Thus, many applications could benefit from secure coding schemes that operate over continuous alphabets.

threshold while the latter tries to ensure that the eavesdropper is bound to operate in a regime of anomalous errors, which greatly impacts the distortion of his estimates. Within the domain of discrete sources we propose a scheme based on randomly punctured LDPC codes. The scheme uses puncturing as a mechanism to introduce artificial noise to create a saturated channel from the eavesdropper's perspective. It also tries to explore the lack of bit-level synchronization at the eavesdropper's side to obtain higher secrecy gains. This is accomplished by allowing the puncturing pattern to be secret. The scheme also takes advantage of the fact that rate-compatible codes enable the adaptation of codes to channel conditions, without the need to design a new code.

The main contributions of this thesis are as follows.

- **Scalar Quantization under Secrecy Constraints:** We propose a joint source-channel coding approach to secrecy that is based on solving an optimization problem. The main idea is to find a joint source-channel code that minimizes the distortion at the legitimate receiver, subject to a distortion constraint on eavesdropper. The process involves finding the boundaries of a scalar quantizer as well as finding the optimal index assignment (channel code) that satisfies the above constraints. Our results demonstrate that such an approach can effectively bound the distortion at which the eavesdropper can operate, even when the channel to the eavesdropper is better than that of the legitimate receiver.

- **Piecewise Torus Layer Spherical Codes for Secrecy:** We propose code constructions for the transmission of continuous sources without the need for quantization. The main technique employed is the transmission of curves over several layers of torus, which are obtained via spherical codes. By exploiting the geometrical properties of this construction, we find the code parameters which, with a desired probability, ensure decoding errors at the eavesdropper's end that induce a large distortion. The construction has the additional advantage of transmitting messages over a dimension that is double the dimension of encoding and decoding. This feature can be used to obtain higher secrecy gains since the noise affecting the eavesdropper possesses more components.

- **Randomly Punctured LDPC Codes for Secrecy:** We propose a coding scheme based on the principles of rate-compatible codes, where random puncturing is used to adapt both to the channel conditions as well as for secrecy purposes. We consider two scenarios: 1) the puncturing pattern is publicly known and 2) the puncturing pattern is a shared secret between the legitimate parties. We analyze the equivocation rate achieved by LDPC codes when the legitimate receiver has access to a *belief propagation* (BP) decoder or a *maximum a posteriori* (MAP) decoder. Our results indicate that using public puncturing patterns while allowing MAP decoding leads to maximum equivocation for the eavesdropper (albeit at an increase in terms of rate - which prevents us from achieving perfect secrecy), while only allowing BP decoding results in a smaller equivocation for the eavesdropper but also a smaller transmission rate. Finally, it is shown that if the puncturing pattern is a shared secret, it is possible to achieve high equivocation rates for the eavesdropper, even for very small block-lengths. The effort to share the puncturing pattern depends on the puncturing probability. Hence, for large enough puncturing probabilities, the required secret rate can be deemed small.

The rest of this thesis is organized as follows. Chapter 2 introduces more formally the wiretap model, its fundamental limits and the state of the art in coding for secrecy. In Chapter 3 we present a methodology for the design of scalar quantizers with secrecy constraints that bound the performance achieved by an eavesdropper. We pose the problem of secrecy as a constrained optimization problem and derive necessary conditions for locally optimal encoders and decoders (under a mean square error distortion criterion). We then present numerical results highlighting the distortion behaviour of the eavesdroppers optimal estimates under several scenarios. Bandwidth expansion mappings are introduced in Chapter 4, as well as a particular construction of these mappings that is based on mapping a source onto a set of curves over several layers of tori. We provide a characterization of the different types of errors that may occur in such construction. Then, assuming the main and wiretap channels are additive white Gaussian noise (AWGN), we use the geometrical properties of these codes to characterize the error probabilities associated with each type of error. Using these probabilities, we find the code parameters that ensure the eavesdroppers will suffer from the decoding errors that induce a distortion of largest magnitude. We then present several numerical results that relate to the code parameters, as well as the distortion behaviour of the eavesdropper. Chapter 5 addresses the design of randomly punctured LDPC codes. We present wiretap channel models that take puncturing into account, derive the eavesdropper's equivocation under these models and characterize their rate-equivocation regions. We also derive bounds on the allowed puncturing probabilities based on the code's thresholds. We characterize the eavesdropper's maximum likelihood decoder and present simulation results for specific code instances based on the derived decoder. We further present numerical results with respect to the eavesdropper's equivocation rate, in particular asymptotic results for public puncturing patterns and finite block-length results for secret puncturing patterns. Chapter 6 presents the conclusions of this thesis, discussing several directions for future work.

Chapter 2

Coding for Secrecy

In this chapter we will introduce some of the notions regarding the theory and practice of secrecy systems based on the information-theoretic security model. We assume familiarity with the basic definitions and results from information theory. For the sake of completeness, a necessary set of results that are used in this thesis are summarized in Appendix A. We will first formally introduce the definitions of wiretap channel and wiretap code, followed by possible definitions of reliability and secrecy constraints. We then move towards the characterization of the fundamental limits of secure communication under some of these constraints. We also review the design of state-of-the-art wiretap codes based on nested structures.

2.1 The Wiretap Channel Model

The basic problem we wish to solve is how to transit some source message to a legitimate receiver that is able to correctly decode such message while keeping it secret from unintended recipients. Thus, we wish to solve a communication problem with two constraints: a reliability constraint for communication between the legitimate party and a secrecy constraint with respect to the eavesdropper's observations.

In the context of physical-layer security, this problem can be modelled using the so-called wiretap channel model, illustrated in its generalized form in Fig. 2.1. It incorporates three users: a sender (Alice), a legitimate receiver (Bob) and an eavesdropper (Eve). Both Bob and Eve receive the messages transmitted by Alice through a broadcast channel, which comprised of two parallel channels. The channel from Alice to Bob is called the *main channel*, while the channel from Alice to Eve is called the *wiretap channel*. For simplicity, we will assume throughout this thesis that both channels are memoryless and the noise is assumed to be independent for Bob and Eve. It is also possible to consider the case where noise the main and wiretap channels do not have independent noise. In these cases, one generally obtains less secrecy, reason for which one should try to use

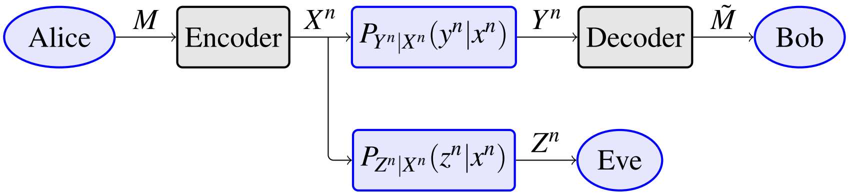


Figure 2.1: Wiretap channel model.

alternative techniques such as interleaving in the attempt to create independent channels. Formally, the wiretap channel can be defined as follows.

Definition 1 (Wiretap channel)**.** A wiretap channel ($\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, $p_{YZ|X}(y,z|x)$) is characterized by a quadruple that consists of one input alphabet $\mathcal{X}$, two output alphabets $\mathcal{Y}$ and $\mathcal{Z}$ and a transition probability matrix $p_{YZ|X}(y,z|x)$.

As noted in Section 1.1, secrecy can be ensured through coding. Hence, to communicate over the wiretap channel, Alice chooses a message M that she wishes to securely transmit to Bob. She then encodes this message onto the channel input vector X^n using some *wiretap code*. Through the main channel, Bob observes a possibly noisy codeword Y^n, while Eve observes also a possibly noisy codeword Z^n through the wiretap channel. Since the main and wiretap channel are memoryless, we have that

$$p_{Y^nZ^n|X^n}(y^n,z^n|x^n) = \prod_{i=1}^{n} p_{YZ|X}(y_i,z_i|x_i).$$

The wiretap code is responsible for ensuring that reliable and secure communication is possible. By reliable it should be understood that Bob can reproduce the source message with negligible error, while by secure it should be understood that Eve's estimates of the source message are erroneous. How one can exactly measure the reliability and secrecy performance of a particular code will be briefly addressed. Let us first formally introduce wiretap codes. A (discrete) wiretap code can be defined as follows.

Definition 2 (Discrete wiretap code)**.** A (2^{nR}, n) code $\mathcal{C}_n$ for a wiretap channel consists of

- A countable message set $\mathcal{M} = [1,\dots,2^{nR}]$;
- An encoding function (possibly stochastic) $f: \mathcal{M} \to \mathcal{X}^n$, mapping source messages onto channel codewords;
- A decoding function $g: \mathcal{Y}^n \to \mathcal{M} \cup \{?\}$, mapping channel observations to the source message set or an error message.

Note that the wiretap code needs to introduce sufficient redundancy so that the legitimate user is able to decode the messages without any errors. This redundancy also provides the eavesdropper useful information. Therefore, allowing the encoder to be stochastic is essential to achieve full secrecy. The introduced randomness provides the means to cancel some information leakage that may occur while using a particular codebook for transmission. On the other hand, as discussed before, unless we have some knowledge about the wiretap channel, it is not clear how one can use this randomness. This problem can be circumvented by designing deterministic secrecy codes, which simply rely on the randomness provided by channel. They do incur in some information leakage, and therefore do not achieve full secrecy. However, if codes are carefully designed, such leakage may be small enough that no meaningful information can be extracted from it.

2.2 Reliability and Secrecy Metrics

Recall that our main objective is to design coding schemes that allow two parties to communicate reliably, while preventing an eavesdropper from acquiring any meaningful information about the transmitted messages. Thus, as mentioned earlier, the system should guarantee two constraints: a reliability constraint and a secrecy constraint. Such constraints may take many forms, although ultimately they aim at the following general goals: an admissible (preferably negligible) error probability for the legitimate party (reliability) and statistical independence between the transmitted messages and the eavesdropper's observations (secrecy). The reason why statistical independence is relevant from a security perspective is that it reduces the best attack strategy of an eavesdropper to random guessing.

2.2.1 Reliability Constraints

The most common measure used for reliability is the average error probability of the wiretap code

$$P_e(\mathcal{C}_n) \triangleq Pr\{\tilde{M} \neq M|\mathcal{C}_n\}, \tag{2.1}$$

which measures the average probability that the legitimate receiver estimates the wrong message. In the discrete case, $P_e(\mathcal{C}_n)$ amounts to

$$P_e(\mathcal{C}_n) = \frac{1}{\lceil 2^{nR} \rceil} \sum_{m=1}^{\lceil 2^{nR} \rceil} Pr\{\tilde{m} \neq m|\mathcal{C}_n\}. \tag{2.2}$$

Commonly, we wish that legitimate parties communicate with negligible error. Then, the reliability constraint to be satisfied is formulated as

$$\lim_{n\to\infty} P_e(\mathcal{C}_n) = 0. \tag{2.3}$$

However, it may be the case that the average error probability is hard to analyze for a given wiretap code. Alternative metrics can be used in such cases, like the average *bit-error rate* (BER) [19, 20]. The BER is an approximate estimate of the bit error probability, thus capturing a similar idea to the average error probability. The reliability constraint can be defined in a similar manner to (2.3), by requiring that BER for the legitimate receiver to approach zero in the limit of large block-lengths.

Finally, distortion can also be used to characterize the reliability of a wiretap code. A distortion formulation of the problem of secure communication was provided in [18] and extended in [21]. The main motivation was to understand how allowing a prescribed level of distortion for the legitimate receiver could provide a positive impact on the secrecy of the system. The reliability constraint to be satisfied can be formulated as

$$\mathbb{E}\{d(M,\tilde{M})\} \leq \tilde{D} + \varepsilon, \tag{2.4}$$

where $\mathbb{E}$ denotes expectation, $d(\cdot,\cdot)$ is a distortion function and $\tilde{D}$ is the prescribed level of distortion. The formulation of reliability is terms of distortions bears an additional challenge, which is to find an appropriate distortion measure, that reflects the cost of choosing the representation of the source message by its reconstruction point. For instance, for some sources squared error distortion may be a good candidate, while for others not.

The choice of a particular measure for reliability does not require an extensive justification. As noted before, the general requirement is a vanishing error probability, be it in any type or form. However, the choice of a particular measure for secrecy should certainly be more judicious.

2.2.2 Secrecy Constraints

The first information-theoretic secrecy metric, introduced by Shannon [9], was unconditional security, also known as *perfect secrecy*. To obtain perfect secrecy exact statistical independence[1] is required with respect to the source message and the eavesdropper's observation. Assuming the wiretap code $\mathcal{C}_n$ is known to all parties, perfect secrecy is defined as follows.

$$H(M|Z^n) = H(M) \tag{2.5}$$

[1] We note that statistical independence could be measured in terms of any distance between joint probability distributions. In this thesis we focus on the Kullback-Leibler divergence, which is equivalent to the mutual information.

or alternatively

$$I(M;Z^n) = 0. \tag{2.6}$$

Systems that provide perfect secrecy have demanding constraints which in practice are very hard to meet. To circumvent this issue, it is possible to relax the secrecy constraint. Rather than requiring exact statistical independence between M and Z^n, consider the case of *asymptotic* statistical independence. The secrecy constraint then becomes

$$\lim_{n\to\infty} I(M;Z^n) = 0. \tag{2.7}$$

This constraint is commonly referred as *strong secrecy* and implies that the total amount of information leaked to the eavesdropper goes to zero as the size of the codewords goes to infinity. While the strong secrecy constraint is less restrictive than perfect secrecy, designing codes for the strong secrecy constraint is still very challenging.

Most practical code constructions adopt an even less restrictive constraint. Instead of requiring a total leakage of zero, they require the *leakage rate* to the eavesdropper to be vanishing, as the size of the codewords goes to infinity. This constraint can be formalized as follows.

$$\lim_{n\to\infty} \frac{1}{n} I(M;Z^n) = 0. \tag{2.8}$$

It should be noted that the same coding rates are achievable under the strong and weak secrecy constraints [22], although current coding schemes still incur in rate losses to ensure strong secrecy [23].

All of the above criteria depend on the ability to analyze the equivocation of $\mathcal{C}_n$. In some cases, most notably when $\mathcal{C}_n$ is a code of finite block-length, it may be hard to exactly analyze the code's equivocation. To circumvent this issue, several researchers have adopted the code's average error probability or the bit error probability as a secrecy criterion. In such cases, it is required that the eavesdropper's estimates of the source message suffer from an arbitrarily high error probability (or alternatively the error probability is bounded above a prescribed threshold). This secrecy formulation was used to analyze the secrecy of punctured LDPC codes [19] or lattice codes [24] over Gaussian wiretap channels. The analysis of the secrecy constraint is simplified by using density evolution techniques in the former case and geometrical arguments on the latter.

We stress that error based metrics do not guarantee secrecy in an information-theoretic sense, i.e. a high error-rate does not imply a high equivocation. That being said, the error-rate could, in fact, be a pointer to the secrecy performance of a particular code. Moreover, since the equivocation of a code can be bounded with respect to the decoding error [25], this constitutes an alternative way to find codes that may be interesting from a secrecy perspective.

Alternatively, it is also possible to use distortion as a secrecy measure. In particular, in [26, 27] the authors have considered a distortion-based approach where the goal is to characterize the fundamental limits of communication when we have a bound for the minimum average distortion for the eavesdropper (which they term as *payoff*). If, from Z^n, the eavesdropper produces an estimate $\hat{M}$ of the source message M, such secrecy criterion can be cast as

$$\mathbb{E}\{d(M,\hat{M})\} \geq \hat{D} - \varepsilon, \tag{2.9}$$

where $\mathbb{E}$ denotes expectation, d is a distortion function and $\hat{D}$ is the minimum distortion allowed at the eavesdropper. This formulation becomes particularly useful for continuous sources, since it is more amenable to analysis than differential entropy. However, its secrecy interpretation is different to a large extent. Rather than measuring the amount of information obtained by the eavesdropper, it measures its ability to correctly estimate the source messages. Therefore, it requires the assumption of some decoder, which may underestimate the amount of information collected by the eavesdropper. That being said, it also offers some advantages. For instance, a distortion metric is able to connect the perceptual quality to the secrecy metric (under a proper choice of distortion function).

The secrecy obtained through the error-rate and distortion formulations is sometimes referred to as *partial secrecy*.

A summary of the described secrecy criteria is provided in Table 2.1.

Table 2.1: Examples of secrecy criteria

Secrecy	**Goal**	Constraint
Perfect Secrecy	exact statistical independence	$I(M;Z^n\|\mathcal{C}_n) = 0$
Strong Secrecy	asymptotic statistical independence	$\lim\limits_{n\to\infty} I(M;Z^n\|\mathcal{C}_n) = 0$
Weak Secrecy	asymptotic statistical independence	$\lim\limits_{n\to\infty} \frac{1}{n} I(M;Z^n\|\mathcal{C}_n) = 0$
Partial Secrecy	Non-decodability	$\mathbb{E}\{P_e(\mathcal{C}_n)\} \approx \frac{\|\mathcal{M}\|-1}{\|\mathcal{M}\|}$
	Bounded error-rate	$\mathbb{E}\{P_e(\mathcal{C}_n)\} \geq P_e^{min}$
	Bounded distortion	$\mathbb{E}\{d(M,\hat{M})\} \geq \hat{D} - \varepsilon$

It is possible to establish an ordering relationship between the secrecy metrics based on the distance between probability distributions [28]. In particular, perfect secrecy is stronger than strong secrecy which is stronger than weak secrecy [28]. Such ordering implies that a metric satisfying a stronger secrecy criterion also satisfies a weaker one. Such ordering is not completely clear with respect a distortion based approach, since there exists a dependency on the choice of distortion function [26].

A final comment is in order with respect to the secrecy metrics. While it is obviously preferable to choose a secrecy metric that is as strong as possible, at this point in time, such choice bears an impact in the code design, be it in terms of rate, delay or complexity. Therefore, for a given application, one should in fact choose a secrecy metric that allows us to trade-off all these quantities while providing a desirable secrecy level. For instance, a streaming application may only require that the distortion of the eavesdropper is high enough to affect is perceptual quality. This would allow an increase in the transmission rate for the legitimate party that could reduce its own distortion (when compared to a more strict secrecy constraint).

2.3 Fundamental Limits of Secure Communication

The notions of reliability and secrecy defined above can be used to establish the fundamental limits of secure communication. These limits answer the question of what is the largest rate at which we can communicate under a given reliability and secrecy constraint (i.e. the secrecy capacity). It is possible to combine any of the criteria presented above. However, we will restrict our attention to the most common characterizations.

2.3.1 Weak and Strong Secrecy

Weak and strong secrecy have a similar characterization. A system is said to operate with weak secrecy if it satisfies conditions (2.2) and (2.8), while a system is said to operate with strong secrecy if it satisfies conditions (2.2) and (2.7). The following definitions are needed for the definition of the weak secrecy capacity.

Definition 3 (Weak rate-equivocation pair)**.** A weak rate-equivocation pair (R, R_e) is said to be achievable for the wiretap channel if there exists a sequence of $(2^{nR}, n)$ codes $\mathcal{C}_n$ such that:

1. $\lim_{n\to\infty} P_e(C_n) = 0$;
2. $\lim_{n\to\infty} \frac{1}{n} H(M|Z^n) \geq R_e$.

Definition 4 (Weak rate-equivocation region)**.** The weak-rate equivocation region of a wiretap channel is given by the closure of all the achievable weak rate-equivocation pairs (R, R_e), i.e.

$$R^{WTC} \triangleq \text{closure}\left(\{(R,R_e) : (R,R_e) \text{ is achievable}\}\right). \tag{2.10}$$

Definition 5 (Weak secrecy capacity)**.** The weak secrecy capacity of a wiretap channel is given by the supremum of all the achievable weak rate-equivocation pairs (R, R_e), such that $R = R_e$, i.e.

$$C_s^{WTC} \triangleq \sup_R \{R : (R, R) \in R^{WTC}\}. \tag{2.11}$$

To define the strong secrecy capacity we need the following definition.

Definition 6 (Strong rate-equivocation pair)**.** A strong rate-equivocation pair (R, R_e) is said to be achievable for the wiretap channel if there exists a sequence of (2^{nR}, n) codes $\mathcal{C}_n$ such that:

1. $\lim_{n\to\infty} P_e(C_n) = 0$;

2. $\lim_{n\to\infty} H(M|Z^n) \geq R_e$.

Then, the definition of the achievable strong rate-equivocation region and strong secrecy capacity are equal to Def. 4 and Def. 5 where the (R, R_e) pairs are strong rate-equivocation pairs.

2.3.2 Weak Secrecy with Lossy Reconstruction

If some distortion is allowed at the side of the legitimate receiver, we can provide a rate-distortion formulation to the weak secrecy problem using conditions (2.4) and (2.8). In this case, we can extend the previous definition of achievable rate to account for lossy reconstruction.

Definition 7 (Weak rate-equivocation pair with lossy reconstruction)**.** A weak rate-equivocation pair (R, R_e) with lossy reconstruction parameter $\tilde{D}$ is said to be achievable for the wiretap channel if there exists a sequence of (2^{nR}, n) codes $\mathcal{C}_n$ such that:

1. $\liminf_{n\to\infty} E\{d(M, \tilde{M})\} \leq \tilde{D} + \varepsilon$;

2. $\lim_{n\to\infty} \frac{1}{n} H(M|Z^n) \geq R_e$.

Then, it is possible to define the admissible region rate-distortion-equivocation region and find the largest rate at which we can communicate under weak secrecy that satisfies our rate-distortion constraint.

2.3.3 Rate-Distortion Secrecy

The above formulation relaxes the reliability constraint, but we may also relax the secrecy constraint by conditioning the eavesdropper to operate under a distortion constraint (rather than equivocation or equivocation rate). The following formulation is particularly useful when sources are continuous. Using constraints (2.4) and (2.9) we can define the following pair.

Definition 8 (Rate-distortion pair)**.** A rate-distortion pair $(R, \hat{D})$ with lossy reconstruction parameter $\tilde{D}$ is said to be achievable for the wiretap channel if there exists a sequence of $(2^{nR}, n)$ codes $\mathcal{C}_n$ such that:

1. $\liminf_{n\to\infty} E\{d(M,\tilde{M})\} \leq \tilde{D}+\varepsilon;$
2. $\limsup_{n\to\infty} E\{d(M,\hat{M})\} \geq \hat{D}-\varepsilon;$

2.4 Secrecy over the Wiretap Channel Model

The definitions provided before allow us to perform suitable choices when we wish to use physical-layer security schemes. Depending on the application at hand we may wish enforce a strict secrecy policy or choose to trade-off reliability and/or secrecy for rate. Moreover, depending on the type of source, one might prefer choosing a certain criterion over other criteria. This can only be accomplished if we are able to characterize these fundamental limits.

For the general wiretap channel introduced in 2.1 the weak rate-equivocation region and weak secrecy capacity are as follows.

Theorem 1. *([14],[3, Corollary 3.3]) Consider a wiretap channel* $(\mathcal{X}, \mathcal{Y}, \mathcal{Z}, p_{YZ|X}(y,z|x))$*. For any joint distribution* p_{UVX} *on* $\mathcal{U}\times\mathcal{V}\times\mathcal{X}$ *that factorizes as* $p_U p_{V|U} p_{X|V}$*, define the set* $R^{WT}(p_{UVX})$ *as*

$$R^{WT}(p_{UVX}) = \left\{ (R,R_e) : \begin{array}{c} 0 \leq R_e \leq R \leq I(V;Y) \\ 0 \leq R_e \leq I(V;Y|U) - I(V;Z|U) \end{array} \right\}.$$

Then, the weak rate-equivocation region for this wiretap channel is the convex set

$$R^{WT} = \bigcup_{p_{UVX}} R^{WT}(p_{UVX}). \tag{2.12}$$

Corollary 1. *([14], [3, Corollary 3.4]) The weak secrecy capacity of the discrete memoryless broadcast wiretap channel is*

$$C_s^{WT} = \max_{p_{VX}} I(V;Y) - I(V;Z).$$

The above characterizations introduce two auxiliary random variables U and V. They both relate to the wiretap channel as follow: U relates to the information decodable by Bob and Eve while V relates to the encoder randomization. Hence, it is not strange that U does not appear in the characterization of the secrecy capacity, since we wish Eve to obtain no information at all. While general, such characterization fails to give a strong intuition with respect to how randomization can affect secrecy without solving the associated maximization problem, which in general is a arduous task. Fortunately, this characterization can be simplified for certain classes of channels, notably those that can be characterized by an explicit advantage.

Definition 9 (Physically degraded channel)**.** A channel ($\mathcal{X}$, $\mathcal{Z}$, $p_{Z|X}$) is said to be physically degraded with respect to a channel ($\mathcal{X}$, $\mathcal{Y}$, $p_{Y|X}$) if, for all triples $(x,y,z) \in \mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$, $p_{YZ|X}(y,z|x)$ factorizes as $p_{Z|Y}(z|y)p_{Y|X}(y|x)$, i.e. if $\mathcal{X} \to \mathcal{Y} \to \mathcal{Z}$.

Definition 10 (Stochastically degraded channel)**.** A channel ($\mathcal{X}$, $\mathcal{Z}$, $p_{Z|X}$) is said to be stochastically degraded with respect to a channel ($\mathcal{X}$, $\mathcal{Y}$, $p_{Y|X}$) if there exists a channel $p_{Z|Y}(z|y)$ such that, for all tuples $(x,z) \in \mathcal{X} \times \mathcal{Z}$, $p_{Z|X}(z,x)$ can be written as $p_{Z|X}(z,x) = \sum_{y \in \mathcal{Y}} p_{Z|Y}(z|y)p_{Y|X}(y|x)$, i.e. if the channel ($\mathcal{X}$, $\mathcal{Z}$, $p_{Z|X}$) has the same marginal distribution as a channel that is physically degraded w.r.t ($\mathcal{X}$, $\mathcal{Y}$, $p_{Y|X}$).

Definition 11 (Noisier channel)**.** A channel ($\mathcal{X}$, $\mathcal{Z}$, $p_{Z|X}$) is said to be noisier than a channel ($\mathcal{X}$, $\mathcal{Y}$, $p_{Y|X}$) if, for every V such that $V \to X \to (Y,Z)$, we have $I(V;Y) \geq I(V;Z)$.

Definition 12 (Less capable channel)**.** A channel ($\mathcal{X}$, $\mathcal{Z}$, $p_{Z|X}$) is said to be less capable than a channel ($\mathcal{X}$, $\mathcal{Y}$, $p_{Y|X}$) if, for every input X, we have $I(X;Y) \geq I(X;Z)$.

Using the above definitions, we refer to *degraded wiretap channel* when a wiretap model is comprised of a wiretap channel that is physically or stochastically degraded w.r.t the main channel. Likewise, we refer to *noisier wiretap channel* when a wiretap model consists of a wiretap channel that is noisier than main channel and by less capable to wiretap model where the wiretap channel is less capable than main channel.

For the degraded wiretap channel, the weak rate-equivocation region and weak secrecy capacity can be characterized as follows.

Theorem 2. *([14], [3, Theorem 3.2]) Consider a wiretap channel ($\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, $p_{YZ|X}(y,z|x)$) such that the wiretap channel is physically or stochastically degraded w.r.t the main channel. Define the set $R^{WT}(p_X)$ as*

$$R^{WT}(p_X) = \left\{ (R,R_e) : \begin{array}{c} 0 \leq R_e \leq R \leq I(X;Y) \\ 0 \leq R_e \leq I(X;Y) - I(X;Z) \end{array} \right\}.$$

Then, the weak rate-equivocation region for this wiretap channel is the convex set

$$R^{WT} = \bigcup_{p_X} R^{WT}(p_X). \tag{2.13}$$

Corollary 2. *([14], [3, Corollary 3.1]) The weak secrecy capacity of the physically or stochastically degraded wiretap channel is*

$$C_s^{WT} = \max_{p_X} I(X;Y) - I(X;Z).$$

These regions hold also in the case of noisier wiretap channels and less capable wiretap channel [3]. In particular, if the same input distribution P_X maximizes both $I(X^n;Y^n)$ and $I(X^n;Z^n)$ we can replace the above terms by the individual channel capacities.

From the above rate-equivocation regions we can see that a strong or weak secrecy formulation has the following implications. If the wiretap channel is less noisy than the main channel, then the secrecy capacity is zero. On the other hand, if the wiretap channel is noisier than the main channel, but if the difference capacities of both channels is very small, then the secrecy capacity will also be very small. Consequently, the allowed transmission rate will be reduced, which may impair the viability of many applications.

Let us now consider the general wiretap model, where we allow the legitimate receiver to have some distortion in his estimates. While solutions to more general models are known [21], we will present the results for the less noisy wiretap channel.

Theorem 3. *([18, Theorem 1]) Consider a wiretap channel ($\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, $p_{YZ|X}(y,z|x)$) such that the main channel is less noisy than the wiretap channel. A rate-equivocation pair (R,R_e) with a lossy reconstruction parameter $\tilde{D}$ is achievable if and only if there exist X, Y and Z such that*

$$\left\{ \begin{array}{c} \frac{1}{n} I(X^n;Y^n)R \geq R(\tilde{D}) \\ 0 \leq R_e \leq \frac{1}{\log|\mathcal{M}|}[H(M) - R(\tilde{D})] + \frac{1}{n} I(X^n;Y^n|Z^n)R \end{array} \right\},$$

where $R(\tilde{D})$ is the ordinary rate-distortion function.

This results can be interpreted as follows: if distortion $\tilde{D}$ is allowed at the legitimate receiver, then a code attaining the rate-distortion function $R(\tilde{D})$ will induce an uncertainty $\frac{1}{\log|\mathcal{M}|}[H(M) - R(\tilde{D})]$, while the channel will induce the remaining part. This further indicates that this pair can be achieved by concatenating an optimal source code and an optimal wiretap code [18].

Intuitively, relaxing the reliability constraint allows us to achieve a larger transmission rate. It also impacts secrecy, as the distortion allowed at the legitimate receiver will also increase the equivocation of the eavesdropper by the difference between the source entropy and the rate-distortion function. The price to pay in this case is an increase in the probability of error of the legitimate receiver (more precisely increased distortion).

Lastly, consider the case where the legitimate receiver are allowed a maximum distortion $\tilde{D}$ and the eavesdropper is imposed a minimum distortion of $\hat{D}$.

Theorem 4. *([27, Corollary 4]) Consider a wiretap channel ($\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, $p_{YZ|X}(y,z|x)$) such that the wiretap channel is physically or stochastically degraded w.r.t the main channel. Define the set $R_d^{WT}(p_X)$ as*

$$R_d^{WT}(p_X) = \left\{ (R, \hat{D}) : \begin{array}{c} R \geq \frac{1}{n} I(X^n; Y^n) \\ \min\limits_{z} E\{d(X,Y,z)\} \geq \hat{D} \end{array} \right\}.$$

Then, the rate-distortion region for this wiretap channel is closure of all the above tuples, i.e.

$$R_d^{WT} = \bigcup_{p_{Y|X}} R_d^{WT}(p_X). \tag{2.14}$$

This last case considers a relaxation of both the reliability and secrecy constraints. As in the previous case, the transmission rate can be increased by allowing a distortion $\tilde{D}$. In particular, the allowed transmission rate is larger than the channel capacity. On the other hand, the eavesdroppers equivocation is no longer bounded by the distortion allowed at the legitimate receiver, but instead a pre-fixed value for distortion is assigned. This may further allow an increase in transmission rate as this condition may be less stringent than the one stated in Theorem 3. In this case, the price to be paid in an increase in the distortion of the legitimate receiver's observations, as well as in the amount of information leaked to the eavesdropper.

Almost all practical code constructions strive to achieve either the weak or strong secrecy capacity. In fact, to the best of our knowledge, there are no practical code constructions designed specifically for regions defined in Theorems 3 and 4. However, with respect to the case of weak/strong secrecy with lossy reconstruction, we point out that the rate-equivocation regions can be achieved by the concatenation of an optimal source code

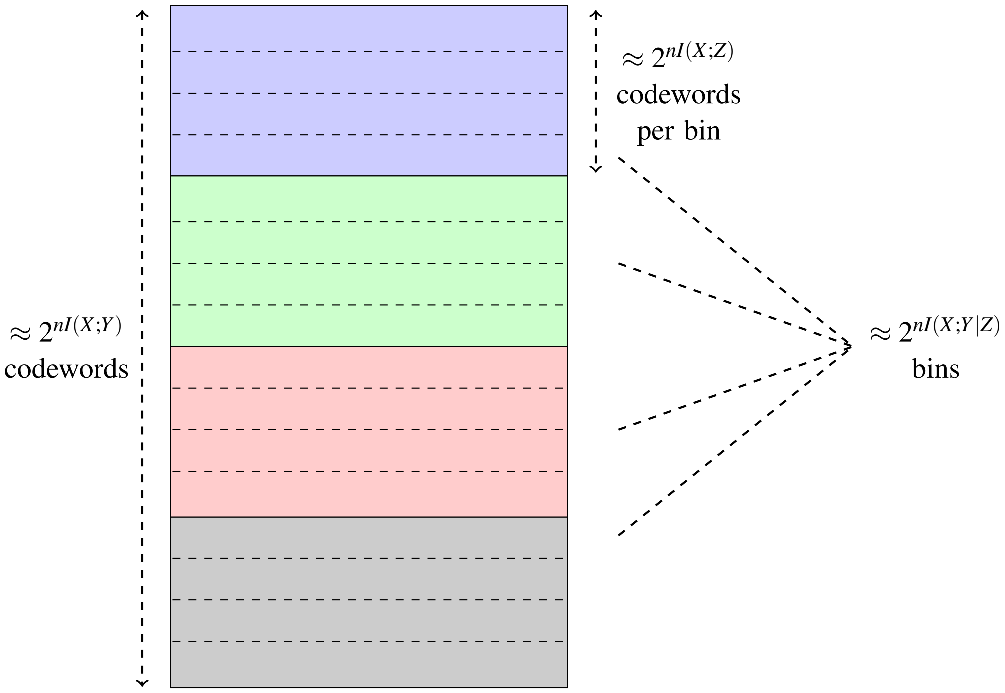


Figure 2.2: Binning structure.

and an optimal wiretap code [18] and, therefore, practical weak/strong secrecy achieving codes can be used within this context. As mentioned in Section 1.2, these code constructions are typically based on nested structures. The following section provides an overview of this design strategy and its connection to the fundamental limits of secure communication.

2.5 Practical Code Constructions for the Wiretap Channel

Most of the practical constructions of secrecy codes draw inspiration from the following random code construction. Let there be $2^{n(R+R_1)}$ codewords with symbols generated independently according to a distribution P_X. Divide the set of codewords into approximately 2^{nR} bins of size greater or equal than 2^{nR_1} and associate a source message to each bin. Then, to securely transmit some source message, select (at random) a codeword from the bin that corresponds to that same source message.

Using typicality arguments [3, Chapter 3.4], it is possible to show that reliability is achieved if $R+R_1 < I(X;Y)-\varepsilon$ and $R_1 < I(X;Z)-\varepsilon$. On the other hand, it is possible to show that the leakage rate $\frac{1}{n}I(M;Z^n)$ of this code construction is upper bounded by $\frac{1}{n}I(M;Z^n) \leq I(X;Z)-R_1+\varepsilon$. Therefore, it suffices to choose $R < I(X;Y)-I(X;Z)$ and $R_1 = I(X;Z)-\varepsilon$ to ensure reliable communications with a vanishing leakage rate (weak secrecy). An example of a binning structure with such an instantiation for R and R_1 is shown in Fig. 2.2.

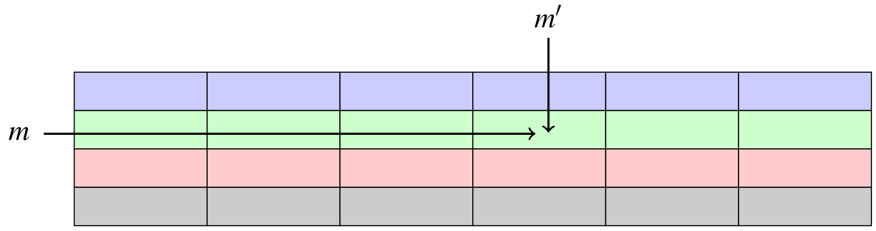


Figure 2.3: Nested code structure.

Such random code constructions are not useful in practice, since they require exponentially large memory for storage. However, it is possible to implement a similar idea using the notion of nested codes. Nested codes can be roughly described as codes that are formed by the union of several sub-codes. More precisely, a nested code $\mathcal{C}$ composed of $\lceil 2^{nR} \rceil$ sub-codes can be defined as $\mathcal{C} = \bigcup_{i=1}^{\lceil 2^{nR} \rceil} \mathcal{C}_i$, where each sub-code has 2^{nR_1} codewords. Hence, nested codes are somewhat analogous to the previously described binning structure (in the sense that we can interpret each bin as a sub-code). Transmission is achieved by choosing a message $m \in [1,\ldots,2^{nR}]$ and an index $m' \in [1,\ldots,2^{nR_1}]$ uniformly at random and transmitting the m'-th codeword in from the sub-code $\mathcal{C}_m$. This strategy is illustrated in Fig. 2.3, where now a particular bin is seen as a row of the nested code.

It is possible to show that the leakage rate of a nested code is bounded by $\frac{1}{n}I(M;Z^n) \leq \frac{1}{n}\left[I(X^n;Z^n) - H(M') + H(M'|MZ^n)\right] \leq \frac{1}{n}\left[nC_e - H(M') + H(M'|MZ^n)\right]$, where C_e denotes the capacity of the wiretap channel (a possible proof is provided in Appendix B). Note that $H(M')$ denotes the rate of each sub-code and $H(M'|MZ^n)$ denotes the uncertainty of the eavesdropper with respect to m' for a given sub-code. Then, it is sufficient to choose sub-codes that are capacity-achieving over the eavesdroppers channel in order to obtain weak secrecy, since in this case we have that $\frac{1}{n}H(M') \approx C_e$ and $\frac{1}{n}H(M'|MZ^n) \approx 0$. This seemingly simple guideline motivated the design of several explicit nested code constructions. These code constructions mostly differ in the way that nesting is implemented, by relying on different properties of the constituent codes. In [15], the general principles of practical code constructions are described as well as detailed constructions of many codes. For the sake of completeness, we will briefly review some of the possible code constructions.

In [29], the authors design nested codes using the cosets of duals of LDPC codes for a wiretap model composed of a noiseless main channel and a binary erasure wiretap channel. The code construction relies on the following property. Let a coset code $\mathcal{C}$ be formed by taking the cosets of a $(n,n-k)$ binary linear code $\mathcal{C}_0$ with generator matrix $\mathbf{G}_0$ and parity check matrix $\mathbf{H}_0$, i.e. $\mathcal{C} = \bigcup_s \mathcal{C}_0(s)$, where $\mathcal{C}_0(s) \triangleq \{\mathbf{x} \in \{0,1\}^n : \mathbf{H}_0\mathbf{x} = s\}$. If any sub-matrix of μ columns of $\mathbf{G}_0$ has rank μ it is possible to show that any sequence $\mathbf{x}'$

of length n with μ unerased positions will be consistent[2] with all cosets of $\mathcal{C}$. Moreover, the number of sequences that are consistent with $\mathbf{x}'$ is the same for all cosets. Thus, a necessary and sufficient condition for perfect secrecy when an eavesdropper observes sequences with μ unerased positions is that any sub-matrix of μ columns of $\mathbf{G}_0$ has rank μ. This property can be leveraged in the following way. If the wiretap channel is a binary erasure channel with erasure probability ε, with high probability, we have $\mu = 1-\varepsilon$. Consider an LDPC code $\mathcal{C}$ with a parity check matrix $\mathbf{H}$, drawn from an ensemble with a *belief propagation* (BP) decoding threshold α^{*}[3]. It is possible to show that, if we randomly select a $n\alpha$ columns of $\mathbf{H}$, with $\alpha < \alpha^{*}$, then, with high probability, the rank of this matrix will be $n\alpha$. Hence, to satisfy the aforementioned conditions on the generator matrix, we can use the parity check matrix of an LDPC code with a BP decoding threshold α^{*} as a generator matrix for our coset code, or in other words, we can use the dual code $\mathcal{C}^{\perp}$ and its cosets, to ensure perfect (weak) secrecy for a binary erasure wiretap channel with erasure probability $\varepsilon > 1-\alpha^{*}$.

In [30], a coset coding solution based on punctured LDPC codes is proposed for the AWGN wiretap model. While in the previous code construction the nested code structure was induced by coset encoding and the code was constructed using code properties inherited from the LDPC decoding thresholds, in [30] the nested structure is induced explicitly by the puncturing operation and the code is constructed directly using the capacity-achieving properties of LDPC codes. The construction is as follows. Consider an (n',l) LDPC code $\mathcal{C}'$, with parity check matrix $\mathbf{H}'$ of the from $\mathbf{H}' = [\mathbf{H}_1, \mathbf{H}_2]$, where $\mathbf{H}_2$ is a $(n'-l) \times (n'-l)$ lower triangular matrix. The codewords of $\mathcal{C}'$ can be thought of as vectors of the form $\mathbf{x} = [\mathbf{m}, \mathbf{m}', \mathbf{c}]$, where $|\mathbf{m}| = k$, $|\mathbf{m}'| = l-k$ and $|\mathbf{c}| = n'-l$, with $k < l$. Moreover, for fixed $\mathbf{m}$ and $\mathbf{m}'$, $\mathbf{c} = [\mathbf{m}, \mathbf{m}']\mathbf{H}_1^{\intercal}(\mathbf{H}_2^{-1})^{\intercal}$. We can induce a nested code structure using $\mathcal{C}'$ by creating an (n,k) code $\mathcal{C}$ that consists of all punctured codewords of the form $[\mathbf{m}'\ \mathbf{c}]$, where $\mathcal{C}$ is partitioned according to the punctured bits $\mathbf{m}$. Hence, $\mathbf{m}$ indicates which sub-code will be used for its transmission. The random choice of a codeword in the sub-code can be performed by randomly choosing the $l-k$ symbols of $\mathbf{m}'$. The transmitted codeword is then given by $\mathbf{x}' = [\mathbf{m}', \mathbf{c}]$ and weak secrecy can be achieved if $\mathcal{C}'$ is designed such that the resulting sub-codes are capacity approaching.

Another possible nested code construction based on two-edge type LDPC codes was proposed in [31]. Two-edge type LDPC codes provide a natural way of implementing a nested structure, since their parity-check matrices are of the form $\mathbf{H} = \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \end{bmatrix}$, where $\mathbf{H}$ is a $n(1-R) \times n$ matrix and $\mathbf{H}_1$ is an $n(1-R_1) \times n$ matrix, with $R_1 > R$. In particular,

[2] A sequence $\mathbf{x}'$ with μ erasures is said to be consistent with a sequence $\mathbf{x}$ if the values of the unerased positions in $\mathbf{x}'$ match with the values of the same positions in $\mathbf{x}$. A sequence $\mathbf{x}'$ is said to be consistent with a coset if the coset possesses at least one sequence that is consistent with $\mathbf{x}'$.

[3] The BP decoding threshold is the largest erasure probability such that a BP decoder can ensure vanishing bit error probability. It will be formally defined in Chapter 5

the linear code $\mathcal{C}$ defined by the matrix $\mathbf{H}$ is a sub-code of the linear code $\mathcal{C}_1$ defined by the matrix $\mathbf{H}_1$, and the distinct cosets of $\mathcal{C}$ in $\mathcal{C}_1$ form a partition of $\mathcal{C}$. Each coset of $\mathcal{C}$ in $\mathcal{C}_1$ consists of the solutions of the equation $\mathbf{Hx} = [\mathbf{H}_1\mathbf{x}\ \mathbf{H}_2\mathbf{x}] = [\mathbf{0}\ \mathbf{m}]$ for some $\mathbf{m}$. To encode a secret message $\mathbf{m}$ we randomly choose a solution $\mathbf{x}$ from all the solutions of the equation above. This can be explicitly accomplished by creating a generator matrix $\mathbf{G}' = \begin{bmatrix} \mathbf{G}^* \\ \mathbf{G} \end{bmatrix}$, where $\mathbf{G}$ is the generator matrix associated with $\mathbf{H}$, $\mathbf{G}^*$ is a matrix composed of linearly independent rows and $\mathbf{G}'$ forms a basis for the code with parity check matrix $\mathbf{H}_1$. Then, $\mathbf{x}$ can be computed as $\mathbf{x} = [\mathbf{m}, \mathbf{m}'] \begin{bmatrix} \mathbf{G}^* \\ \mathbf{G} \end{bmatrix}$, where $\mathbf{m}'$ is a vector of nR random bits. Consequently, to achieve weak secrecy we only need for the cosets charaterized by $\mathbf{H}$ to be capacity-achieving for the eavesdropper's channel.

Finally, polar codes have also been used to design nested structures [32, 33, 34]. The basic idea behind polar codes is to use a specific linear transformation to encode the messages, such that when each bit is transmitted over its respective bit channel, it polarizes (i.e. it becomes either almost noise free or almost noisy). Moreover, these bit channels always polarize in the same direction, hence if the quality of the channel is measured, one can identify the set of bits that will become noise free (also known as good bit channels) and the set of bits that will become noisy (bad bit channels). Additionally, it can be shown that the fraction of good bit channels converges to the channel capacity. Now suppose that the wiretap channel is degraded with respect to main channel. Using the above results its possible to identify three sets of channels: the set of bits channels that are only decodable by the legitimate receiver, the set of bit channels that are decodable by both receivers, and the set of bit channel that are not decodable by the legitimate receiver. Then, the secret bits can be sent over the set of channels that is decodable only by the legitimate receiver, while random bits are sent over the bit channels that are decodable by both and *frozen bits* are sent onto the channels that are not decodable by any. The nested code structure is implicit in the choice of the set of channels, since partitioning the bit channels induce a coset code.

While in general, the above constructions only achieve weak secrecy, it is possible to show that under additional constraints they may achieve also strong secrecy. For instance, [23] shows that the duals of LDPC codes with large girth are able to ensure strong secrecy. However, this is achieved at the cost of the achievable rate. If the main channel is noiseless, polar codes can also offer strong secrecy in [32]. Table 2.2 (from [15]) summarizes the state-of-the-art in coding for secrecy.

Table 2.2: Comparison of code families

Constituent codes	Secrecy	Main channel	Eavesdropper's channel
Duals of LDPC	weak [29]	noiseless	erasure
Duals of LDPC	strong [35, 23]	noiseless	erasure
Two-edge LDPC codes	weak [31, 36]	erasure	erasure and degraded
Polar codes	weak [32, 34, 33]	binary symmetric	binary symmetric and degraded
Polar codes	strong [32]	noiseless	symmetric

2.6 Discussion

In this chapter we reviewed the basic principles underlying the information-theoretic security model. In particular, we reviewed the definitions of the generalized wiretap channel and wiretap codes. We provided the characterization of the fundamental limits of secure communications for this channel model under multiple reliability and secrecy constraints, as well as examples of common secrecy code constructions for achieving weak secrecy capacity. In the light of these results, let us revisit some of the design choices stated in Section 1.2.

Throughout this thesis, the proposed wiretap codes are deterministic. We have seen that stochastic encoding can achieve a leakage rate $\frac{1}{n}I(M;Z^n) \leq \frac{1}{n}\big[I(X^n;Z^n) - H(M') + H(M'|MZ^n)\big]$. In this case, the randomization over the choice of M' provided a simple guideline to design codes with vanishingly small leakage: choose a code such that $\frac{1}{n}H(M' \approx \frac{1}{n}I(X^n;Z^n)$ to cancel the leakage of information to the eavesdropper and such that $\frac{1}{n}H(M'|MZ^n) \approx 0$. If codes are deterministic, the leakage rate $\frac{1}{n}I(M;Z^n) \leq \frac{1}{n}I(X^n;Z^n)$. This means that the leakage rate of a deterministic code will be less or equal to the channel capacity of the wiretap channel. From an operational perspective, it is still possible to achieve a vanishingly small leakage rate with a deterministic code if somehow one is able to reduce the wiretap channel to a channel with a vanishingly small capacity. On the other hand, $\frac{1}{n}I(M;Z^n) = \frac{1}{n}\big[H(M) - H(M|Z^n)\big]$. This definition suggests the following obvious observation: reducing the rate of source messages $H(M)$ also reduces the leakage to the eavesdropper. Therefore, the rate of source messages can be used to bound the leakage rate. Of course that reducing the rate of source messages affects negatively the legitimate party. However, it suggests that it is possible to control the leakage rate to some extent, if the difference between $H(M|Z^n)$ and $H(M|Y^n)$ is exploited properly. These two aspects support the intuition behind the proposed code designs. More precisely, the coding scheme presented in Chapter 3 uses the idea of restricting the source rate to ensure that an eavesdropper has a lower bound on his distortion, while allowing the legitimate receiver to lower its own distortion when the channel to the eavesdropper becomes poor. The coding scheme in Chapter 4 uses the idea of emulating a poor wiretap channel by designing a code that is unfit for that channel. In particular, it uses the noise of the main channel to design a code that operates with negligible error only below this noise threshold. If the wiretap channel in noisier than the main channel, then it is possible to parametrize the code to induce large errors on the eavesdroppers estimates. Chapter 5 also uses the idea of emulating a poor channel to the eavesdropper by considering a random puncturing strategy, where puncturing is essentially used to introduce erasures. Therefore, a new (artificial) wiretap channel is created, trying to ensure this channel has a vanishingly small capacity.

Another aspect that is present in this thesis is the separation of codes according to the source type (continuous or discrete). With this respect, we note that wiretap codes (possibly stochastic) can also be defined for continuous random variables. They have the same structure as the codes defined above, though defined over continuous sets. For example, assuming that sources take values from the set of the real numbers $\mathbb{R}$, the message set would be defined over the support set of the source outputs (i.e. $\mathcal{M} \subseteq \mathbb{R}$). The encoding function would operate over these continuous variables and the decoding function would map channel outputs also onto continuous variables (i.e. $g: \mathcal{Y}^m \to \mathcal{M} \subseteq \mathbb{R} \cup \{?\}$). As noted before, such construction can be avoided if separation theorems hold and assume the existence of practical optimal source and wiretap codes [37, 38]. For this reason, practical code constructions for continuous sources are almost non-existent in their own. For instance, [37] proposed the use an optimal vector quantizer, whose outputs are coded using a wiretap code. To achieve a graceful SNR degradation when there is an SNR mismatch, the authors also superimpose the coded message with a scaled version of the quantization error. In [38], the authors propose a scheme that does not use an explicit quantization of the source, but a rather pre-coding stage where the encoded message is added with a properly scaled version of the source message and then encoded using a wiretap code. In both cases, a fixed leakage rate is assumed and the authors study the impact of channel mismatch on the distortion of the legitimate user. In contrast with this approach, our goal is to design codes with finite block-lengths, where the above separation arguments may not hold. Consequently, we explore two strategies for designing codes for continuous sources. The first is a (digital) joint source-channel code based on scalar quantizers and the second is a (fully) continuous code based on bandwidth expansion mappings. The objectives of each construction are different. While the first tries to explore quantization (and channel) noise for secrecy, the second tries to explore the mapping between the source and the channel space to guarantee secrecy. Clearly, the performance of these codes cannot be assessed in the same manner as discrete codes. While for discrete codes we can measure efficiency through the code rate $R = \frac{1}{n} \log |\mathcal{M}|$, in the continuous case the efficiency of the code should to be measured by other characteristics, such as the bandwidth expansion. Furthermore, the metrics that are used in assessing the secrecy performance of wiretap codes for discrete sources may lose their meaning when moving towards continuous sources. For instance, the continuous representation of equivocation is the differential entropy, which does not have the same operational meaning has its discrete counterpart (and in fact could be negative). In particular, for continuous sources we adopt distortion as a secrecy metric as it is a measurable quantity and provides some operational meaning to secrecy.

Recall that Chapters 3 and 4 address the case of continuous sources, while Chapter 5 addresses the case of discrete sources.

Chapter 3

Scalar Quantization under Secrecy Constraints

Scalar quantization generally refers to the process of partitioning a continuous interval onto a finite set of disjoint sub-intervals, indexed in an arbitrary manner. The main idea is that one can represent the values in each sub-interval through its associated index, thus compressing the representation of an infinitely large set of values to a finite (or countably infinite) one. This partition can be characterized by a set of *thresholds* (or quantizer boundaries) defining each of the sub-intervals. Obviously, scalar quantization can be used as a building block to a communication system where the source is continuous. In particular, it is possible to partition the support set of the source into index sub-intervals for which some channel codeword is assigned. Then, each source value falling in a given sub-interval will be mapped onto the channel codeword associated with the sub-interval's index. If the quantizer is of fixed rate (meaning that each index can be represented by a binary sequence of the same length), then the rate of the quantizer is a function of the number of partitions though the relationship $R = \lceil \log N \rceil$, where N denotes the number of partitions. Decoding can then be accomplished as usual by estimating which source value was transmitted given that a particular codeword was observed.

The quality of a quantizer can be assessed through a distortion measure comparing the original source value with its estimate. Then, designing a good scalar quantizer is the matter of finding a good trade-off between the allocated rate and a prescribed distortion. In general, the distortion associated with a scalar quantizer is a function of the quantizer boundaries. If this quantizer is used in a communication system subject to noise, it is also a function of the chosen index assignments. This contrasts with the case where scalar quantization is used in a source coding context, which essentially sees channels as being noiseless. Thus, designing a scalar quantizer that performs optimally for a given channel amounts to finding the optimal partitions as well as the optimal index assignments. Consequently, scalar quantizers for channels of different quality may differ in both these

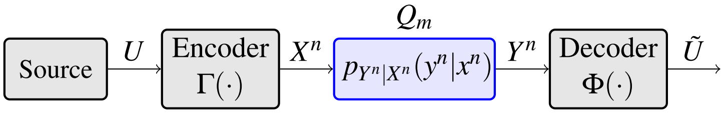


Figure 3.1: Communication system for transmission of continuous sources using with a joint source-channel code based on a scalar quantizer.

parameters. This feature is what we will explore in the context of secrecy. More precisely, we wish to find scalar quantizers, i.e. a set of partitions and index assignments, such that the distortion to the legitimate receiver is minimized while placing a lower bound on the eavesdropper's distortion. In particular, we extend the work of Farvardin and Vaishampayan [39], which considered the problem of scalar quantizer design for noisy channels without secrecy constraints, to account for a third malicious user who obtains noisy versions of the transmitted messages. While both problems can be formalized as optimization problems, their nature is different since [39] deals with an unconstrained minimization problem and we deal with a constrained optimization problem.

3.1 Channel-Optimized Scalar Quantizers

Consider the communication system illustrated in Fig. 3.1. A continuously-valued, time-discrete, memoryless source U with outputs $u \in \mathbb{R}$, drawn according to the probability density function (PDF) $p_U(u)$ is to be transmitted to some receiver over a noisy channel. This can be achieved by using an encoder that operates in a single-letter fashion by mapping each source output u onto a vector of n channel (input) symbols $X^n = (X_1, X_2, \dots, X_n)$.

We will consider a two-stage encoding procedure. First a quantization function $q : \mathbb{R} \to \mathcal{I}$ maps the output of the continuous-valued source U onto a discrete-valued *quantization index* I with realizations $i \in \mathcal{I} = \{0, 1, \dots, L-1\}$, such that $i = q(u)$ and where $L = |\mathcal{I}|$ denotes the number of quantization intervals. This operation can be described in terms of the boundaries (or *quantization thresholds*) $B^L = (B_1, B_2, \dots, B_{L-1})$ that partition the support of the PDF $p_U(u)$ into L disjoint and exhaustive regions $\mathcal{B}(i) =]B_i, B_{i+1}]$, $i = 0, 1, \dots, L-1$, with $B_0 = -\infty$ and $B_L = +\infty$. A source sample u is mapped onto the quantization index i, if $u \in \mathcal{B}(i)$, i.e. if $B_i < u \leq B_{i+1}$. The *channel encoder* $f : \mathcal{I} \to \mathcal{X}^n_{\mathcal{I}} \subseteq \mathcal{X}^n$ then maps the quantization indices I onto a vector of channel symbols X^n such that $X_i^n = f(i)$. The overall encoding procedure is expressed in terms of the *encoding function* $\Gamma : \mathbb{R} \to \mathcal{X}^n_{\mathcal{I}}$ such that $x^n = \Gamma(u) = f(q(u))$, where $x^n = (x_1, x_2, \dots, x_n)$ is a particular realization of the vector X^n and each of its members is chosen within the set $\mathcal{X}$ such that $x^n \in \mathcal{X}^n_{\mathcal{I}}$. The transmission rate is defined as $R = \log_2(|\mathcal{X}^n_{\mathcal{I}}|)$.

Subsequently, the vector of channel input symbols X^n is transmitted over the memoryless channel Q_m, where the vector of channel output symbols $Y^n = (Y_1, Y_2, \dots, Y_n)$ is obtained. A particular realization of Y^n is given by $y^n = (y_1, y_2, \ldots, y_n)$, where each of its members is drawn from the set $\mathcal{Y}$ such that $y^n \in \mathcal{Y}^n$. The channel Q_m is thus fully defined by $(\mathcal{X}_{\mathcal{I}}^n, p(y^n|x^n), \mathcal{Y}^n)$. Given the channel outputs, the receiver forms the estimate of the source message $\tilde{U}$, with realizations $\tilde{u} \in \mathbb{R}$. In particular, the *decoding function* is defined as $\Phi : \mathcal{Y}^n \to \mathbb{R}$, such that $\tilde{u} = \Phi(y^n)$. Alternatively, Φ can be described in terms of the quantizer *reconstructions levels* taking values in $\mathbb{R}$. The vector of reconstruction levels is denoted by $\tilde{U}^M = (\tilde{U}_1, \tilde{U}_2, \ldots, \tilde{U}_M)$, where $M = |\mathcal{Y}^n|$ represents the number of reconstruction levels. Under the distortion metric $d(u, \tilde{u}) : \mathbb{R} \times \mathbb{R} \to [0, \infty[$, the fidelity of the estimates of the receiver is quantified by the overall end-to-end distortion $D(\Gamma, \Phi) = E\{d(U, \tilde{U})\}$, where $E\{\cdot\}$ denotes expectation.

If one knows the statistics of the channel, it is possible to design jointly the source and channel encoders that minimize the end-to-end distortion under a certain criterion. The design of channel-optimized scalar quantizers without security considerations was addressed, among others, by Fine [40] and Farvardin and Vaishampayan [39]. In particular, [39] develops the necessary conditions of an optimal system using the mean square error (MSE) as the distortion criterion. For convenience, we will summarize the derivation of the necessary conditions for optimality from [39] with a slightly different notation that will be subsequently useful.

Let us for now assume that we have a fixed decoder Φ. Our problem is to minimize the distortion $D(\Gamma, \Phi)$ at the receiving end. We have that

$$\begin{aligned} D(\Gamma, \Phi) &= \int_{-\infty}^{+\infty} p_U(u) E\{(U - \tilde{U})^2 | U = u\} du \\ &= \int_{-\infty}^{+\infty} p_U(u) E\{(U - \tilde{U})^2 | U = u\} du \\ &= \int_{-\infty}^{+\infty} \sum_{x_i^n \in \mathcal{X}_{\mathcal{I}}^n} p_U(u) p_{X^n|U}(X^n = x_i^n | U = u) E\{(U - \tilde{U})^2 | U = u, X^n = x_i^n\} du \end{aligned}$$

Since $p_U(u)$ is always non-negative and since that for a given x_i^n, $p_{X_i^n|U}(x_i^n|u) = \mathbb{1}(\Gamma(u) = x_i^n)$, where $\mathbb{1}(\cdot)$ is the indicator function, it is sufficient to consider the minimization over the conditional expectation $E\{(u - \tilde{U})^2 | X^n = x_i^n\}$. Consequently, we need to find the mappings between all the possible values of u and the respective codewords x_i^n, such that this conditional expectation is minimized. This can be achieved by finding the region $\mathcal{B}(i)$ associated with x_i^n, such that, for $u \in \mathcal{B}(i)$ the aforementioned conditional expectation is minimized with respect to any other codeword x_j^n, $j \neq i$. Formally, we can define $\mathcal{B}(i)$ as

$$\mathcal{B}(i) \;=\; \{u : E\{(u-\tilde{U})^2|X^n = x_i^n\} \leq E\{(u-\tilde{U})^2|X^n = x_j^n\},\ \text{for all } j \neq i\}.$$

However, a more explicit characterization can be obtained. Define the region $\mathcal{B}(i,j)$ as the set on the real line for which mapping its elements on codeword x_i^n leads to a lower MSE when compared to a mapping onto codeword x_j^n. We have that

$$\mathcal{B}(i,j) \;=\; \{u : 2u\big(E\{\tilde{U}|X^n = x_j^n\} - E\{\tilde{U}|X^n = x_i^n\}\big) \leq E\{\tilde{U}^2|X^n = x_j^n\} - E\{\tilde{U}^2|X^n = x_i^n\}\}.$$

Then, the region $\mathcal{B}(i)$ is given by $\mathcal{B}(i) = \bigcap_{\substack{j=1\\ j\neq i}}^{M} \mathcal{B}(i,j)$.

Define the following auxiliary variables $\alpha_{i,j}$, $\beta_{i,j}$ and $\vartheta_{i,j}$ respectively as

$$\begin{aligned}
\alpha_{i,j} &\triangleq E\{\tilde{U}^2|X^n = x_j^n\} - E\{\tilde{U}^2|X^n = x_i^n\}\\
\beta_{i,j} &\triangleq E\{\tilde{U}|X^n = x_j^n\} - E\{\tilde{U}|X^n = x_i^n\}\\
\vartheta_{i,j} &\triangleq \frac{1}{2}\frac{\alpha_{i,j}}{\beta_{i,j}},\ \text{with } \beta_{i,j} \neq 0.
\end{aligned}$$

The set $\mathcal{B}(i,j)$ can be found by solving the inequality $2u\beta_{i,j} \leq \alpha_{i,j}$, whose solution is given by

$$\mathcal{B}(i,j) = \begin{cases} \emptyset & ,\text{if } \beta_{i,j} = 0 \text{ and } \alpha_{i,j} < 0\\ \mathbb{R} & ,\text{if } \beta_{i,j} = 0 \text{ and } \alpha_{i,j} \geq 0\\ ]-\infty, \vartheta_{i,j}] & ,\text{if } \beta_{i,j} > 0\\ [\vartheta_{i,j}, \infty[& ,\text{if } \beta_{i,j} < 0. \end{cases}$$

It is possible to see that $\mathcal{B}(i,j)$ is an interval and consequently so is $\mathcal{B}(i)$ since it is the finite intersection of multiple intervals. Moreover, when $\mathcal{B}(i,j)$ is non-empty, it is either unbounded ($\beta_{i,j} = 0$), left-bounded ($\beta_{i,j} < 0$) or right-bounded ($\beta_{i,j} > 0$). Hence, we can define lower and upper endpoints ϑ_i^l and ϑ_i^u as

$$\begin{aligned}
\vartheta_i^l &\triangleq \max_{j:\beta_{i,j}<0} \{\vartheta_{i,j}\}\\
\vartheta_i^u &\triangleq \min_{j:\beta_{i,j}>0} \{\vartheta_{i,j}\}.
\end{aligned}$$

If we assume that $\vartheta_i^l \leq \vartheta_i^u$, we have that

$$\mathcal{B}(i) = \begin{cases} \emptyset & ,\text{if } \exists j : \beta_{i,j} = 0 \text{ and } \alpha_{i,j} < 0\\ \mathbb{R} & ,\text{if } \forall j\ \ \beta_{i,j} = 0 \text{ and } \alpha_{i,j} \geq 0\\ [\vartheta_i^l, \vartheta_i^u] & ,\text{otherwise.} \end{cases}$$

Remark 1. In certain cases it is possible that $\vartheta_i^l > \vartheta_i^u$, when $\beta_{i,j} \neq 0$ for all j. A simple example consists of having the $\beta_{i,j}$ to be all left or right bounded intervals, for which the right-bounded intervals are all to the left of the left bounded intervals. This situation generally occurs in very noisy channels and leads to empty intersections (in practice it implies that x_n^i should not be used).

Remark 2. Without loss of generality endpoint ambiguities under adjacent intervals are assumed to be solved by considering the intervals to be right-open.

Following the definition of $\mathcal{B}(i)$, Lemma 1 summarizes the necessary conditions for optimal encoding.

Lemma 1 ([39])**.** *The optimal encoder Γ^* for a fixed decoder Φ is a mapping from u to $\mathcal{X}_{\mathcal{I}}^n$ s.t.*

$$\Gamma^* \triangleq \Gamma_\Phi(u) = x_i^n \quad , if\, u \in \mathcal{B}(i), i = 1, 2, \ldots, |\mathcal{X}_{\mathcal{I}}^N| \tag{3.1}$$

where x_i^n is the i-th channel input vector.

Let us now assume that we have a fixed encoder Γ and wish to find the optimal decoder Φ^*. This problem is recurrent in estimation theory. If we consider the distortion metric to be the MSE, then we are looking for the minimum mean square error (MMSE) estimator which is given by the conditional mean estimates of the source message. In our particular context, we obtain the following description of the optimal decoder.

Lemma 2 ([39])**.** *The optimal decoder Φ^* for a fixed encoder Γ is given by*

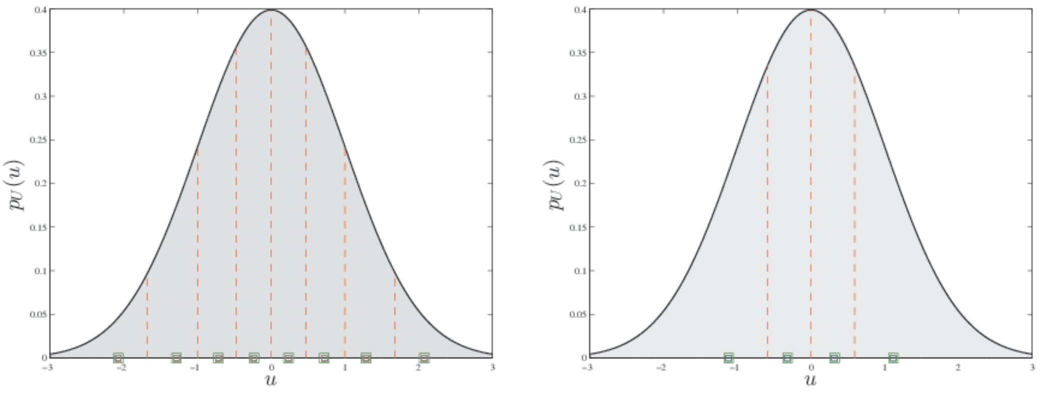
$$\Phi^* \triangleq \Phi_\Gamma(y_i^n) = E\{U|Y^n = y_i^n\}, i = 1, 2, \ldots, |\mathcal{Y}^n|, \tag{3.2}$$

where y_i^n is the i-th channel output vector.

The developed necessary conditions can be used within the context of an iterative algorithm, which successfully optimizes the encoder according to (3.1) and the decoder according to (3.2), assuming that $\vartheta_i^u < \vartheta_i^l$ for all i. Since these conditions lead to decreasing values of MSE, the algorithm converges. This strategy is similar to the generalized Lloyd-Max algorithm [41], [42].

Remark 3. Individually, (1) and (2) satisfy the necessary and sufficient conditions for optimality [39]. However, the iterative application of the two conditions does not necessarily satisfy the sufficient conditions for optimality of the system. Therefore, the application of such iterative algorithm will lead to a locally optimal solution, rather an a global optimal one.

One important characteristic of channel-optimized scalar quantizer is that, depending on channel conditions, the number of regions that compose the quantizer may be reduced



(a) LM-SC.

(b) CO-SC for a BSC(0.1).

Figure 3.2: Example of thresholds and reconstruction values of two scalar quantizers with $n = 3$ for a Gaussian source $U \sim \mathcal{N}(0,1)$.

Table 3.1: Performance of Lloyd-Max and channel-optimized scalar quantizers for a BSC(δ)

SNR (dB)	$\delta = 10^{-3}$	$\delta = 10^{-2}$	$\delta = 10^{-1}$
OPTA	17.86	16.60	9.59
LM-SC	13.82	9.86	1.56
CO-SC	13.86	10.55	4.67

with respect to the maximum number of quantization regions allowed by a given rate. The following example illustrates this property.

Example 1. Consider a Gaussian source $U \sim \mathcal{N}(0,1)$ to be encoded onto sequences of $n = 3$ bits and transmitted over a binary symmetric channel (BSC) with crossover probability δ. Fig 3.2 shows the thresholds and reconstruction values associated with a scalar quantizer based on the Lloyd-Max algorithm (LM-SC) and a channel-optimized scalar quantizer (CO-SC) for a BSC(δ), where $\delta = 10^{-1}$. It can be seen that channel-optimized scalar quantizers may lead to a quantizer that uses less quantization regions, a consequence of elevated channel-noise. Table 3.1 further summarizes the performance of both scalar quantizer designs for channel crossover probabilities $\delta \in \{10^{-}3, 10^{-2}, 10^{-1}\}$, comparing them with the optimum performance theoretically attainable (OPTA). In general, channel-optimized scalar quantizers offer an advantage with respect to the standard Lloyd-Max construction, although the difference is pronounced only when the channel-crossover probabilities are large enough. Still, both approaches are far from the optimum attainable SNR. On one hand this is due to the small value of n, while on the other hand, the form of error correction introduced by the index assignment stage is somewhat rudimentary.

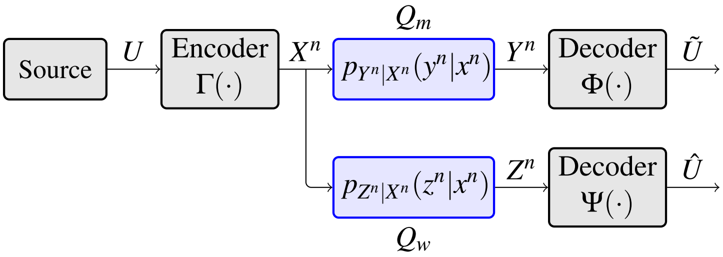


Figure 3.3: Wiretap model for analog sources.

3.2 Scalar Quantizers under Security Constraints

In the previous section we saw channel-optimized scalar quantizers could be designed when two users communicate over a noisy channel. Let us now extend this model to account for a third (malicious user). This scenario is depicted in Fig. 3.3. This model is an instance of the broadcast wiretap channel presented in Section 2.1 when the source U is continuous. However, we now consider an explicit decoder for the eavesdropper, mostly because we wish to make explicit statements with respect to the minimum distortion achieved by the eavesdropper.

Again, a sender is trying to transmit a confidential message to a given receiver. Let us assume a continuous-valued, time-discrete, memoryless source U with outputs $u \in \mathbb{R}$ having a PDF $p_U(u)$. The source message is quantized and channel coded into a sequence $X^n \in \mathcal{X}_{\mathcal{I}}^n$ as described in Section 3.1. The legitimate receiver obtains the vector of channel output symbols $Y^n \in \mathcal{Y}^n$ and forms source estimates $\tilde{U} \in \mathbb{R}$ through the decoding function Φ.

The unintended recipient has access to the transmitted sequence via the *wiretap channel* Q_w, which provides a vector of channel outputs $Z^n = (Z_1, Z_2, \ldots, Z_n)$. A particular realization of the vector Z^n is given by $z^n = (z_1, z_2, \ldots, z_n)$, where each of its members is drawn from the set $\mathcal{Z}$ such that $z^n \in \mathcal{Z}^n$. The eavesdropper forms source estimates $\hat{U}$ with realizations $\hat{u} \in \mathbb{R}$.

Using the distortion metric $d(u,\hat{u}) : \mathbb{R} \times \mathbb{R} \to [0,\infty[$, the fidelity of the estimates of the eavesdropper is quantified by the overall end-to-end distortion $D(\Gamma,\Psi) = E\{d(U,\hat{U})\}$, where Ψ is the eavesdroppers *decoding function* and is defined as $\Psi : \mathcal{Z}^n \to \mathbb{R}$ such that $\hat{u} = \Psi(\mathbf{z})$. The decoder Ψ is assumed to produce optimal estimates for the eavesdropper in the sense of minimizing the distortion $D(\Gamma,\Psi)$. This configures a worst-case scenario.

As with any physical-layer security scheme, we wish to exploit the nature of the communication channel to our benefit. In particular, our goal is to design a scalar quantizer that ensures that the eavesdroppers distortion lies above a prescribed level, while minimizing the distortion of the legitimate receiver. Hence, distortion is used both as a reliability and secrecy metric.

In this context, we provide a methodology for the design of a scalar quantizers with secrecy constraints in the spirit of [39]. More precisely, we formulate the problem of quantizer design as an optimization problem, where the goal is to minimize the legitimate receiver's distortion subject to a lower bound on the distortion of the eavesdropper, i.e. $D(\Gamma,\Psi) > \Delta$. This secrecy constraint controls the rate-secrecy trade-off and ultimately defines the secrecy level of the system.

In the following, we will assume that channel codewords have binary representations, i.e. $x_n^i \in \mathbb{F}_2^n$ for all i, and we will focus on the case where the distortion criterion amounts to the mean square error, i.e. $D(\Gamma,\Phi) = E\{(\tilde{u}-u)^2\}$ and $D(\Gamma,\Psi) = E\{(\hat{u}-u)^2\}$, which is a widely accepted distortion metric. However, we note that the problem formulation is sufficiently general to allow for other channel input alphabets as well as distortion metrics[1].

3.2.1 Problem Statement

The general problem we aim to solve is a non-linear constraint optimization problem of the form

$$\begin{aligned} &\text{minimize} \quad D(\Gamma,\Phi) \\ &\text{subject to} \quad D(\Gamma,\Psi) > \Delta. \end{aligned} \tag{3.3}$$

One way to solve (3.3) is to translate it into an unconstrained optimization problem using Lagrange multipliers [43, Chapter 5]. To this purpose we define the following Lagrangian function

$$L(\Gamma,\Phi,\Psi,\lambda) = D(\Gamma,\Phi) - \lambda(D(\Gamma,\Psi) - \Delta), \tag{3.4}$$

where λ is the *Lagrange multiplier* as well as the Lagrange dual function $g:\ \mathrm{dom}(\lambda) \to \mathbb{R}$ such that

$$g(\lambda) = \min_{\Gamma,\Phi,\Psi}\{L(\Gamma,\Phi,\Psi,\lambda)\}, \tag{3.5}$$

Let us denote the optimal solution (if it exists) of the problem with tuple $(\Gamma^*,\Phi^*,\Psi^*,\lambda^*)$, giving rise to the solution $L(\Gamma^*,\Phi^*,\Psi^*,\lambda^*)$. Under certain conditions it is possible to directly obtain the necessary conditions for optimality. For instance, if $L(\Gamma,\Phi,\Psi,\lambda)$, $D(\Gamma,\Phi)$ and $D(\Gamma,\Psi)$ are differentiable and $D(\Gamma,\Phi) - g(\lambda) = 0$, i.e. if the duality gap is zero, then the Karush-Kuhn-Tucker (KKT) conditions can be employed [43, Chapter 5].

[1] The reason for stating such assumptions at this point is due their implications with respect to what we may state about the optimality of the proposed approach.

In particular, if $(\Gamma^*, \Phi^*, \Psi^*, \lambda^*)$ is an optimal solution, then we have that

$$\begin{aligned} D(\Gamma^*, \Psi^*) - \Delta &\geq 0 \\ \lambda^* &\geq 0 \\ \lambda^*(D(\Gamma^*, \Psi^*) - \Delta) &= 0 \\ \nabla D(\Gamma^*, \Phi^*) - \lambda^* \nabla (D(\Gamma^*, \Psi^*) - \Delta) &= 0, \end{aligned}$$

where ∇ is the gradient function.

However, in the context of our quantization problem, the Lagrangian function is not differentiable in general. For instance, if the MSE is used as a distortion criteria, $L(\Gamma, \Phi, \Psi, \lambda)$ is not differentiable with respect to the encoder parameters (boundary thresholds)[2]. Thus, if we are to use the method of Lagrange multipliers, we are bound to obtain a sub-optimal solution. Nevertheless, such solution still satisfies our secrecy constraint (it is a local minimum with respect to the distortion at the legitimate receiver). If we wish to minimize the objective function given by (3.3) we can find the solution to the following problem

$$\underset{\lambda}{\operatorname{argmax}} \underset{\Gamma, \Phi}{\operatorname{argmin}} L(\Gamma, \Phi, \Psi, \lambda), \tag{3.6}$$

where $\lambda \in [0, \infty[$. We do not need to consider optimization of Ψ since we assume the eavesdropper will always use its optimal decoder. We will employ an iterative optimization strategy, similar to that of the Lloyd-Max algorithm [41], [42], in which the encoder and decoder are alternately optimized until convergence (or some stopping criterion is met). An overview of the strategy adopted to solve (3.6) is provided next.

3.2.2 Overview of the optimization strategy

Starting with an encoder/decoder pair Γ and Φ we iteratively compute the optimal encoder Γ^* (assuming a fixed decoder) and the optimal decoder Φ^* (assuming the previously found encoder). This procedure is repeated until convergence is achieved, at which point Γ^* and Φ^* are output.

The encoder optimization procedure makes use of the Lagrange dual principle (as described in Section 3.2.3) and tackles the problem of finding the optimal encoder as a function of the Lagrange multiplier λ. To achieve this, the optimal encoder is found for a fixed λ and then the optimal value of λ is found numerically through the Lagrange dual function. This two-stage procedure may incur in a loss of global optimality, since the solution to the Lagrangian dual function only provides, in general, a lower bound to the optimal solution.

[2] Changing the encoder parameters leads to an effective change on the size of the associated intervals. Since this change might lead to the disappearance of some other boundary, a discrete change in the number of quantization intervals occurs, which reflects as a non-differentiable points in the Lagrangian function.

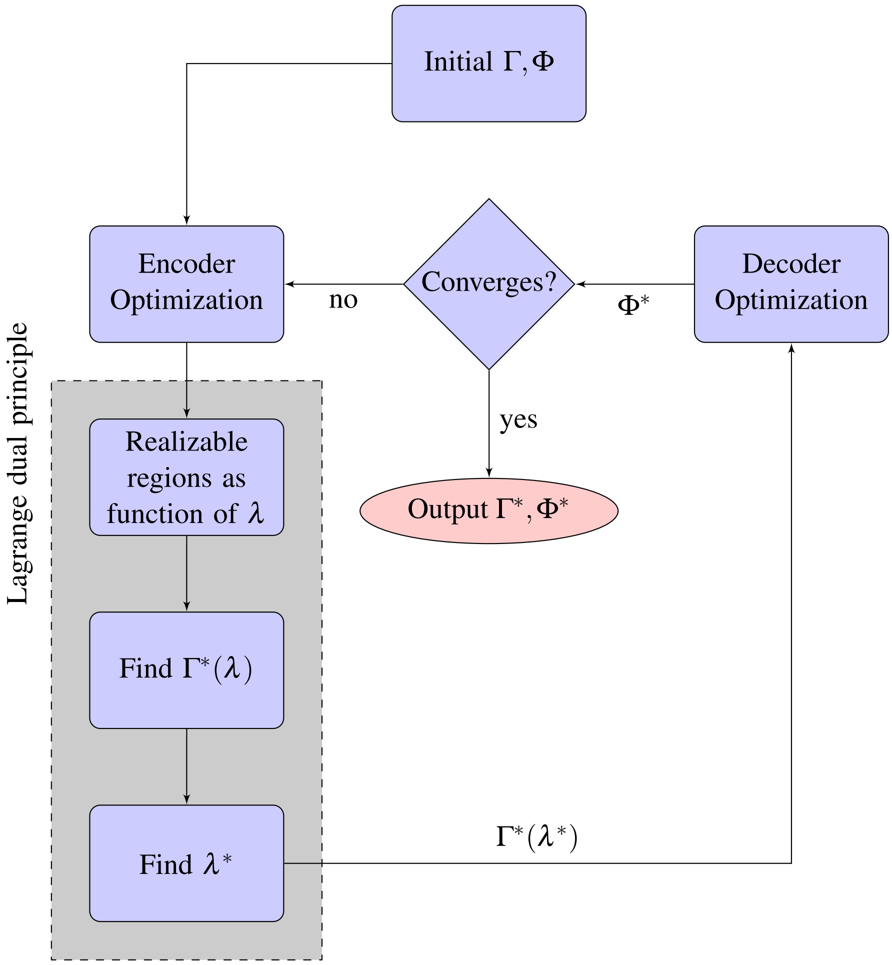


Figure 3.4: Overview of optimization procedure.

In the context of our problem, there is a further aspect that has to be taken into account. The encoder is a function of λ. In particular, the value of λ affects the encoder structure, in the sense that different values of λ may change for instance the number of quantizations intervals, as noted before. Accordingly, before finding the optimal encoder as a function of λ, we first find the regions for which the encoder structure is not changed as a function of λ, denoted as realizable regions. Then, the above two-stage procedure is used for each of these realizable regions. Consequently, the optimal encoder and the solution to the Lagrangian dual function must be found among all the individual solutions for each realizable region. An illustration of the flow for the complete optimization procedure is depicted in Fig 3.4.

3.2.3 Encoder Optimization

Following the principles from [39] we wish to develop the necessary optimality conditions for the encoder Γ, given a fixed decoder Φ, i.e. for the problem

$$\underset{\lambda}{\operatorname{argmax}}\,\underset{\Gamma}{\operatorname{argmin}}\, L(\Gamma,\Phi,\Psi,\lambda). \tag{3.7}$$

To reduce the complexity of the problem (which involves optimizing both Γ and λ) we can decouple the optimization in two stages, which can be approached subsequently. The encoder can be written as a function of of λ, i.e. $\Gamma = \Gamma(\lambda)$ and (3.5) can be simplified as

$$g(\lambda) = \min_{\Gamma}\{L(\Gamma,\Phi,\Psi,\lambda)\}, \tag{3.8}$$

where Φ and Ψ are considered to be given. From the Lagrange dual principle, we have that

$$L^* \geq \max_{\lambda \geq 0}\{g(\lambda)\}. \tag{3.9}$$

If equality holds in (3.9), the global optimal solution (Γ^*,λ^*) is given by

$$\Gamma^* = \Gamma^*(\lambda^*), \tag{3.10}$$

where

$$\Gamma^*(\lambda) = \underset{\Gamma}{\operatorname{argmin}}\{L(\Gamma,\Phi,\Psi,\lambda)\} \tag{3.11}$$

and

$$\lambda^* = \underset{\lambda \geq 0}{\operatorname{argmax}}\{g(\lambda)\}. \tag{3.12}$$

Consequently, the solution to (3.7) is $L^* = L(\Gamma^*,\lambda^*)$. If equality does not hold, then (3.9) merely presents a valid lower bound for L^* and the derived solution reflects, at most, a local optimal solution. Equations (3.10)- (3.12) suggest the following two-step procedure: 1) derive the encoder setting $\Gamma^*(\lambda)$ and 2) derive λ^* leading to the solution $\Gamma^* = \Gamma^*(\lambda^*)$.

3.2.3.1 Encoder Optimality Conditions for Fixed λ

Let us assume that λ is fixed. Using the definition of the distortion function we can define the objective function as a function of λ such that $L(\lambda) = L(\Gamma,\Phi,\Psi,\lambda)$. In particular, we have that

$$\begin{aligned} L(\lambda) &= D(\Gamma,\Phi)-\lambda(D(\Gamma,\Psi)-\Delta) \\ &= E\{(\tilde{U}-U)^2\}-\lambda(E\{(\hat{U}-U)^2\}-\Delta) \\ &= E\{(\tilde{U}-U)^2-\lambda((\hat{U}-U)^2-\Delta)\} \\ &= \int\limits_{u=-\infty}^{\infty} E\{(\tilde{U}-u)^2-\lambda((\hat{U}-u)^2-\Delta)|U=u\}\cdot p_U(u)\,du. \end{aligned}$$

The quantization step is deterministic. Therefore, $E\{(\tilde{U}-u)^2-\lambda((\hat{U}-u)^2-\Delta)|U=u\}=E\{(\tilde{U}-u)^2-\lambda((\hat{U}-u)^2-\Delta)|X^n=x_i^n\}$. Since $p_U(u)$ is non-negative, then $L(\lambda)$ is minimized if $E\{(\tilde{U}-u)^2-\lambda((\hat{U}-u)^2-\Delta)|X^n=x_i^n\}$ is also minimized, for all $i\in\mathcal{I}$. Thus, to find the optimal quantization regions and the corresponding channel code we need to find the optimal partition of u and the respective quantization indices, for all $i\in\mathcal{I}$. We will denote as $\mathcal{B}(i,\lambda)$, the optimal partition such that a source symbol u minimizes $E\{(\tilde{U}-u)^2-\lambda((\hat{U}-u)^2-\Delta)|X^n=x_i^n\}$, if $u\in\mathcal{B}(i,\lambda)$. Consider a pair of quantizer indices $j,k\in\mathcal{I}$ with $j\neq k$ and assume that λ is fixed. The region $\mathcal{B}(j,k,\lambda)$ of all u's that should be encoded onto j rather than k is given by

$$\begin{aligned} \mathcal{B}(j,k,\lambda) &= \{u: E\{(\tilde{U}-u)^2-\lambda((\hat{U}-u)^2-\Delta)|X^n=x_j^n\} \\ &\leq E\{(\tilde{U}-u)^2-\lambda((\hat{U}-u)^2-\Delta)|X^n=x_k^n\}\}. \end{aligned} \tag{3.13}$$

Define $\varepsilon_{j,k,\lambda}$, $\delta_{j,k,\lambda}$ and $\vartheta_{j,k,\lambda}$ respectively as

$$\varepsilon_{j,k,\lambda} \triangleq E\{\tilde{U}^2-\lambda\hat{U}^2|X^n=x_k^n\}-E\{\tilde{U}^2-\lambda\hat{U}^2|X^n=x_j^n\} \tag{3.14}$$

$$\delta_{j,k,\lambda} \triangleq E\{\tilde{U}-\lambda\hat{U}|X^n=x_k^n\}-E\{\tilde{U}-\lambda\hat{U}|X^n=x_j^n\} \tag{3.15}$$

$$\vartheta_{j,k,\lambda} \triangleq \frac{1}{2}\cdot\frac{\varepsilon_{j,k,\lambda}}{\delta_{j,k,\lambda}}. \tag{3.16}$$

Then, $\mathcal{B}(j,k,\lambda)$ can be found by solving the inequality $2u\delta_{j,k,\lambda}\leq\varepsilon_{j,k,\lambda}$. As before, $\vartheta_{j,k,\lambda}$ represents the threshold on the source samples u for the set associated with the pair of indices j and k. The quantization region $\mathcal{B}(j,k,\lambda)$ is given by

$$\mathcal{B}(j,k,\lambda)=\begin{cases}]-\infty,\vartheta_{j,k,\lambda}] & \text{, if } \delta_{j,k,\lambda}>0 \\ [\vartheta_{j,k,\lambda},\infty[& \text{, if } \delta_{j,k,\lambda}<0 \\ \mathbb{R} & \text{, if } \delta_{j,k,\lambda}=0 \text{ and } \varepsilon_{j,k,\lambda}\geq 0 \\ \emptyset & \text{, otherwise.} \end{cases}$$

Note that $\mathcal{B}(j,k,\lambda)$ is either a one-sided open interval, the real line or the empty set. Therefore, the overall quantization regions $\mathcal{B}(j,\lambda)$ can then be derived by subsequently intersecting the intervals given by $\mathcal{B}(j,k,\lambda)$ according to

$$\mathcal{B}(j,\lambda) = \bigcap_{k\in\mathcal{I}:k\neq j} \mathcal{B}(j,k,\lambda). \tag{3.17}$$

Since $\mathcal{B}(j,\lambda)$ is obtained by a subsequent intersection of intervals, it will be a single interval or the empty set. In case the intersection results in a single interval, we define the lower endpoint $\vartheta^{l}_{j,\lambda}$ and the upper endpoint $\vartheta^{u}_{j,\lambda}$ such that

$$\mathcal{B}(j,\lambda) = [\vartheta^{l}_{j,\lambda}, \vartheta^{u}_{j,\lambda}]. \tag{3.18}$$

It is important to point out that there might be cases where $\vartheta^{l}_{j,\lambda} \leq \vartheta^{u}_{j,\lambda}$ does not hold. For such cases it is possible to conclude that there is actually no value of u that should be encoded onto x^n_j, thus $\mathcal{B}(j,\lambda) = \emptyset$. The optimal quantization regions for our scalar quantizer are then given by

$$\mathcal{B}(j,\lambda) = \begin{cases} \emptyset & \text{, if } \exists k: \delta_{j,k,\lambda} = 0 \text{ and } \varepsilon_{j,k,\lambda} < 0 \\ \mathbb{R} & \text{, if } \forall k \;\; \delta_{j,k,\lambda} = 0 \text{ and } \varepsilon_{j,k,\lambda} \geq 0 \\ [\vartheta^{l}_{j,\lambda}, \vartheta^{u}_{j,\lambda}] & \text{, otherwise} \end{cases}$$

The partition induced by $\mathcal{B}(j,\lambda)$ provides us with the optimal encoder subject to our constraints, for a particular value of λ. Thus, the optimal encoder is given as follows.

Lemma 3. *Let the quantization step be an injective function. For a fixed λ and decoder Φ, the optimal encoder $\Gamma^*(\lambda)$ is given by*

$$\Gamma^*(\lambda) \triangleq \Gamma(u,\lambda) = x^n_j, u \in \mathcal{B}(j,\lambda), j = 1,2,\ldots,|\mathcal{X}^N_{\mathcal{I}}|, \tag{3.19}$$

where x^n_j is the j-th channel input vector.

The encoder given in Lemma 3 satisfies the necessary conditions for optimality for a fixed decoder and λ. To find the final quantizer design, we should solve (3.12) with some care since the value of λ may affect the configuration of the encoder. In particular, some regions of λ may result in configurations that are not possible (upper boundaries are below lower boundaries) or may be contained in some other region of λ and therefore do not necessarily need to be accounted for. As noted in Section 3.2.2, it is possible to identify these changes as a function of λ, and therefore efficiently obtain the realizable quantization regions, for which the aforementioned maximization problem can be solved individually.

3.2.3.2 Realizable Quantization Regions

To simplify the problem it is possible to partition the range of λ into smaller intervals for which the encoder setup, i.e., the number of thresholds and index assignments, remains unchanged for a smaller λ interval. In particular, the quantizer arrangements can be identified as a function of λ and the intervals can be computed in a structured way, by considering the ordering relationships between the index assignments (and the respective quantization regions). The following definitions and propositions aim at providing the basis for finding the λ intervals in an efficient manner.

Definition 13 (Well-defined Quantization Region)**.** A quantization region $\mathcal{B}(j,\lambda)$ is well-defined if $\mathcal{B}(j,\lambda)$ is non-empty, for $j \in \mathcal{I}$.

Definition 14 (Index Ordering)**.** We say that the index $j \in \mathcal{I}$ lies to the left of $k \in \mathcal{I}$ if

$$E\{\tilde{U} - \lambda\hat{U}|X^n = x^n_j\} \leq E\{\tilde{U} - \lambda\hat{U}|X^n = x^n_k\}.$$

We denote this relationship by $j < k$.

Definition 15 (Quantization Region Ordering)**.** The quantization regions $\mathcal{B}(j,\lambda)$ and $\mathcal{B}(k,\lambda)$ are ordered if $\mathcal{B}(j,\lambda)$ lies to the left of $\mathcal{B}(k,\lambda)$.

Let there be an arrangement of indices in $\mathcal{I}$ such that each index $j = 0,1,\ldots,|\mathcal{I}|-2$ is left of the index k according to Definition 14. Additionally, let the regions $\mathcal{B}(j,\lambda)$ and $\mathcal{B}(k,\lambda)$ be well-defined according to Definition 13, i.e. $\mathcal{B}(j,\lambda) = [\vartheta^l_{j,\lambda}, \vartheta^u_{j,\lambda}]$ with $\vartheta^l_{j,\lambda} \leq \vartheta^u_{j,\lambda}$ and $\mathcal{B}(k,\lambda) = [\vartheta^l_{k,\lambda}, \vartheta^u_{k,\lambda}]$ with $\vartheta^l_{k,\lambda} \leq \vartheta^u_{k,\lambda}$.

Proposition 1. *If the index $j \in \mathcal{I}$ lies left of the index $k \in \mathcal{I}$ and $\mathcal{B}(j,\lambda)$, $\mathcal{B}(k,\lambda)$ are well-defined, then the quantization region $\mathcal{B}(j,\lambda)$ lies to the left of $\mathcal{B}(k,\lambda)$.*

Proof. If $\mathcal{B}(j,\lambda)$ and $\mathcal{B}(k,\lambda)$ are well-defined, then $\mathcal{B}(j,\lambda)$ is to the left of $\mathcal{B}(k,\lambda)$ if $\vartheta^u_{j,\lambda} \leq \vartheta^l_{k,\lambda}$. Let $l \in \mathcal{I}$ be an arbitrary index such that $\mathcal{B}(l,\lambda)$ is well-defined. To find the upper threshold $\vartheta^u_{j,\lambda}$ we simply need to consider those indices such that $l > j$, i.e. the ones for which $\delta_{j,l,\lambda} > 0$. Hence, we have that $\vartheta^u_{j,\lambda} = \min_{l>j} \vartheta_{j,l,\lambda} \leq \vartheta_{j,k,\lambda}$. To find the lower threshold $\vartheta^l_{k,\lambda}$ we simply need to consider those indices such that $l < k$, i.e. the ones for which $\delta_{k,l,\lambda} < 0$. Consequently, we have that $\vartheta^l_{k,\lambda} = \max_{l<k} \vartheta_{k,l,\lambda} \geq \vartheta_{k,j,\lambda}$. By definition we have that $\vartheta_{j,k,\lambda} = \frac{1}{2} \cdot \frac{\varepsilon_{j,k,\lambda}}{\delta_{j,k,\lambda}} = \frac{1}{2} \cdot \frac{-\varepsilon_{j,k,\lambda}}{-\delta_{j,k,\lambda}} = \frac{1}{2} \cdot \frac{\varepsilon_{k,j,\lambda}}{\delta_{k,j,\lambda}} = \vartheta_{k,j,\lambda}$. From the above expressions we can see that $\vartheta^u_{j,\lambda} \leq \vartheta^u_{j,k,\lambda} = \vartheta^u_{k,j,\lambda} \leq \vartheta^l_{k,\lambda}$ and hence, $\mathcal{B}(j,\lambda)$ is to the left of $\mathcal{B}(k,\lambda)$.

□

Proposition 2. *If the region $\mathcal{B}(j,\lambda)$ is well-defined, index j lies to the left of index k and $\vartheta^u_{j,\lambda} = \vartheta_{j,k,\lambda}$ for some $k \in \{j+1,\ldots,L-1\}$, then*

1. *For all indices $l \in \{j+1, j+2, \ldots, k-1\}$ the regions $\mathcal{B}(l,\lambda)$ are not well-defined;*
2. *The region $\mathcal{B}(k,\lambda)$ is well-defined;*
3. *The upper endpoint of $\mathcal{B}(j,\lambda)$ corresponds to the lower endpoint of $\mathcal{B}(k,\lambda)$ and is equal to $\vartheta_{j,k,\lambda}$, i.e. $\vartheta^u_{j,\lambda} = \vartheta^l_{k,\lambda} = \vartheta_{j,k,\lambda}$*

Proof. To prove point 1) note that the indices j, k and l are ordered such that $j < l < k$. Consider an arbitrary index $m \neq k$ such that $m > j$. By definition $\vartheta^u_{j,\lambda} = \min_{m>j} \vartheta_{j,m,\lambda}$ and by the proposition conditions we have that $\vartheta^u_{j,\lambda} = \vartheta_{j,k,\lambda}$. This implies that $\vartheta_{j,k,\lambda} \leq \vartheta_{j,m,\lambda}$ for any m and, in particular, that $\vartheta_{j,k,\lambda} \leq \vartheta_{j,l,\lambda}$ for $l \in \{j+1,\ldots,k-1\}$. Now consider an index $m' \neq k$ such that $m' \geq l > j$. The upper threshold $\vartheta^u_{l,\lambda}$ is such that $\vartheta^u_{l,\lambda} = \min_{m'>l} \vartheta_{l,m',\lambda} \leq \vartheta_{j,k,\lambda} = \vartheta^u_{j,\lambda}$ (otherwise $\vartheta^u_{j,\lambda}$ would be $\vartheta_{j,m',\lambda}$). On the other hand, the lower threshold $\vartheta^l_{l,\lambda} \geq \vartheta^u_{j,\lambda}$ since j is to the left of l and if $\mathcal{B}(l,\lambda)$ was well-defined it would be to the right of $\mathcal{B}(j,\lambda)$ (Proposition 1). Consequently, we have that $\vartheta^l_{l,\lambda} \geq \vartheta^u_{j,\lambda} \geq \vartheta^u_{l,\lambda}$ and we can conclude that $\mathcal{B}(l,\lambda)$ is not well-defined. To prove point 2) consider an arbitrary index m. We have that $\vartheta^u_{k,\lambda} = \min_{m>k} \vartheta_{k,m,\lambda} \geq \min_{m>k} \vartheta_{j,m,\lambda} \geq \vartheta_{j,k,\lambda}$ by the same arguments as above. On the other hand, we have that $\vartheta^l_{k,\lambda} = \max_{m<k} \vartheta_{k,m,\lambda} \leq \vartheta_{k,j,\lambda} = \vartheta_{j,k,\lambda}$. Hence, we have that $\vartheta^l_{k,\lambda} \leq \vartheta_{j,k,\lambda} \leq \vartheta^u_{k,\lambda}$ and the region $\mathcal{B}(k,\lambda)$ is well-defined. Finally, to prove 3) consider again an arbitrary index m. We have that $\vartheta^l_{k,\lambda} = \max_{m<k} \vartheta_{k,m,\lambda} \geq \vartheta_{j,k,\lambda} = \vartheta^u_{j,\lambda}$. At the same time, we know from above that $\vartheta^l_{k,\lambda} = \max_{m<k} \vartheta_{k,m,\lambda} \leq \vartheta_{j,k,\lambda} = \vartheta^u_{j,\lambda}$, which implies that $\vartheta^l_{k,\lambda} = \vartheta^u_{j,\lambda}$. □

Propositions 1 and 2 allow us to identify the order of the quantization regions $\mathcal{B}(j,\lambda)$ from left to right along the real line and efficiently sort-out the regions $\mathcal{B}(j,\lambda)$ that are not well-defined. Moreover, we know that starting with a well-defined region, it is possible to derive all other well-defined regions to the right. Thus, we are able to identify the quantizer thresholds that will effectively be used in the final quantizer design.

Considering the aforementioned relationships, we can now identify the ranges of λ that do not change the encoder setup, i.e. the ranges of λ such that the ordering of indices is preserved.

Assume that $j < k$. According to Definition 14 we have that $E\{\tilde{U} - \lambda\hat{U}|X^n = x^n_j\} \leq E\{\tilde{U} - \lambda\hat{U}|X^n = x^n_k\}$. Then, with respect to λ we have $\lambda[E\{\hat{U}|X^n = x^n_k\} - E\{\hat{U}|X^n = x^n_j\}] \leq E\{\tilde{U}\ |X^n = x^n_k\} - E\{\tilde{U}|X^n = x^n_j\}$ and therefore $j < k$ if

$$\lambda \leq \frac{E\{\tilde{U}|X^n = x^n_k\} - E\{\tilde{U}|X^n = x^n_j\}}{E\{\hat{U}|X^n = x^n_k\} - E\{\hat{U}|X^n = x^n_j\}}. \tag{3.20}$$

Let $\eta_{j,k} = E\{\tilde{U}|X^n = x_k^n\} - E\{\tilde{U}|X^n = x_j^n\}$ and $\gamma_{j,k} = E\{\hat{U}|X^n = x_k^n\} - E\{\hat{U}|X^n = x_j^n\}$. The set of values of λ for which an index $j \in \mathcal{I}$ lies to the left of $k \in \mathcal{I}$ is then defined as

$$\Lambda(j,k) = \begin{cases} \left[0, \frac{\eta_{j,k}}{\gamma_{j,k}}\right] & \text{, if } \gamma_{j,k} > 0, \\ \left[\frac{\eta_{j,k}}{\gamma_{j,k}}, \infty\right[& \text{, if } \gamma_{j,k} < 0, \\ \mathbb{R} & \text{, if } \gamma_{j,k} \text{ and } \eta_{j,k} \geq 0 \\ \emptyset & \text{, otherwise.} \end{cases}$$

To obtain the ranges of λ that do not change the encoder setup consider the following definition.

Definition 16 (Sequence of Indices)**.** A sequence of all indices in $\mathcal{I}$ is a vector $\mathbf{i}_{\mathcal{I}} = (i_1, i_2, \ldots, i_{|\mathcal{I}|})$ containing each index in $\mathcal{I}$ exactly once, i.e. $\mathbf{i}_{\mathcal{I}}$ is a permutation of $(0, 1, \ldots, |\mathcal{I}| - 1)$.

Definition 17 (Ordered Sequence of Indices)**.** A sequence of indices $\mathbf{i}_{\mathcal{I}}$ is *ordered* if the position of the indices in $\mathbf{i}_{\mathcal{I}}$ also reflect the order defined in Definition 14, i.e., for all $j \in \{1, \ldots, |\mathcal{I}|\}$, $i_j < i_{j+1}$ where i_j and i_{j+1} are the j-th and $(j+1)$-th indices in $\mathbf{i}_{\mathcal{I}}$.

Given a sequence $\mathbf{i}_{\mathcal{I}}$ of all indices in $\mathcal{I}$, let the set of all λ's for which $\mathbf{i}_{\mathcal{I}}$ is ordered be denoted as $\Lambda(\mathbf{i}_{\mathcal{I}})$. Then, $\Lambda(\mathbf{i}_{\mathcal{I}})$ can be derived by subsequently intersecting the intervals for which adjacent indices are ordered:

$$\Lambda(\mathbf{i}_{\mathcal{I}}) = \bigcap_{j=1}^{|\mathcal{I}|-1} \Lambda(i_j, i_{j+1}), \tag{3.21}$$

Since $\Lambda(\mathbf{i}_{\mathcal{I}})$ is obtained by subsequent intersection of the intervals, we conclude that $\Lambda(\mathbf{i}_{\mathcal{I}})$ is a single interval in cases where $\mathbf{i}_{\mathcal{I}}$ is ordered, or the empty set, otherwise. This provides us with the means to test if a certain sequence is ordered and, at the same time, provides us with the values of λ for which the ordering of the sequence is preserved.

To solve (3.12) we need to find a set of sequences which covers the complete range of λ. Let $\mathcal{S}$ denote a set of subsequences $\mathbf{i}_{\mathcal{I}}$. Then, $\mathcal{S}$ covers the whole range of λ if

$$\bigcup_{\mathbf{i}_{\mathcal{I}} \in \mathcal{S}} \Lambda(\mathbf{i}_{\mathcal{I}}) = [0, \infty[. \tag{3.22}$$

Such a set can be found recursively by exploiting (3.21) with the method described next.

Consider a single-index sequence $\mathbf{i}_{\mathcal{I}} = i_a$, with $i_a \in \mathcal{I}$ and $a \in \{1, \ldots, |\mathcal{I}| - 1\}$. Clearly, for such sequence we have that $\Lambda(\mathbf{i}_{\mathcal{I}}) = [0, \infty[$. Now consider the possibility of adding to the sequence $\mathbf{i}_{\mathcal{I}}$ an index i_b such that $i_b \in \mathcal{I} \backslash \mathbf{i}_{\mathcal{I}}$. Let us denote the new subsequence $\mathbf{i}'_{\mathcal{I}} = (\mathbf{i}_{\mathcal{I}}, i_b)$. If $\Lambda(\mathbf{i}'_{\mathcal{I}}) = \emptyset$ it means that this sequence is not ordered, and thus adding further indices will also result in unordered sequences. On the other hand, if $\Lambda(\mathbf{i}'_{\mathcal{I}}) \neq \emptyset$ it means that this new sequence is ordered for some value of λ. We can now repeat the process by taking $\mathbf{i}_{\mathcal{I}} = \mathbf{i}'_{\mathcal{I}}$ as the basis for the next step. By recursively applying the procedure until all the indices have been added or we reach some unordered sequence and repeating the procedure with the initial single-index sequence to take all possible values from $\mathcal{I}$ we can obtain all the valid index arrangements for the quantization regions together with the respective values of λ for which they remain valid.

Having established the realizable quantization regions it is useful to explicitly find the quantization thresholds from the ordering relationships defined above.

3.2.3.3 Quantizer Thresholds

We know from the previous discussion that we can find the thresholds by comparison. Let us start by considering a well defined region $\mathcal{B}(j,\lambda), j \in \mathcal{I}$. The upper endpoint of $\mathcal{B}(j,\lambda)$ takes the form of $\vartheta_{j,k,\lambda}$, where $j < k$. All the regions that are to the left of $\mathcal{B}(j,\lambda)$ should have their respective thresholds to the left of $\mathcal{B}(j,\lambda)$. Then, we have to determine for which values of λ the following inequality holds:

$$\vartheta_{j,k,\lambda} \leq \vartheta_{j,l,\lambda}. \tag{3.23}$$

If the quantization indices are sorted according to Definition 14 such that $j < k < l$, equation (3.23) can be written as a quadratic inequality

$$\kappa\lambda^2 + \tau\lambda + \upsilon \leq 0, \tag{3.24}$$

where

$$\begin{aligned}
\kappa \;=\; & E\{\hat{U}_2^2|X^n = x_k^n\}E\{\hat{U}_2|X^n = x_l^n\} - E\{\hat{U}_2^2|X^n = x_j^n\}E\{\hat{U}_2|X^n = x_l^n\} \\
& -E\{\hat{U}_2^2|X^n = x_k^n\}E\{\hat{U}_2|X^n = x_j^n\} - E\{\hat{U}_2^2|X^n = x_l^n\}E\{\hat{U}_2|X^n = x_k^n\} \\
& +E\{\hat{U}_2^2|X^n = x_j^n\}E\{\hat{U}_2|X^n = x_k^n\} + E\{\hat{U}_2^2|X^n = x_l^n\}E\{\hat{U}_2|X^n = x_j^n\},
\end{aligned}$$

$$\begin{aligned}
\tau \;=\; & E\{\hat{U}_1^2|X^n = x_j^n\}E\{\hat{U}_2|X^n = x_l^n\} + E\{\hat{U}_2^2|X^n = x_j^n\}E\{\hat{U}_1|X^n = x_l^n\} \\
& +E\{\hat{U}_1^2|X^n = x_k^n\}E\{\hat{U}_2|X^n = x_j^n\} + E\{\hat{U}_2^2|X^n = x_k^n\}E\{\hat{U}_1|X^n = x_j^n\} \\
& +E\{\hat{U}_1^2|X^n = x_l^n\}E\{\hat{U}_2|X^n = x_k^n\} + E\{\hat{U}_2^2|X^n = x_l^n\}E\{\hat{U}_1|X^n = x_k^n\} \\
& -E\{\hat{U}_1^2|X^n = x_j^n\}E\{\hat{U}_2|X^n = x_k^n\} - E\{\hat{U}_2^2|X^n = x_j^n\}E\{\hat{U}_1|X^n = x_k^n\} \\
& -E\{\hat{U}_1^2|X^n = x_k^n\}E\{\hat{U}_2|X^n = x_l^n\} - E\{\hat{U}_2^2|X^n = x_k^n\}E\{\hat{U}_1|X^n = x_l^n\} \\
& -E\{\hat{U}_1^2|X^n = x_l^n\}E\{\hat{U}_2|X^n = x_j^n\} - E\{\hat{U}_2^2|X^n = x_l^n\}E\{\hat{U}_1|X^n = x_j^n\}
\end{aligned}$$

and

$$\begin{aligned}\upsilon \;=\;& E\{\hat{U}_1^2|X^n=x_k^n\}E\{\hat{U}_1|X^n=x_l^n\}-E\{\hat{U}_1^2|X^n=x_j^n\}E\{\hat{U}_1|X^n=x_l^n\}\\ &-E\{\hat{U}_1^2|X^n=x_k^n\}E\{\hat{U}_1|X^n=x_j^n\}-E\{\hat{U}_1^2|X^n=x_l^n\}E\{\hat{U}_1|X^n=x_k^n\}\\ &+E\{\hat{U}_1^2|X^n=x_j^n\}E\{\hat{U}_1|X^n=x_k^n\}+E\{\hat{U}_1^2|X^n=x_l^n\}E\{\hat{U}_1|X^n=x_j^n\}.\end{aligned}$$

Let Δ_ϑ be the discriminant of (3.24) and λ_1 and λ_2, $\lambda_1 \le \lambda_2$ be the solutions to the quadratic equality with coefficients κ, τ and υ. The solution to (3.23), which we denote as $\Pi(j,k,l)$ can then be characterized by

$$\Pi(j,k,l)=\Lambda(j,k,l)\cap\Theta(j,k,l), \tag{3.25}$$

where

$$\Theta(j,k,l)=\begin{cases}[\lambda_1,\lambda_2] & \text{, if } \kappa>0,\ \Delta_\vartheta\ge 0,\\ ]-\infty,-\frac{\upsilon}{\tau}] & \text{, if } \kappa=0,\ \tau>0,\\ [-\frac{\upsilon}{\tau},\infty[& \text{, if } \kappa=0,\ \tau<0,\\ ]-\infty,\lambda_1]\cup[\lambda_2,\infty[& \text{, if } \kappa<0,\ \Delta_\vartheta\ge 0,\\ \mathbb{R} & \text{, if } \kappa<0,\ \Delta_\vartheta<0 \text{ or } \kappa=0,\ \tau=0,\ \upsilon\le 0,\\ \emptyset & \text{, otherwise.}\end{cases}$$

Unlike the previous cases, $\Pi(j,k,l)$ is either a union of two intervals, a single interval or the empty set. We can now proceed to define the set of λ's for which a certain threshold $\vartheta_{j,k,\lambda}$, $j,k\in\mathcal{I}$, $j<k$, is smaller than (or equal to) all other thresholds $\vartheta_{j,l,\lambda}$, $l\in\mathcal{I}$, $j\neq l$, $l\neq k$ using the intersection of the intervals defined by $\Pi(j,k,l)$. Let us denote this set as $\Pi(j,k)$. Then, $\Pi(j,k)$ can be derived by subsequently intersecting the intervals in (3.25) such that

$$\Pi(j,k)=\Lambda(j,k)\bigcap_{l\in\mathcal{I}:k<l}\Theta(j,k,l)\bigcap_{l\in\mathcal{I}:j<l<k}\Theta(j,l,k) \tag{3.26}$$

We can now extend this result to a sequence of ordered indices $\mathbf{i}_\mathcal{I}=(i_1,i_2,\ldots,i_{|\mathcal{I}|})$. We have that the set of all λ's satisfying

$$\vartheta_{i_j,i_{j+1},\lambda}\le\vartheta_{i_{j+1},i_{j+2},\lambda} \tag{3.27}$$

with $i=1,2,\ldots,|\mathcal{I}|-2$, denoted as $\Pi(\mathbf{i}_\mathcal{J})$, can be derived by successive intersection as follows

$$\Pi(\mathbf{i}_\mathcal{I})=\Lambda(\mathbf{i}_\mathcal{I})\bigcap_{j=1,2,\ldots,|\mathcal{I}|-1}\Theta(i_j,i_{j+1}), \tag{3.28}$$

where $\Theta(j,k)=\bigcap_{k<l}\Theta(j,k,l)\bigcap_{j<l<k}\Theta(j,l,k)\bigcap_{l<j}\Theta(l,j,k)$, with $l\in\mathcal{I}$. Using this result

together with Proposition 2, we can now derive the final encoder by using a recursive procedure similar to the one described in the previous section.

Consider an ordered sequence $\mathbf{j}_{\mathcal{J}}$ derived by the successive intersections in (3.21). Now consider a single index sequence $\mathbf{i}_{\mathcal{I}}$ that is composed by the first entry of $\mathbf{j}_{\mathcal{J}}$, i.e. $\mathbf{i}_{\mathcal{I}} = j_1$ and a sequence of indices $\mathbf{k}_{\mathcal{K}}$ that contains the rest of the indices in $\mathbf{j}_{\mathcal{J}}$, i.e. $\mathbf{k}_{\mathcal{K}} = (j_2, \dots, j_{\mathcal{I}})$. At each step of the procedure we will add an index from $\mathbf{k}_{\mathcal{K}}$ to $\mathbf{i}_{\mathcal{I}}$. The added index is essentially the one that guarantees that (3.27) holds. To achieve this we determine for each $j_k \in \mathbf{k}_{\mathcal{K}}$ the region $\Pi((\mathbf{i}_{\mathcal{I}}, j_k))$ according to (3.28) to find out if there exists any value of λ for which the threshold $\vartheta_{j_{|\mathbf{i}_{\mathcal{I}}|}, j_k, \lambda}$ is smaller than any other threshold $\vartheta_{j_{|\mathbf{i}_{\mathcal{I}}|}, j_l, \lambda}$, with $j_l \in \mathbf{k}_{\mathcal{K}}$, $l \neq k$. After choosing j_k, we add it to $\mathbf{i}_{\mathcal{I}}$ which will be the basis for the next step. Meanwhile, we can neglect all indices to the left of j_k in $\mathbf{k}_{\mathcal{K}}$, so they can be removed from $\mathbf{k}_{\mathcal{K}}$. We repeat the procedure until there are no more indices in $\mathbf{k}_{\mathcal{K}}$ to consider and at the end we obtain the threshold sequences $\mathbf{i}_{\mathcal{I}}$ and the set of all λ's for which the corresponding arrangement is valid. Like in the previous case, the described procedure is also repeated for every possible arrangement of initially ordered sequences $\mathbf{j}_{\mathcal{J}}$ obtained from (3.21).

The previously described procedures enable us to tackle the optimization problem in (3.9), as the computed sequences $\mathbf{i}_{\mathcal{I}}$ fully describe the optimal configuration of the quantizer q as well as the optimal index assignment function s for all λ's that fall in the region $\Pi(\mathbf{i}_{\mathcal{I}})$. More precisely, the optimal quantizer q is obtained by setting $b_0 = -\infty$, $b_j = \vartheta(i_j, i_{j+1}, \lambda)$, $j = 1, 2, \dots, |\mathcal{I}| - 1$, $b_{|\mathcal{I}|} = \infty$ and the optimal symbol assignment s is chosen such that the j-th quantization index i is mapped onto the i_j-th codeword vector x_j^n.

Thus, the encoder setting $\Gamma^*(\lambda)$, required in (3.9), can be represented for all λ's in $\Pi(\mathbf{i}_{\mathcal{I}})$ by a single analytic expression. If the source PDF $p_U(u)$ is differentiable, then $D(\Gamma^*(\lambda), \Phi)$, $D(\Gamma^*(\lambda), \Psi)$, $L(\Gamma^*(\lambda), \Phi, \lambda)$ and, thus, $g(\lambda)$ in (3.9) are also differentiable. This allows for gradient-based methods or an efficient numerical optimization to solve the maximization problem in (3.9). At the end we obtain the best value for λ, i.e. λ^*, and, thus, the optimal encoder $\Gamma^* = \Gamma^*(\lambda^*)$.

3.2.4 Decoder Optimization

The problem of finding the optimal decoder for a fixed encoder Γ is a well-known problem in the literature. In particular, assuming the MSE as the distortion criterion, it is known that the optimal decoder simply performs the conditional mean estimation [44, Chapter IV.B]. Thus, for the legitimate receiver we have that the optimal decoder Φ^* for a fixed encoder Γ is given as before by

$$\Phi^* \triangleq \Phi_\Gamma(y_i^n) = E\{U | Y^n = y_i^n\}, i = 1, 2, \dots, |\mathcal{Y}^n|, \tag{3.29}$$

where y_i^n is the i-th main channel output vector. Hence, the legitimate receiver makes estimates $\tilde{u}_i = \Phi_\Gamma(y_i^n)$. As for the eavesdropper, the same results hold. Hence, the eavesdropper's optimal decoder Ψ^* for a fixed encoder Γ is given by

$$\Psi^* \triangleq \Psi_\Gamma(z_i^n) = E\{U|Z^n = z_i^n\}, i = 1,2,\ldots,|\mathcal{Z}^n|, \tag{3.30}$$

where z_i^n is the i-th wiretap channel output vector.The eavesdropper then estimates $\hat{u}_i = \Psi_\Gamma(z_i^n)$.

3.3 Numerical Results

To highlight the properties of the proposed quantizer in terms of secrecy we evaluate the performance of our scheme for the case of a Gaussian source $U \sim \mathcal{N}(\mu, \sigma^2)$, with mean $\mu = 0$ and variance $\sigma^2 = 1$. The performance is evaluated by recurring to the SNR of the estimates of both the legitimate receiver and the eavesdropper, defined as SNR $= 10\log_{10}\frac{\sigma^2}{D}$, where D is the MSE distortion. We will refer to the difference between the SNR of the legitimate receiver and the eavesdropper as *secrecy gain*.

Before explicitly considering the secrecy gains of a quantizer designed with secrecy constraints, we consider using the channel optimized scalar quantizers from [39]. Suppose our encoder is optimized for the legitimate receiver. It is highly unlikely that an eavesdropper will have access to a channel with the same properties as the main channel. Therefore, the encoding boundaries of the scalar quantizer do not reflect the optimal encoding boundaries for the eavesdropper. This channel mismatch will introduce some distortion at the eavesdropper, even when he uses an optimal decoder. This effect is illustrated in Fig. 3.5, where we consider a binary symmetric main channel with a constant crossover probability $\delta = 10^{-2}$ and a binary symmetric wiretap channel with crossover probability ε, and find the channel-optimized encoders for a quantizer using channel codewords of length $n = 3$ bits. For reference, we plot the signal-to-noise ratio (SNR) of the legitimate receiver, represented by the solid horizontal line. The dashed green line shows the performance of the eavesdropper when we employ a scalar quantizer that is optimized for the main channel, whereas the dashed red line (with markers) plots the SNR for an eavesdropper when the scalar quantizer is optimized for the wiretap channel. From the plot we observe (as expected) that the SNR of the eavesdropper degrades with respect to a scalar quantizer optimized for the wiretap channel whenever there is a channel mismatch, irrespective of the wiretap channels quality. E.g. if the wiretap channel has a crossover probability $\varepsilon = 10^{-3}$, then an eavesdropper that is forced to use an encoder that is optimized for the main channel loses about 0.6dB when $n = 3$. When $\varepsilon = 10^{-1}$, we have an SNR difference of about 1dB. It is possible to see that a scalar quantizer designed without security considerations brings little secrecy benefits for small resolutions[3]. Moreover, the legitimate receiver only has an advantage over the eavesdropper when the main channel is *better* than the wiretap channel.

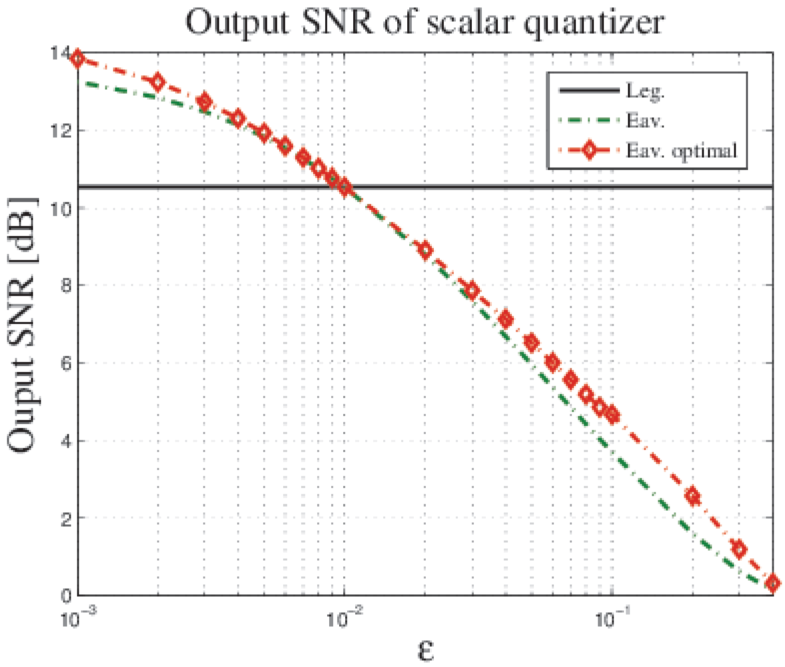


Figure 3.5: SNR of legitimate receiver and eavesdropper for a channel optimized scalar quantizer without secrecy constraints. The quantizer resolution is $Q = 3$ bits (8 levels).

Let us now consider the iterative optimization approach developed earlier. We conducted experiments with different initial conditions that consisted in varying the initial index assignments and reconstructions levels. The number of iterations considered in the iterative optimization procedure was three, which already provided good convergence properties for relevant system designs. The results presented in this section use distortion thresholds $\Delta \in \{0.3981, 0.1585, 0.0631, 0.0251\}$, which correspond respectively to signal-to-noise ratios of 4, 8, 12 and 16dB.

Additionally, we consider three instances of the wiretap model defined in Section 3.2. In the first, we consider a binary symmetric main channel with constant crossover probability $\delta = 10^{-5}$ (i.e. almost noiseless) and binary symmetric wiretap channel with varying crossover probability ε, such that $\varepsilon \in [10^-3, 4 * 10^{-1}]$. Hence, this configures a scenario where the main channel is always better than the wiretap channel. In the second case, we allow for the wiretap channel to be better than the main channel. Here, we consider a binary symmetric main channel with constant crossover probability $\delta = 10^{-2}$ and varying crossover probability ε for the wiretap channel, where $\varepsilon \in [10^-3, 4 * 10^{-1}]$. The third scenario consists of a binary symmetric main channel is characterized by a varying crossover

[3] Increasing the block-length results in a larger gap between the SNR of the legitimate receiver and the eavesdropper. However, this comes at the cost of a better SNR for the eavesdropper as well.

probability δ and a binary symmetric wiretap channel characterized by a crossover probability $\varepsilon = 2\delta(1-\delta)$). This scenario emulates a stochastically degraded channel where the main and wiretap channel have the same crossover probabilities.

The performance of our wiretap scalar quantizer under the aforementioned wiretap scenarios is summarized in Figs. 3.6, 3.7 and 3.8. In each figure, the solid black lines represent the SNR for the legitimate receiver, while the dashed red lines plot the SNR for the eavesdropper. A first observation with respect to every figure is that the proposed scalar quantizer effectively meets the requirement of enforcing an upper bound on the SNR of the eavesdropper (lower bound on the distortion). A second observation is that the distortion of the legitimate receiver is tightly connected to the distortion of the eavesdropper. Thus, even if the legitimate receiver enjoys a *good* channel, it may be necessary to reduce the legitimate receiver signal quality in order to obtain secrecy. This shows there exists a fundamental trade-off between the choice of Δ and the obtained performance for both receivers.

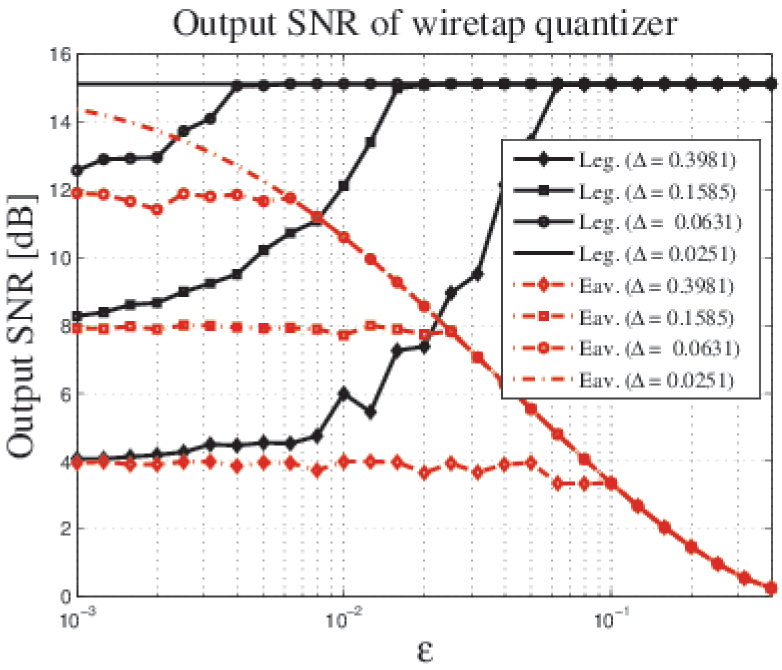


Figure 3.6: SNR of legitimate receiver and eavesdropper for a scalar quantizer with secrecy constraints when the main channel has a crossover probability of $\delta = 10^{-5}$. The quantizer resolution is $Q = 3$.

Another interesting consequence of our design is that two performance regions can be identified. In a first region, the performance is dominated by the quantization parameters which limit the eavesdropper's SNR (and consequently the legitimate receiver's SNR) until a given threshold on ε. Above this threshold the performance of the eavesdropper becomes dominated by the channel quality and the legitimate receiver is able to achieve optimum performance.

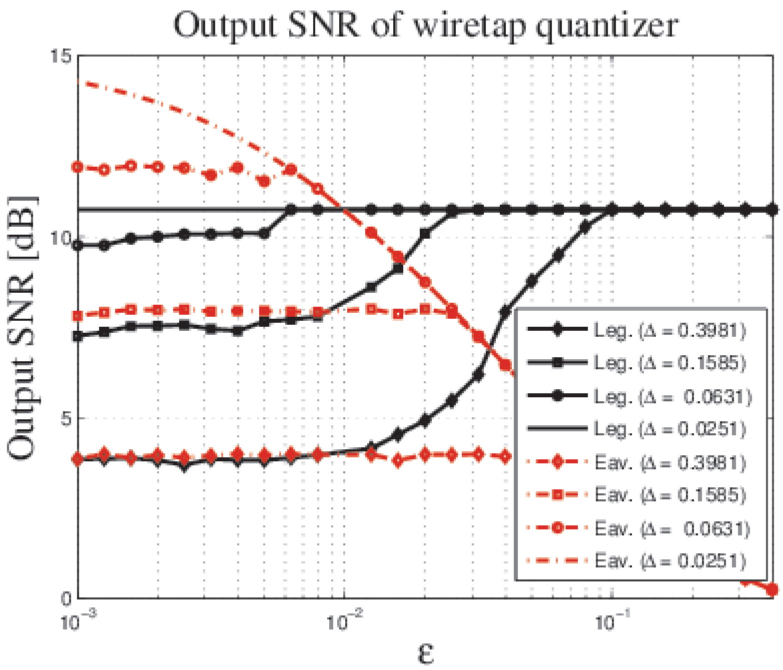


Figure 3.7: SNR of legitimate receiver and eavesdropper for a scalar quantizer with secrecy constraints when the main channel has a crossover probability of $\delta = 10^{-2}$. The quantizer resolution is $Q = 3$.

We note that the largest secrecy gains appear when the eavesdropper's crossover probability approaches the threshold where its performance is bounded by the channel parameters. When the main channel is almost noiseless (Fig. 3.6), secrecy gains of approximately 3dB, 7dB and 11dB are obtained when we approach this point, respectively for $\Delta = 0.0631$, 0.1585 and 0.3981. On the other hand, the more strict the value of Δ, the higher the eavesdropper's crossover probability needs to be in order to achieve these secrecy gains. For the degraded scenario (Fig. 3.8), secrecy gains of about 1dB ($\Delta = 0.0631$) and 2dB ($\Delta = 0.1585$ and 0.3981) can be achieved. Fig. 3.8 allows us to understand what happens when the wiretap channel is better than the main channel. Here, the main channel is characterized by a constant crossover probability of $\delta = 10^{-2}$. If the distortion constraint is loose, then the eavesdropper will outperform the legitimate receiver. However, if we design our encoder with a distortion constraint such that the performance is bounded by the constraint instead of the channel, the performance of both users goes on level. Thus, the advantage of the eavesdropper is mitigated by our encoder design. This is better understood when we look at the number of levels of the final quantizer (Fig. 3.9). We can observe that for the scenario where the main channel is almost noiseless, we often use all the quantization levels available (the performance of the eavesdropper is limited by its own channel). On the other hand, if the channel to the legitimate receiver is worse than the eavesdropper (which happens in the second scenario whenever $\varepsilon < 10^{-2}$), we are forced to use less quantization levels. When we are in the presence of a degraded

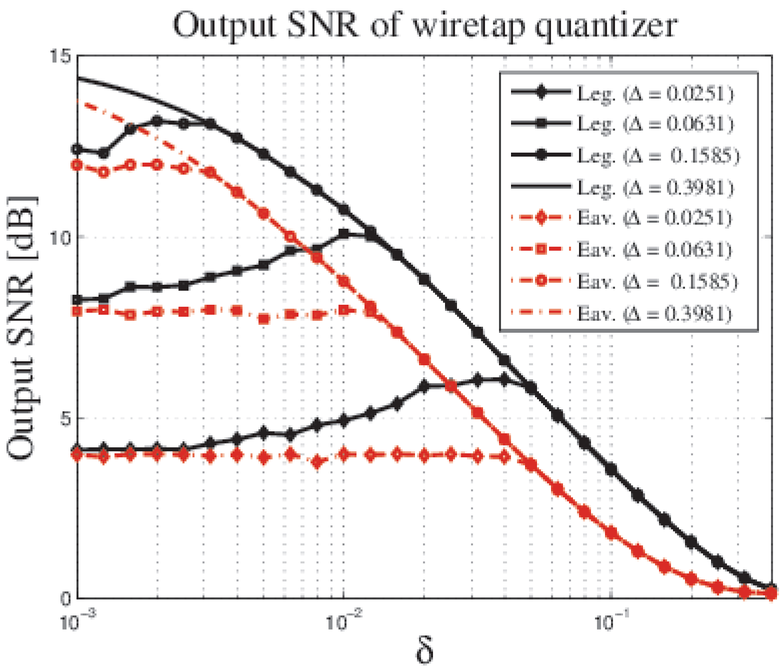


Figure 3.8: SNR of legitimate receiver and eavesdropper for a scalar quantizer with secrecy constraints for the degraded scenario. The quantizer resolution is $Q = 3$.

wiretap channel (third scenario), then we start be reducing the number of levels until the eavesdroppers performance is dominated by its channel properties. That is the point at which we are able to fully use the available resolution to the advantage of the legitimate receiver.

3.4 Discussion

In this chapter we considered the design of channel-optimized scalar quantizers for secure communications, by extending the work of Farvardin and Vaishampayan [39] to a system with three users comprised by a sender, a legitimate receiver and a wiretapper. To the best of our knowledge, this work represents the first approach to the design of scalar quantizers with secrecy constraints. The proposed scheme looks at the problem of designing a scalar quantizer as an optimization problem where the goal is to minimize the distortion of the legitimate receiver, subject to a lower bound on the eavesdroppers distortion. We note that the proposed strategy can be employed for other secrecy constraints of similar type. We derived the necessary conditions for a locally-optimum system and proposed a methodology for the quantizer design under an MSE criteria. Numerical results for quantizers obtained via the iterative optimization procedure assuming binary symmetric main and wiretap channels were presented. The results highlight several properties that can be obtained by our system when employed both over degraded and non-degraded channels.

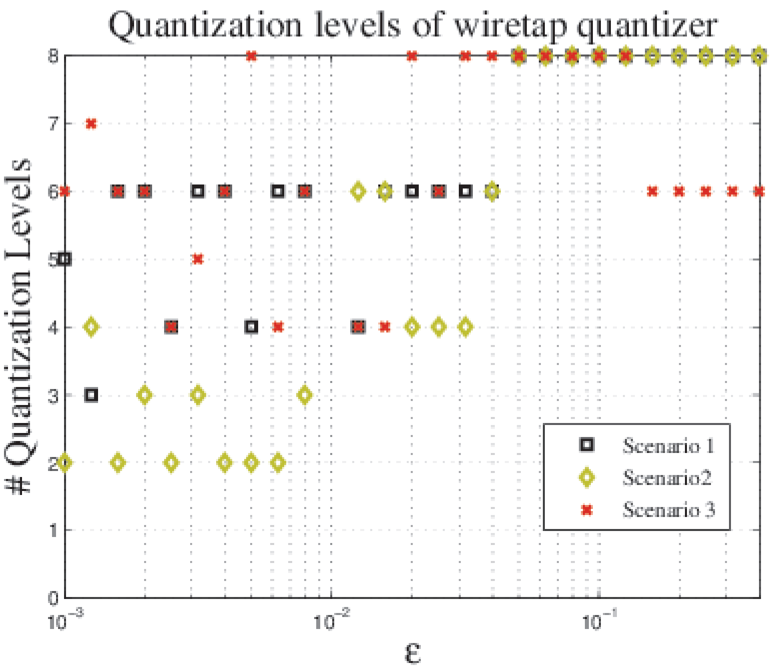


Figure 3.9: Number of levels of final quantizer design for the three wiretap instances when $\Delta = 0.3981$.

In particular, our design ensures SNR advantage for the legitimate parties while bounding the quality of the eavesdroppers channel when the main channel is better and ensures a leveraged SNR for both users when the channel statistics of the eavesdropper are *better* than those of the legitimate receiver. The problem formulation allows to fine tune the design of the scalar quantizer to specific secrecy levels, which is a useful property for many applications. Given the nature of our results, the proposed scheme may find applications within the domain of wireless sensor networks and near field communications, since these are applications where quantization is generally a requirement and the physical proximity of the communicating entities practically allows for the legitimate party to enjoy a better channel than the eavesdropper.

Chapter 4

Continuous Spherical Codes for Secrecy

In the previous chapter we have seen how to design channel-optimized scalar quantizers to meet some distortion constraint on a third-party. Such construction involved creating a set of discrete points to be mapped onto discrete channel input sequences. However, it is possible to communicate discrete time continuously-valued sources without requiring source discretization. More precisely, it is possible to project the (continuous) source space onto a (continuous) lower dimensional subspace [45] or a (continuous) higher dimensional space [46]. The former mappings constitute a form of compression (or source coding), whereas the latter constitute a form of channel coding. Hence, the encoding operations can be defined through linear or non-linear functions that map source samples onto the channel space. In particular, non-linear functions project source samples onto curves defined over the channel space, providing a geometrical interpretation to the problem of communication (as will be seen in Example 2). These mappings are also known as *Shannon-Kotel'nikov* mappings [47].

Formally, both source and channel coding can be defined as follows. Consider a continuously-valued discrete-time memoryless source $u \in \mathbb{R}$. If we wish to compress this source (dimension reduction) we may take a vector of m source samples and take the projection of this m-dimensional vector onto the channel space $\mathbb{R}^n$, with $n < m$, using a mapping $\mathcal{S}: \mathbb{R}^m \to \mathbb{R}^n$. $\mathcal{S}$ can be seen as an n-dimensional locally euclidean manifold embedded in $\mathbb{R}^m$. On the other hand, if we wish to perform error control coding (dimension expansion), we may take a vector of m source samples and map it onto the channel space $\mathbb{R}^n$, with $n > m$, using a similar mapping $\mathcal{S}: \mathbb{R}^m \to \mathbb{R}^n$, such that this mapping is injective. In both cases, we may see $\mathcal{S}$ as a continuous or piecewise continuous linear or non-linear transformation between the spaces $\mathbb{R}^m$ and $\mathbb{R}^n$, which can be realized through a parametric function.

Let us introduce a simple example of a 1:2 bandwidth expansion mapping, which will be of help in determining some important characteristics of these mappings.

Example 2. Suppose we wish to transmit a source u that takes values in the interval

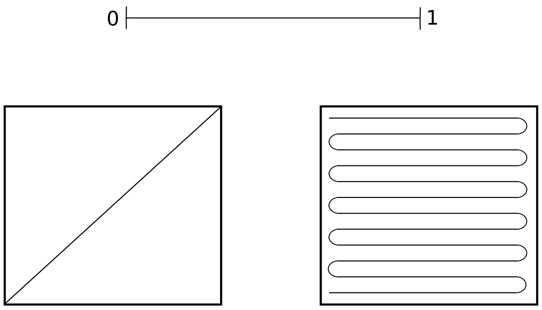


Figure 4.1: Examples of 1:2 bandwidth expansion mapping.

$[0,1]$. Since we are performing a 1:2 bandwidth expansion we are considering mapping of a one-dimensional line onto a two-dimensional square. One possible way of doing this would be a linear map of the source values such that u is mapped onto the point (u,u). This mapping is represented on the left side square of Fig. 4.1. On the other hand, we can use a non-linear mapping similar to the one on the right side square of Fig. 4.1 (this picture appears originally in [45]). In the latter case, the original line was stretch and twisted to occupy a larger portion of the channel space. Suppose that these mappings are used to communicate u over a noisy channel. For simplicity, assume that we are operating in the high-SNR regime. Then, a decoder that minimizes the Euclidean distance between the received vector and the line or curve is approximately optimal in a mean-square sense [47]. With respect to a linear mapping, the magnitude of the error vector will remain unchanged after decoding, as the mapping is simply a rotation of the source vector. With respect to the non-linear mapping, the magnitude of the error vector will change after decoding. More precisely, the magnitude of the error vector can decrease if the original error vector does not send the received vector closer to another curve fold, or increase otherwise. The increase and decrease in the error magnitude is mainly due to the need for re-scaling after decoding, which depends on how much the line is stretched.

The above example is useful to illustrate the virtues and limitations of linear and non-linear mappings. While linear mappings have a *constant* error profile, non-linear mappings introduce a sort of threshold where below that threshold the magnitude of error decreases and above that threshold the magnitude of error increases. Hence, the existence of these *anomalous* errors that induce a threshold implies that non-linear mappings have some limits as to how much they can stretch the source space. It also illustrates the two main criteria involved in the design of specific curves. On the one hand, the length of the curves should be maximized (for a given power constraint) in order to reduce the magnitude of the error vector. On the other hand, the stretching achieved by the mapping must be limited to avoid any anomalous errors (which induce a high distortion). These two constraints are at odds with each other.

When compared to other strategies for communicating continuous sources such as channel-optimized scalar quantizers, bandwidth expansion mappings compare favourably in terms of performance and complexity, when operating in the high-SNR regime [47]. Additionally, these techniques incur in low complexity and delay in comparison with other schemes performing error correction, while allowing control over the bandwidth expansion (or reduction). From a secrecy perspective, they allow us to take advantage of other types of code characteristics. In particular, they introduce some geometrical meaning to the process of communication which may be used from a secrecy perspective.

In this chapter we will focus on a particular construction of these bandwidth expansion mapping called *Piecewise Torus Layer Spherical Codes*, introduced in [48] for discrete sources. In essence, these mappings are curves defined over several flat tori. They provide a basis for efficient encoding/decoding while guaranteeing a good bandwidth expansion performance. Moreover, their geometrical properties allow us to control the distances between folds, thus providing means to easily control the code's error thresholds. In this context, we aim at designing codes guaranteeing, with high probability, that the eavesdroppers error vector will be above the noise threshold defined by the anomalous errors, described in Example 2. If one can achieve this, the distortion of the eavesdropper is bound to be small. With respect to reliability, we wish that the legitimate party operates below this error threshold, and therefore does not incur in fold errors. Thus, the proposed codes are based on finding a suitable parametrization that satisfies the reliability and secrecy constraints. Formally, we define these constraints achieving an arbitrarily small probability of anomalous errors (reliability) and an arbitrarily large probability of anomalous errors (secrecy). However, these conditions can be relaxed to allow for a specific fraction of anomalous errors. This allows us to find a balance between the distortion experienced by the legitimate receiver and the eavesdropper explicitly.

4.1 Torus Layer Spherical Codes For Continuous Alphabets

Torus Layer Spherical Codes (TLSC) were recently introduced in [48] as a new class discrete spherical codes and were extended in [49] to account for continuous sources. The codes in [49] essentially perform a $1:2n$ bandwidth expansion by mapping source values $u \in \mathbb{R}$ onto several curves that are defined over a flat tori contained in the unit sphere $\mathbb{S}^{2n-1}$. Fig. 4.2 serves as an informal illustration of how these codes generally operate[1]. The signal interval is divided into M partitions. Each of these partitions is then mapped onto a curve defined over a flat torus. Hence, the channel space is divided onto non-intersecting hyper-surfaces which, ideally, are densely packed. If we choose a proper distance between torus (a distance such that the probability of anomalous errors

[1] For an easier visualization the depicted torus is not necessarily flat, since embedding a flat torus in 3 dimensions requires repeatedly corrugating a regular torus [50]. For our purposes, it is sufficient to illustrate the construction and its properties on non-flat tori.

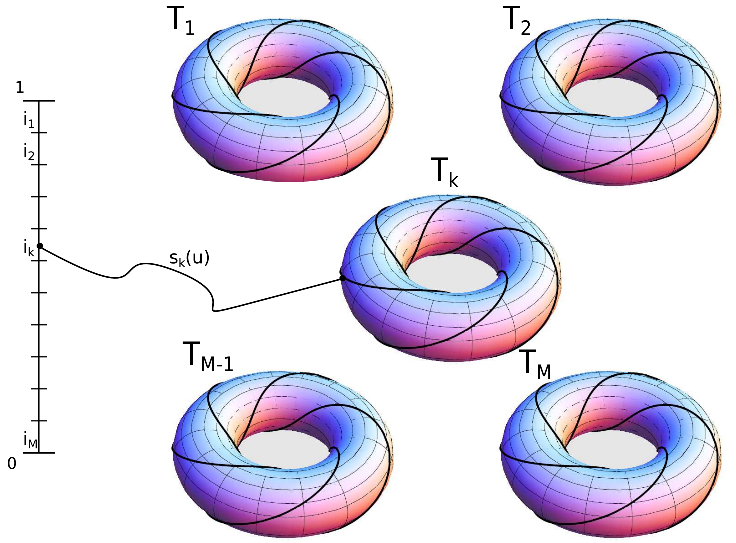


Figure 4.2: Example of a mapping between the line $[0,1]$ and curves over several tori.

is arbitrarily small), the error of the estimates will be bounded by the size of the signal partition. A dense packing of tori implies smaller partitions and, consequently, smaller decoding errors. The considered curves are $(v_1,\dots,v_n)$-type knots over the torus. These knots have maximal length for a pre-defined fold distance (so it is possible to fulfil the principles previously discussed). These notions will be formalized next.

4.1.1 Flat Tori

An n-dimensional flat torus T is a closed surface defined as the Cartesian product of n circles $\mathbb{S}^1$ in $\mathbb{R}^2$, where the sum of the squares of the radii of these circles sum up to one. A particular flat torus can be defined through Cartesian coordinates as follows. Let $c^n = (c_1,\dots,c_n) \in \mathbb{R}^n$, such that $c_i \geq 0$, $i = 1,\dots,n$. Then, the flat torus T_{c^n} is the subset of points on the unit sphere $\mathbb{S}^{n-1}$ that is given by

$$T_{c^n} = \{(x_1,\dots,x_{2n}) \in \mathbb{R}^{2n} : x_{2i-1}^2 + x_{2i}^2 = c_i^2, 1 \leq i \leq n\} \tag{4.1}$$

Alternatively, we can define a flat torus through parametric equations. Consider the application $\Phi_{c^n} : \mathbb{R}^n \to \mathbb{R}^{2n}$, defined as

$$\Phi_{c^n}(v^n) = \left(c_1 \cos\frac{v_1}{c_1}, c_1 \sin\frac{v_1}{c_1}, \dots, c_n \cos\frac{v_n}{c_n}, c_n \sin\frac{v_n}{c_n}\right), \tag{4.2}$$

with $v^n = (v_1,\dots,v_n) \in \mathbb{R}^n$. The torus T_{c^n} can then be seen as the image by Φ_{c^n}.

T_{c^n} is also the image of an injective n-dimensional hyperbox

$$P_{c^n} \triangleq \{v^n \in \mathbb{R}^n : 0 \leq v_i < 2\pi c_i\}. \tag{4.3}$$

This can be seen by establishing classes of equivalence between points in $\mathbb{R}^n$. In particular, two points $x=(x_1,\dots,x_n)$ and $y=(y_1,\dots,y_n)$ are equivalent by Φ if $\Phi(x)=\Phi(y)$. Consequently, for Φ_{c^n}, x and y are equivalent if $x_i=y_i \mod 2\pi c_i$. Now note that P_{c^n} is the fundamental parallelotope of a lattice Λ generated by the vectors $v_i=2\pi c_i e_i, 1\leq i\leq n$, where $\{e_i\}$ is the canonical basis of $\mathbb{R}^n$. Then $x_i=y_i \mod 2\pi c_i$ is equivalent to $x-y\in\Lambda$. Since, the image by Φ_{c^n} is a flat torus T_{c^n} and Φ_{c^n} is well defined in the quotient R^n/Λ, the flat torus T_{c^n} is also the image by Φ_{c^n} restricted to the hyperbox P_{c^n} [51]. Finally, we note that Φ_{c^n} is a local isometry between $\mathbb{R}^n$ and T_{c^n}, implying that distances are preserved by Φ_{c^n}, whenever Φ_{c^n} is injective.

Both representations are useful for code construction. For instance, the canonical representation of the torus by Φ is useful to characterize the code's distance properties, while the hyperbox representation is useful to design the curves over the torus.

Geometrical distances will play an important role in code design, as they directly relate to the existence of anomalous errors. The following properties will be useful to our analysis.

Proposition 3 ([48]). *The* minimum distance *between two points in different flat tori T_{c^n} and T_{b^n} is given by*

$$d_{\min}(T_{c^n},T_{b^n})=\|c^n-b^n\|=\left(\sum_{i=1}^{n}(c_i-b_i)^2\right)^{\frac{1}{2}}. \tag{4.4}$$

Proposition 4 ([48]). *The distance between two points x and y in the same torus T_{c^n} is given by*

$$d(\Phi_{c^n}(x),\Phi_{c^n}(y))=2\left(\sum_{i=1}^{n}c_i^2\sin^2\left(\frac{x_i-y_i}{2c_i}\right)\right)^{\frac{1}{2}}. \tag{4.5}$$

Proposition 5 ([48]). *Let $c_\xi=\min_{1\leq i\leq n} c_i$ and suppose that $0<||x-y||\leq\pi\frac{c_\xi}{2}$. The distance in (4.5) is bounded in terms of the pre-image distance $||x-y||$ by*

$$d(\Phi_{c^n}(x),\Phi_{c^n}(y)) \ \ \geq \ \ \sin\left(\frac{\|x-y\|}{2c_\xi}\right)2c_\xi. \tag{4.6}$$

Having defined flat tori and some of their properties, we will describe how these can be used to define a piecewise continuous code.

4.1.2 Piecewise Torus Layer Spherical Codes

Consider a collection of flat tori $T=\{T_1,\dots,T_M\}$, where each torus is defined over $\mathbb{S}^{2n-1}$ using M non-negative n-dimensional unit vectors $c_i^n=(c_{i,1},\dots,c_{i,n})$, $1\leq i\leq M$. These M

unit vectors equivalently define a spherical code $\mathcal{SC} \subset \mathbb{S}^{n-1}$ with M non-negative codewords c_i^n, $1 \leq i \leq M$. From herein, without loss of generality, we will consider only non-degenerate tori, i.e. tori generated by vectors whose coordinates are non-zero[2].

These flat tori can be used to design both discrete and continuous spherical codes. For instance, we may fill each flat tori with a suitable n-dimensional code (e.g. a lattice code) and take its image by (4.2) to obtain a spherical code in $\mathbb{R}^{2n}$ [48]. The minimum distance of this discrete spherical code is given by the minimum distance between any two tori in T. A similar strategy can be used to design a piecewise continuous code, where instead of a discrete set of points, we fill each hyperbox $P_{c_i^n}$ with continuous curves [53, 49].

A curve can be defined as follows. Let $s: \ [a,b] \to \mathbb{R}^n$, $a,b \in \mathbb{R}$, be a mapping of a real valued signal u within the interval $[a,b]$, onto an n-dimensional point $s(u)$ defined over the real numbers. If s is a continuous mapping, then s represents a curve in $\mathbb{R}^n$. The stretch $\mathcal{S}(u)$ of s is the function $\|\dot{s}(u)\|$, where $\dot{s}(u)$ is the derivative of $s(u)$, $u \in [a,b]$. The length of the curve is given by $L = \int_a^b \mathcal{S}(u)du$. For a given point $s(u)$ in the curve s, the Voronoi region $V(u)$ of $s(u)$ is the set of all points of $\mathbb{R}^n$ such that these points are closer to $s(u)$ than any other point in the curve. We can define the *small-ball radius* of a curve s as the largest radius $r > 0$ such that $B_r(s(u)) \cap H(u) \subset V(u)$, where $B_r(s(u))$ is an Euclidean ball of radius r centred at $s(u)$ and $H(u)$ is the hyperplane orthogonal to s at $s(u)$. A pictorial representation of a small-ball radius would be an n-dimensional cylinder of radius r that is placed along the curve and does not intersect itself. Hence, the small-ball radius can be seen as a measure of the minimum distance between the folds of a curve [54].

Evoking the previous insights for good bandwidth expansion mappings, a code for transmission of continuous sources should be a mapping of substantial length, capable of guaranteeing, with high probability, that curve folds are sufficiently apart as to avoid anomalous errors. For our piecewise codes, this translates onto finding a curve of maximum length on the unit sphere, such that the small ball radius is greater than a given δ_{SB}, which relates to the noise affecting the communication channel[3].

The encoding and decoding operations associated with a piecewise TLSC are as follows. For simplicity, assume that the support set of u is restricted to the interval $[0,1]$. Split the interval $[0,1]$ into M sub-intervals I_k, $1 \leq k \leq M$. Each of these sub-intervals I_k is then stretched and mapped onto a curve on the i-th torus T_i. We will consider uniformly spaced intervals and equally stretched sub-intervals. These can be obtained using

[2] Degenerate tori can also be seen as embeddings in lower dimensional boxes, where the reduction in dimensions is equal to the number of zero coordinates [52].

[3] Alternatively, one could consider mappings of a certain resolution, i.e. consider a curve of fixed length L, and try to maximize the small-ball radius for this given length.

a bijective function f_k such as

$$\begin{aligned} f_k &: I_k \to [0,1) \\ f_k(u) &= \frac{u - \sum_{j=1}^{k-1} l_j/L}{l_k/L}, \end{aligned}$$

where $I_k = \left[\frac{\sum_{j=1}^{k-1} l_j}{L}, \frac{\sum_{j=1}^{k} l_j}{L}\right)$ and $L = \sum\limits_{k=1}^{M} l_k$, with $k = 1, \ldots, M$ and where l_k is the length of the curve defined over the torus $T_{c_k^n}$. Note that other mappings can be considered. The mapping between the stretched sub-intervals and the torus curves can be described as follows. Define $\hat{v}_k^n = c_k^n \circ v_k^n$, where $\circ$ represents the Hadamard product and $v_k^n \in \mathbb{R}^n$. The full encoding map s can be defined composing $f(\cdot)$ and $\Phi(\cdot)$ as

$$s_k(u) := \Phi_{c_k^n}(f_k(u) 2\pi \hat{v}_k^n), \text{ for } u \in I_k. \tag{4.7}$$

In principle, v_k^n could be any real-valued vector. However, not having a constraint could lead to knots that have multiple components, meaning a non-negligible probability of anomalous errors. We are interest in curves that form torus knots, i.e. knots that have a link with one component only. A sufficient condition for such curves is to guarantee that the elements of v_k^n are co-prime. This ensures that the curve defined over the torus does not have any self-intersections.

The full encoding process is illustrated in Fig. 4.3 for a dimension of $n = 2$. The upper line illustrates the partition of the signal interval, while below we show the re-stretched interval associated with I_k. After mapping u on the stretched (and normalized) line through $f_k(\cdot)$, the final value for s_k is computed according to (4.7), by considering the application of $\Phi_{c_k^n}$ restricted to the chosen curve v_k^n. The hyperbox in the bottom of the figure illustrates an hyperbox for the torus T_k defined by $c_k = (c_{k,1}, c_{k,2})$ and the associated curve defined by $v_k = (v_{k,1}, v_{k,2})$. Note that v_k is a $(v_{k,1}, v_{k,2})$ torus-knot. Hence, the curve s_k will turn $v_{k,1}$ times around the axis of rotational symmetry of the torus and $v_{k,2}$ times around a circle in the interior of the torus, which can be seen through the number of intersections of the image of the curve with the sides of the hyperbox.

On the other hand, maximum likelihood decoding of a piecewise TLSC (in the high SNR regime) attempts at minimizing the Euclidean distance between the received point and any other point on the curves of the considered set of tori. Let the vector $x = (x_1, \ldots, x_{2n}) \in \mathbb{R}^{2n}$ be the channel input that results from encoding u and let $y = (y_1, \ldots, y_{2n}) \in \mathbb{R}^{2n}$ be the received vector that is corrupted by channel noise. In particular, if the channel is an Additive White Gaussian Noise (AWGN) channel with zero mean and variance σ^2, the likelihood function is defined as $f_{\hat{U}|u}(\hat{U}|u) = \left(\frac{1}{2\pi\sigma^2}\right)^n \exp^{\frac{\|y - s_T(u)\|}{2\sigma^2}}$ [55]. The ML

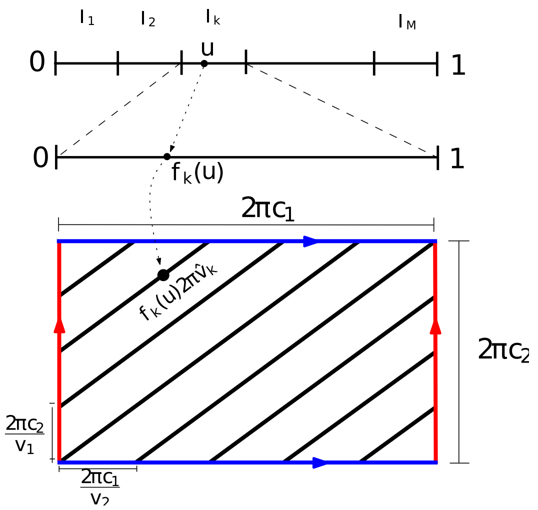


Figure 4.3: Illustration of the encoding operations.

decoder is thus given by

$$
\begin{aligned}
\hat{u} &= \max_{u\in[0,1]} f_{\hat{U}|u}(\hat{U}|u) \\
&= \max_{u\in[0,1]} \left(\frac{1}{2\pi\sigma^2}\right)^n \exp^{\frac{\|y-s_T(u)\|}{2\sigma^2}} \\
&= \operatorname*{argmin}_{u\in[0,1]} \|y-s_T(u)\|,
\end{aligned}
$$

where $s_T(u) = \bigcup_{k=1}^{M} s_k(u)$.

The aforementioned decoding strategy may be computationally expensive as it involves a search over curves. Moreover, specific algorithms to solve this search problem may be hindered by the fact that multiple local minima exist. This can be avoided using a modified decoder that employs a technique that is specific to torus decoding [53]. The idea is to find the set of tori that are closest to the received point and project the received point onto the curves of these tori to find the closest solution [53].

Torus decoding can be accomplished as follows[51]. First re-write the received vector in a similar form to the parametric equations. Let $\gamma_i = \sqrt{y_{2i-1}^2 + y_{2i}^2}$. The received vector y can be expressed as follows

$$
\begin{aligned}
y &= \left(\gamma_1\left(\frac{y_1}{\gamma_1}, \frac{y_2}{\gamma_1}\right), \ldots, \gamma_n\left(\frac{y_{2n-1}}{\gamma_n}, \frac{y_{2n}}{\gamma_n}\right)\right) \\
&= \left(\gamma_1\left(\cos\frac{\theta_1}{\gamma_1}, \sin\frac{\theta_1}{\gamma_1}\right), \ldots, \gamma_n\left(\cos\frac{\theta_n}{\gamma_n}, \sin\frac{\theta_n}{\gamma_n}\right)\right),
\end{aligned}
$$

where $\theta_i = \arccos\left(\frac{y_{2i-1}}{\gamma_i}\right)\gamma_i$. Now compute the projection of y on all tori belonging to

T. Let $\bar{y}_i$ be the projection on torus $T_{c_i^n}$. Then we have that $\|y-\bar{y}_i\| \leq \|y-z\|, \forall z \in T_{c_i^n}$. Defining $\Delta_i = \|y-\bar{y}_i\|$ we may find the minimum distance between the received point and the torus $T_{c_i^n}$. Hence, we may order the tori that are candidates for decoding by this minimum distance. Note that, with high probability, the first candidate will contain the curve which has the point that is closest to the received vector. Moreover, only tori which are, at most, at distance $\frac{d}{2}$ need to be considered, where d denotes the code's minimum distance. For each of these tori, say $T_{c_i^n}$, we may now decode an estimate of the source value by projecting the received vector on the corresponding curve. This can be done by using a modified approach to the shortest vector algorithm, which finds the minimum distance between the received vector and the signal curves based on the torus generating vectors c_i^n and v_i^n [53]. Thus, decoding can be done in the hyperbox $P_{c_i^n}$.

While torus decoding has a performance very close to that of ML decoding, they are only useful in the high-SNR regime, where they are near optimal (in the mean square sense). Nevertheless, we should note that under the low to medium SNR regime, one can use a minimum mean square error (MMSE) decoder to obtain optimal estimates in the mean square sense. In this case, the estimates $\hat{u}_{MMSE} = E\{U|y\} = \int_u up(u|Y)du$, i.e. the estimator is the conditional mean estimator [56]. In particular, the MMSE decoder is not bound to estimate a point in the curves, but estimates directly any point in the source space. Therefore, the geometrical arguments used latter in this chapter can no longer be applied. On the other hand, one may formalize a secrecy problem similar to that of Chapter 3 and optimize a spherical code for secrecy purposes. While in scalar quantization the encoder optimization involved finding the boundaries and the index assignment function of a scalar quantizer, one would now need to find the optimal distance between folds and tori, which are sufficient to define our encoder. However, we do not pursue this avenue here.

Herein, we will assume that we are operating in the high-SNR regime and ML decoding is used by all system users.

4.1.3 Decoding Errors

From a geometrical perspective, ML decoding errors from a piecewise TLSC can be categorized in three types. The first type of error is an estimate of the wrong torus. Errors of this type translate into estimating the wrong source sub-interval. Consequently, they may lead to a high distortion. The second type of errors is the wrong estimate of the curve fold on the correct torus. In this case, the error will be bound by the size of the source sub-intervals. Depending on the size of these sub-intervals, the associated distortion can be high or low. The last type of error is the estimate of the wrong point on the correct torus and curve fold. This generally represents small error and low added distortion if the curve has a large length. The three types of errors are illustrated in Fig. 4.4.

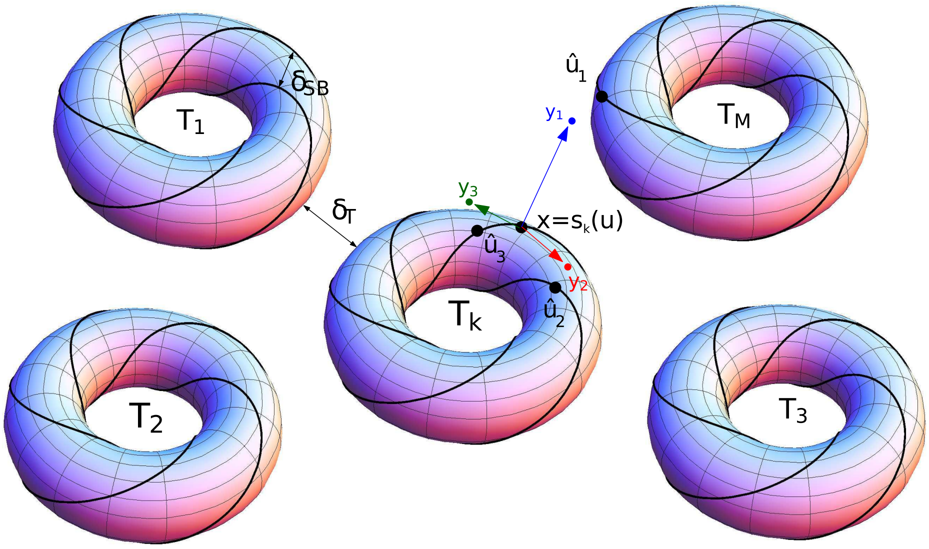


Figure 4.4: Example of the decoding operations with the possible associated errors.

The figure illustrates a transmitted point $x = s_k(u)$ in a torus T_k, and three received vectors y_1, y_2 and y_3. The estimates obtained from each of the vectors represent different types of error. The estimate $\hat{u}_1$ of the point y_1 is located on the wrong torus. The estimates $\hat{u}_2$ and $\hat{u}_3$ are located on the correct torus but in the case of $\hat{u}_2$ on the wrong fold and $\hat{u}_3$ on the correct fold. The impact (in terms of distance between the source value and the estimate) is illustrated in Fig. 4.5. As described above, $\hat{u}_1$ is in a different sub-interval than the source message, and therefore will the resulting distortion will be high. Both $\hat{u}_2$ and $\hat{u}_3$ are restricted to the correct sub-interval. However, $\hat{u}_3$ is closer to u than $\hat{u}_2$, which lead to a lower distortion value.

4.2 Piecewise Torus Layer Spherical Codes for Secrecy

Consider the following wiretap scenario depicted in Fig. 4.6. A sender wishes to reliably transmit a real valued signal $u \in \mathbb{R}$ to a receiver, while preventing an eavesdropper from correctly estimating u. Both the main channel and the wiretap channel are AWGN channels subject to an input average power constraint P. The wiretap channel is degraded with respect to the main channel, i.e., $\sigma_w^2 > \sigma_m^2$, where σ_m^2 and σ_w^2 are the noise variances associated with the main and wiretap channels. To transmit the source value u, the sender employs a piecewise TLSC as described in the previous section, i.e. he employs an encoder that maps u onto a codeword $x \in \mathbb{R}^{2n}$. The codeword x is then transmitted

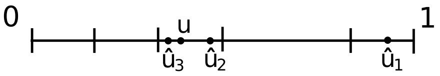


Figure 4.5: Example of the impact of errors on the estimates of the source message.

to the destination over the main channel and corrupted by the additive noise vector n_b, where $n_b = (n_{b,1}, \ldots, n_{b,2n})$, with $n_{b,i} \sim \mathcal{N}(0, \sigma_m^2)$. Similarly, the eavesdropper observes the transmission of x over the wiretap channel, which is corrupted by the noise vector n_e, where $n_e = (n_{e,1}, \ldots, n_{e,2n})$, $n_{e,i} \sim \mathcal{N}(0, \sigma_w^2)$. The legitimate receiver obtains the main channel output sequence $y = x + n_b$, while the eavesdropper obtains the wiretap channel output sequence $z = x + n_e$. Then both receivers estimate the source message using the ML decoder described in the previous section. The legitimate user estimates the point $\tilde{u}$, while the eavesdropper estimates the point $\hat{u}$.

We have seen in the previous section that there exists a large impact in distortion when the decoder chooses the wrong torus. Hence, it is desirable that the eavesdropper makes such mistakes. On the other hand, it should be sufficient for the legitimate receiver to guess the correct torus and fold in order to obtain estimates with small distortion. These observation allows us to provide guidelines for code design, that provide a reliability/secrecy trade-off. More precisely, let δ_{SB} and δ_T be the small-ball radius and the minimum distance between tori ($\delta_{SB} < \delta_T$) of a given spherical code. Then, our goal is to find codes that satisfy the following constraints:

$$P(\|n_b\| \leq \delta_{SB}/2) \geq 1 - \alpha, \tag{4.8}$$

$$P(\|n_e\| > \delta_T/2) \geq 1 - \varepsilon, \tag{4.9}$$

In the limit of large n, α represents the fraction of noise vectors which we allow to go outside a $2n$-dimensional sphere of radius $\delta_{SB}/2$. On the other hand, ε represents the fraction of noise vectors which we allow to live inside a $2n$-dimensional sphere of radius $\delta_T/2$. Ideally, one should consider α and ε arbitrarily small.

Both these probabilities allow for close form expressions that are efficiently computable. Let $n_{AWGN} = (n_1, \ldots, n_{2n})$, $n_i \sim \mathcal{N}(0, \sigma^2)$, $1 \leq i \leq 2n$, be a random vector of additive gaussian noise. Consider the following auxiliary r.v. $T = \sqrt{\sum_{i=1}^{2n} \left(\frac{n_i' - \mu_i}{\sigma_i}\right)^2}$, with $n_i' \sim \mathcal{N}(\mu_i, \sigma_i^2)$. Then, T follows chi distribution that is parametrized by $2n$. The proba-

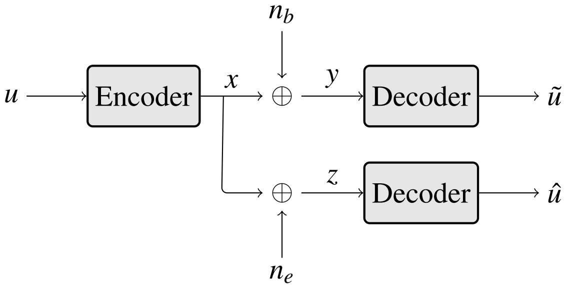


Figure 4.6: The AWGN wiretap channel model.

bility density function (*pdf*) of T is given by

$$f_T(t) = \frac{2^{1-n}t^{2n-1}}{\Gamma(n)e^{\frac{t^2}{2}}}, \tag{4.10}$$

where $\Gamma(\cdot)$ is the gamma function. Its cumulative density function (*cdf*) is given by

$$P(T \leq a) = \frac{2^{1-n}}{\Gamma(n)} \int_0^a t^{2n-1} e^{\frac{-t^2}{2}} \mathrm{d}t, \tag{4.11}$$

Consequently, we have that

$$\begin{aligned} P\left(\|n_{AWGN}\| \leq d\right) &= P\left(\sqrt{\sum_{i=1}^{2n} \frac{n_i^2}{\sigma^2}} \leq \frac{d}{\sigma}\right) \\ &= \frac{2^{1-n}}{\Gamma(n)} \int_0^{\frac{d}{\sigma}} t^{2n-1} e^{\frac{-t^2}{2}} \mathrm{d}t. \end{aligned} \tag{4.12}$$

Similarly, we have that

$$P\left(\|n_{AWGN}\| > d\right) = 1 - \frac{2^{1-n}}{\Gamma(n)} \int_0^{\frac{d}{\sigma}} t^{2n-1} e^{\frac{-t^2}{2}} \mathrm{d}t. \tag{4.13}$$

For both cases, the integral $\int t^{2n-1} e^{\frac{-t^2}{2}}$ evaluates to

$$\int t^{2n-1} e^{\frac{-t^2}{2}} \mathrm{d}t = -e^{\frac{-t^2}{2}} \sum_{i=1}^{n} \left(\prod_{j=1}^{i-1} (2n-2j)\right) t^{2n-2i}. \tag{4.14}$$

Solving (4.8) and (4.9) is a simple matter of substituting d with $\delta_{SB}/2$ in (4.13) and with $\delta_T/2$ in (4.12).

From (4.8) we can obtain a lower bound on the minimum required distance between folds $d_{SB,\min}$, that satisfy this reliability constraint. Similarly, (4.9) provide us with an upper bound on the maximum distance $d_{T,\max}$ allowed between tori that respect the secrecy constraints. Therefore, the values of δ_{SB} and δ_T should be chosen such that $\delta_{SB} \geq d_{SB,\min}$ and $\delta_T \leq d_{T,\max}$, with $\delta_{SB} < \delta_T$.

It should be noted that, depending on the code dimension and channel noise, it may not be possible to find a parametrization that satisfies the reliability and secrecy constraints. For instance, if the number of dimensions is very small (e.g. $n = 2$) and the constraints are very tight (e.g. $\alpha \approx 0, \varepsilon \approx 0$), solutions can only be found for larger values of σ_w^2. Consequently, it may be necessary to relax either the secrecy or reliability constraints (i.e.

consider larger values for α and ε) or increasing the number of dimensions (increasing the bandwidth expansion). The latter solution also allows an increase in the number of tori to be packed onto the channel space, hence a larger number of subdivision of the source interval. Finally, note that additional constraints similar to those described in (4.8) and (4.9) can be imposed to the eavesdropper and legitimate receiver. For instance, we may wish to lower bound the probability of an eavesdropper of having fold errors or lower bound the probability that a legitimate receiver does not have torus errors. This allows for more degrees of freedom in the code design process.

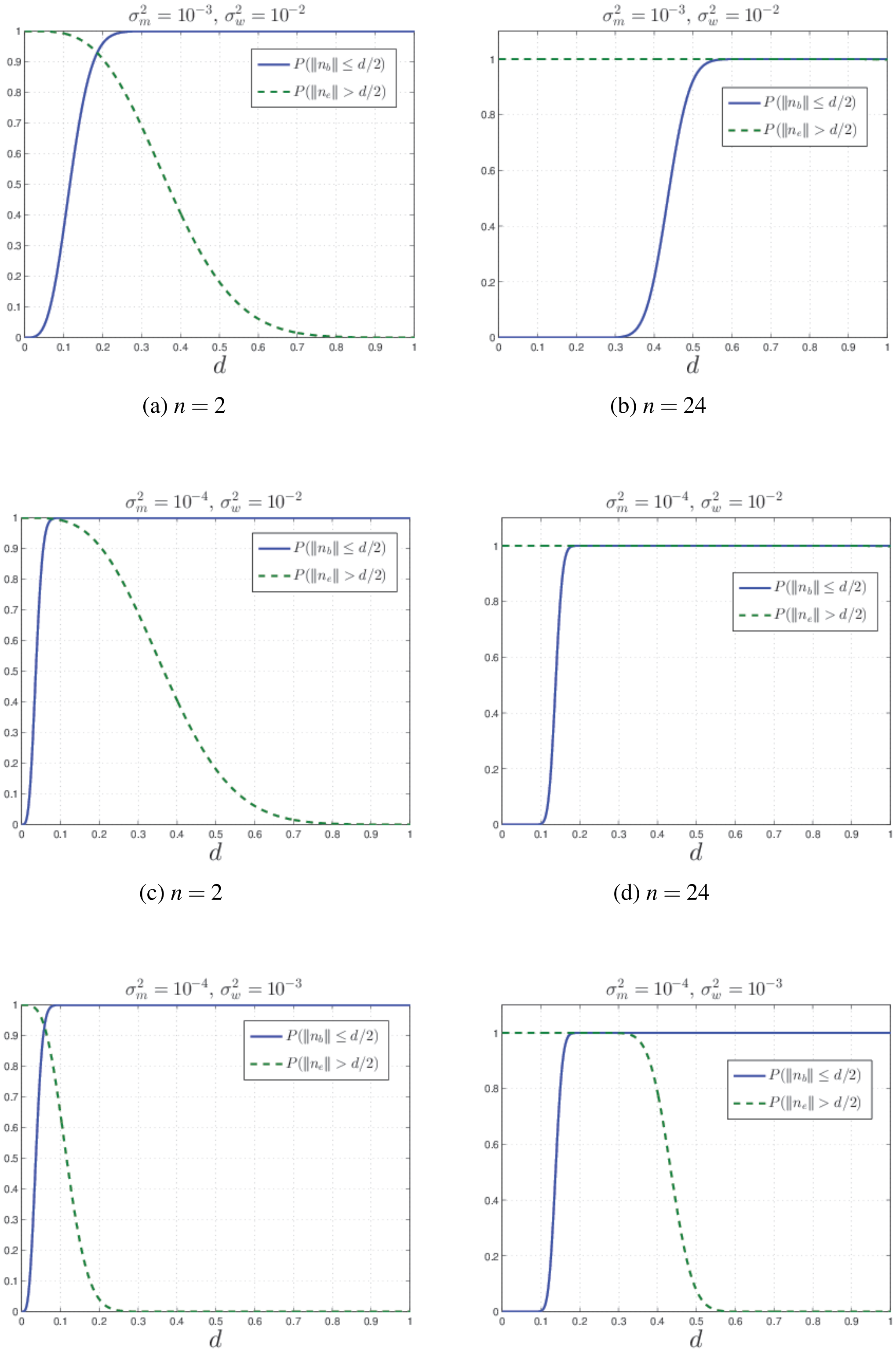
4.3 Numerical Results

Before analyzing the secrecy perfomance of the proposed code construction, let us first develop an intuition with respect to satisfiability of the constraints defined in the previous section. Consider the three following instances of an AWGN wiretap model: i) $\sigma_m^2 = 10^{-3}$ and $\sigma_w^2 = 10^{-2}$, ii) $\sigma_m^2 = 10^{-4}$ and $\sigma_w^2 = 10^{-2}$ and iii) $\sigma_m^2 = 10^{-4}$ and $\sigma_w^2 = 10^{-3}$. Fig. 4.7 plots the cumulative noise distributions $P(\|n_b\| \leq d/2)$ and $P(\|n_e\| > d/2)$ for a varying distance d for the three cases mentioned above when $n = 2$ (Figs. 4.7a, 4.7c and 4.7e on the left side of the figure) and $n = 24$ (Figs. 4.7b, 4.7d and 4.7f on the right side of the figure).The cumulative distributions allows us to directly understand if there exists a solution that jointly satisfies the reliability and secrecy constraints. The admissible values for δ_{SB} can be found by drawing an horizontal line at $1-\alpha$ and taking the values of d to the right of the intersection. Similarly, the admissible values for δ_T can be found by drawing an horizontal line at $1-\beta$ and taking the values of d to the left of the intersection. There exists a solution to the parametrization problem if the intersection of these intervals is non-empty.

The immediate observation is that, for small dimensions, is is impossible to jointly satisfy strict reliability and secrecy constraints, i.e. $\alpha \approx 0$ and $\varepsilon \approx 0$. The second scenario is the one that best approximates these conditions, i.e. both requirements can be satisfied for small (but non-negligible) values of α and ε. Nevertheless, this requires a channel advantage of 20dB. However, an increase in the number of dimensions almost guarantees an existence of a solution. For instance, for $n = 24$ all the scenarios have a satisfiable solution to this problem.

It is interesting to note that, for increasing n, larger distances are required to satisfy the reliability constraint. On the other hand, the secrecy constraints also hold for longer distances. Note that the reliability and secrecy conditions are still at odds. When we increase dimensions, we are essentially shifting and stretching the cumulative noise distributions, and that is why it is easier to find solutions to the parametrization problem.

As mentioned previously, we may also opt to relax the reliability and secrecy constraints. This amounts to considering larger values for α and ε. Recall that α controls



(a) $n = 2$

(b) $n = 24$

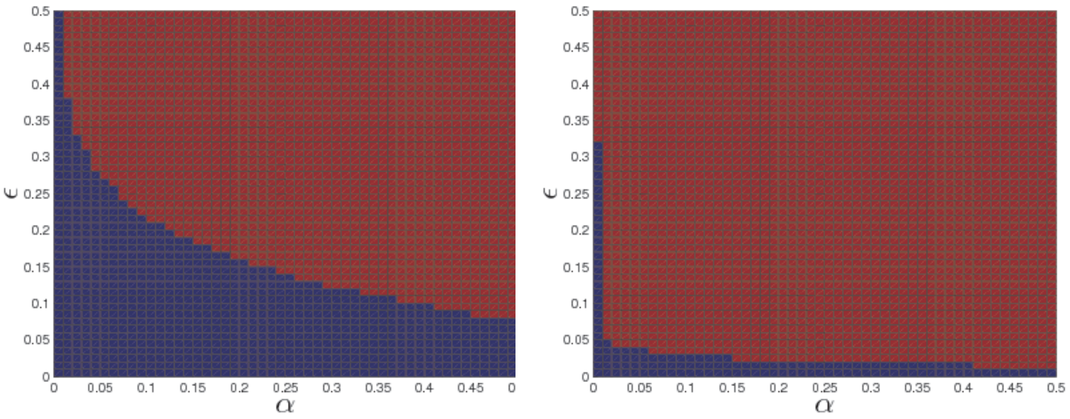
(c) $n = 2$

(d) $n = 24$

(e) $n = 2$

(f) $n = 24$

Figure 4.7: $P(\|n_b\| \leq d/2)$ and $P(\|n_e\| > d/2)$ as a function of distance d, for dimensions $n = 2$ and $n = 24$.



(a) $\sigma_m^2 = 10^{-3}$ and $\sigma_m^2 = 10^{-2}$

(b) $\sigma_m^2 = 10^{-4}$ and $\sigma_m^2 = 10^{-3}$

Figure 4.8: Map of parametrizations for which there exists a solution to the reliability and secrecy constraints with $n = 2$.

the distance between folds and ε the distance between tori. Thus, it is expected that increasing α does not provide much impact with respect to the legitimate receivers distortion, whereas increasing ε may reduce the distortion of the eavesdropper by a considerable amount. Let us focus on the case of $n = 2$, which is the case that does not allow a parametrization with vanishing α and ε. Fig. 4.8 draws a map of the parametrizations that allow for a solution as a function of α and ε for the first and third cases (10dB of channel SNR advantage). The red area shows the pairs where a solution is found, whereas the blue area shows the pairs where a solution is not found. We can see that an increase in the channel conditions for both users actually allows for a broader selection of parameters. The reason is that both cumulative noise distribution curves are shifted in opposite directions, which allows to cover a broader range of parameters.

Table 4.1: Performance of spherical codes for the wiretap channel with $n = 2$.

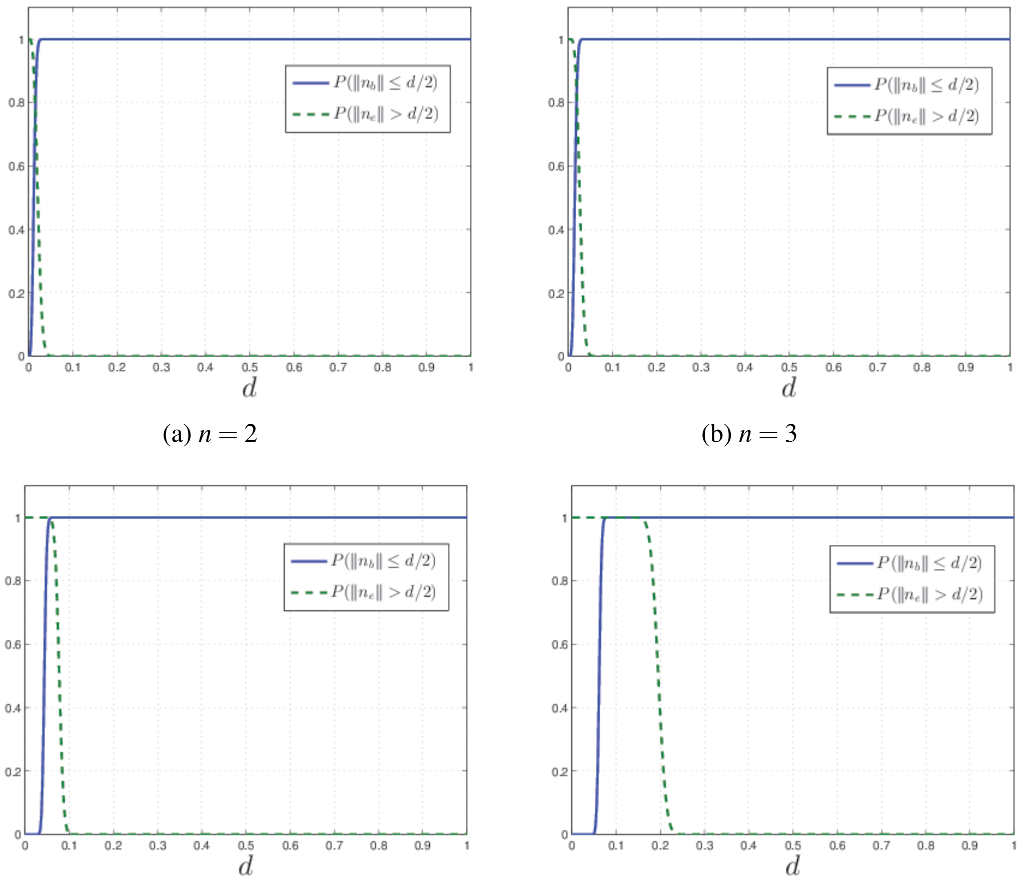
$CSNR_E$	20	25	30	30	35	35	40	40	40	45	45	45
α	0	0	0	0.01	0	0.09	0.01	0.1	0.3	0.1	0.25	0.5
ε	0.02	0.02	0.08	0.01	0.4	0.02	0.19	0.08	0.04	0.42	0.27	0.15
δ_{SB}	0.058	0.058	0.058	0.024	0.058	0.018	0.024	0.018	0.014	0.018	0.015	0.012
δ_T	0.131	0.073	0.061	0.034	0.059	0.019	0.025	0.019	0.015	0.019	0.016	0.013
SNR_B	108	85	112	59	69	58	75	57	32	40	35	31
SNR_E	10	13	15	16	21	18	21	20	20	26	25	24

In Table 4.1 we show the performance of spherical codes for several parametrizations of σ_w^2 (reflected on the channel SNR of the wiretap channel, $CSNR_E$), as well as α and ε for a dimension of $n = 2$. Throughout the table we fix $\sigma_m^2 = 10^{-5}$, i.e. a channel SNR of $50dB$. The mapping strategy above is used to choose the reported parameters α and

ε. For smaller value of $CSNR_E$ we see that both α and ε may take small values. On the other hand, for values of $CSNR_E$ closer to the main channel CSNR, a relaxation of ε (or α) is required. It is interesting to see how the distances δ_{SB} and δ_T vary according to these parameters. While for the cases of lower $CSNR_E$ it is possible to obtain distances δ_{SB} and δ_T that are already some distance apart, the same is not true with respect to the cases where $CSNR_E$ increases. The reason is that for such values of wiretap CSNR, the cumulative curves have a very sharp decay. Moreover, as $CSNR_E$ approximates the main channel CSNR, the cumulative noise distributions become almost complementary curves. Thus, in such cases one must increase the code's dimensions in order to be able to exploit the distance diversity. Fig. 4.9 illustrates this point. Here we fix the main channel CSNR at 50dB and the wiretap channel CSNR at 45dB and let the dimension increase. As it increases, we see that the cumulative noise distributions have a less sharp decay. Moreover, the difference between channels becomes more noticeable, as the cumulative distributions are longer complements of each other. Under such conditions it is now possible to use α and β to provide for trade-offs that have the desired consequence of ensuring that δ_{SB} and δ_T are sufficiently far apart. Table 4.1 also shows that the eavesdropper's output SNR, although increasing with the wiretap channel SNR, is constantly kept small. We note that there are two effects under play. When δ_{SB} and δ_T are not similar, the eavesdropper's distortion is mostly affected by torus errors. However, when both are similar, there is also a large distortion contribution from fold errors. This can be seen in the largest values of $CSNR_E$, where we allow a larger value of ε, but still the eavesdropper as a low output SNR. From the legitimate receiver's perspective, we see that when δ_{SB} and δ_T are closer, the distortion is greatly impacted. While the fraction of torus errors for the legitimate receivers is residual, the fact that the considered values for δ_{SB} are very small impacts negatively on its distortion, especially in such low dimensions.

4.4 Discussion

We have proposed the use of spherical codes based on flat tori as the foundation of a coding scheme for the Gaussian wiretap channel with continuous inputs. The scheme inherits the advantages of spherical codes: efficient encoding/decoding, good performance in the high SNR regime and bandwidth expansion. We show that a careful parametrization of these codes (which takes into account their geometrical properties) enables legitimate users to communicate under a small distortion, while forcing the eavesdropper to operate at larger distortions. Moreover, the proposed construction provides a simple mechanism to trade-off reliability with secrecy.



(a) $n = 2$

(b) $n = 3$

(c) $n = 24$

(d) $n = 48$

Figure 4.9: $P(\|n_b\| \leq d/2)$ and $P(\|n_e\| > d/2)$ as a function of distance d, for dimensions $n = 2, 3, 24$ and 48.

Chapter 5

Randomly Punctured LDPC Codes for Secrecy

Rate-compatible coding [57, 58, 59] is an error-control strategy that allows to adapt the rate of a given code to the channel statistics. The design principle behind these codes is to use a low-rate code whenever channel conditions are not favourable and a high-rate code embedded in the low-rate code whenever channel conditions improve. Hence, rate-compatible codes use the same underlying structure regardless of channel conditions.

One of the most efficient ways of implementing rate compatible codes is through puncturing. This technique consists in selecting only a subset of the encoded bits from the original codewords for transmission. The bits that are not selected are said to be punctured. Typically, the puncturing operation is deterministic in the sense that puncturing patterns are agreed upon a priori for the desired rates. However, the puncturing operation can also be stochastic, i.e. bits can be randomly punctured. The latter approach however, requires some sort of mechanism to inform the receiver of the positions of punctured bits.

In the context of physical-layer security, the first secrecy strategy that used puncturing was developed in the context of the Gaussian wiretap channel [19]. The authors showed that, under belief propagation decoding, the eavesdropper will experience bit error rates (BER) close to 0.5 if its signal to noise ratio is lower than a given threshold. Consequently, the eavesdropper's observations contained nearly *i.i.d* errors, and non-decodability could be ensured. With respect to secrecy metrics, [19] introduces a new secrecy metric called the *security gap*, which attempts at measuring the point at which decoding is successful for the legitimate receiver and fails for the eavesdropper. Then, puncturing distributions are optimized in order to reduce the security gap. The idea is that, if the security gap can be reduced to zero, then any stochastically degraded channel will be enough to ensure non-decodability. Puncturing has also been used in the context of packet erasure channels with authenticated feedback [60]. The availability of an authenticated feedback channel enables only the legitimate receiver to request retransmissions for missing packets, thus

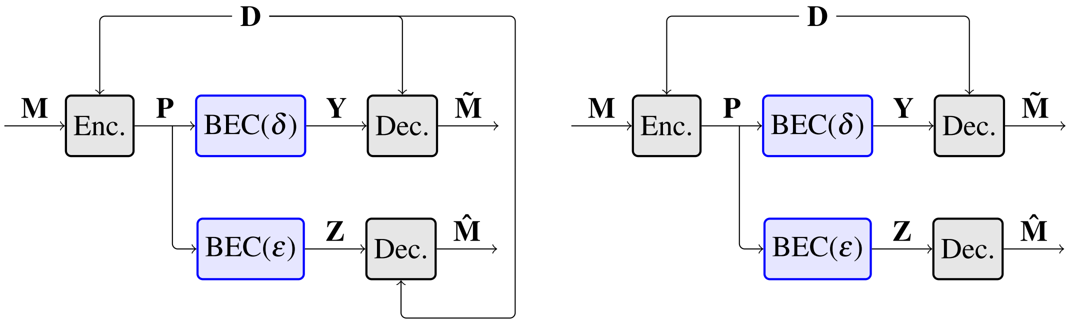


Figure 5.1: Wiretap model of a coding scheme that uses puncturing to obtain secrecy. Two cases are considered: to the left the puncturing pattern is public, whereas to the right the the puncturing pattern is a shared secret between the legitimate parties.

limiting the opportunities of the attacker to eavesdrop on a particular packet. Secrecy is achieved by making use of a stopping-set based puncturing strategy that adds new degrees of freedom for each missing packet. Thus, the diversity of erasure patterns experienced by the legitimate receiver and the eavesdropper is sufficient to ensure secrecy, even when the eavesdropper has channel advantage. As mentioned in Section 2.5, punctured LDPC codes have also been used to design nested codes for the Gaussian wiretap channel [30] that achieve the weak secrecy capacity.

The aforementioned code constructions make use of puncturing patterns (or distributions) that are agreed upon a priori. This means that puncturing is used to hide information from the eavesdropper, neglecting it as a mechanism that can be used to adapt the code to channel conditions. However, this does not mean that puncturing cannot take both roles at the same time.

In this chapter, we propose a new framework for coding for the binary erasure wiretap channel (BEWC) that is based on random puncturing. Within this framework, we analyze two cases. First, we consider the case of publicly known puncturing patterns, i.e. known by the legitimate party and the eavesdropper. We then consider the case where the puncturing pattern is secret, known by the legitimate party but not by the eavesdropper. Both models are depicted in Fig. 5.1, where $\mathbf{M}$ represents the source message, $\mathbf{D}$ represents the puncturing pattern, $\mathbf{P}$ represents the punctured output of the encoder, $\mathbf{Y}$ and $\mathbf{Z}$ the respective main and wiretap channel outputs, and $\tilde{\mathbf{M}}$ and $\hat{\mathbf{M}}$ the respective estimates of the source message by the legitimate receiver and eavesdropper.

While the coding scheme is the same in both cases, its operational interpretation is quite different with respect to the knowledge of the secret key. If a given user is aware of the puncturing pattern, then the puncturing operation can be seen as a mechanism that introduces erasures. In particular, if the encoder output bits are independently punctured with the same probability, we may model the puncturing operation as passing the outputs of the encoder through a binary erasure channel. Ergo, random puncturing with a public

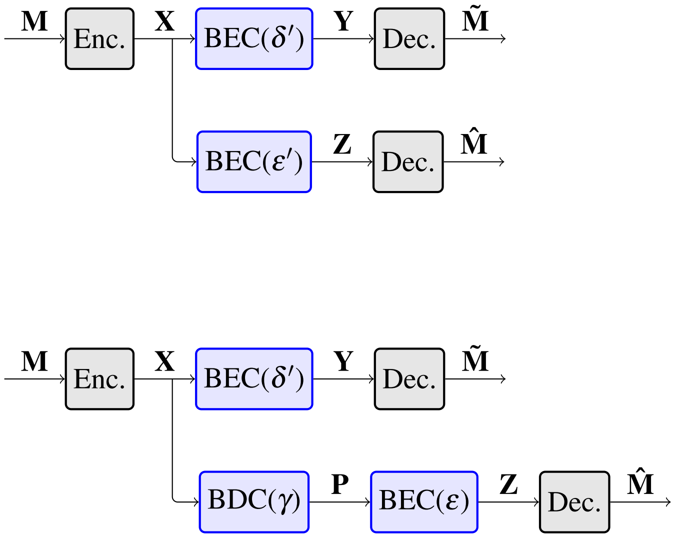


Figure 5.2: Operational interpretation of random puncturing, when the pattern is public (top figure) or secret (bottom figure).

puncturing pattern is essentially a way to introduce artificial noise in the form of erasures w.r.t. the transmitted messages. On the other hand, if a given user is not aware of the puncturing pattern, his observations of the transmitted messages will be lacking bit-level synchronization, i.e. he does not know the positions of the received bits w.r.t to the original unpunctured codeword. Again, if the encoder output bits are independently punctured with the same probability, the puncturing operation can now be modelled as passing the outputs of the encoder through a binary deletion channel [61]. Fig. 5.2 illustrates these operational models, where now $\mathbf{X}$ represents the unpunctured encoder output, and where δ' and ε' are erasure probabilities that take into account the puncturing probability.

From a security perspective, the second model is more appealing, since the equivocation associated with a deletion channel is higher than that of an erasure channel (the deletion channel can be seen as a genie-aided erasure channel). Hence, hiding the puncturing pattern from the eavesdropper would result in a better secrecy performance. However, there is an added cost associated with secretly sharing the puncturing pattern which must not be disregarded.

Herein, we will consider a coding scheme that uses randomly punctured LDPC codes, which are known by their efficiency and high performance. Moreover, there are many established techniques to analyze the performance of LDPC codes over binary erasure channels, which will be useful to our analysis.

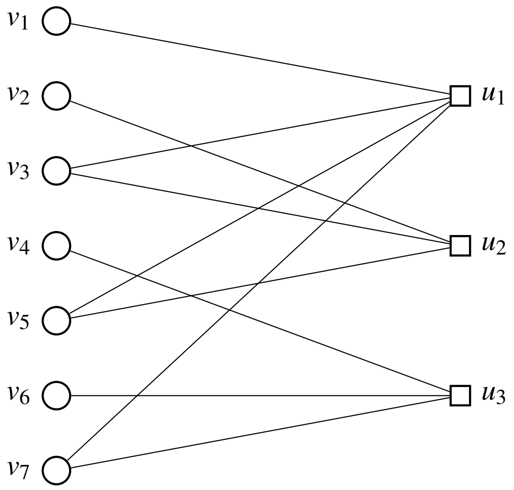


Figure 5.3: Bipartite graph with $n = 7$ variable nodes (represented by circles) and $n-k = 3$ check nodes (represented by squares).

5.1 Randomly Punctured LDPC codes

An LDPC ensemble [62] of length n can be defined as an ensemble of bipartite graphs comprised of variable nodes and parity check nodes (illustrated in Fig 5.3). These bipartite graphs can be characterized by polynomial degree distributions [62] either from an edge perspective or from a node perspective. As usual, we will denote the variable node and parity check node degree distributions from an edge perspective by $\lambda(x)$ and $\rho(x)$, respectively. In particular, we define $\lambda(x) = \sum_{l=2}^{c} \lambda_l x^{l-1}$ and $\rho(x) = \sum_{l=2}^{d} \rho_l x^{l-1}$, where λ_l and ρ_l represent the fraction of edges connected to variable nodes with degree l and the fraction of edges connected to parity check nodes with degree l and c and d represent the maximum degree of variable and check nodes. Then, we denote an LDPC ensemble as LDPC(n, λ, ρ).

Similarly, we will denote the degree distributions from a node perspective by $\Lambda(x)$ (variable node) and $\Gamma(x)$ (parity check node). We define $\Lambda(x) = \sum_{l=2}^{c} \Lambda_l x^l$ and $\Gamma(x) = \sum_{l=2}^{d} \Gamma_l x^l$, where Λ_l and Γ_l represent, respectively, the fraction of variable nodes with degree l and the fraction of parity check nodes with degree l and c and d are defined as above. The conversion between the degree distributions from the edge perspective to the node perspective (and vice-versa) are given by the following two relations:

$$\lambda(x) = \frac{\Lambda'(x)}{\Lambda'(1)}, \rho(x) = \frac{\Gamma'(x)}{\Gamma'(1)}.$$

From the degree distributions it is possible to compute the design rate R of an LDPC code. We have that $R = 1 - \frac{\Gamma}{\Lambda}$, where $\Gamma = \sum_{l=2}^{d} \frac{\rho_l}{l}$ and $\Lambda = \sum_{l=2}^{c} \frac{\lambda_l}{l}$.

A puncturing distribution for an LDPC code is similar to that of a code's degree distribution. We denote such distribution by $\gamma(x) = \sum_{l=2}^{c} \gamma_l x^{l-1}$, where γ_l represents the fraction of variable nodes of degree l that are punctured. The resulting punctured code has a design rate R^* given by $R^* = \frac{R}{1-\frac{\Lambda_p}{\Lambda}}$, where R is the initial code rate, $\Lambda_p = \sum_{l=2}^{c} \frac{\gamma_l \lambda_l}{l}$, and Λ is defined as before. As noted before, it is useful to think of the puncturing operation in terms of an erasure channel. Therefore, we will specifically consider puncturing distributions of the form $\gamma(x) = \sum_{l=2}^{c} \gamma x^{l-1}$, i.e. each bit is punctured independently and uniformly at random with probability γ.

5.1.1 Encoding and Decoding of LDPC Codes over Binary Erasure Channels

From an encoding and decoding perspective, LDPC codes are more useful when instanced as linear codes. An (n,k) linear binary block code is a set $\mathcal{C} \subset \mathbb{F}_2^n$ composed of 2^k codewords of length n. The encoder is a bijective map between messages of k bits and codewords of n bits, while a decoder is a surjective map between the set of all binary sequences of length n and the set $\mathcal{C}$. Moreover, one of the code's properties is that it is a subspace of $\mathbb{F}_2^n$ with dimension k. In particular, $\mathcal{C}$ is closed under addition, i.e. the sum of any two codewords in $\mathcal{C}$ also belongs to $\mathcal{C}$.

Encoding with linear block codes can be performed by multiplying $\mathbf{m}$ by a $(k \times n)$ generator matrix $\mathbf{G}$, whose rows form a basis of the linear code, i.e. $\mathbf{x} = \mathbf{mG}$. It is possible to associate with $\mathcal{C}$ an $(n-k) \times n$ parity check matrix $\mathbf{H}$, with the property that $\mathbf{xH}^\intercal = 0$, for any $\mathbf{x} \in \mathcal{C}$. Then, for any sequence $\mathbf{y} \in \mathbb{F}_2^n$, it is possible to compute a quantity called *syndrome* which is given by $\mathbf{s} = \mathbf{yH}^\intercal$. Thus, $\mathbf{s} = \mathbf{0}$ if and only if $\mathbf{y} \in \mathcal{C}$. While decoding of a binary linear code may take many forms, these are in general related to the parity check matrix. We will describe some possible decoding techniques latter in this section.

The connection between the definition of LDPC codes via degree distributions and via a linear block code formulation is almost straightforward. Consider a code $\mathcal{C} \in \mathrm{LDPC}(n,\lambda,\rho)$. $\mathrm{LDPC}(n,\lambda,\rho)$ is a set of bipartite graphs that satisfy the constraints indicated by λ and ρ and $\mathcal{C}$ is a specific instance of such bipartite graph. Each of the bipartite graphs in $\mathrm{LDPC}(n,\lambda,\rho)$ can be represented (and fully defined) in terms of the parity check matrix $\mathbf{H}$ for specific values of k and n, where k is the length of the source message and n is the desired codeword block-length n. The parity-check matrix $\mathbf{H}$ is an $(n-k) \times n$ matrix, whose rows are associated with the check nodes of the bipartite graph and whose columns are associated with the variable nodes. Denote the set of variable nodes by denoted $V = (v_1, v_2, \ldots, v_n)$, and the set of check nodes by denoted $U = (u_1, u_2, \ldots, u_{n-k})$. Then, $\mathbf{H}_{i,j} = 1$ if and only if there is an edge between check node u_i and variable node v_j, where $\mathbf{H}_{i,j}$ denotes the entry in $\mathbf{H}$ that corresponds to the i-th row and j-th column. For the bipartite graph in Fig. 5.3 we have the following parity check matrix.

$$\mathbf{H} = \begin{bmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

While the definition of an LDPC code is intrinsically connected to parity check matrices, such connection is not obvious with respect to the generator matrix. However, we can obtain the generator matrix from the parity check matrix as follows. Assuming that $\mathbf{H}$ has full row rank, it is possible to use Gauss-Jordan elimination to put $\mathbf{H}$ into the form $\mathbf{H} = [\mathbf{A}, \mathbf{I}_{n-k}]$, from where we take $\mathbf{G} = [\mathbf{I}_k, A^\intercal]$[1].

Decoding with LDPC codes can be performed using the general *maximum a posteriori* (MAP) block decoder obtained for linear block codes. If $\mathbf{x} \in \mathcal{C}$ is the transmitted codeword and $\mathbf{y}$ is the received codeword and if codewords have a uniform prior distribution, then the MAP decoder is given by

$$\hat{\mathbf{x}}(\mathbf{y}) = \underset{\mathbf{x}}{\operatorname{argmax}} P(\mathbf{x}|\mathbf{y}). \tag{5.1}$$

Now let $\mathcal{E}$ and $\overline{\mathcal{E}}$ denote the set of indices of known bits and erased bits in $\mathbf{y}$, respectively. Define the matrices $H_{\mathcal{E}}$ and $H_{\overline{\mathcal{E}}}$ to be matrices including only the columns of H indexed by $\mathcal{E}$ and $\overline{\mathcal{E}}$, respectively. Similarly, let $x_{\mathcal{E}}$ and $x_{\overline{\mathcal{E}}}$ be vectors obtained from $\mathbf{x}$ through indexing of $\mathcal{E}$ and $\overline{\mathcal{E}}$. Since the received bits are either correct or erased, we can write $0 = Hx^\intercal = H_{\mathcal{E}}x_{\mathcal{E}}^\intercal + H_{\overline{\mathcal{E}}}x_{\overline{\mathcal{E}}}^\intercal$. In particular, since $x_{\mathcal{E}} = y_{\mathcal{E}}$, we simply need to solve for the channel erased bits, i.e.

$$H_{\overline{\mathcal{E}}}x_{\overline{\mathcal{E}}}^\intercal = s^\intercal, \tag{5.2}$$

where $s^\intercal = H_{\mathcal{E}}x_{\mathcal{E}}^\intercal$.

There may be more than one solution to (5.2) if it represents an under-determined system of equations. However, it is possible to identify the set of compatible codewords that represent solutions to the above problem. Consider the following set $\mathcal{X}^{MAP}(\mathbf{y}) = \{\mathbf{x} \in \mathcal{C} : s^\intercal = H_{\mathcal{E}}x_{\mathcal{E}}^\intercal\}$. Since all the elements in $\mathcal{X}^{MAP}(\mathbf{y})$ are equally likely, we may define the following rule for the MAP decoder (5.1)

$$\hat{\mathbf{x}}(\mathbf{y}) = \begin{cases} \mathbf{x} \in \mathcal{X}^{MAP}(\mathbf{y}) & \text{, if } |\mathcal{X}^{MAP}(\mathbf{y})| = 1 \\ ? & \text{, otherwise,} \end{cases}$$

[1] It is also possible to use the parity check matrix for encoding [62, Appendix A] by putting $\mathbf{H}$ into an upper triangular form and considering codewords of the form $x = (p, m)$, where m are the message bits and p are the parity bits, which can be found by back substitution from $\mathbf{H}$.

where ? denotes a decoding error. In particular, the MAP decoder provides a solution only when $rank(H_{\mathcal{E}}) = |\mathcal{E}|$, i.e. the sub-matrix associated with the erased positions is full rank[2].

We have seen that MAP decoding can be accomplished by solving a system of linear equations. However, such strategy has a cubic polynomial complexity in general. For efficiency reasons, LDPC decoding is generally accomplished using a *belief-propagation* (BP) decoder. The BP decoder [62] operates by passing messages (or beliefs) between variable nodes and check nodes iteratively until errors are corrected (or a prescribed number of iterations is reached). This process is summarized in algorithmic form in Algorithm 1, where sums are taken modulo 2. Note that BP decoders are sub-optimal with respect to MAP decoders. The reason stems from the iterative nature of the algorithm that can get stuck in the so called *stopping sets*, i.e. subsets of variable nodes such that all neighbours of the subset are connected to it at least twice. In fact, after BP decoding takes place, the remaining variable nodes containing erasures is the maximum stopping set contained in the original set of erasures. In particular, for the possible erasure patterns, it is more frequent for a stopping set to occur than for the erasure pattern to be the same as the support set of a codeword. Consequently, BP decoding fails in cases where MAP decoding does not. Still, belief propagation decoders are known to approximate bit-wise maximum a posteriori decoding and therefore, this decoding strategy generally achieves a good performance.

Algorithm 1 Belief propagation decoding over erasure channels.

```
1:  Initialize: For $y_j \neq ?$, set each variable node $v_j = y_j$ and send $v_j$ along the outgoing edges;
2:  if All variable nodes are known then
3:      Output the codeword.
4:  end if
5:  Processing at check nodes: At each check node $u_i$, if there is an outgoing edge to variable
    node $v_j$, send a message to $v_j$ containing:
6:    1) an erasure if there was some incoming message from $v_k$, $k \neq j$ that was an erasure;
7:    2) a message containing the parity $p = \sum_{v_k \neq v_j} v_k$, where $v_k$ are the set variable nodes con-
    nected to check node $u_i$
8:  Processing at variable nodes: At each variable node $v_j$, if there is an incoming message from
    check node $u_i$ that is not an erasure, set $v_j = u_i$ and send $v_j$ along the outgoing edges;
9:  if All variable nodes are known or no more progress can be made then
10:     Output the (possibly partial) codeword.
11: else
12:     Go to Processing at check nodes (line 5).
13: end if
```

While there are many possibilities for improving BP decoding over erasure channels,

[2] It is possible to further refine the estimate of the transmitted codeword by considering bit-wise map decoding, which leads to estimates of individual bits, rather than codewords.

there is a conceptual decoder that is useful for characterizing the difference in terms of performance of BP and MAP decoding. Consider the following variation of the BP described earlier. In the check node processing stage accumulate the values of the incoming messages and delete all the edges that stem from variable nodes whose bit is already known. Then randomly select a check node of degree 1 and send its accumulate values onto the remaining outgoing edge. This determines the value of the respective variable node that is connected to that check node. Now, select this variable node and transmit the new found value along the remaining outgoing edges of this node and iteratively repeat the procedure. When there are no more paths connecting either a variable node or an edge node, these can be discarded. At the end of the iterative procedure, a residual graph will be obtained, containing the variable nodes that could not be resolved. If this graph is empty, decoding was successful. Otherwise, the remaining variable nodes form a maximum stopping set, just like in the previous case. This decoder is usually denoted as *peeling decoder*.

Let us illustrate this decoding process. Suppose that, using the LDPC code whose bipartite graph is depicted in Fig. 5.3, we transmit the codeword $\mathbf{x} = 1010111$ and the channel erases the second and fifth bits (see Fig. 5.4). In the first iteration, all non-erased variable nodes transmit their respective values along the outgoing edges. These values are accumulated at the check nodes and the respective edges are erased. From the remaining graph, it is possible to see that check node u_1 has an induced degree of 1 and check node u_2 has an induced degree of 2. Thus, we select check node u_1 and transmit the accumulated value along its outgoing edge, i.e. we send a message bit of 1 to variable node v_5. Since we send the value 1, the accumulated value on u_1 is now 0. Furthermore, the edge connecting u_1 and v_5 can be removed. Upon setting $v_5 = 1$, v_5 sends its value along the remaining outgoing edges, which at this point is only one edge to u_2. The accumulated value at u_2 is updated to zero, the edge is erased, leaving u_2 with an induced degree of 1. Therefore u_2 can send its accumulated value to v_2, determining the last erased bit. The residual graph in this case is empty, and the codeword was completely recovered.

Now consider the same code and transmitted codeword, where the channel erased bits are now the third and fifth (depicted in Fig. 5.5). In this case, after setting the bit values of the variable nodes, sending them through the outgoing edges and removing these edges, it can be seen that no check node as induced degree 1 (both u_1 and u_2 have induced degree of 2). The peeling decoder thus gets stuck (it may output the partially decoded codeword) and the obtained residual graph is non-empty. In this case, it can be seen that v_2 and v_5 form a non-empty stopping set, and in particular, a MAP decoder cannot also decode this codeword, since the set of erasures includes a support set of the codeword (the sequences 1000011 and 1010111 are valid codewords of the code). It should be noted that, in the limit of an infinite number of iterations, both BP and peeling decoders have the same performance, as they get stuck in exactly the same structures.

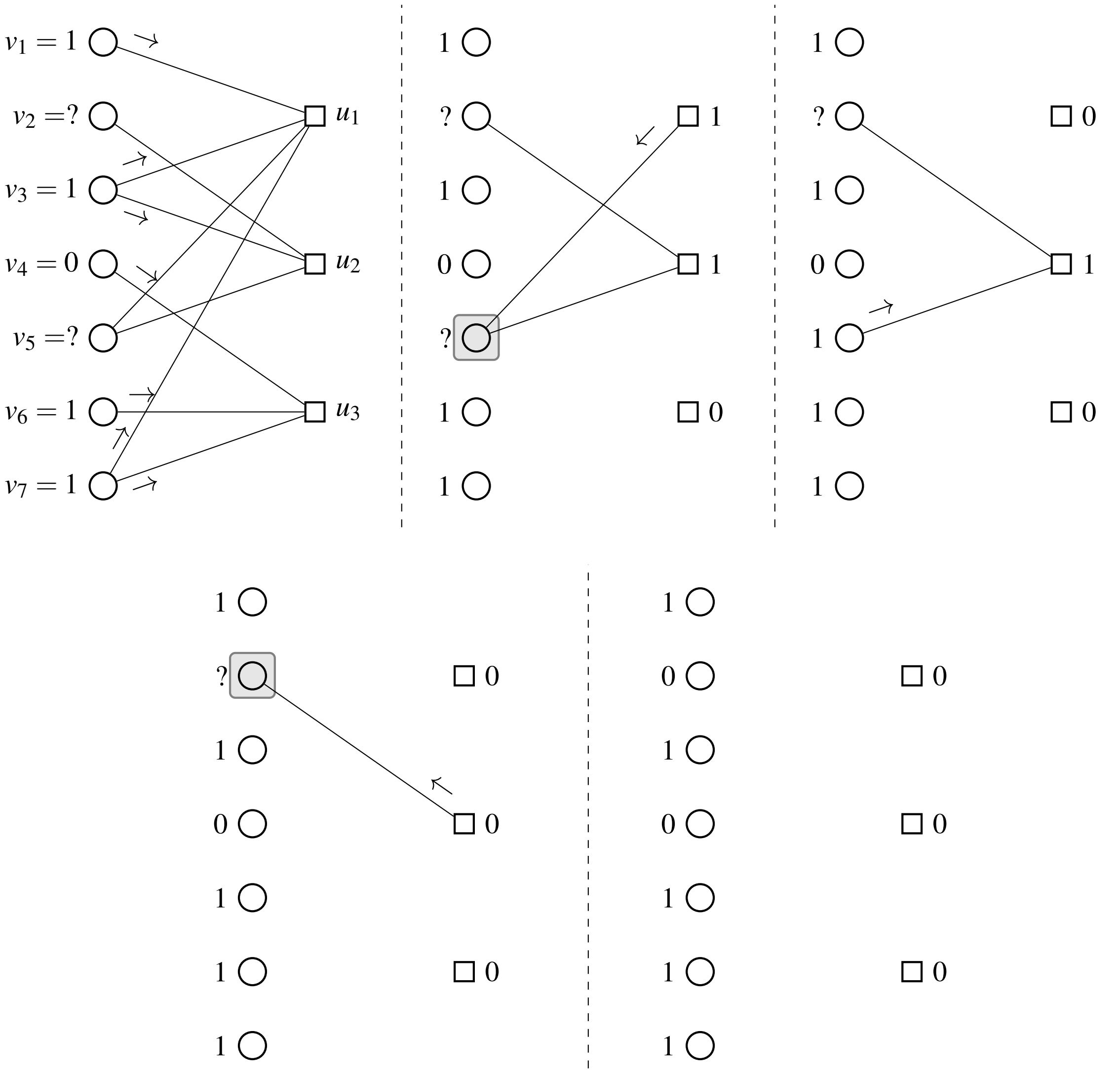


Figure 5.4: Example of a peeling decoder that is able to recover the transmitted codeword.

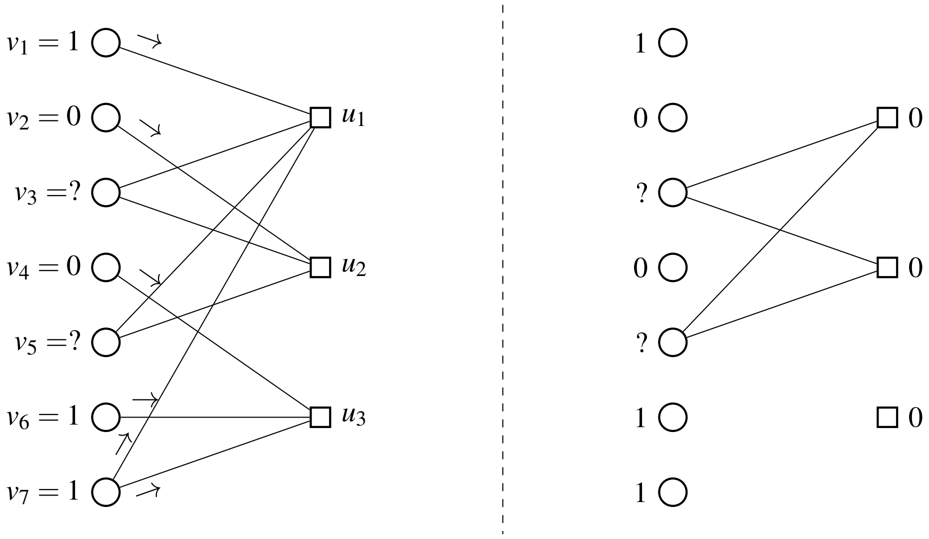


Figure 5.5: Example of a peeling decoder that is not able to recover the transmitted codeword.

The connection between the performance of MAP decoders and BP decoders arises when we consider a decoder known as *Maxwell decoder*. This decoder is a modification of a peeling decoder in the following sense: whenever the peeling decoder gets stuck in a non-empty stopping set a symbolic variable, say s_i, is chosen to represent a value of a certain unresolved variable node v_i. The decoder then proceeds as if this value is known. Thus, the messages passed are not only binary symbols but they may be equations. At some point, it is possible that such symbolic variable is connected to a check node of induced degree 1, which allows us to solve the symbolic variable s_i and therefore determine the actual bit-value of the variable node v_i. Moreover, we may possibly resolve other equations that involve s_i. If all the introduced variables are resolved, we obtain the only solution that is compatible with the codeword. On the other hand, if some variables are not resolved, we obtain a set of equations whose multiple solutions result in compatible codewords. Thus, the Maxwell decoder in fact performs MAP decoding by employing a peeling decoder with some sort of *guessing device*. Moreover, since it performs list decoding, we may also relate the number of unresolved introduced variables to the *conditional entropy* of the code. These two aspects are what enable the performance analysis of MAP decoding and their connection to BP decoding.

5.1.2 LDPC Decoding Thresholds over Binary Erasure Channels

One way to assess the performance of LDPC codes is through the notion of decoding thresholds. Although the properties behind these thresholds are different in nature, they capture the idea of characterizing the largest channel parameter that allows for reliable communication, given a certain code ensemble and decoding technique. In particular, they rely on concentration results, which essentially say that, for large enough block-lengths ($n \to \infty$) most matrices in an LDPC ensemble will exhibit the same properties.

For belief propagation decoding, the BP decoding threshold ε^{BP} is defined as the largest channel erasure probability such that the bit erasure probability goes to zero as the block-length and the number of decoding iterations goes to infinity. Formally, the BP decoding threshold can be defined as follows:

$$\lim_{t\to\infty} P_b(n,\varepsilon,t) = 0, \text{ if } \varepsilon < \varepsilon^{BP} \tag{5.3}$$

and

$$\lim_{t\to\infty} P_b(n,\varepsilon,t) > 0, \text{ if } \varepsilon > \varepsilon^{BP}. \tag{5.4}$$

Here, P_b is the bit erasure probability, n is the code block-length, t is the number of decoding iterations and ε is the channel erasure probability.

For a given LDPC ensemble characterized by the polynomial degree distributions $\lambda(x)$ and $\rho(x)$, it is possible to obtain and analyze the BP decoding threshold with density evolution techniques [62, 63]. Density evolution techniques track the evolution of the probability density functions associated with the messages passing from check nodes to variable nodes and vice-versa. For binary erasure channels, they track the probability that these messages are erasure messages. For non-punctured ensembles, the density evolution equation $F(x,\varepsilon)$ is given by [62]

$$F(x,\varepsilon) = x - \varepsilon\lambda(1-\rho(1-x)). \tag{5.5}$$

Thus, the decoding threshold ε^{BP} is the maximum value of $\varepsilon \in [0,1]$, such that $F(x,\varepsilon) = 0$ has no solution in $x \in (0,1]$. For punctured LDPC ensembles, the density evolution equation is similar. Define the polynomials $\lambda_p(x)$ and $\overline{\lambda}_p(x)$ as

$$\lambda_p(x) = \sum_{l=2}^{c} \gamma_l \lambda_l x^{l-1}$$

and

$$\overline{\lambda}_p(x) = \sum_{l=2}^{c} (1-\gamma_l)\lambda_l x^{l-1}.$$

Then, the density evolution equation is given by [63]

$$F_p(x,\varepsilon) = x - \lambda_p(1-\rho(1-x)) - \varepsilon\overline{\lambda}_p(1-\rho(1-x)). \tag{5.6}$$

Consequently, the decoding threshold for puncture codes ε_p^{BP} is the maximum value of $\varepsilon \in [0,1]$, such that $F_p(x,\varepsilon) = 0$ has no solution in $x \in (0,1]$.

For maximum a posteriori decoding, the MAP decoding threshold ε^{MAP} is defined has the largest erasure probability such that the normalized conditional entropy of the code converges to zero [64]. The general method for computing the MAP threshold of a certain ensemble uses the EXIT curves of a BP decoder. The idea is as follows. Given the BP EXIT curve take a vertical line starting at ε_p^{BP} and shift it to the right. When the area under the BP curve to the left of the line is equal to the area under the BP curve to the right of the line, the abscissa of this line marks the MAP threshold, giving rise to a generalized area theorem for erasure channels [64].

While determining the MAP threshold may be complicated for many ensembles (BP EXIT curves may have many discontinuities, for certain ensembles it is possible to obtain a straightforward computation of the MAP threshold using the notion of a peeling decoder. It can be shown that the residual graph obtained by a peeling decoder is uniformly distributed conditioned on its degree profile [65] and that its degree distribution pair is sharply concentrated around its expected value [64]. More precisely, consider an LDPC ensemble (n, λ, ρ) which is used for transmission over a BEC(ε). A peeling decoder gives rise (w.h.p) to a residual ensemble (Λ_ε, Γ_ε) with the following distribution [64]

$$\Lambda_\varepsilon(z) \triangleq \varepsilon\Lambda(zy) \tag{5.7}$$

$$\Gamma_\varepsilon(z) \triangleq \Gamma(\bar{x}+zx)-\Gamma(\bar{x})-zx\Gamma'(\bar{x}), \tag{5.8}$$

where x is the fixed point of the density evolution equation of the BP decoder, $\bar{x} \triangleq 1-x$ and $y \triangleq 1-\rho(1-x)$.

The normalized equivocation is then given by the average rate of the residual ensemble [64]. Moreover, if the design rate of the residual ensemble is equal to its average rate, the normalized equivocation can be computed from this quantity. The following concentration lemma [64] provides us a way to identify which ensembles satisfy such constraint.

Lemma 4 ([64], Lemma 7)**.** *Consider an LDPC ensemble (n,λ,ρ) with a design rate* $r \triangleq 1-\frac{\Lambda'(1)}{\Gamma'(1)}$. *Let* $\phi(x)=\log_2(1+x)$ *and consider the function* $\Psi(u)$ *defined as*

$$\Psi(u) = -\Lambda'(1)[\phi(uv)-\phi(v)]+\sum_l \Lambda_l\phi(u^l)+(1-r)\sum_l \Gamma_l\phi\left(\left(\frac{1-v}{1+v}\right)^l\right)-\Lambda(1),$$

where $v=\left(\sum_l \frac{\lambda_l}{1+u^l}\right)^{-1}\left(\sum_l \frac{\lambda_l u^{l-1}}{1+u^l}\right)$.

Let G be a code picked uniformly at random from the ensemble LDPC(n,λ,ρ) with a rate r_G*. If* $\Psi(u)$ *takes on its global maximum at* $u=1$*, for* $u\in[0,\infty)$*, then there exists* $B>0$ *such that, for any* $\xi>0$ *and* $n>n_0(\xi,\Lambda,\Gamma)$*,*

$$Pr\{|r_G-r|>\xi\}\leq e^{-Bn\xi}.$$

Thus, if the condition from Lemma 4 on $\Psi(u)$ is met for the ensemble LDPC(n,λ,ρ), with high probability, the design rate of the code will be asymptotically close to the average code rate.

The following theorem provides the basis to compute the average normalized equivocation for a randomly chosen code for an ensemble LDPC(n,λ,ρ).

Theorem 5 ([64], Theorem 10)**.** *Consider the LDPC ensemble (n,λ,ρ). Let G be a code picked at random from this ensemble, (Λ_ε,Γ_ε) be the corresponding residual ensemble with respect to the transmission over a BEC(ε) and let the conditions of Lemma 4 hold for the residual ensemble. Then,*

$$\lim_{n\to\infty} \frac{1}{n}\mathbb{E}[H_G(\mathbf{X}|\mathbf{Z})] = \Lambda'(1)x(1-y) - \frac{\Lambda'(1)}{\Gamma'(1)}[1-\Gamma(1-x)] + \varepsilon\Lambda(y),$$

where Λ and Γ are the degree distributions of the ensemble from a node perspective, x is the fixed point of the density evolution equation of the BP decoder and $y \triangleq 1-\rho(1-x)$.

Since by definition the MAP threshold is the largest channel erasure probability such that the normalized conditional entropy of the code converges to zero, Theorem 5 can be used to numerically find the code's MAP threshold.

5.1.3 LDPC Decoding over Binary Deletion Channels with Erasures

As noted in the beginning of this chapter, if the puncturing pattern is not known to the eavesdropper, the observations of the eavesdropper can be described as the outputs of a deletion channel concatenated with a binary erasure channel (recall that we are considering a binary erasure wiretap channel model). Therefore, it is useful to derive the MAP decoder associated for the eavesdropper's estimates under this setting. This MAP decoder can be described as follows.

Let $\mathbf{x} \in \mathbf{X}$ be a randomly chosen codeword to be transmitted from a uniformly distributed source. Let also $\mathbf{p} \in \mathbf{P}$ be the resulting sequence from puncturing $\mathbf{x}$ using the puncturing pattern $\mathbf{d} = (d_1,\ldots,d_n) \in \mathbf{D}$ and $\mathbf{z} \in \mathbf{Z}$ be the sequence that results from erasing the bits from $\mathbf{p}$ using the erasure pattern $\mathbf{e} = (e_1,\ldots,e_{n-n_d}) \in \mathbf{E}$, where $n_d = \sum_{i=1}^{n} d_i$. Assuming that bits are punctured and erased uniformly, we can derive the conditional probability $P(\mathbf{p}|\mathbf{x})$ as follows:

$$\begin{aligned} P(\mathbf{p}|\mathbf{x}) &= \sum_{\mathbf{d}} P(\mathbf{p}|\mathbf{x},\mathbf{d})P(\mathbf{d}|\mathbf{x}) \\ &= \sum_{\mathbf{d}} P(\mathbf{p}|\mathbf{x},\mathbf{d})P(\mathbf{d}). \end{aligned}$$

We know that

$$P(\mathbf{p}|\mathbf{x},\mathbf{d}) = \begin{cases} 1, & \text{if } \Pi_D(\mathbf{x},\mathbf{d}) = \mathbf{p} \\ 0, & \text{otherwise,} \end{cases}$$

with $\Pi_D(\mathbf{x},\mathbf{d})$ denoting the sequence obtained by puncturing $\mathbf{x}$ with the pattern $\mathbf{d}$. Thus, summing over $\mathbf{d}$, we can group all the puncturing patterns that originate the same $\mathbf{p}$. Noting that $|\mathbf{p}| = n - n_d$, we can write

$$P(\mathbf{p}|\mathbf{x}) \quad = \quad f(\mathbf{p},\mathbf{x})(1-\gamma)^{|\mathbf{p}|}\gamma^{|\mathbf{x}|-|\mathbf{p}|}, \tag{5.9}$$

where $f(\mathbf{p},\mathbf{x})$ denotes the number of times $\mathbf{p}$ appears as a subsequence of $\mathbf{x}$. In particular, since the generation of $\mathbf{p}$ by puncturing of $\mathbf{x}$ implies patterns that always have the same weight, we have that $P(\mathbf{d}) = (1-\gamma)^{|\mathbf{p}|}\gamma^{|\mathbf{x}|-|\mathbf{p}|}$ and (5.9) follows. Then, using the fact that $(\mathbf{X},\mathbf{D}) \rightarrow (\mathbf{P},\mathbf{E}) \rightarrow \mathbf{Z}$ forms a Markov chain, we can write

$$\begin{aligned} P(\mathbf{z}|\mathbf{x}) &= \sum_{\mathbf{p}} P(\mathbf{z}|\mathbf{p})P(\mathbf{p}|\mathbf{x}) \\ &= \sum_{\mathbf{p}} P(\mathbf{z}|\mathbf{p})f(\mathbf{p},\mathbf{x})(1-\gamma)^{|\mathbf{p}|}\gamma^{|\mathbf{x}|-|\mathbf{p}|} \end{aligned}$$

Now note that for a given $\mathbf{p}$, any erasure pattern generates a different sequence $\mathbf{z}$ with the same size as $\mathbf{p}$. Hence, we have that

$$P(\mathbf{z}|\mathbf{x}) = \sum_{\mathbf{p}:|\mathbf{p}|=|\mathbf{z}|} f(\mathbf{p},\mathbf{x})(1-\gamma)^{|\mathbf{p}|}\gamma^{|\mathbf{x}|-|\mathbf{p}|}(1-\varepsilon)^{|\mathbf{p}|-n_e}\varepsilon^{n_e}, \tag{5.10}$$

where n_e is the number of erasures in $\mathbf{z}$. Thus, $P(\mathbf{z}|\mathbf{x})$ can be found by computing the number of ways a subsequence $\mathbf{p}$ can be generated from $\mathbf{x}$ and summing over all the subsequences that are compatible with $\mathbf{z}$ through erasures.

Now, let $\mathbf{x}$ be a randomly chosen codeword to be transmitted from a uniformly distributed source and let $\mathbf{z}$ be the sequence observed by the eavesdropper after puncturing and channel erasures. The MAP estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ is given by

$$\begin{aligned} \hat{\mathbf{x}} &= \operatorname*{argmax}_{\mathbf{x}} P(\mathbf{x}|\mathbf{z}) \\ &= \operatorname*{argmax}_{\mathbf{x}} P(\mathbf{z}|\mathbf{x})\frac{P(\mathbf{x})}{P(\mathbf{z})} \\ &= \operatorname*{argmax}_{\mathbf{x}} P(\mathbf{z}|\mathbf{x}), \end{aligned}$$

where $P(\mathbf{z}|\mathbf{x})$ can be computed from (5.10). signifies There are a couple of aspects to retain from the above derivation. First, the conditional probability $P(\mathbf{z}|\mathbf{x})$ depends on the number of times an erased subsequence $\mathbf{z}$ is compatible with a given codeword $\mathbf{x}$. Unfortunately, there are no known bounds on such distribution for arbitrary lengths of $\mathbf{x}$.

Thus, in principle, these have to be computed on a code basis. Hence, for the case of secret puncturing patterns, we may use a very efficient decoder at the legitimate receiver (e.g. BP decoding), while the eavesdropper's optimal decoder will be very inefficient (counting the number of sub-sequences can be solved with polynomial complexity [66]; however this problem must be solved for every possible codeword). On the other hand, since the computation of equivocation of the eavesdropper requires this conditional probability, the same problem will be present. Consequently, an exact equivocation analysis can only be done for small block-lengths.

5.2 Puncturing for Secrecy over the BEWC

Consider the wiretap models depicted Fig. 5.2. The transmitter (Alice) wishes to send a message $\mathbf{M} \in \{0,1\}^k$ to the legitimate receiver (Bob), while preventing an eavesdropper (Eve) from obtaining a correct copy of that message. To achieve this, Alice encodes $\mathbf{M}$ into the codeword $\mathbf{X} \in \{0,1\}^n$ using a code from an ensemble LDPC(n,λ,ρ). The outputs of the LDPC encoder are further punctured according to the distribution $\gamma(x) = \sum_{l=2}^{c} \gamma x^{l-1}$, where γ represents the probability of a variable node being punctured (irrespective of its degree). Let $\mathbf{D}$ be a random variable that represents the puncturing pattern, such that $\mathbf{D} \in \{0,1\}^n$, where a 0 in the i-th entry of vector $\mathbf{D}$ means that the i-th message bit remains unpunctured, whereas a 1 determines that the i-th message bit is punctured. Then, the channel input $\mathbf{P}$ is given by taking the values of $\mathbf{X}$ that are indexed by 0-entries in $\mathbf{D}$. Upon transmission of the punctured message, Bob and Eve observe (noisy) copies of $\mathbf{P}$, respectively through the main channel Q_m and the wiretap channel Q_w. Both channels are assumed to be binary erasure channels with erasure probabilities δ for Q_m and ε for Q_w. Bob receives $\mathbf{Y} \in \{0,1,?\}^{n-n_d}$, where ”?” represents an erasure and n_d is the number of punctured bits, and makes an estimate $\tilde{\mathbf{M}}$ of the source message. Eve obtains $\mathbf{Z} \in \{0,1,?\}^{n-n_d}$, which she also uses to make an estimate $\hat{\mathbf{M}}$ of the source message.

By definition, this coding scheme is deterministic (and bijective). We are interested in understanding how using this simple encoding procedure can be enough to ensure reliable and secure communication and compare it with a more sophisticated approach, such as using nested codes. In this context, we say that reliable communication is possible if the probability of error is vanishing. On the other hand, we will measure secrecy through the code's equivocation. Note that we do not set (a priori) a particular secrecy constraint. The reason is that we are interested in measuring the secrecy associated with the coding scheme, rather than setting up an initial goal such as achieving weak secrecy capacity. That being said, our objective is to maximize the equivocation experienced by the eavesdropper (similar to a *best effort* approach to secrecy).

With respect to the nature of the puncturing pattern (public or secret), there is a clear impact in terms of the secrecy analysis, as it is affected by the side information possessed by the eavesdropper. On the other hand, the reliability analysis is essentially the same, since in both cases the puncturing pattern is known by the legitimate party.

5.2.1 Reliability

The coding scheme under consideration has a single parameter that can be adjusted, namely the puncturing probability. It is possible to obtain very simple bounds on the maximum puncturing probability, for a given code ensemble, such that it allows vanishing error probability is obtained. Moreover, for the case of LDPC codes, it is also possible to connect this puncturing probability to the type of decoder one wishes to use at the legitimate receiver through the decoding thresholds.

In general, let ε^* denote the decoding threshold associated with an ensemble LDPC(n,λ,ρ). Thus, with high probability, $\lim_{n\to\infty} P_e(\mathcal{C}_n) = 0$, if $\delta < \varepsilon^*$, where $\mathcal{C}_n$ is an instance of the LDPC ensemble.

When modelling random puncturing as an erasure channel, our coding scheme can be seen as using the original LDPC ensemble to transmit over a binary erasure channel charaterized by an erasure probability of $\delta' = \gamma + (1-\gamma)\varepsilon$. Since reliable communication is only possible if $\delta' < \varepsilon^*$, we immediately obtain that the puncturing probability is bounded by

$$\gamma \leq \frac{\varepsilon^* - \delta}{1-\delta}. \tag{5.11}$$

In particular, for BP and MAP decoding, the thresholds can be computed according to the descriptions given in Section 5.1.2. Clearly, with increasing decoding thresholds, larger admissible puncturing probabilities are obtained. Thus, the choice of a particular decoding strategy bears an impact in terms of secrecy, since higher puncturing probabilities imply a larger equivocation with respect to the eavesdropper. This introduces a trade-off between decoding complexity and secrecy, which may be useful when designing a particular system.

5.2.2 Secrecy

The secrecy performance of the proposed coding scheme will be measured in terms of the equivocation of the eavesdropper's observations. For the considered model, it is given by the following lemma.

Lemma 5. *Let* **M**, **X**, **Z**, **D** *be random variables that represent respectively, the source message, the unpunctured encoded message, the eavesdropper's observation and the* *puncturing pattern. Then,* $H(\mathbf{M}|\mathbf{Z}) = H(\mathbf{M}|\mathbf{X},\mathbf{Z}) + H(\mathbf{X}|\mathbf{Z},\mathbf{D}) + H(\mathbf{D}|\mathbf{Z}) - H(\mathbf{D}|\mathbf{X},\mathbf{Z}) - H(\mathbf{X}|\mathbf{M},\mathbf{Z})$.

Proof. The proof is obtained by multiple applications of the chain rule of entropy.

$$\begin{aligned}
H(\mathbf{M}|\mathbf{Z}) &= H(\mathbf{M},\mathbf{Z}) - H(\mathbf{Z}) \\
&= H(\mathbf{M},\mathbf{X},\mathbf{Z}) - H(\mathbf{X}|\mathbf{M},\mathbf{Z}) - H(\mathbf{Z}) \\
&= H(\mathbf{M}|\mathbf{X},\mathbf{Z}) + H(\mathbf{X},\mathbf{Z}) - H(\mathbf{X}|\mathbf{M},\mathbf{Z}) - H(\mathbf{Z}) \\
&= H(\mathbf{M}|\mathbf{X},\mathbf{Z}) + H(\mathbf{X}|\mathbf{Z}) - H(\mathbf{X}|\mathbf{M},\mathbf{Z}) \\
&= H(\mathbf{M}|\mathbf{X},\mathbf{Z}) + H(\mathbf{X},\mathbf{D}|\mathbf{Z}) - H(\mathbf{D}|\mathbf{X},\mathbf{Z}) - H(\mathbf{X}|\mathbf{M},\mathbf{Z}) \\
&= H(\mathbf{M}|\mathbf{X},\mathbf{Z}) + H(\mathbf{X}|\mathbf{Z},\mathbf{D}) + H(\mathbf{D}|\mathbf{Z}) - H(\mathbf{D}|\mathbf{X},\mathbf{Z}) - H(\mathbf{X}|\mathbf{M},\mathbf{Z}) \\
&= H(\mathbf{M}|\mathbf{X},\mathbf{Z}) + H(\mathbf{X}|\mathbf{Z},\mathbf{D}) + I(\mathbf{X};\mathbf{D}|\mathbf{Z}) - H(\mathbf{X}|\mathbf{M},\mathbf{Z})
\end{aligned}$$

□

The next two corollaries specify the eavesdropper's equivocation to the case of a public or secret puncturing pattern using the proposed coding scheme, which consists of deterministic and bijective encoder.

Corollary 3. *Let there be a* one-to-one mapping *between* $\mathbf{M}$ *and* $\mathbf{X}$ *(and vice-versa). If* $\mathbf{D}$ *is publicly known, then* $H(\mathbf{M}|\mathbf{Z}) = H(\mathbf{X}|\mathbf{Z},\mathbf{D})$.

Corollary 4. *Let there be a* one-to-one mapping *between* $\mathbf{M}$ *and* $\mathbf{X}$ *(and vice-versa). If* $\mathbf{D}$ *is only known by the legitimate receiver, then* $H(\mathbf{M}|\mathbf{Z}) = H(\mathbf{X}|\mathbf{Z},\mathbf{D}) + H(\mathbf{D}|\mathbf{Z}) - H(\mathbf{D}|\mathbf{X},\mathbf{Z})$.

Corollary 3 states that if the puncturing pattern is public, the eavesdropper's equivocation is the equivocation associated with the transmission of codewords over a binary erasure channel with erasure probability $\gamma + (1-\gamma)\varepsilon$. On the other hand, Corollary 4 states that hiding the puncture pattern from the eavesdropper results in added equivocation since $H(\mathbf{D}|\mathbf{Z}) - H(\mathbf{D}|\mathbf{X},\mathbf{Z}) = I(\mathbf{D};\mathbf{X}|\mathbf{Z})$, which is always a non-negative quantity. In particular, this term is associated with the loss of bit-level synchronization at the eavesdropper's decoder. Both are a consequence of having $H(\mathbf{M}|\mathbf{X},\mathbf{Z}) = H(\mathbf{X}|\mathbf{M},\mathbf{Z}) = 0$, while for Corollary 3 it is further implied that $H(\mathbf{D}|\mathbf{Z})$ and $H(\mathbf{D}|\mathbf{X},\mathbf{Z})$ are zero, as $\mathbf{D}$ is known.

As noted before, computing the equivocation of certain ensemble can be done straightforwardly, provided that the ensemble obeys certain criteria. On the other hand, computing the equivocation of a code where $\mathbf{D}$ is unknown, requires the computation of conditional probabilities over that may take an exponential time to compute. This means that an exact analysis of the equivocation in this case can only be done for small block-lengths. Alternatively, it is possible to bound the equivocation of the eavesdropper by considering the connection between the MAP decoder error probability and the respective conditional entropy. While the MAP decoder itself has exponential complexity, it may be reduced with respect to the complexity of computing the equivocation. The reason is that the conditional probabilities to be computed in the MAP decoder only need to be computed over sequences of a certain length, meaning that a certain observation defines the weight of the puncturing patterns. This means that one compute the conditional probabilities with respect to these patterns, thus reducing the amount of computations that need to be performed.

In [25], the authors provide expressions for bounding the equivocation using the MAP error probability. Let P_e be the expected MAP error probability and consider the two following functions $\Phi(P_e)$ and $\Phi^*(P_e)$ as given in [25]

$$\Phi(P_e) = (1-P_e) i \log_2 i + h(iP_e - (i-1)), \tag{5.12}$$

and

$$\Phi^*(P_e) = a_i \left(P_e - \frac{i-1}{i} \right) + b_i, \tag{5.13}$$

with $\frac{i-1}{i} \leq P_e \leq \frac{i}{i+1}$, $i = 1, \dots, M-1$ and where $h(\cdot)$ is the binary entropy function, M is the alphabet size, $a_i = i(i+1)\log_2((i+1)/i)$ and $b_i = \log_2(i)$. Then, the equivocation of the eavesdropper is bounded according to the following:

Theorem 6 ([25], Theorem 1)**.** *Let $P_e(\mathbf{X}|\mathbf{Z})$ denote the MAP error probability and $H(\mathbf{X}|\mathbf{Z})$ denote the equivocation. Then,*

$$\Phi(P_e(\mathbf{X}|\mathbf{Z})) \geq H(\mathbf{X}|\mathbf{Z}) \geq \Phi^*(P_e(\mathbf{X}|\mathbf{Z})) \tag{5.14}$$

5.2.3 Rate-Equivocation Regions

While there are several ways in which the puncturing pattern can be shared among the legitimate parties (discussed in more detail in Section 5.4), it is possible to do it purely from an information-theoretic point of view. From this perspective, we need to derive the rate-equivocation regions, as our wiretap model requires some side information. This side information can be provided either by an external source[3] (genie-aided) or can be generated and transmitted by the legitimate receiver. The disclosure strategy and nature of the puncturing pattern may define a variant of the wiretap model, which may result in a modified rate-equivocation region. For instance, if we assume that the puncturing pattern is obtained via a genie, a model which considers a public puncturing pattern is simply the

[3] For instance using information-theoretic secret-key agreement schemes [3, Chapter 4].

wiretap model we have considered so far and introduced in Section 2.1. On the other hand, if the puncturing pattern is a shared secret we are dealing with a wiretap with a shared key [67]. In the case the puncturing pattern is to be transmitted, considering a puncturing pattern public can be equivalent to have a broadcast wiretap channel with common and confidential messages (BCC) [14], where the common rate is the rate allocated to the transmission of the puncturing pattern. This would be a worst case scenario, since in practice we do not need to require that the eavesdropper decodes de puncturing pattern, and consequently the rate-equivocation region can be further extended. In the event that we assume that the puncturing pattern is to be transmitted and kept secret, we have again a simple wiretap channel where now the secret rate has to be split to account for the messages and patterns.

Table 5.1: Equivalence of rate-equivocation regions as a function of the disclosure strategy and nature of the puncturing pattern.

	public **D**	secret **D**
genie-aided **D**	WTC	WTC-SK
transmitted **D**	BCC	WTC

In the following, we will consider that the puncturing pattern is obtained via a genie. Therefore, the two rate-equivocation regions of interest are the rate-equivocation regions for the wiretap channel and the wiretap channel with a shared key (a proof is provided in Appendix C).

Theorem 7. *([3, Corollary 3.3]) Consider a wiretap channel ($\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, $p_{YZ|X}(y,z|x)$). For any joint distribution p_{UVX} on $\mathcal{U}\times\mathcal{V}\times\mathcal{X}$ that factorizes as $p_U p_{V|U} p_{X|V}$, the weak rate-equivocation region for this wiretap channel is the convex set*

$$R^{WT} = \bigcup_{p_{UVX}} R^{WT}(p_{UVX}), \tag{5.15}$$

where

$$R^{WT}(p_{UVX}) = \left\{ (R,R_e) : \begin{array}{c} 0 \le R_e \le R \le \frac{1}{n} I(V^n;Y^n) \\ 0 \le R_e \le \frac{1}{n}\left[I(V^n;Y^n|U^n) - I(V^n;Z^n|U^n) \right] \end{array} \right\}.$$

Theorem 8. *Consider a wiretap channel ($\mathcal{K}$, $\mathcal{X}$, $\mathcal{Y}$, $\mathcal{Z}$, $p_{YZ|X}(y,z|x)$), where $\mathcal{K}$ is the key alphabet. Moreover, assume that the key has a fixed rate R_k. For any joint distribution p_{UVX} on $\mathcal{U}\times\mathcal{V}\times\mathcal{X}$ that factorizes as $p_U p_{V|U} p_{X|V}$, the weak rate-equivocation region for* *this wiretap channel is the convex set*

$$R^{SK} = \bigcup_{p_{UVX}} R^{SK}(p_{UVX}), \tag{5.16}$$

where

$$R^{SK}(p_{UVX}) = \left\{ (R,R_e) : \begin{array}{c} 0 \leq R_e \leq R \leq \frac{1}{n} I(V^n;Y^n) \\ 0 \leq R_e \leq \frac{1}{n}\Big[I(V^n;Y^n|U^n) - I(V^n;Z^n|U^n) \Big] + R_k \end{array} \right\}.$$

For noisier wiretap channel models the above regions can be simplified by taking $V^n = X^n$ and letting U^n to be independent of (V^n,X^n,Y^n,Z^n). It is not surprising that a shared secret extends the rate equivocation region (in the sense that the maximum equivocation of the eavesdropper saturates at a higher value), since the key-rate is given for free in this case. However, it is not clear whether there is an advantage in terms of simplifying the coding process, since coding using a secret key may be an easier task than its keyless counterpart. Note that secret-key agreement is generally considered to be easier than coding for secrecy, and thus there is no added complexity in obtaining this model to start a priori.

5.3 Numerical Results

We have seen in Section 5.2.2 that the eavesdropper's equivocation depends simply on the code's performance over the binary erasure channel when the puncturing pattern is public. When instead we consider a secret puncturing pattern, the eavesdropper's equivocation is further affected by the shared information between the codewords and the puncturing pattern, given the eavesdropper's observation. Given that this is the case, a simple strategy to increase the eavesdropper's equivocation is to consider codes that allow for a large puncturing probability, which corresponds to the creation of a wiretap channel with a high erasure probability. Note that the manageable puncturing probability is associated with the decoder being used. Let us first consider the secrecy performance of the proposed coding scheme with public puncturing patterns and then move to secret puncturing patterns.

5.3.1 Asymptotic Performance with Public Puncturing Patterns

Consider the ensembles defined by the degree distribution pairs in Table 5.2, which is comprised of three irregular codes. Note that all ensembles have small rates. This is due to the fact that we wish to have a large puncturing probability and, therefore, enough redundancy must be added in order to account for reliability. The BP thresholds ε^{BP} and MAP thresholds ε^{MAP} are also listed in the same table. In particular, the BP thresholds of codes C_2 and C_3 are very similar, while the MAP thresholds of C_1 and C_3 are also very similar.

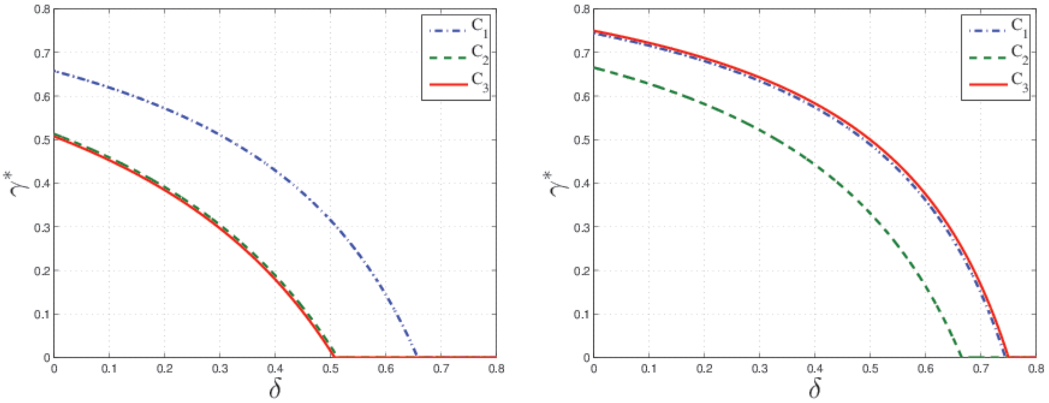
Table 5.2: Degree distributions of LDPC code ensembles C_1, C_2 and C_3.

	C_1	C_2	C_3
λ_2	0.057143	-	-
λ_3	0.942857	0.06383	-
λ_4	-	0.93617	0.067797
λ_5	-	-	0.932203
ρ_3	0.085714	-	-
ρ_4	0.914286	-	-
ρ_5	-	0.106383	-
ρ_6	-	0.893617	0.40678
ρ_7	-	-	0.59322
R	0.25	0.3333	0.25
ε^{BP}	0.6576	0.5129	0.5075
ε^{MAP}	0.7444	0.6654	0.7499

To understand how the puncturing probability affects the reliability limits of our code, we will first fix the main channel crossover probability δ. From (5.11), we may obtain the largest admissible puncturing probability as a function of the decoding threshold. Fig. 5.6 plots the largest puncturing probability γ^* as a function of the main channel erasure probability δ. There exists a symmetry between the admissible puncturing probabilities and the channel parameter. Obviously, almost noiseless channels allow for very large puncturing probabilities.

Having found the admissible puncturing probabilities, we may turn our attention to the equivocation rate experienced by the eavesdropper. In particular, the ensemble average equivocation rate is given by Theorem 5 whenever the required conditions hold, (which is the case for the considered codes) and can be computed through (5.9) for channel parameters above the MAP threshold (by definition the equivocation evaluates to zero bellow the MAP threshold).

Figures 5.7 and 5.8 plot the normalized equivocation, as a function of the eavesdropper's channel erasure probability ε, for a noisy main channel with erasure probability $\delta = 0.25$, for the respective puncturing probability γ^*. Fig. 5.7 illustrates that, puncturing up to the BP threshold limit, leads to a constant gap from the maximum achievable equivocation (solid black line) that is a function of the MAP threshold. Thus, if an LDPC



(a) Belief propagation decoding.

(b) Maximum a posteriori decoding.

Figure 5.6: Maximum puncturing probability γ^* as a function of the channel erasure probability δ for the ensembles C_1, C_2 and C_3 when using a BP and a MAP decoder.

code presents a reasonable gap between the BP and MAP thresholds, using puncturing with a BP decoding scheme leads to non-negligible leakage to the eavesdropper. On the other hand, if the legitimate receiver is allowed to use a MAP decoder as in Fig. 5.8, the eavesdropper will have near maximum equivocation (the small gap is a function of the gap between the MAP threshold and the Shannon limit [62]). This is a consequence of having artificially saturated the eavesdropper's channel with erasures until decoding is no longer possible.

It is interesting to see that, unlike nested codes that require the design of codes that are capacity achieving for the eavesdroppers channel, puncturing can have a similar effect simply by artificially creating a noisier channel to the eavesdropper. However, like nested codes, this requires the eavesdropper to have a channel that is *worse* than the main channel to allow for reliable communication and forces the legitimate receiver to use a maximum-likelihood decoder. Consequently, in order to obtain secure communication with random public puncturing patterns, these two constraints have to be taken into account.

Another interesting fact is that no care in the code design itself was needed, as long as the puncturing probability approaches the limit imposed by the MAP threshold. Consequently, there is no need to know the statistics of the eavesdropper's channel in order to maximize its equivocation.

Finally, note that, due to the linearity of the equivocation, if the BP and MAP thresholds of a certain code are very close, the MAP decoder can be replaced by a BP decoder and obtain a similar secrecy performance, enabling more efficient decoding to take place.

Let us now consider how puncturing strategies can be placed within the rate-equivocation region. For simplicity, let us assume that the channel is degraded and the input distribution

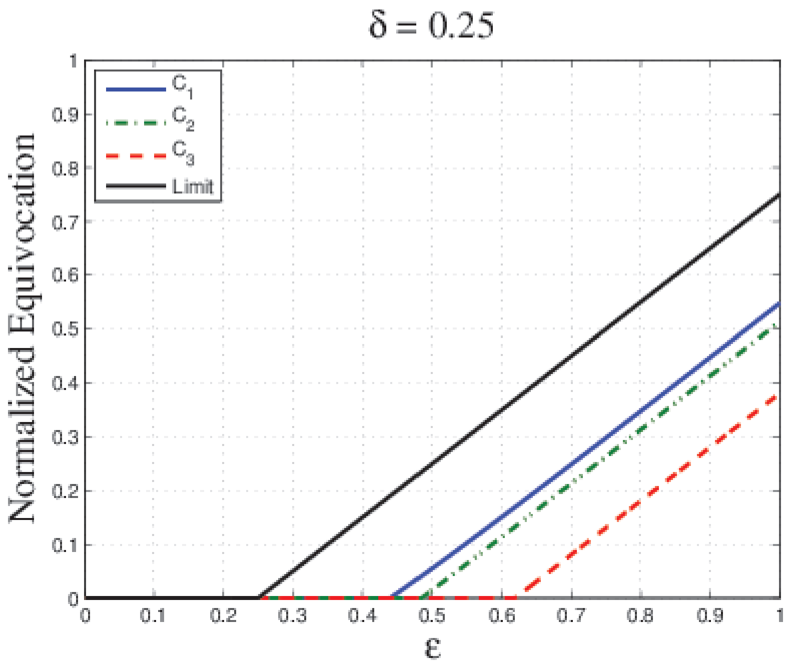


Figure 5.7: Normalized equivocation rate for a publicly known puncturing pattern as a function of the wiretap channel erasure probability ε using a BP decoder.

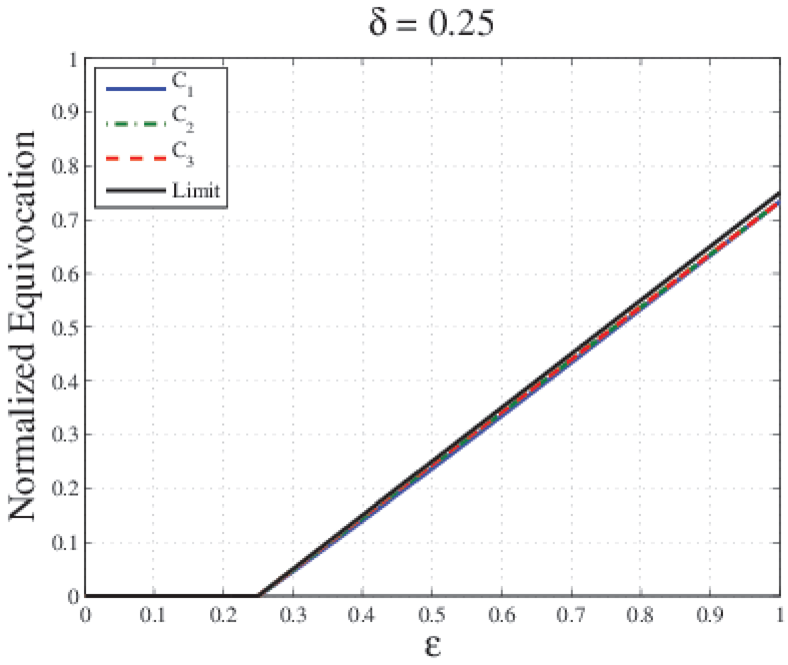


Figure 5.8: Normalized equivocation rate for of a publicly known puncturing pattern as a function of the wiretap channel erasure probability ε using a MAP decoder.

is fixed. Thus, the rate equivocation region in 7 simplifies to

$$R^{WT} = \left\{ (R, R_e) : \begin{matrix} 0 \leq R_e \leq R \leq 1 - \delta \\ 0 \leq R_e \leq \varepsilon - \delta \end{matrix} \right\}.$$

Fig. 5.9 illustrates, for the case of $\delta = 0.25$, the rate-equivocation region (solid line) and the achievable rate-equivocation pairs for the considered ensembles (markers). The achievable (R,R_e) points are normalized to account for the puncturing probability, i.e., R is the rate of the punctured code and R_e is the equivocation normalized by the number of unpunctured symbols. The plots shows that punctured LDPC codes, where the puncturing probability is the largest admissible for a MAP decoder, operate near maximum equivocation at the eavesdropper. Naturally, the punctured code has also near maximum rate. Consequently, the code operates far from secrecy capacity. However, the code rate can be artificially reduced by transmitting dummy bits. On the other hand, puncturing up to the BP threshold leads in general to considerable leakage rates. For instance, it can be seen that all codes provide no secrecy at all if $\varepsilon = 0.30$ (Fig. 5.9a), while little secrecy is provided even when the channel advantage is large (Fig. 5.9b and Fig. 5.9b).

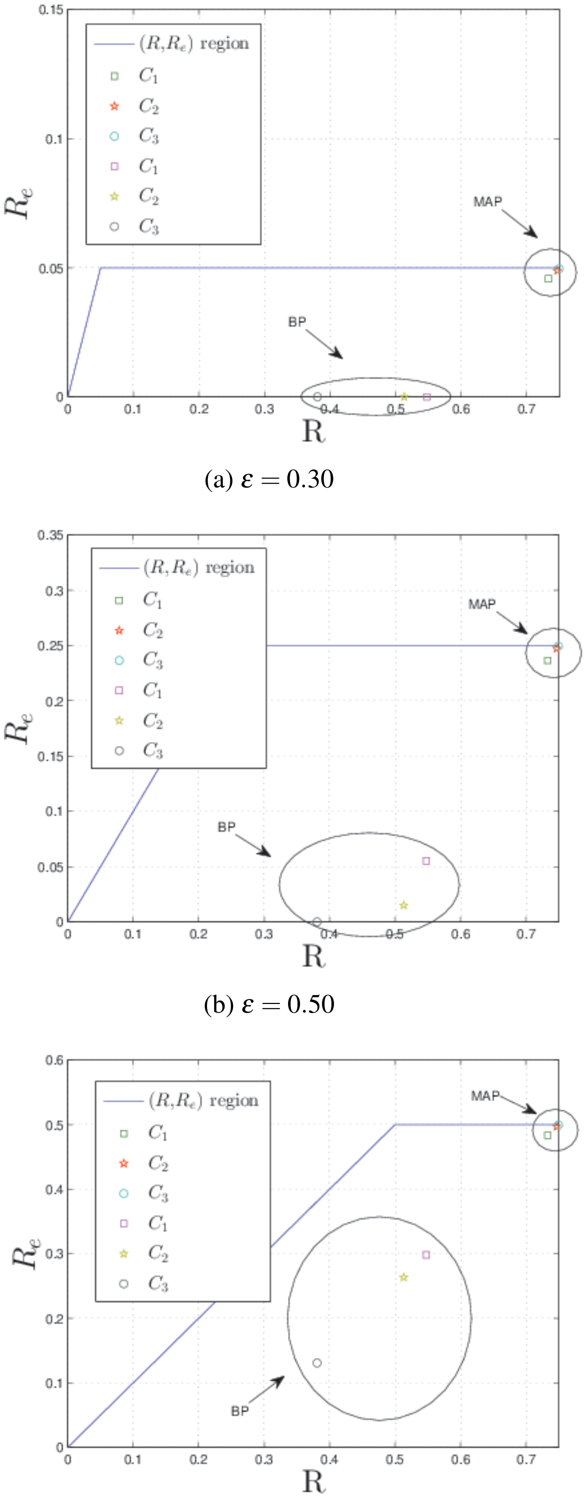
5.3.2 Finite-Length Performance with Secret Puncturing Patterns

While an asymptotic analysis is possible when the puncturing pattern is public (by virtue of being able to model the system using only erasure channels), when the puncturing pattern is secret, the assessment of the eavesdroppers equivocation is a cumbersome process, since it depends on the realization of the codebook (in particular of the sub-sequences of codewords).

In this section we sample the ensembles defined in Table 5.2 to obtain codes with a block-length $n = 12$. Since at such short block-lengths it is not possible to obtain a vanishing error probability, we resort to the analysis of the exact code equivocation to find the admissible puncturing probabilities. Fig. 5.10 provides a comparison between the equivocation of the considered codes when $n = 12$ and for the asymptotic case. As expected, there is a large gap between the values of δ for which the legitimate receiver can communicate with near zero equivocation. Consequently, a reduction on the puncturing probability is necessary. To allow for a reasonable puncturing probability, we allow the legitimate receiver to incur in a small equivocation. This is represented in Fig. 5.10 by the dashed horizontal lines, where we mark the points at which the legitimate receiver makes estimates with an equivocation of a prescribed level.

Table 5.3: Puncturing probabilities of LDPC code ensembles C_1, C_2 and C_3 with prescribed level of equivocation.

	C_1	C_2	C_3
ε^{BP}	0.6576	0.5129	0.5075
ε^{MAP}	0.7444	0.6654	0.7499
γ_1	0.3510	0.2645	0.3680
γ_2	0.4965	0.4085	0.5150



(a) $\varepsilon = 0.30$

(b) $\varepsilon = 0.50$

(c) $\varepsilon = 0.75$

Figure 5.9: Rate-equivocation region and achievable rate-equivocation pairs for the ensembles C_1, C_2 and C_3, when using the largest admissible puncturing probabilities for varying values of ε.

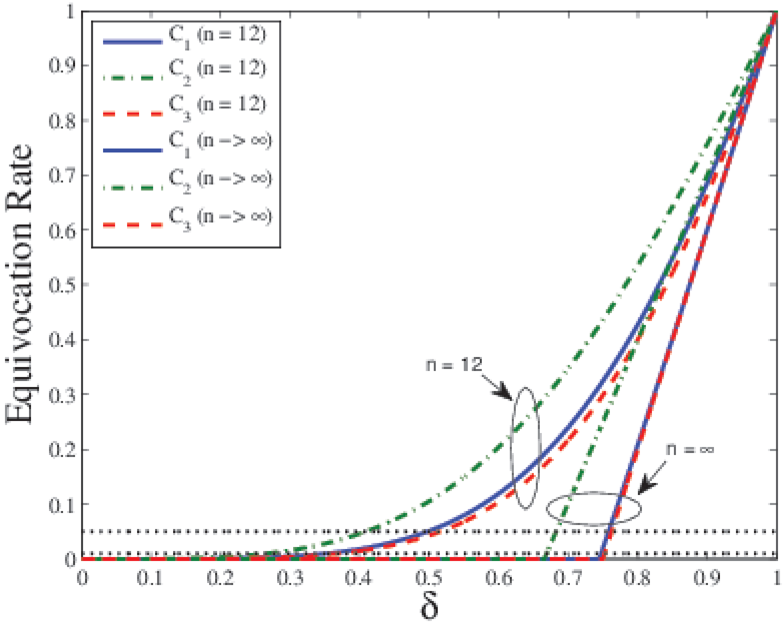


Figure 5.10: Equivocation rate for the legitimate receiver for codes C_1, C_2 and C_3 with block-length $n = 12$ as a function of the main channel erasure probability δ.

Table 5.3 summarizes the admissible puncturing probabilities γ_1 and γ_2 such that the allowed equivocation rate of the legitimate receiver 0.01 and 0.05, respectively. While greatly reduced with comparison to the puncturing probabilities given by the BP and MAP thresholds, these still allow for a relatively large puncturing probability.

In Fig. 5.11 we plot the eavesdropper's normalized equivocation rate for the codes C_1, C_2 and C_3 with block-length $n = 12$, assuming a noiseless main channel. The considered puncturing probabilities γ are ε^{MAP}, γ_1 and γ_2 listed in Table 5.3. With respect to the equivocation rate obtained by schemes where the puncturing pattern is publicly known, we obtain much higher values for the eavesdropper's equivocation overall, even when the eavesdropper's channel is noiseless. This is mostly due to the inability of the eavesdropper to correctly distinguish sub-sequences due to the large values of puncturing probabilities.

It should be noted that the case of $\gamma = \varepsilon^{MAP}$ is merely representative of the performance of a code with very short block-length that is highly punctured. In practice, such code would have a very poor performance in terms of reliability.

While such finite block-length interpretations do not carry to the asymptotic domain, it certainly motivates the used of the puncturing pattern as a shared secret since, even for very modest block-lengths, high equivocation rates are achieved. It is of note that the scheme is intended to use very large block-lengths which allow us to approximate the puncturing limits given by the corresponding thresholds. Unfortunately, for such large block-lengths an exact equivocation analysis is intractable. Hence, the limitations of the

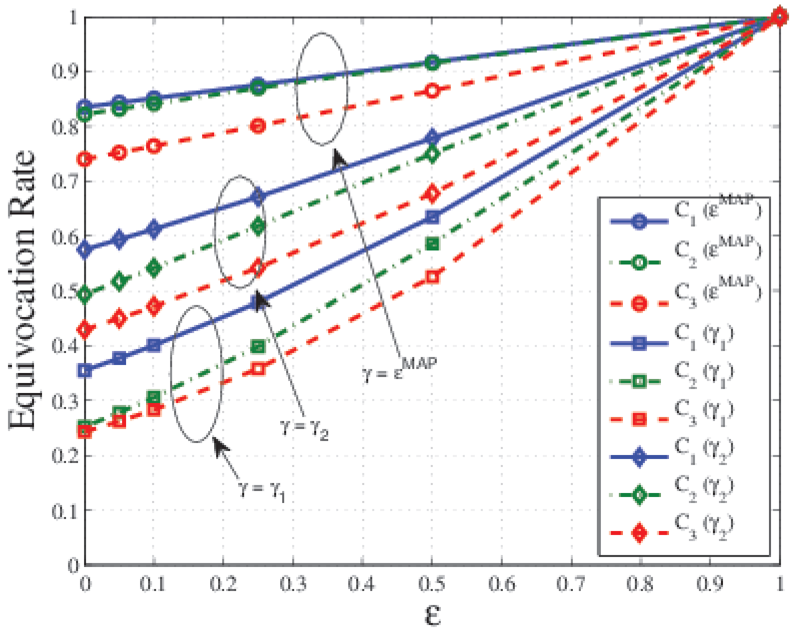


Figure 5.11: Normalized equivocation rate for the eavesdropper for codes C_1, C_2 and C_3 with block-length $n = 12$, as a function of the wiretap channel erasure probability ε. The considered puncturing probabilities γ are equal to ε^{MAP}, γ_1 and γ_2.

analysis are due to fact that we cannot compute the eavesdropper's equivocation rather than on establishing reliable communication.

Let us put the achievable rate-equivocation pairs in perspective. Consider the above mentioned ensembles C_1, C_2 and C_3. Let the main channel be a noiseless channel and the eavesdropper's channel have an erasure probability $\varepsilon = 0.25$. Let us further assume that we have a key of rate $R_k \geq 0.75$ available, in the scenario where the puncturing patterns are secret.

Fig. 5.12 illustrates the rate-equivocation regions of both models, where the model without a shared key is represented by a dashed blue line and the model with a shared key is represented by the solid black line. For the randomly drawn codes of block-length $n = 12$, consider a puncturing probability $\gamma = \gamma_1$. It is possible to see that the achievable equivocation rates are already approaching the rate-equivocation region of the general wiretap channel without a shared key (the three leftmost points of the plot). This should be interpreted within the respective context, since we are allowing the legitimate receiver to have a small equivocation (as opposed to zero equivocation) and we have a shared key of a large rate. Nevertheless, it is surprising that, for such small block-lengths, one can almost reach the rate-equivocation region of the regular wiretap channel. For illustrative purposes, we also plot the case where we have a block-length of $n = 12$ and puncture up to the MAP threshold (which implies a large equivocation on the legitimate receiver side). From the plot we can observe that puncturing up to the MAP thresholds pushes the eavesdropper's equivocation to near maximum equivocation. This suggests that for increasing block-lengths, one can rapidly approach the secrecy capacity of this channel model. As a further comment, it should be noted that, even considering modest block-lengths, punctured codes with a secret puncturing pattern have a performance that is not very far from its theoretical limit, considering the code construction in question is using deterministic and bijective encoders.

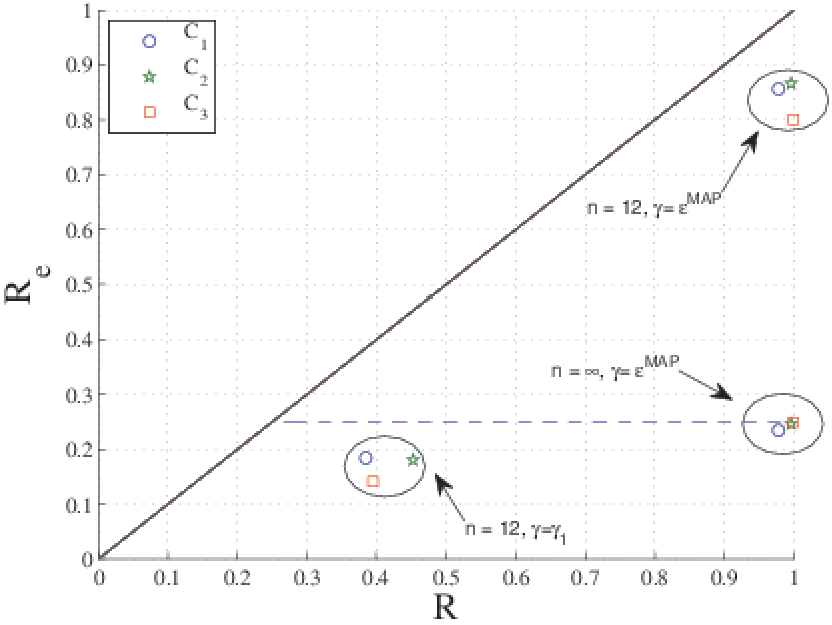


Figure 5.12: Rate-equivocation regions of the considered models and the rate-equivocation pairs for codes C_1, C_2 and C_3 for both the asymptotic and finite block-length case. Wiretap model parameters are $\delta = 0$ and $\varepsilon = 0.25$.

As noted before, it is possible to manage slightly larger block-lengths by focusing on the eavesdroppers MAP decoder. To this purpose, we recur to simulation and compute the average MAP error probability, using Theorem 6 to bound the equivocation obtained by the simulated average error probability. The employed codes have block-length $n = 20$ and their degree distribution are defined in Table 5.4, along with the BP threshold. The considered puncturing probability is bounded by the BP decoding threshold and the transmission of 10000 codewords is considered. We considered both noiseless and noisy main channels, with the main channel erasure probabilities $\delta \in \{0, 0.1, 0.25\}$ and varying wiretap channel erasure probabilities ε. Note that code C_4 is a regular code, while code C_5 is an irregular code. Moreover, the puncturing probability is the maximum allowed by BP decoding.

Figs. 5.13 and 5.14 show the average bit error probability experienced by the eavesdropper. For this particular case, it can be seen that the irregular code is able to constantly achieve a high bit error rate, even when the channel to the eavesdropper is noiseless. On

Table 5.4: Degree distributions for the LDPC code ensembles C_4, and C_5.

	C_4	C_5
λ_2	-	0.25105
λ_3	1	0.30938
λ_4	-	0.00104
λ_{10}	-	0.43853
ρ_6	1	-
ρ_7	-	0.63676
ρ_8	-	0.36324
ε^{BP}	0.4294	0.4701

the other hand, the regular code performs well, but is more affected when the main channel error probability increases, i.e. when the puncturing probability needs to be reduced.

The simulated equivocation bounds are plotted in Figs. 5.15 and 5.16, the regular and irregular code, respectively. In each figure, the dashed lines represent the lower bonds on the equivocation obtained via simulation of the average error probability, while the solid lines represent the upper bounds. Each line color is associated with a particular erasure probability of the main channel.

For a noiseless main channel, the bounds are tight. With increasing erasure probabilities over the main channel, the puncturing limit decreases, hence the eavesdropper's error probability also decreases and the bounds become loose. For this particular case, even though the BP decoding thresholds are just slightly apart, the irregular code presents much tighter bounds that the regular code.

In summary, it is possible to see that hiding the puncturing pattern results in a high error rate, even when the eavesdropper's estimates are optimal. Thus, even for modest block-lengths, puncturing can provide secrecy benefits.

5.4 System Aspects

In this section we discuss several questions that pertain to the assumptions behind the proposed model as well as other system aspects.

5.4.1 Sharing Secret Keys

We have seen that using the puncturing pattern as a shared secret may help in increasing the eavesdropper's equivocation, due to the lack of synchronization. However, this requires either the transmission or agreement of a secret key. In practice this can be done in several ways. First, it is possible to use cryptographic methods such as the Diffie-Hellman key agreement scheme [68]. Hence, a cross-layer approach may be taken in the system design. Of course that the derived keys do not obey any information-theoretic secrecy

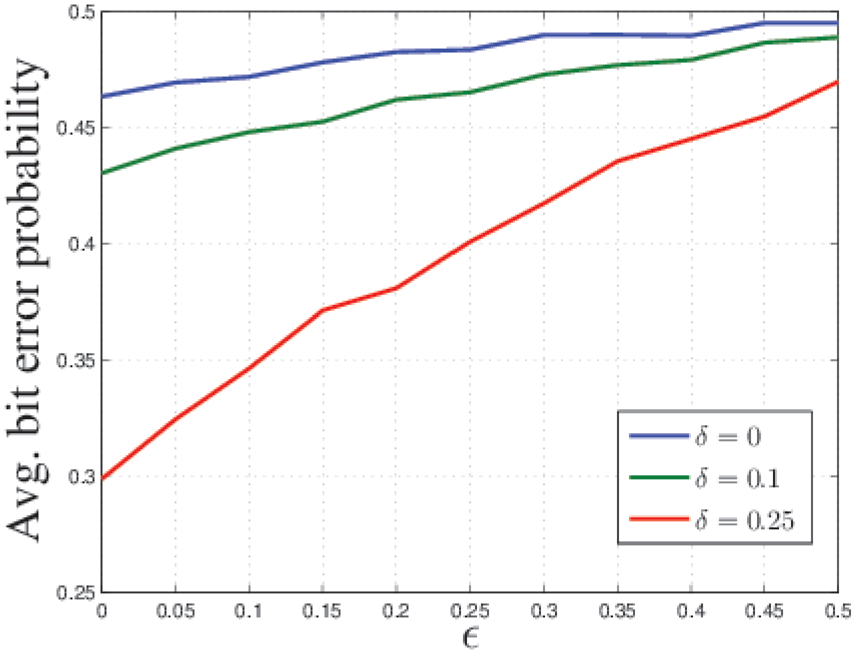


Figure 5.13: Simulated average bit error rate for the code C_4 as a function of the wiretap channel erasure probability ε, when the main channel erasure probability takes values from $\delta \in \{0, 0.1, 0.25\}$.

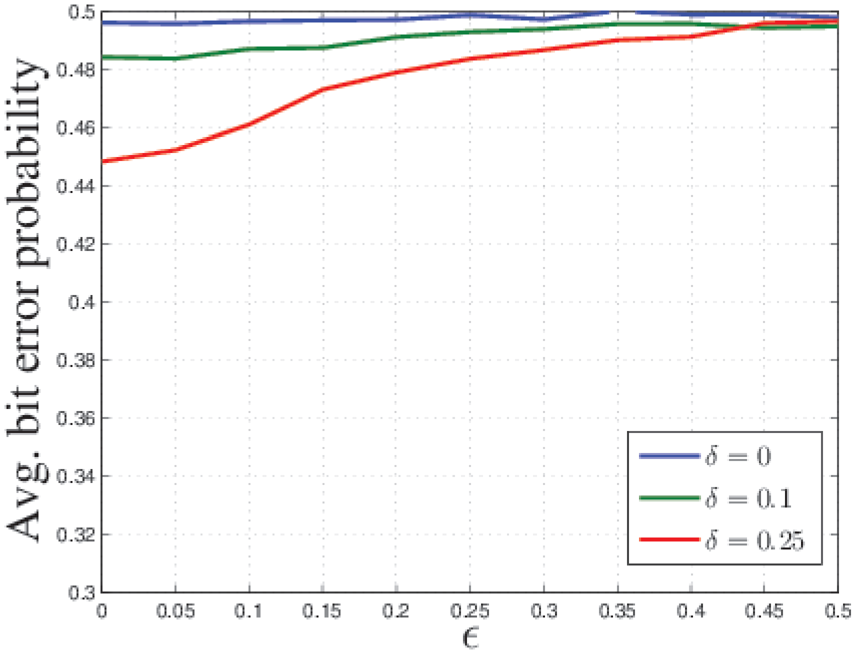


Figure 5.14: Simulated average bit error rate for the code C_5 as a function of the wiretap channel erasure probability ε, when the main channel erasure probability takes values from $\delta \in \{0, 0.1, 0.25\}$.

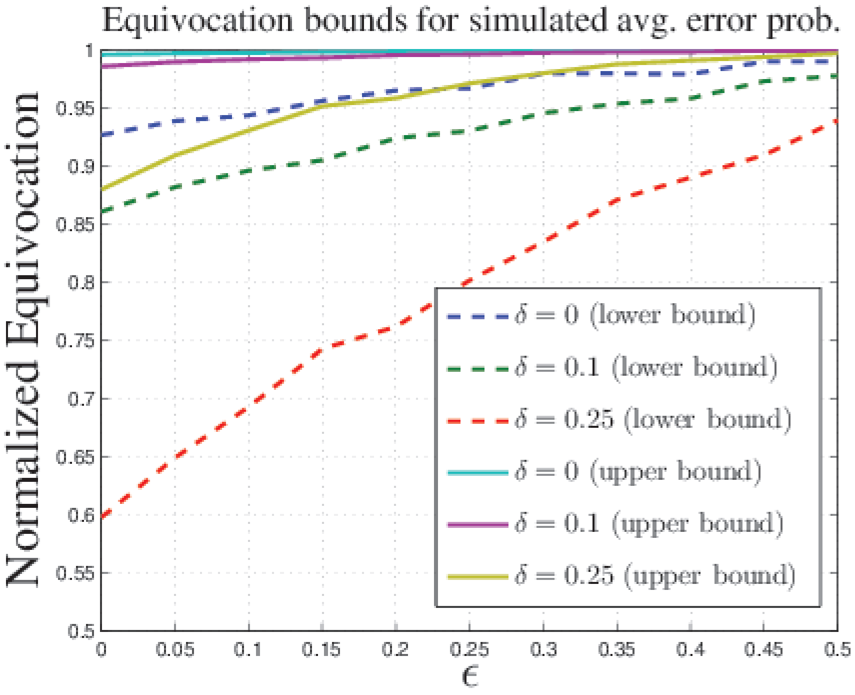


Figure 5.15: Bounds on the equivocation of simulated error probability for the code C_4 as a function of the wiretap channel erasure probability ε, when the main channel erasure probability takes values from $\delta \in \{0, 0.1, 0.25\}$.

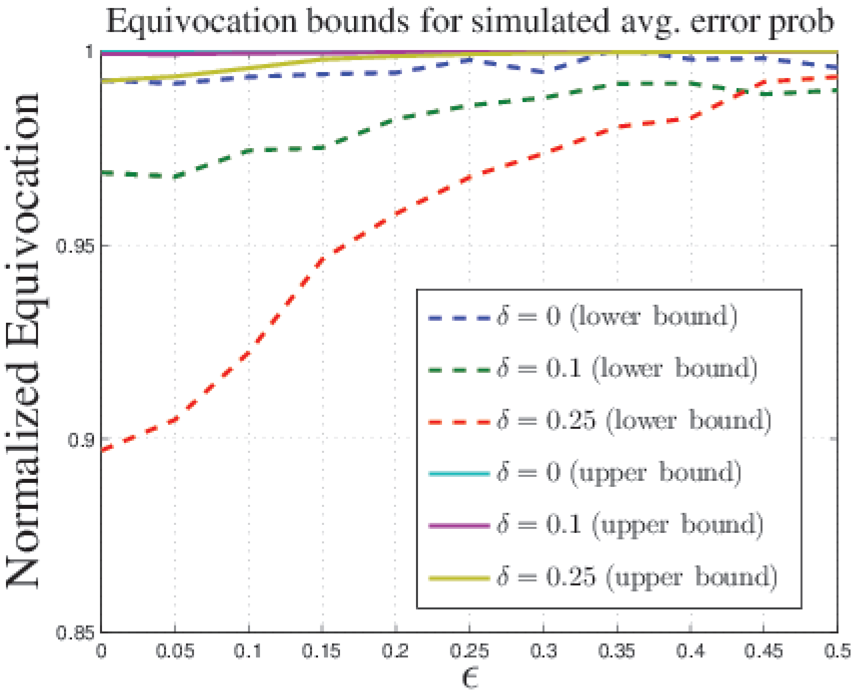


Figure 5.16: Bounds on the equivocation of simulated error probability for the code C_5 as a function of the wiretap channel erasure probability ε, when the main channel erasure probability takes values from $\delta \in \{0, 0.1, 0.25\}$.

criterion, and therefore, the equivocation of the eavesdropper could in general be less, as he may obtain some information about the puncturing pattern. Nevertheless, even in a worst case scenario where the eavesdropper obtains a perfect copy of the secret key, we have seen that puncturing can provide for maximum equivocation (if we puncture up to the MAP threshold). Thus, from a practical standpoint, using a cross-layer approach may be sufficient for secrecy purposes.

On the other hand, one may use information-theoretic secret key agreement schemes [3, Chapters 3 and 4]. A possibility is to use a regular wiretap code to share this key. While apparently this would defeat the motivation for our scheme, this is not necessarily true, as the wiretap code can be used in a conservative way (meaning that we use a wiretap code assuming the quality of the wiretap channel that is very close to the quality of the main channel). Consequently, the secure rate would be small. However, if the code allows for a large enough puncturing probability, the rate required for the secret key is also small, and therefore such strategy may be sufficient. A second possibility that is more appealing is to use sequential key distillation strategies [3, Chapter 4.3], [69], either using one-way or two way communications. Unlike wiretap codes, sequential key distillation strategies do not require a *better* main channel, but they do require an external source. Lastly, it is also possible to use a parallel secure channel of limited rate, if such channel is available.

5.4.2 Secret Key Rates

The examples used in the previous section may suggest a large key rate is always required for the scheme to be effective. Once again, this limitation is imposed in the examples due to the fact that we may only compute the eavesdropper's equivocation for small block-lengths. In fact, in the limit of large block-lengths, one expects to find codes of very low rate and large MAP threshold, which would translate onto a reduced key rate. However, codes with very low rate require a very large block-length.

In particular, the proposed scheme requires $R_k = H_b(\varepsilon^{MAP})$, where $H_b(\cdot)$ is the binary entropy function. Thus, for increasingly larger MAP thresholds (where reliability can be achieved by letting $n \to \infty$) we would need keys of very low rate. On the other hand, the puncturing operation ensures that we actually communicate at an increased rate, therefore the usage of such codes does not force us to operate with a low communication rate.

5.4.3 Comparison with the One-Time Pad Cryptosystem

Consider the system depicted in Fig. 5.17, where we have a BEWC and the encoder performs one-time pad encryption. Assume that the messages are represented by an n-dimensional random variable $\mathbf{M}$ and chosen according to a i.i.d. uniform distribution. Let the employed key $\mathbf{K}$ be also an i.i.d. n-dimensional random variable that follows a Binomial distribution with probability γ. The transmitted message is given by $\mathbf{X} = \mathbf{M} \oplus$

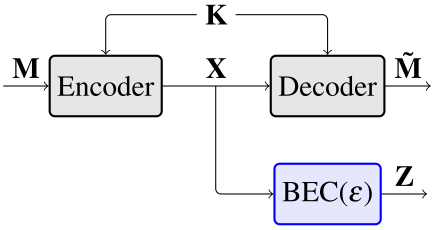


Figure 5.17: Shannon Cipher System with erasures.

$\mathbf{K}$, where $\oplus$ represents the modulo 2 addition operation. The normalized equivocation $\frac{1}{n}H(\mathbf{M}|\mathbf{Z})$ can be upper bounded as follows:

$$\begin{aligned}
\frac{1}{n}H(M^n|Z^n) &= \frac{1}{n}H(M^n,K^n|Z^n) - \frac{1}{n}H(K^n|M^n,Z^n) \\
&= \frac{1}{n}H(K^n|Z^n) + \frac{1}{n}H(M^n|K^n,Z^n) - \frac{1}{n}H(K^n|M^n,Z^n) \\
&\leq \frac{1}{n}H(K^n) + \frac{1}{n}H(M^n|K^n,Z^n) - \frac{1}{n}H(K^n|M^n,Z^n) \\
&= (1-\varepsilon)H(K) + \varepsilon H(M).
\end{aligned}$$

Since we have assumed a uniform input distribution, this expression simplifies to $\frac{1}{n}H(M^n|Z^n) = H(K) + \varepsilon(1-H(K))$. Obviously, if the wiretap channel is noiseless ($\varepsilon = 0$), the eavesdropper's equivocation will amount to the key equivocation. Thus, if we have access to a truly random key we will be able to operate at secrecy capacity with a one-time pad. However, if we only have access to a very biased key, the eavesdropper's equivocation will decrease. The same relationship is not true in our scheme. A biased key (of lower rate) will lead to an increase in equivocation, since more bits can be punctured (provided that we satisfy the reliability constraint). Moreover, if by some reason an adversary obtains the key used in the system, the eavesdropper's equivocation will amount to the erasure probability of the wiretap channel when the one-time pad scheme is used, while in the proposed scheme the eavesdropper's equivocation will amount the modified channel erasure probability, which is always larger than the original wiretap channel erasure probability.

5.5 Discussion

In this chapter we have proposed the use of random puncturing for secrecy over the erasure wiretap channel. We have shown that, to achieve high equivocation rates using a public puncturing pattern and a bijective deterministic encoder, the legitimate user is required to use a MAP decoder and puncture bits with a probability up to the MAP threshold. If the puncturing pattern is used as a shared secret, higher secrecy rates can be achieved. If the secret key is derived using physical-layer security methods, then the equivocation rate analysis has to take this fact into account. However, this shared secret can also be derived by public key cryptographic schemes, providing a cross-layer solution to the problem of confidential data transmission. Among the benefits provided by random puncturing is an easy adaptation of the code rate to the main channel, as well as avoidance of the need for channel statistics of the eavesdropper. It also provides easy guidelines for code design, as the only requirement is to puncture up to the permitted thresholds. An interesting point is also worth noting. The use of a puncturing pattern as a shared secret essentially creates a wiretap channel model that is fundamentally different from the main channel. This difference implies that the optimal source distribution with respect to the main channel (which is the one used in practice), may not be optimal for the wiretap channel (for instance, the uniform distribution is not the optimal input distribution for the deletion [61]). This can be seen as a further advantage in using such schemes.

Chapter 6

Conclusions

Under the paradigm of physical-layer security, we develop several coding schemes to provide for confidential data transmission. The focus is essentially mostly on practical code constructions. Our work differs from other works in this field in several ways. First, we explicitly design codes for finite block-lengths, which contrasts with the traditional asymptotic code designs. Second, we provide code constructions for continuous sources. Hence, our work does not inherit the assumptions required for separation theorems to hold. Third, we consider deterministic code constructions, thus circumventing the need for wiretap channel state information, as well as the need to optimize the use of local randomness. These differences also translate onto several limitations that can be assigned to our code constructions. The greatest perhaps is that, in general, our coding schemes do not achieve the fundamental limits of secrecy. On the other hand, their performance depends on the chosen block-length. In particular, for very small block-lengths, we have shown a relaxation is generally required, with respect to the reliability and secrecy criteria.

In summary, our main contributions can be stated as follows. In Chapters 3 and 4 we addressed code designs for continuous sources. In the former, a joint-source channel code is proposed which hinders the eavesdropper by forcing him to operate at a desired distortion level. This is accomplished by solving an optimization problem over the code parameters which constitutes in finding the scalar quantizer boundaries as well as the corresponding channel code that satisfies the aforementioned secrecy constraint. In the latter chapter, we use a particular instance of bandwidth expansion mappings which allows us to enforce anomalous errors for the eavesdropper, resulting in a large distortion. In particular, the code is parametrized based on its geometrical properties to ensure this constraint is satisfied. As a further advantage, these codes allow for very efficient encoding and decoding schemes. Chapter 5 proposes the use of random puncturing in order to create a wiretap channel of very poor quality to the eavesdropper, ensuring that his equivocation is highly increased. Using the insight that the loss of bit-level synchronization generally creates a channel with reduced capacity, we propose the use of the puncturing pattern as a shared secret, showing that, in such scenarios, the eavesdropper is bound to have a high equivocation, even for very small block-lengths.

6.1 Future Research

The coding schemes and the general framework under which these schemes are treated can be extended in multiple ways.

- **Nested Code Constructions:** While we avoid the use of nested code constructions to circumvent the issues of the lack of wiretap channel state information, this is not necessarily true in all cases. Therefore, there it could be interesting to extend these code constructions to use nested structures. While this has been largely done in the context of codes for discrete sources (see e.g.Chapter 2), the design of continues nested codes is practically non-existent. For instance, the code construction in Chapter 3 can be extended to the use of multiple scalar quantizers, each one with non-overlapping channel codes and optimized boundaries. The code construction in Chapter 4 could be extended to account for a multiple correspondence between source intervals and tori or source intervals and parallel curves over a given torus.

- **Secrecy from the absence of synchronization:** In Chapter 5 we have studied the secrecy performance of LDPC codes when the puncturing pattern was a shared secret. Consequently, this strategy induced a wiretap channel with deletions. Little is known with respect to the deletion channel. Thus, advances in the understanding of this channel model could benefit a more fundamental understanding of the performance of the secrecy codes proposed here. In particular, it could interesting to understand if it is possible to induce a deletion channel of zero capacity using this channel model, which essentially would mean that there is no leakage to the eavesdropper, possibly paving the way to the design of strong secrecy achieving schemes. On the other hand, it would be interesting to understand if expurgating some codewords of the randomly punctured ensemble could be useful from a secrecy perspective.

- **Fundamental Limits of Secure Communication over the Finite Block-length Regime:** In this thesis we avoided the formalization of these fundamental limits. In this regime one inevitably needs to assume non-vanishing error probability for the legitimate party. We have placed these assumptions heuristically, but analysed the performance of our constructions with respect to the fundamental limits in the asymptotic regime. However, this comparison is not necessarily fair. Hence, there is a need to establish the fundamental limits of secure communication under the finite block-length regime. The recent works of Yury Polyanskiy, Vincent Poor and Sergio

Verdú [70] have addressed this problem in the context of communications without secrecy requirements, and perhaps can be used to formalize the fundamental limits of secure communication under the finite block-length regime.

- **Cross-Layer Security:** The codes proposed in this thesis have the general goal of inducing a high distortion or high equivocation to the eavesdropper, with a focus on partial secrecy. In this context, the proposed schemes may be very useful with respect to a cross-layer implementation of secrecy. However, there is a need to formalize/understand how cryptanalysis is actually impacted by the errors at the lower layers. In this context, the work of Harrison [16] may be seen as a starting point for an information-theoretic perspective on the impact of errors achieved by physical-layer security with respect to cryptanalysis techniques.

Appendix A

Basic Notions on Information Theory

In this appendix we summarize several basic definitions and useful theorems from information theory. Unless noted otherwise, proofs can be found in [71].

A.1 Information Sources

In information theory it is common to model information sources as random variables [72]. Here we are interested in sources that are discrete in time, but could be either discrete or continuously-valued. Therefore, when we mention to a discrete source we are referring to a discrete random variable X defined over an alphabet $\mathcal{X}$ with an associated *probability mass function* (p.m.f.) $p_X(x) = Pr\{X = x\}, x \in \mathcal{X}$. Similarly, when we mention a continuous source we are referring to a continuous random variable X defined over a support set $\mathcal{X}$ with an associated *probability density function* (p.d.f.) $p_X(x) = \frac{\partial}{\partial x} F_X(x), x \in \mathcal{X}$, where $F_X(x)$ is the *cumulative distribution function* (c.d.f). and $F_X(x) = Pr\{X \leq x\}$[1].

A.2 Entropy, Joint Entropy and Conditional Entropy

For discrete sources, we can define the entropy, joint entropy and conditional entropy as follows[2].

Definition 18 (Entropy)**.** Let $X \in \mathcal{X}$ be a discrete random variable. The *entropy* of X is given by

$$H(X) \triangleq - \sum_{x \in \mathcal{X}} p(x) \log p(x). \tag{A.1}$$

It is also possible to obtain the mutual information between several random variables by mapping the measures defined above onto the so-called I-measures [73, Chapter 6].

[1] For simplicity, we may omit the subscript X when it is clear from the context.

[2] Unless specified otherwise, we assume that all logarithms are taken base 2 and therefore the corresponding unit of information is called "bit". By convention $0 \cdot \log 0 \triangleq 0$

I-measures are signed measures on a field $\mathcal{F}_n$, generated by sets $\tilde{X}_1, \ldots, \tilde{X}_n$. $\mathcal{F}_n$ can be obtained be any sequence of set operation, i.e., union, intersection, complement and difference, on $\tilde{X}_1, \ldots, \tilde{X}_n$. In particular, it can be shown that a signed measure μ on $\mathcal{F}_2$ can be completely defined by the values $\mu(\tilde{X}_1 \cap \tilde{X}_2)$, $\mu(\tilde{X}_1^c \cap \tilde{X}_2)$, $\mu(\tilde{X}_1 \cap \tilde{X}_2^c)$ and $\mu(\tilde{X}_1^c \cap \tilde{X}_2^c)$, where $\tilde{X}_i^c$ denotes the complement of $\tilde{X}_i^c$.

Definition 19 (Joint entropy)**.** Let $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ be two discrete random variables. The *joint entropy* of X and Y is given by

$$H(XY) \triangleq -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x,y) \log p_{XY}(x,y). \tag{A.2}$$

Definition 20 (Conditional entropy)**.** Let $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ be two discrete random variables. The *conditional entropy* of Y given X is given by

$$H(Y|X) \triangleq -\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p_{XY}(x,y) \log p_{Y|X}(y|x). \tag{A.3}$$

The interpretation of these information theoretic quantities is rather intuitive. The entropy of a random variable measures the uncertainty associated with guessing the realizations of a random variable (equivalently the amount of information contained in the random variable). Similarly, the joint entropy measures the uncertainty associated with guessing the joint realizations of two random variables and the conditional entropy the uncertainty associated with guessing the realizations of a random variable when another random variable is known.

A.3 Mutual Information and Conditional Mutual Information

For discrete sources, we can define the mutual information and conditional mutual information as follows.

Definition 21 (Mutual information)**.** Let $X \in \mathcal{X}$ and $Y \in \mathcal{Y}$ be two discrete random variables. The *mutual information* between X and Y is given by

$$I(X;Y) \triangleq H(X) - H(X|Y). \tag{A.4}$$

Definition 22 (Conditional mutual information)**.** Let $X \in \mathcal{X}$, $Y \in \mathcal{Y}$ and $Z \in \mathcal{Z}$ be three discrete random variables. The *conditional mutual information* between X and Y given Z

is given by

$$I(X;Y|Z) \triangleq H(X|Z) - H(X|YZ). \tag{A.5}$$

The mutual information measures the amount of information X and Y share in common. Alternatively, we can interpret the mutual information as the uncertainty of a random variable that is not resolved by knowing another random variable. Similarly, the conditional mutual information measures the amount of information X and Y share in common when a third random variable Z is given (or alternatively the uncertainty of a random variable that is not resolved by knowing Y when Z is given).

A.4 I-measures

The information measures defined above are also known as Shannon measures. It is possible to extend some of the above notions to several random variables (in particular the mutual information). However, the obtained expressions are typically very involved. This issue can be circumvented by relating these measures with set theory. This can be done by mapping the Shannon measures onto the so-called I-measures [73, Chapter 6]. I-measures are signed measures on a field $\mathcal{F}_n$, generated by sets $\tilde{X}_1,\ldots,\tilde{X}_n$. $\mathcal{F}_n$ can be obtained be any sequence of set operation, i.e., union, intersection, complement and difference, on $\tilde{X}_1,\ldots,\tilde{X}_n$. In particular, it can be shown that a signed measure μ on $\mathcal{F}_2$ can be completely defined by the values $\mu(\tilde{X}_1 \cap \tilde{X}_2)$, $\mu(\tilde{X}_1^c \cap \tilde{X}_2)$, $\mu(\tilde{X}_1 \cap \tilde{X}_2^c)$ and $\mu(\tilde{X}_1^c \cap \tilde{X}_2^c)$, where $\tilde{X}_i^c$ denotes the complement of $\tilde{X}_i^c$. A one-to-one correspondence between Shannon measures and I-measures can be obtained. Let $\mu^\star$ be a signed measure on $\mathcal{F}_2$. Define $\mu^\star$ the following correspondence.

1. $\mu^\star(\tilde{X}_1 - \tilde{X}_2) = H(X_1|X_2)$;
2. $\mu^\star(\tilde{X}_2 - \tilde{X}_1) = H(X_2|X_1)$;
3. $\mu^\star(\tilde{X}_1 \cap \tilde{X}_2) = I(X_1;X_2)$;

Using set operations, we can obtain the remaining measures: $\mu^\star(\tilde{X}_1) = H(X_1)$, $\mu^\star(\tilde{X}_2) = H(X_2)$ and $\mu^\star(\tilde{X}_1 \cup \tilde{X}_2) = H(X_1,X_2)$.

Generalization to n random variables can be obtained by constructing the I-measure $\mu^\star$ on $\mathcal{F}_n$ by defining $\mu^\star(\tilde{X}_G) = H(X_G)$, where G is a non-empty subset of $\mathcal{N}_n = \{1,\ldots,n\}$. Then, the mutual information of several random variables $X_1,\ldots,X_n$ can be seen the measure $\mu^\star$ of the intersection of the sets $\tilde{X}_1,\ldots,\tilde{X}_n$, i.e. $I(X_1;X_2;\ldots;X_n)$ can be thought of as $\mu^\star(\tilde{X}_1 \cap \tilde{X}_2 \cap \ldots \cap \tilde{X}_n)$.

More details regarding I-measures can be found in [73, Chapter 6].

A.5 Properties of Entropy and Mutual Information

Theorem 9 (Chain rule of entropy)**.** *Let $X_1, X_2, \ldots, X_n$ be n discrete random variables with a joint distribution $p(x_1, x_2, \ldots, x_n)$. Then, we have that*

$$H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i|X^{i-1}), \tag{A.6}$$

where $X^u = (X_1, \ldots, X^u)$.

Theorem 10 (Conditioning reduces entropy)**.** *Let X and Y be two discrete random variables. Then, we have that*

$$H(X|Y) \leq H(X), \tag{A.7}$$

with equality if and only if X and Y are independent

Corollary 5 (Non-negativity of entropy and mutual information)**.** *Let X and Y be two discrete random variables. Then,*

$$H(X) \geq 0, \tag{A.8}$$

with equality when X has only one or no possible outcome, and

$$I(X;Y) \geq 0, \tag{A.9}$$

with equality if and only if X and Y are independent.

Corollary 6 (Non-negativity of conditional mutual information)**.** *Let X, Y and Z be three discrete random variables. Then,*

$$I(X;Y|Z) \geq 0, \tag{A.10}$$

with equality if and only if X and Y are independent given Z.

Corollary 7 (Conditional mutual entropy of a Markov Chain with four random variables)**.** *Let $(X_1, X_2) \to X_3 \to X_4$ form a Markov chain. Then, $I(X_1;X_2;X_4|X_3) + I(X_1;X_4|X_2,X_3) + I(X_2;X_4|X_1,X_3) = I(X_1,X_2;X4|X_3) = 0$.*

A.6 Distortion

In information theory, distortion generally refers to a cost function that measures some distance between a random variable and its estimate. Formally, it can be defined as follows.

Definition 23 (Distortion)**.** Let $x \in \mathcal{X}$ and $\tilde{x} \in \tilde{\mathcal{X}}$ be a continuous random variables, where $\tilde{x}$ is the reproduction point of x. A distortion function $d(\cdot,\cdot)$ is a mapping $d(x,\tilde{x}) : \mathcal{X} \times \tilde{\mathcal{X}} \rightarrow [0,\infty[$ that measures the cost of representing x by $\tilde{x}$. The distortion measure is bounded if $d(x,\tilde{x}) < \infty$.

Common examples of distortion functions are the squared error distortion, instantiated as $d(x,\tilde{x}) = (x-\tilde{x})^2$ and the Hamming distortion, instantiated as

$$d(x,\tilde{x}) = \begin{cases} 1 & \text{, if } x \neq \tilde{x} \\ 0 & \text{, otherwise.} \end{cases}$$

This definition can be extended to sequences of random variables. We will defined the distortion between two sequences as the average of the per letter distortion[3].

Definition 24 (Distortion between two sequences)**.** Let $x^n \in \mathcal{X}^n$ and $\tilde{x}^n \in \tilde{\mathcal{X}}^n$ be two sequences of continuous random variables, where $\tilde{x}^n$ is the reproduction point of x^n. The distortion between the sequences x^n and $\tilde{x}^n$ is given by

$$d(x^n,\tilde{x}^n) = \frac{1}{n}\sum_{i=1}^{n} d(x_i,\tilde{x}_i). \tag{A.11}$$

[3] There may be other ways to define the distortion between two sequences

Appendix B

Leakage Bound for Wiretap Codes

We wish to show that $I(M;Z^n) \leq nC_e - H(M') + H(M'|MZ^n)$.

Proof. Let $(M, M') \to X^n \to Z^n$ form a Markov chain and M and M' be independent. We know that $I(MM';Z^n|X^n) = 0$ (see [73, Eq. 6.86]), which implies that $I(M;Z^n|X^n) = 0$ and $I(M';Z^n|X^n) = 0$.

$$\begin{aligned} I(MM';Z^n|X^n) &= H(MM'|X^n) - H(MM'|X^nZ^n) \\ &\overset{(a)}{=} H(M|X^n) + H(M'|X^n) - [H(M|X^nZ^n) + H(M'|X^nZ^n)] \\ &= H(M|X^n) - H(M|X^nZ^n) + H(M'|X^n) - H(M'|MX^nZ^n) \\ &= I(M;Z^n|X^n) + I(M';Z^n|X^n), \end{aligned}$$

where (a) follows from the independence of M and M'. Since we have that $I(M;Z^n|X^n) \geq 0$, $I(M';Z^n|X^n) \geq 0$ and $I(MM';Z^n|X^n) = I(M;Z^n|X^n) + I(M';Z^n|X^n) = 0$, it must be that $I(M;Z^n|X^n) = 0$ and $I(M';Z^n|X^n) = 0$. Additionally, we have that $H(X^n|M) = H(M'|M) = H(M')$ and $H(X^n|MZ^n) = H(M'|MZ^n)$. The first equality can be easily seen from the fact that $H(X^n|M) = H(M'|M) - H(M'|X^nM) + H(X^n|M'M) = H(M'|M) = H(M')$, where in the second equality we use the fact $H(M'|X^nM) = H(X^n|M'M) = 0$ since two of the random variables completely determine the third and the last equality follows from the independence between M and M'. The proof of the second equality follows from the same principles. Finally we have that

$$\begin{aligned}
\frac{1}{n}I(X^n;Z^n) &= \frac{1}{n}\Bigg[I(Z^n;X^nM) - I(X^n;Z^n|M)\Bigg] \\
&= \frac{1}{n}\Bigg[I(Z^n;X^n) + I(M;Z^n|X^n) - I(X^n;Z^n|M)\Bigg] \\
&= \frac{1}{n}\Bigg[I(Z^n;X^n) - H(X^n|M) + H(X^n|MZ^n)\Bigg] \\
&= \frac{1}{n}\Bigg[I(Z^n;X^n) - H(M') + H(M'|MZ^n)\Bigg] \\
&\leq \frac{1}{n}\Bigg[nC_e - H(M') + H(M'|MZ^n)\Bigg].
\end{aligned}$$

□

Appendix C

Achievable Rate-Equivocation Region for the Wiretap Channel With a Shared Key

A derivation of the secrecy capacity of the wiretap channel model with a shared key is given in [74]. However, in [74], the rate-equivocation region is not explicitly established. While it is straightforward to obtain the rate-equivocation region from [74], we provide a simplified proof of the achievability, that does not require a separate analysis based on the key rate. For the converse, we re-direct the reader to [74]. We wish to prove the existence of $(2^{nR}, n)$ codes $\{\mathcal{C}_n\}_{n\geq 1}$, such that $\lim_{n\to\infty} P_e(\mathcal{C}_n) \leq \delta_\varepsilon(n)$ and $\lim_{n\to\infty} \frac{1}{n} I(M;Z^n) \leq \delta_\varepsilon(n)$, where $\delta_\varepsilon(n)$ represents a function of ε and n such that $\lim_{n\to\infty} \delta_\varepsilon(n) = 0$.

Proof. Let there be three message sets, M, M_k and M_d, where $M \in [1, 2^{nR}]$, $M_k \in [1, 2^{nR_k}]$ and $M_d \in [1, 2^{nR_d}]$. In particular, M represents the set of messages for transmission, the M_k the set of possible keys and M_d a set of dummy messages used to randomize the encoder.

Consider the following random code construction. First, generate codewords $u^n(m_k)$, for $m \in [1, 2^{nR_k}]$ by generating symbols $u_i(m_k)$, with $i \in [1, n]$ and $m \in [1, 2^{nR_k}]$ independently according to $p_U(u)$. Then, for every generated $u^n(m_k)$, generate codewords $x^n(m, m_k, m_d)$, for $m \in [1, 2^{nR}]$, $m_d \in [1, 2^{nR_d}]$, by generating symbols $x_i(m, m_k, m_d)$ with $i \in [1, n]$, $m \in [1, 2^{nR}]$, $m_d \in [1, 2^{nR_d}]$ independently according to $p_{X|U=u_i(m_k)}$.

The encoding procedure is as follows. Given m, m_k and m_d the sender transmits $x^n(m, m_k, m_d)$. The considered decoder is essentially a typical set decoder, which can be described as follows.

1. Given y^n and m_k the legitimate receiver outputs $(\tilde{m}, \tilde{m}_d)$ if it is the unique tuple such that $(u^n(m_k), x^n(\tilde{m}, m_k, \tilde{m}_d), y^n) \in T^n_\varepsilon(UXY)$.

2. Given z^n, m_k and m, the virtual receiver outputs $\hat{m}_d$ if it is the unique message such that $(u^n(m_k), x^n(m, m_k, \hat{m}_d), z^n) \in T^n_\varepsilon(UXZ)$.

Let us now analyze the error probability of this random code construction. We have that

$$\begin{aligned} E[P_e(\mathcal{C}_n)] &= E_{\mathcal{C}_n}\left[P[(\tilde{M},\tilde{M_d}) \neq (M,M_d) \text{ or } \hat{M}_d \neq M_d|\mathcal{C}_n\right] \\ &\overset{(a)}{=} E_{\mathcal{C}_n}\left[P[(\tilde{M},\tilde{M_d}) \neq (M,M_d) \text{ or } \hat{M}_d \neq M_d|M=1,M_d=1,K=1,\mathcal{C}_n]\right], \end{aligned}$$

where (a) follows from the symmetry of the random-coding construction. Therefore, without loss of generality, we can assume that $M=1$, $M_k=1$ and $M_d=1$. Define the two following events:

1. $\mathcal{E}_{ij} = (u^n(1), x^n(i,1,j), y^n) \in T^n_\varepsilon(UXY)$

2. $\mathcal{F}_j = (u^n(1), x^n(1,1,j), z^n) \in T^n_\varepsilon(UXZ)$

We can write $E[P_e(\mathcal{C}_n)]$ as a function of $\mathcal{E}_{ij}$ and $\mathcal{F}_j$ as follows:

$$\begin{aligned} E[P_e(\mathcal{C}_n)] &= P\left[\mathcal{E}^c_{11} \cup \bigcup_{(i,j)\neq(1,1)} \mathcal{E}_{ij} \cup \mathcal{F}^c_1 \cup \bigcup_{j\neq 1} \mathcal{F}_j\right] \\ &\leq P[\mathcal{E}^c_{11}] + \sum_{(i,j)\neq(1,1)} P\left[\mathcal{E}_{ij}\right] + P[\mathcal{F}^c_1] + \sum_{j\neq 1} P\left[\mathcal{F}_j\right] \end{aligned}$$

By the AEP we know that $P\left[\mathcal{E}^c_{11}\right] \leq \delta_\varepsilon(n)$ and $P\left[\mathcal{F}^c_1\right] \leq \delta_\varepsilon(n)$. Additionally, we have that for $(i,j) \neq (1,1)$, $x^n(i,1,j)$ is conditionally independent of y^n given $u^n(1)$ and for $j \neq 1$, $x^n(1,1,j)$ is is conditionally independent of z^n given $u^n(1)$. Therefore, we have that $P\left[\mathcal{E}_{ij}\right] \leq 2^{-n(I(X;Y|U)-\delta_\varepsilon(n))}$ and $P\left[\mathcal{F}_j\right] \leq 2^{-n(I(X;Z|U)-\delta_\varepsilon(n))}$.

Consequently, we have that

$$\begin{aligned} E[P_e(\mathcal{C}_n)] &\leq \delta_\varepsilon(n) + 2^{n(R+R_d)}2^{-n(I(X;Y|U)-\delta_\varepsilon(n))} + \delta_\varepsilon(n) + 2^{nR_d}2^{n(I(X;Z|U)-\delta_\varepsilon(n))} \\ &= \delta_\varepsilon(n) + 2^{n(R+R_d-I(X;Y|U)+\delta_\varepsilon(n))} + 2^{n(R_d-I(X;Z|U)+\delta_\varepsilon(n))}. \end{aligned}$$

Thus, a sufficient condition for having $E[P_e(\mathcal{C}_n)] \leq \delta_\varepsilon(n)$ is to choose R and R_d such that

$$\begin{cases} R+R_d \leq I(X;Y|U) - \delta_\varepsilon(n) \\ R_d \leq I(X;Z|U) - \delta_\varepsilon(n). \end{cases}$$

With respect to the leakage, it can be upper bounded as follows.

$$
\begin{aligned}
\frac{1}{n}I(M;Z^n) &\leq \frac{1}{n}I(M;Z^nM_k)\\
&= \frac{1}{n}\Big[I(MX^n;Z^nM_k)-I(X^n;Z^nM_k|M)\Big]\\
&= \frac{1}{n}\Big[I(X^n;Z^nM_k)+I(M;Z^nM_k|X^n)-I(X^n;Z^nM_k|M)\Big]\\
&= \frac{1}{n}\Big[I(X^n;M_k)+I(X^n;Z^n|M_k)+I(M;Z^nM_k|X^n)-I(X^n;Z^nM_k|M)\Big]\\
&= \frac{1}{n}\Big[I(X^n;M_k)+I(X^n;Z^n|M_k)+I(M;Z^n|X^n)+I(M;M_k|X^nZ^n)-I(X^n;Z^nM_k|M)\Big]\\
&\overset{(a)}{=} \frac{1}{n}\Big[I(X^n;Z^n|U^n)+I(X^n;M_k)+I(M;Z^n|X^n)+I(M;M_k|X^nZ^n)-I(X^n;Z^nM_k|M)\Big]\\
&= \frac{1}{n}\Big[I(X^n;Z^n|U^n)+I(X^n;M_k)+I(M;Z^n|X^n)+I(M;M_k|X^nZ^n)-I(X^n;Z^n|M)\\
&\quad -I(X^n;M_k|MZ^n)\Big]\\
&= \frac{1}{n}\Big[I(X^n;Z^n|U^n)+H(X^n)-H(X^n|M_k)+H(M|X^n)-H(Z^n|MX^n)+H(M|X^nZ^n)\\
&\quad -H(M_k|MX^nZ^n)-H(X^n|M)+H(Z^n|MX^n)-H(X^n|MZ^n)+H(M_k|MX^nZ^n)\Big]\\
&= \frac{1}{n}\Big[I(X^n;Z^n|U^n)+H(X^n)-H(X^n|M_k)+H(M|X^n)+H(M|X^nZ^n)-H(X^n|M)\\
&\quad -H(X^n|MZ^n)\Big]\\
&\leq \frac{1}{n}\Big[I(X^n;Z^n|U^n)-H(X^n|M_k)-H(X^n|M)+H(X^n)+H(M|X^n)+H(M|X^nZ^n)\\
&\quad -H(X^n|MZ^n)\Big]\\
&\leq \frac{1}{n}\Big[I(X^n;Z^n|U^n)-H(X^n|M_k)-H(X^n|M)+H(X^n|MZ^n)+H(M|X^n)\\
&\quad +H(M|X^nZ^n)-H(X^n|MZ^n)\Big]\\
&\leq \frac{1}{n}\Big[I(X^n;Z^n|U^n)-H(X^n|M_k)-H(X^n|M)\Big]\\
&\leq \frac{1}{n}\Big[I(X^n;Z^n|U^n)-H(X^n|MM_k)-H(X^n|MM_d)\Big]\\
&= \frac{1}{n}\Big[I(X^n;Z^n|U^n)-H(M_d)-H(M_k)\Big]\\
&\overset{(b)}{=} \frac{1}{n}I(X^n;Z^n|U^n)-R_d-R_k,\\
&\leq I(X;Z|U)-R_d-R_k,
\end{aligned}
$$

where (a) comes from the fact that there is a one-to-one mapping between M_k and U^n and (b) comes from the code construction. Let us choose $R_d = I(X;Z|U) - R_k - \delta_\varepsilon(n)$ and $R = I(X;Y|U) - I(X;Z|U) + R_k$. Then, we have that $\frac{1}{n}I(M;Z^n) \leq I(X;Z|U) - R_d - R_k = I(X;Z|U) - I(X;Z|U) + R_k + \delta_\varepsilon(n) - R_k \leq \delta_\varepsilon(n)$ and thus, the secrecy constrain holds. At the same time, we have that $R + R_d = I(X;Y|U) - I(X;Z|U) + R_k + I(X;Z|U) - R_k - \delta_\varepsilon(n) = I(X;Y|U) - \delta_\varepsilon(n)$ and $R_d = I(X;Z|U) - R_k - \delta_\varepsilon(n) \leq I(X;Z|U) - \delta_\varepsilon(n)$, thus satisfying the reliability conditions.

□

References

[1] D. Kahn. *The Codebreakers: The Story of Secret Writing*. Macmillan Publishing Co., 1967.

[2] Alfred J. Menezes, Scott A. Vanstone, and Paul C. Van Oorschot. *Handbook of Applied Cryptography*. CRC Press, Inc., Boca Raton, FL, USA, 1st edition, 1996.

[3] Matthieu Bloch and João Barros. *Physical-Layer Security: From Information Theory to Security Engineering*. Cambridge University Press, 2011.

[4] William Stallings. *Cryptography and Network Security*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 4th edition, 2005.

[5] Oded Goldreich. *The Foundations of Cryptography - Volume 1, Basic Techniques*. Cambridge University Press, 2001.

[6] R.L. Rivest, A. Shamir, and L. Adleman. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications of the ACM*, 21:120–126, February 1978.

[7] Taher El Gamal. A Public Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms. In *Proceedings of CRYPTO 84 on Advances in Cryptology*, pages 10–18, 1985.

[8] Bruce Schneier. *Applied Cryptography*. John Wiley & Sons, 1996.

[9] Claude E Shannon. Communication theory of secrecy systems. *Bell Systems Technical Journal*, 28(4):656–715, 1949.

[10] Serge Vaudenay. *A classical introduction to cryptography: Applications for communications security*. Springer-Verlag New York Incorporated, 2005.

[11] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, October 1997.

[12] Dan Boneh, Ron Rivest, Adi Shamir, and Len Adleman. Twenty years of attacks on the RSA cryptosystem. *Notices of the American Mathematical Society*, 46(2):203–213, 1999.

[13] A. D. Wyner. The Wiretap Channel. *Bell Systems Technical Journal*, 54:1355–1387, October 1975.

[14] I. Csizár and J. Körner. Broadcast Channels with Confidential Messages. *IEEE Transactions on Information Theory*, 24:339–348, May 1978.

[15] W.K. Harrison, J. Almeida, M. Bloch, J. Barros, and S.W. McLaughlin. Coding for Secrecy: An Overview of Error-Control Coding Techniques for Physical-Layer Security. *IEEE Signal Processing Magazine*, 30(5):41–50, September 2013.

[16] W. K. Harrison. *Physical-layer security: practical aspects of channel coding and cryptography*. PhD thesis, Georgia Institute of Technology, 2012.

[17] Ueli Maurer. Secret Key Agreement by Public Discussion from Common Information. *IEEE Transactions on Information Theory*, 39(3):733–742, May 1993.

[18] H. Yamamoto. Rate-distortion theory for the Shannon cipher system. *IEEE Transactions on Information Theory*, 43(3):827–835, May 1997.

[19] D. Klinc, Jeongseok Ha, S. W. McLaughlin, J. Barros, and Byung-Jae Kwak. LDPC Codes for the Gaussian Wiretap Channel. *IEEE Transactions on Information Forensics and Security*, 6(3):532–540, September 2011.

[20] M. Baldi, M. Bianchi, and F. Chiaraluce. Coding With Scrambling, Concatenation, and HARQ for the AWGN Wire-Tap Channel: A Security Gap Analysis. *IEEE Transactions on Information Forensics and Security*, 7(3):883–894, June 2012.

[21] N. Merhav. Shannon's Secrecy System With Informed Receivers and its Application to Systematic Coding for Wiretapped Channels. *IEEE Transactions on Information Theory*, 54(6):2723–2734, June 2008.

[22] Ueli Maurer and Stefan Wolf. Information-Theoretic Key Agreement: From Weak to Strong Secrecy for Free. In *Advances in Cryptology — EUROCRYPT 2000*, volume 1807 of *Lecture Notes in Computer Science*, pages 351–368, May 2000.

[23] A. T. Suresh, A. Subramanian, A. Thangaraj, M. Bloch, and S. W. McLaughlin. Strong secrecy for erasure wiretap channels. In *Proceedings of the IEEE Information Theory Workshop*, pages 1–5, Dublin, Ireland, August-September 2010.

[24] J. C Belfiore and F. Oggier. Secrecy gain: A wiretap lattice code design. In *International Symposium on Information Theory and its Applications*, pages 174–178, 2010.

[25] M. Feder and N. Merhav. Relations between entropy and error probability. *IEEE Transactions on Information Theory*, 40(1):259–266, January 1994.

[26] C. Schieler, E.C. Song, P. Cuff, and H.V. Poor. Source-Channel Secrecy with Causal Disclosure. In *Proceedings of the 50th Annual Allerton Conference on Communication, Control, and Computing*, pages 968–973, 2012.

[27] C. Schieler and P. Cuff. Rate-distortion theory for secrecy systems. In *Proceedings of the IEEE International Symposium on Information Theory*, pages 2219–2223, 2013.

[28] M.R. Bloch and J.N. Laneman. Strong Secrecy From Channel Resolvability. *IEEE Transactions on Information Theory*, 59(12):8077–8098, 2013.

[29] A. Thangaraj, S. Dihidar, A. R. Calderbank, S. W. McLaughlin, and J-M. Merolla. Applications of LDPC Codes to the Wiretap Channel. *IEEE Transactions on Information Theory*, 53(8):2933–2945, August 2007.

[30] Chan Wong Wong, T.F. Wong, and J.M. Shea. Secret-Sharing LDPC Codes for the BPSK-Constrained Gaussian Wiretap Channel. *IEEE Transactions on Information Forensics and Security*, 6(3):551–564, September 2011.

[31] V. Rathi, M. Andersson, R. Thobaben, J. Kliewer, and M. Skoglund. Performance Analysis and Design of Two Edge-Type LDPC Codes for the BEC Wiretap Channel. *IEEE Transactions on Information Theory*, 59(2):1048–1064, 2013.

[32] H. Mahdavifar and A. Vardy. Achieving the Secrecy Capacity of Wiretap Channels Using Polar Codes. *IEEE Transactions on Information Theory*, 57(10):6428–6443, 2011.

[33] O. Ozan Koyluoglu and Hesham El Gamal. Polar Coding for Secure Transmission and Key Agreement. In *Proceedings of the IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, pages 2698–2703, Istambul, Turkey, September 2010.

[34] E. Hof and S. Shamai. Secrecy-achieving polar-coding. In *Proceedings of the IEEE Information Theory Workshop*, pages 1–5, Dublin, Ireland, September 2010.

[35] A. Subramanian, A. Thangaraj, M. Bloch, and S. W. McLaughlin. Strong Secrecy on the Binary Erasure Wiretap Channel Using Large-Girth LDPC Codes. *IEEE Transactions on Information Forensics and Security*, 6(3):585–594, September 2011.

[36] V. Rathi, R. Urbanke, M. Andersson, and M. Skoglund. Rate-equivocation optimal spatially coupled LDPC codes for the BEC wiretap channel. In *Proceedings of the IEEE International Symposium on Information Theory*, pages 2393–2397, St. Petersburg, Russia, August 2011.

[37] M.P. Wilson and K. Narayanan. Transmitting an analog Gaussian source over a Gaussian wiretap channel under SNR mismatch. In *Proceedings of the 17th IEEE International Conference on Telecommunications*, pages 44–47, Doha, Qatar, April 2010.

[38] G. Bagherikaram and K.N. Plataniotis. Secure hybrid digital-analog Wyner-Ziv coding. In *Proceedings of the 22nd IEEE International Symposium on Personal Indoor and Mobile Radio Communications*, pages 1161–1166, Toronto, ON, September 2011.

[39] N. Farvardin and V. Vaishampayan. Optimal quantizer design for noisy channels: An approach to combined source - channel coding. *IEEE Transactions on Information Theory*, 33(6):827–838, November 1987.

[40] T. Fine. Properties of an optimum digital system and applications. *IEEE Transactions on Information Theory*, 10(4):287–296, October 1964.

[41] S. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, March 1982.

[42] J. Max. Quantizing for minimum distortion. *IRE Transactions on Information Theory*, 6(1):7–12, March 1960.

[43] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.

[44] H.V. Poor. *An Introduction to Signal Detection and Estimation*. Springer-Verlag, 1994.

[45] C. E. Shannon. Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21, January 1949.

[46] Kotel'nikov, V.A. *The theory of optimum noise immunity*. McGraw-Hill, 1959.

[47] F. Hekland, P.A. Floor, and T.A. Ramstad. Shannon-Kotel'nikov mappings in joint source-channel coding. *IEEE Transactions on Communications*, 57(1):94–105, 2009.

[48] C. Torezzan, S. I R Costa, and V.A. Vaishampayan. Spherical codes on torus layers. In *Proceedings of the IEEE International Symposium on Information Theory*, pages 2033–2037, Seoul, South Korea, 2009.

[49] A. Campello, C. Torezzan, and S. I R Costa. Curves on torus layers and coding for continuous alphabet sources. In *Proceedings of the IEEE International Symposium on Information Theory Proceedings*, pages 2127–2131, Cambridge, MA, 2012.

[50] Vincent Borrelli, Saïd Jabrane, Francis Lazarus, and Boris Thibert. Flat tori in three-dimensional space and convex integration. *Proceedings of the National Academy of Sciences*, 2012.

[51] Cristiano Torezzan. *Codigos esféricos em toros planares (Spherical codes on flat torus)*. PhD thesis, Universidade Estadual de Campinas . Instituto de Matemática, Estatística e Computação Científica, 2009.

[52] Sueli I. R. Costa, Cristiano Torezzan, Antonio Campello, and Vinay A. Vaishampayan. Flat tori, lattices and spherical codes. In *Proceedings of the Information Theory and Applications Workshop (ITA)*, pages 1–8, San Diego, CA, February 2013.

[53] V.A. Vaishampayan and S.I.R. Costa. Curves on a sphere, shift-map dynamics, and error control for continuous alphabet sources. *IEEE Transactions on Information Theory*, 49(7):1658–1672, July 2003.

[54] V.A. Vaishampayan, N. Sloane, and S. I R Costa. Dynamical systems, curves and coding for continuous alphabet sources. In *Proceedings of the IEEE Information Theory Workshop*, pages 111–114, 2002.

[55] J. M. Wozencraft and I. M. Jacobs. *Principles of communication engineering*. John Wiley & Sons, 1965.

[56] Yichuan Hu, J. Garcia-Frias, and M. Lamarca. Analog Joint Source-Channel Coding Using Non-Linear Curves and MMSE Decoding. *IEEE Transactions on Communications*, 59(11):3016–3026, 2011.

[57] David M. Mandelbaum. An adaptive-feedback coding scheme using incremental redundancy. *IEEE Transactions on Information Theory*, 20(3):388–389, 1974.

[58] J. Cain, G. Clark, and J. Geist. Punctured Convolutional Codes of Rate and Simplified Maximum Likelihood Decoding. *IEEE Transactions on Information Theory*, 25(1):97–100, January 1979.

[59] J. Hagenauer. Rate-compatible punctured convolutional codes (RCPC codes) and their applications. *IEEE Transactions on Communications*, 36(4):389–400, 1988.

[60] W.K. Harrison, J. Almeida, S.W. McLaughlin, and J. Barros. Coding for Cryptographic Security Enhancement Using Stopping Sets. *IEEE Transactions on Information Forensics and Security*, 6(3):575–584, September 2011.

[61] M. Mitzenmacher. A survey of results for deletion channels and related synchronization channels. *Probability Surveys*, 6:1–33, 2009.

[62] T. Richardson and R. Urbanke. *Modern Coding Theory*. Cambridge University Press, 2008.

[63] R. Urbanke and I. Andriyanova. Waterfall region performance of punctured LDPC codes over the BEC. In *Proceedings of the IEEE International Symposium on Information Theory*, pages 2644–2648, Seoul, South Korea, 2009.

[64] C. Measson, A. Montanari, and R. Urbanke. Maxwell Construction: The Hidden Bridge Between Iterative and Maximum a Posteriori Decoding. *IEEE Transactions on Information Theory*, 54(12):5277–5307, December 2008.

[65] M.G. Luby, M. Mitzenmacher, M.A. Shokrollahi, and D.A. Spielman. Improved low-density parity-check codes using irregular graphs. *IEEE Transactions on Information Theory*, 47(2):585–598, February 2001.

[66] Cees Elzinga, Sven Rahmann, and Hui Wang. Algorithms for subsequence combinatorics. *Theoretical Computer Science*, 409(3):394–404, December 2008.

[67] Wei Kang and Nan Liu. Wiretap channel with shared key. In *Proceedings of the Information Theory Workshop*, pages 1–5, 2010.

[68] W. Diffie and M. Hellman. New directions in cryptography. *IEEE Transactions on Information Theory*, 22(6):644–654, November 1976.

[69] R.A. Chou and M.R. Bloch. One-way rate-limited sequential key-distillation. In *Proceedings of the IEEE International Symposium on Information Theory*, pages 1777–1781, Cambridge, MA, 2012.

[70] Yury Polyanskiy. *Channel coding: non-asymptotic fundamental limits*. PhD thesis, Princeton University, 2010.

[71] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2006.

[72] Robert G. Gallager. *Information Theory and Reliable Communication*. John Wiley & Sons, Inc., New York, NY, USA, 1968.

[73] Raymond W. Yeung. *A First Course in Information Theory (Information Technology: Transmission, Processing and Storage)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[74] Wei Kang and Nan Liu. Wiretap channel with shared key. In *Proceeding of the IEEE Information Theory Workshop*, pages 1–5, Dublin, Ireland, 2010.