Coding and Signal Processing for Secure Wireless Communication

Li Chia Choo

Thesis submitted for the degree of

Doctor of Philosophy

of

University College London

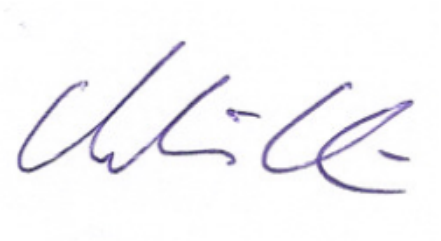
Department of Electronic and Electrical Engineering

University College London

2012

Declaration

I, Li Chia Choo, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.



_____

Abstract

Wireless communication networks are widely deployed today and the networks are used in many applications which require that the data transmitted be secure. Due to the open nature of wireless systems, it is important to have a fundamental understanding of coding schemes that allow for simultaneously secure and reliable transmission. The information theoretic approach is able to give us this fundamental insight into the nature of the coding schemes required for security.

The security issue is approached by focusing on the confidentiality of message transmission and reception at the physical layer. The goal is to design coding and signal processing schemes that provide security, in the information theoretic sense. In so doing, we are able to prove the simultaneously secure and reliable transmission rates for different network building blocks.

The multi-receiver broadcast channel is an important network building block, where the rate region for the channel without security constraints is still unknown. In the thesis this channel is investigated with security constraints, and the secure and reliable rates are derived for the proposed coding scheme using a random coding argument.

Cooperative relaying is next applied to the wiretap channel, the fundamental physical layer model for the communication security problem, and signal processing techniques are used to show that the secure rate can be improved in situations where the secure rate was small due to the eavesdropper enjoying a more favorable channel condition compared to the legitimate receiver.

Finally, structured lattice codes are used in the wiretap channel instead of unstructured random codes, used in the vast majority of the work so far. We show that lattice coding and decoding can achieve the secrecy rate of the Gaussian wiretap channel; this is an important step towards realizing practical, explicit codes for the wiretap channel.

Acknowledgements

Firstly, I wish to thank my supervisor, Dr. Kit Wong, for kindly allowing me to join his research group. I am thankful for his kind understanding, support and encouragement over these years; this thesis would have been much harder to complete without his help.

I would like to thank my collaborators, Dr. Ling Cong and Dr. Zheng Gan, who not only kindly shared many insights during our discussions, contributing much to my understanding of our research together, but shared their friendship as well. I would also like to thank Prof. Frederique Oggier who kindly arranged for an attachment with her research group where we had many insightful and helpful discussions that gave me a better understanding of coding theory.

This thesis is dedicated to my beloved wife Anne Chan and our daughter Ying Hui. Thank you for your love, patience and support; I thank God that we have managed to end this marathon journey stretching over so many years together. I also wish to thank our extended family for their wonderful support, and especially my in-laws the Chan family, for their fantastic help and support in looking after our daughter when I was away; our extended family definitely has a large part to play in the completion of this thesis.

I also wish to thank my friends and colleagues at UCL Adastral Park Campus, Dr. Jin Shi, Dr. Zhong Caijun, Dr. Chen Jia, Dr. Elsheikh Elsheikh and Dr. Sara Vicente for many fruitful and helpful discussions. Thanks also go to my colleagues at the Gower Street Lab, Agis and Hai Xia, for their help. Last but not least, thanks must go to friends at Bethesda Baptist Church in Ipswich for their care and friendship over the past four years.

Abbreviations

AF	Amplify and Forward
AWGN	additive white Gaussian noise
BC	broadcast channel
BCC	broadcast channel with confidential messages
BEC	binary erasure channel
BSC	binary symmetric channel
CJ	Cooperative Jamming
DF	Decode and Forward
DMC	discrete memoryless channel
DMS	degraded message sets
IC	infinite constellation
i.i.d.	independent and identically distributed
KKT	Karush-Kuhn-Tucker
LDPC	low density parity check
MAC	multiple access channel
ML	maximum likelihood
MLAN	modulo lattice additive noise
MRC	maximum ratio combining
MMSE	minimum mean square error
MSE	mean square error
p.d.f.	probability distribution function
p.m.f.	probability mass function
RC	relay channel
RHS	right hand side
r.v.	random variable

SNR	signal to noise ratio
TWRC	two way relay channel
VNR	volume to noise ratio
ZF	zero forcing

Notation

Ball(r)	n-dimensional ball of radius r
$\mathrm{conv}(\mathcal{R})$	the convex hull of set $\mathcal{C}$
dom(f)	domain of the function f
$\mathcal{E}^c$	the complement of event $\mathcal{E}$
$E_P(.)$	Poltyrev exponent
$E_P^r(.)$	random coding Poltyrev exponent
$E_P^x(.)$	expurgated Poltyrev exponent
G_n^*	normalised second moment of an n-dimensional sphere
$G(\Lambda)$	normalised second moment of lattice Λ
$h_\rho(Z)$	Renyi entropy of order ρ for random variable Z
$\mathbb{I}(.)$	indicator function
$\mathbf{I}_n$	$n \times n$ identity matrix
$\log$	logarithm to base 2, unless stated otherwise
$\Lambda^{(n)}$	sequence of lattices which are in the set of real vectors of length n
μ	normalised volume to noise ratio of lattice Λ
$\mu^*(\Lambda^{(n)})$	unnormalized volume to noise ratio of sequence of lattices $\Lambda^{(n)}$ in the set of real vectors of length n
$\mathrm{mod}_\Omega\ \Lambda$	modulo-lattice operation with respect to lattice Λ and a fundamental region Ω
$\mathrm{mod}\ \Lambda$	modulo-lattice operation with respect to lattice Λ and the fundamental Voronoi region $\mathcal{V}$
∇f	gradient of function f
$\nabla^2 f$	Hessian or second derivative of function f
Ω	a fundamental region of a lattice
P_e	error probability
$\mathcal{P}_k(\mathcal{A})$	k-subset of the set $\mathcal{A}$, the subset of $\mathcal{A}$ having exactly k elements

$Q_{\mathcal{V}}(.)$	quantizer associated with the Voronoi region $\mathcal{V}$ of lattice Λ
R_u	covering radius of a lattice
R_l	effective radius of a lattice
$\mathbb{R}$	set of real numbers
$\mathbb{R}^n$	set of real vectors of length n
$\sigma^2(\mathcal{V})$	minimized second moment of lattice Λ with Voronoi region $\mathcal{V}$
$\sim$	according to the distribution
$\mathcal{V}$	Voronoi region of a lattice
X	random variable X
x	realization of random variable X
$\mathcal{X}$	alphabet set of X
$\mathcal{X}^n$	n-th Cartesian power of $\mathcal{X}$, the set of n-length sequences of elements of $\mathcal{X}$
$\mathbf{X}$	sequence of n random variables by $(X_1, \ldots, X_n)$
$\mathbf{x}$	realization of the sequence of n random variables $(X_1, \ldots, X_n)$, $x_i \in \mathcal{X}$ for $i = 1, 2, \ldots, n$
$\mathbf{X}^i$	the subsequence of $\mathbf{X}$ defined as $(X_1, X_2, \ldots, X_i)$
$\tilde{\mathbf{X}}^i$	the subsequence of $\mathbf{X}$ defined as $(X_i, \ldots, X_n)$
$(.)^T$	transpose operation on the argument
$\|.\|$	Euclidean norm of the argument
$\succeq, \preceq$	generalized inequality; for vectors it denotes the element wise inequality, for symmetric matrices, it denotes matrix inequality

Contents

List of Figures

Chapter 1

Introduction

During the summer of 2006, police in Washington in the United States, tracked down a suspected criminal to the home of an elderly lady whose wireless router was 'hijacked' by the real culprit for his criminal activities [4]. Due to the rapidly increasing proliferation of mobile devices and services today, including such sensitive services as mobile online banking, this case truly sums up the fact that our wireless devices are very open to attack. The attacks faced by the mobile terminal include active attacks, where the malicious party takes steps to change the messages sent and received, such as our unfortunate case above. We also have passive attacks where the malicious party simply listens to and then decodes the sent messages, known as eavesdropping or wiretapping.

Common security issues include active attacks such as denial of service attacks by jamming, impersonation attacks, integrity attacks; and passive attacks such as eavesdropping and traffic analysis. In denial of service attacks, the attacker may use jamming signals derived from its knowledge of the legitimate transmitters' and receivers' codes and signals, or simply noise to occupy the transmitted signal band, and so disrupt communication. In impersonation attacks, the attacker will pretend to be a legitimate user and attempt to deceive the authentication system, and usually tries to capture the authentication codes. The attacker may also attempt to modify confidential messages or pass on confidential messages to other colluding nodes in an integrity attack. In eavesdropping, the attacker intercepts and tries to decode confidential messages sent over the channel, to either discover the sent messages or to determine the communication patterns of the legitimate users in the network. By analyzing the communication traffic, the attacker can obtain useful information on the legitimate users in the network.

As can be seen from the security issues mentioned above, we can identify several

common security requirements and how these can be achieved. Firstly, for resistance to jamming, we note that it is well known that some of the communications systems in wide use today already have inherent anti-jamming capabilities. These include the Global System for Mobile Communications (GSM), Code Division Multiple Access (CDMA) based systems and General Packet Radio Service (GPRS) systems in cellular mobile communications, and Orthogonal Frequency Division Multiplex (OFDM) systems used in wireless local area networks. The GSM, CDMA and GPRS systems all use spread spectrum techniques to combat multipath fading; at the same time, this gives the system an inherent resistance to jamming due to the wideband nature of the spread spectrum signal. In OFDM systems, the wideband nature of the transmitted signal again gives anti-jamming resistance. To counter jammers further, we should also design signals that have a low probability of detection or interception, which can be done using pseudorandom spreading sequences to modulate the useful information, since these spreading sequences are difficult to distinguish from white noise.

Secondly, message confidentiality can be secured by symmetric encryption. Here, it is assumed that the legitimate transmitter-receiver pair have a shared key (hence the name) to be kept secret from the attacker, which is used to convert the message into a cryptogram and *vice versa*, while the attacker cannot break the encryption in the time when the message can be intercepted. To achieve the encryption, operations depending on the secret key are applied to the message one symbol at a time or on a block of symbols. Examples of such encryption methods are the Data Encryption Standard and the Advanced Encryption Standard. A disadvantage that arises from using such encryption is that there are no mathematical proofs to exactly quantify the level of security. However, there are ways of testing the randomness of the transmitted encrypted message, the goal being to make the encrypted message close to perfectly random.

Thirdly, key distribution can be achieved using public key cryptography, which allows for a two-way conversation between the legitimate transmitter and receiver without them sharing a secret key. Here, each of the legitimate parties use a private key and a public key known to all parties, including any potential attackers. The transmitter encrypts the confidential message using the public key, which is decrypted by the legitimate receiver using his own private key. The public key is designed so that, without the knowledge of a private key, it is computationally hard for the attacker to decrypt the

message. The transmitter and legitimate receiver also perform authentication using the public key. An example of public key cryptography in practice is the well known RSA scheme. If we have a restriction on the receivers such that only a one-way transmission from the transmitter is desired, we can use broadcast encryption [82] instead. In this case a trusted agency produces special blocks of data called session key blocks to the transmitter and also assigns every receiver a set of keys. The transmitter processes a session key block to obtain a session key, then encrypts the message with the session key and sends the encrypted message, along with the session key block to the receiver. The receiver processes the session key block to obtain the session key and decrypt the message, and this processing is unique to each receiver. Secure communication is then established without a two-way communication between the legitimate parties. A disadvantage of both public key cryptography and broadcast encryption schemes' dependence on computational hardness for security is that modern computers with increasing computational power can conceivably break the cryptographic primitives used.

Fourthly, keys may need to be re-used several times, and this can be done by processing the original key with simple operations or modulating the original key with pseudorandom sequences. The disadvantages are that we might need additional protocols or architecture and sometimes trusted third parties; the probability of a successful attack increases every time a key is re-used.

Fifthly, authentication and integrity of messages can be ensured using hash functions. For the authentication problem, it is possible to design authentication tags, which are mappings of the message by a member of the universal family of hash functions, secure (with certain probability) against an attacker with infinite computing power. Another class of hash functions called one-way hash functions are used to encrypt messages into much shorter message digests which have very low correlation (as a function of the message) and are difficult to reverse, so as to ensure message integrity.

While security issues span a wide spectrum, the confidentiality of information exchanged by mobile terminals is an important one, since the information exchanged may be sensitive information such as bank or credit card details or protocol control information in the medium access layer of the system, which, when captured by the eavesdropper, can lead to further attacks. In wireless systems, an eavesdropper will be able to intercept and decode any messages exchanged by the legitimate users as long

as the eavesdropper is within range of the transmission. In principle, a well hidden eavesdropper is very hard to detect. Thus wireless systems are extremely vulnerable to the passive attacks. In our work, we want to characterize the simultaneously secure and reliable information rates for wireless systems in the presence of passive attacks, with the confidentiality of messages as the primary concern.

To maintain message confidentiality, a common approach is to use cryptographic encryption, where a transmitter uses a key to encrypt the message and the legitimate receiver uses the key to extract the message. The eavesdropper cannot extract the message if it does not have access to the key. Usually the eavesdropper is assumed to have limited time or computational resources to enable it to discover the key. The encryption code partly relies on computational security, so that it is 'unbreakable' in a limited time.

In encryption, key encryption algorithms are used, which in some form or other requires the secure key storage and distribution to the different users in the network (see for example, the articles in the Special Issue on Cryptography and Security of the Proceedings of the IEEE [87]). In wireless systems key encryption algorithms add complexity and computational resources, yet are vulnerable to the keys being intercepted by eavesdroppers. Furthermore the move towards mobile ad hoc networks and decentralized networks pose more challenges for key distribution and management. We also note that the encryption approach is designed to be insensitive to the characteristics of the communication channel and relies on computational hardness to provide security.

An information theoretic approach to the cryptographic security problem was introduced by Shannon [106], where the notion of provable information theoretic secrecy was introduced. The model considered by Shannon, shown in Figure 1.1, assumed noiseless links and an eavesdropper with unlimited resources. In Figure 1.1, the message W is encoded and transmitted as $\mathbf{Y}$, and since we have noiseless links, $\mathbf{Y}$ is received at the decoder and the eavesdropper as well. The secret key, K, is shared between encoder and decoder while the eavesdropper has no knowledge of the key. The secret key K represents the advantage that the legitimate transmitter-receiver pair has over the eavesdropper in the worst case. Shannon also showed that for perfect secrecy (no information leaking to the eavesdropper) the size of the key should be at least as large as the size of the message.

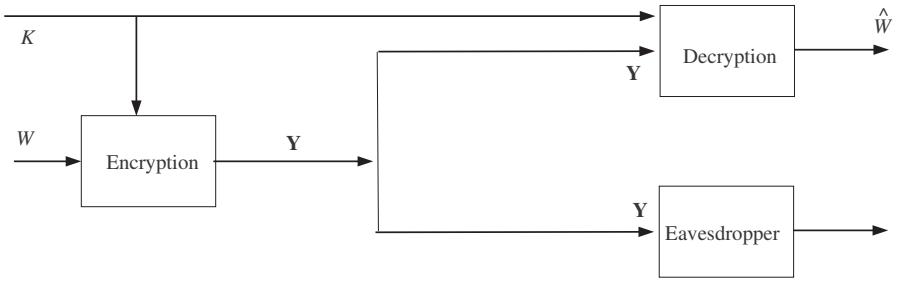


Figure 1.1: Shannon's secrecy system.

In the information-theoretic approach, the seminal works by Wyner [115] and Csiszár and Körner [30] later showed that by using stochastic encoding, secure communication was in fact possible without using key encryption in the presence of the eavesdropper with unlimited resources. The unlimited resources of the eavesdropper includes both computational resources as well as the knowledge of the coding, decoding and signal processing algorithms. The difference between the models in [30, 115] and Shannon's secrecy system in Figure 1.1 is that there is no shared key and communication takes place over noisy channels, and encryption and decryption should be replaced by encoding and decoding. This model is known as the wiretap channel and is shown in Figure 1.2. Thus, using channel codes and signal processing at the physical layer, it is possible to prove as shown in [30] and [115], that secure communication can be achieved, with simultaneous reliable communication at the legitimate receiver. The maximum transmission rate that is achievable is known as the secrecy capacity, which has been shown to be strictly positive whenever the eavesdropper's observation is 'noisier' than the legitimate receiver's.

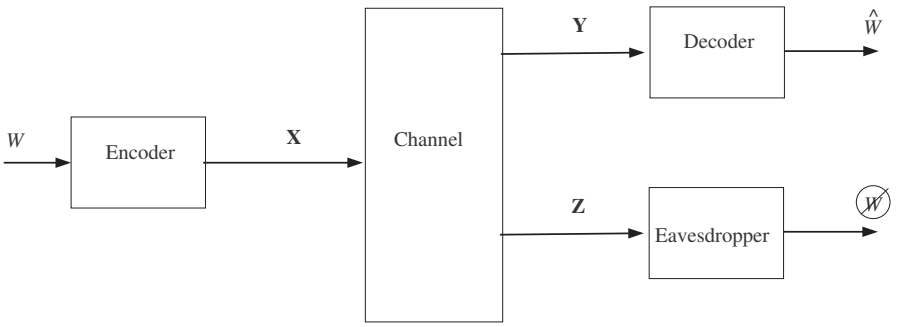


Figure 1.2: Wiretap channel.

We note the following assumptions in the approach of [30, 115]. Firstly, knowledge of the legitimate receiver's and the eavesdropper's channels are needed in the code design. For the case of the eavesdropper, this may not be a realistic assumption. However, there are ways to model the effects of imperfections in the channel estimate and its effects on the secure communication rate. Secondly, the attacker is assumed to be passive. For an active attacker, different coding techniques are required. Thirdly, uniform random bit sequences are assumed, which may not be available in practice. The information leaked to the eavesdropper is increased when bit sequences are not uniform. Fourthly, the legitimate transmitter-receiver pair is assumed to be authenticated to begin with; codes providing security in this context cannot provide security against impersonation attacks, but it is expected that authentication can be provided in the upper layers of the protocol stack.

The advantages of the channel coding information theoretic approach are that there are no computational restrictions on the eavesdropper and the information leakage can be quantified. The disadvantages are that, firstly, the analyses are based on average information measures, so although we can design a coding scheme that is secure with high probability, we may not be able to guarantee security with probability one. Secondly, the assumptions made about the communication channels may not be realistic in practice.

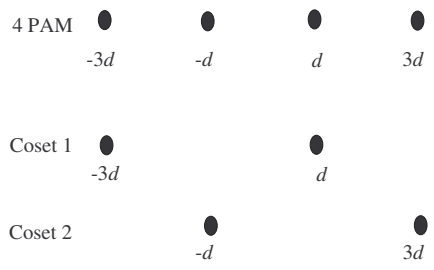


Figure 1.3: 4 PAM constellation and its two cosets.

Example on channel coding approach: We now give an example of how this channel coding might be achieved in practice in a communications system. We use a simple one-dimensional example, and consider the transmission of points from a constellation over an additive Gaussian noise eavesdropper's channel. The metric we will use is the probability of correct decoding $P_{c,e}$ for the eavesdropper, which is a weaker metric than

the information leakage, but will help us to to observe the benefits of secure channel coding. Naturally we want $P_{c,e}$ to go to zero for security. Specifically, the secure channel coding is known as coset coding, as proposed in [115]. Instead of a one-to-one mapping from the message symbols to the transmitted codeword (constellation point), we map each message symbol to a set of codewords (a coset), and randomly select one of the codewords in the coset to map to a constellation point for transmission.

We now assume we want to transmit a 1-symbol message, an element from the set $\{0,1\}$. A one-dimensional constellation is the Pulse Amplitude Modulation (PAM) constellation shown in Figure 1.3 for 4 PAM, where d is the distance from the origin. In the case of normal transmission, we map the set $\{0,1\}$ to the 2 PAM constellation (or binary phase shift keying), given by the central two points $[-d, d]$. For coset coding, we want to map each of 0 or 1 to a coset. In this case, we let each coset have two elements. So now we have to decompose a 4 PAM constellation into two cosets, as shown in Figure 1.3. The reasoning for the coset decomposition is as follows: for the set of integers $\mathbb{Z}$, its cosets are $2\mathbb{Z}$ and $2\mathbb{Z} + 1$, which means that the cosets are the set of integers scaled by 2, and the set of integers scaled by 2 and translated by one. Mapping this onto a 4 PAM constellation, we should have the first coset as the alternate points in the constellation, and the second coset as the first coset shifted by one constellation point. Suppose we now map $0 \mapsto$ Coset 1 and $1 \mapsto$ Coset 2, which means that $0 \mapsto \{-3d, d\}$ and $1 \mapsto \{-d, 3d\}$. Then we choose a point at random in the coset and transmit it. We now analyze the probability of correct detection at the eavesdropper. We consider the following model for the Gaussian wiretap channel

$$z = x + n_e, \tag{1.1}$$

where x is the transmitted symbol, n_e is the Gaussian noise at the eavesdropper, with zero mean and two-sided power spectrum density $N_0/2$. We let the energy per bit be E_b. The energy per bit (symbol) for the constellation mapped to cosets 1 and 2 are both $\frac{d^2}{2}(1+9) = 5d^2$. For the symbol 0 (from coset 1), the constellation points $-3d$ or d are in error if they are decoded to $-d$ or $3d$. That is, they are decoded to the constellation points in coset 2. Similarly, for the symbol 1, the constellation points $-d$ or $3d$ are in error if they are decoded to d or $-3d$. We focus on symbol 0 (coset 1). From Figure 1.3, the constellation point $-3d$ is in error if the noise exceeds d or $5d$, but is not in

error if the noise exceeds $3d$. Thus, following [19], the bit error probability for this constellation point is

$$\begin{aligned} P_{e,0,1} &= \frac{1}{2}\text{erfc}\left(\frac{d}{\sqrt{N_0}}\right) + \frac{1}{2}\text{erfc}\left(\frac{5d}{\sqrt{N_0}}\right) - \frac{1}{2}\text{erfc}\left(\frac{3d}{\sqrt{N_0}}\right) \\ &= \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E_b}{5N_0}}\right) + \frac{1}{2}\text{erfc}\left(5\sqrt{\frac{E_b}{5N_0}}\right) - \frac{1}{2}\text{erfc}\left(3\sqrt{\frac{E_b}{5N_0}}\right). \end{aligned} \tag{1.2}$$

Here the complementary error function is defined as

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}}\int_x^\infty e^{-u^2}du. \tag{1.3}$$

Next, the constellation point d in coset 1 is in error if the noise exceeds d in either direction, but is not in error if the noise exceeds $3d$. The error probability for this constellation point is

$$\begin{aligned} P_{e,0,2} &= 2.\frac{1}{2}\text{erfc}\left(\frac{d}{\sqrt{N_0}}\right) - \frac{1}{2}\text{erfc}\left(\frac{3d}{\sqrt{N_0}}\right) \\ &= \text{erfc}\left(\sqrt{\frac{E_b}{5N_0}}\right) - \frac{1}{2}\text{erfc}\left(3\sqrt{\frac{E_b}{5N_0}}\right). \end{aligned} \tag{1.4}$$

The error probability for symbol 0 mapping to coset 1 is then $P_{e,0} = \frac{1}{2}P_{e,0,1} + \frac{1}{2}P_{e,0,2}$. By symmetry, the error probability for symbol 1 mapping to coset 2 has a similar expression. Thus the probability of correct detection at the eavesdropper for the coset code is

$$P_{c,e} = 1 - \frac{1}{4}\left[3\text{erfc}\left(\sqrt{\frac{E_b}{5N_0}}\right) + \text{erfc}\left(5\sqrt{\frac{E_b}{5N_0}}\right) - 2\text{erfc}\left(3\sqrt{\frac{E_b}{5N_0}}\right)\right]. \tag{1.5}$$

If we did not use coset coding, but sent the information symbols using normal 2-PAM or BPSK instead, then the probability of correct detection at the eavesdropper will be the well known

$$P_{c,e} = 1 - \frac{1}{2}\text{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right). \tag{1.6}$$

To illustrate the benefit of using coset coding, we plot the $P_{c,e}$ for symbols sent using coset coding and BPSK in Figure 1.4. We can observe a region where $P_{c,e}$ is lowered using coset decoding, so we can already see the benefits of coset coding using this simple one-dimensional example. We also observe that when E_b/N_0 is too high or too low, there is no benefit in using coset coding in terms of lowering $P_{c,e}$. Also, $P_{c,e}$ is still quite high for reasonable values of E_b/N_0, and this can be lowered (ideally to 0)

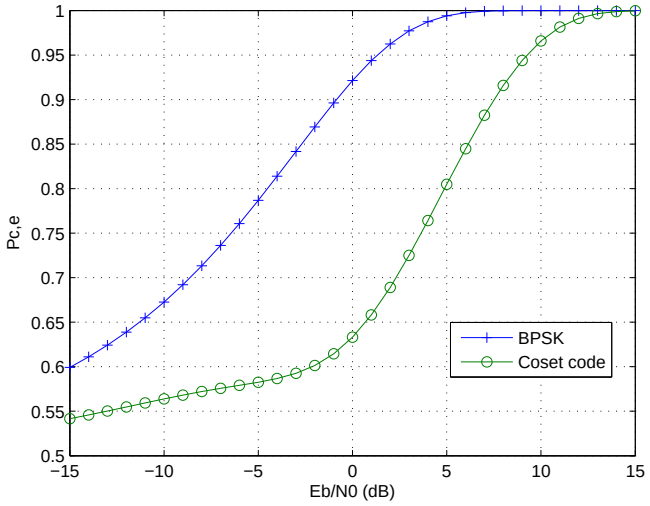


Figure 1.4: $P_{c,e}$ for coset code $\mathbb{Z}/2\mathbb{Z}$ and 2-PAM (BPSK).

by using multidimensional lattice coding in high dimensions, as shown in [97], or later in Chapter 5.

Using the combined key encryption-channel coding information theoretic approach, Ahlswede and Csiszár [1] and Maurer [85] showed that legitimate users in a (possibly wireless) network can agree on a secret key used for encryption later, and this key can be kept secret from an eavesdropper with unlimited resources. Specifically, the legitimate receiver and the eavesdropper observe outputs of a discrete memoryless BC which has inputs controlled by the transmitter. The legitimate transmitter-receiver pair have sources of local randomness and communicate with each other over a public noiseless channel, while the BC communicates messages and is used to generate randomness. It was shown in [1, 85] that a wiretap code of [115] is an optimal key distillation strategy and we can send secret keys uniformly to the legitimate receiver without using the public channel. Also, the key may be interpreted as a secret message, and feedback improves the secret key (message) capacity even if the feedback is known by the eavesdropper. This means that the need for an advantage over the eavesdropper in the models of [30, 115] are due to the limits of the coding schemes.

The wireless medium can be used as a source of randomness and exploited in an opportunistic way using signal processing schemes to guarantee secrecy even if the eavesdropper has, on average, a better signal to noise ratio (SNR) than the legitimate receiver. We can use multiple antennas at the legitimate transmitter-receiver pair and

beamforming in the direction that the eavesdropper either does not receive any useful signal, or where the eavesdropper gets a lower SNR than the legitimate receiver. In this case the secrecy capacity is positive as long as the number of antennas at the eavesdropper does not exceed the number of antennas at the legitimate transmitter-receiver pair. Another scheme is the artificial noise strategy, where we send information in the direction of the nonsingular values of legitimate user's channel matrix and noise in all other directions. This strategy is semi-blind in that we need knowledge of the eavesdropper's channel matrix in the code design. Lastly, we can use cooperative jamming where perhaps more than one trusted transmitters send coded jamming signals to increase the confusion at the eavesdropper, and signal processing schemes can be used to increase the secrecy capacity.

In a block fading fading channel environment where each codeword experiences several channel realizations, secure communication is determined by the instantaneous fading realization. We can have positive secrecy capacity for any transmit power and channel statistics, provided that the probability that the legitimate receiver's channel gain exceeds the eavesdropper's channel gain, is positive. The coding or transmission scheme is for the transmitter to send messages during the time instants when the eavesdropper has a lower SNR than the legitimate receiver. Thus, even if the eavesdropper has an average SNR better than the legitimate receiver, it is possible to attain a positive secrecy capacity, illustrating that fading actually helps in securing communications. The secrecy capacity is limited by the knowledge of channel state information. In the absence of knowledge of the eavesdropper's channel state information, we can design a coding or transmission scheme that allocates power to the transmitter according to the probability that the legitimate receiver's channel gain is above some threshold to be adjusted. In this case, we can achieve a reduced, but still positive secure rate. In a quasi-static fading environment, where the channel fading coefficients are constant over a codeword, but change from one codeword to another, then we can only set a target secure rate and design a coding scheme for it, but have to accept that only at the times when the legitimate transmitter-receiver pair enjoys a clear advantage over the eavesdropper, then we can have positive secrecy capacity.

Finally, we should remark that information theoretic physical layer security can be deployed in networks as an additional security in conjunction with existing cryp-

tographic schemes to enhance the overall security of the system. We can replace the usual channel codes and modulation schemes, which ensure reliability but not secrecy, with ones based on wiretap codes of [30, 115]. Such codes can, at the cost of some loss of transmission rate and increased complexity, achieve both reliability and security, and can significantly degrade the eavesdropper's observation of the useful message when the legitimate transmitter-receiver pair have a better SNR than the eavesdropper. In the case when this condition cannot be met, we could jam the eavesdropper in the areas where he might be located. By implementing wiretap codes in a modular fashion in the network architecture independently of the higher layer cryptographic schemes, we can also enhance the overall security by ensuring that the eavesdropper does not have access to an error free copy of the cryptogram.

In our work, we shall follow the information-theoretic approach of [115] and [30] to characterize simultaneously secure and reliable information rates for wireless systems without using key agreement with different types of network architecture. Thus the aims of our research are as follows:

1. To find fundamental limits on the reliable and secure (confidential) information rates for multi-user wireless network building blocks and models;

2. To discover channel coding schemes that achieve the reliable and secure (confidential) information rates, using random, practical or structured codes;

3. To use signal processing schemes to enhance security where a very low or no secure rate is possible, or to mitigate the effects of the eavesdropper.

The broadcast channel (BC) and the relay channel (RC) are ubiquitous network building blocks in wireless networks today. In Figure 1.5 we show these building blocks at the node level.

The BC (depicted in Figure 1.5(a) for a source $\mathtt{S}$ and 2 receiver nodes $\mathtt{D}_1$ and $\mathtt{D}_2$) models the downlink from a base station to mobile terminals, and is therefore a key element in the study of multi-user networks, which helps us to see how we can carry out the channel coding. The extension of channel coding schemes to the BC with three or more receivers is still an open problem, with or without security constraints. The RC is yet another important network building block, and it is depicted in Figure 1.5(b).

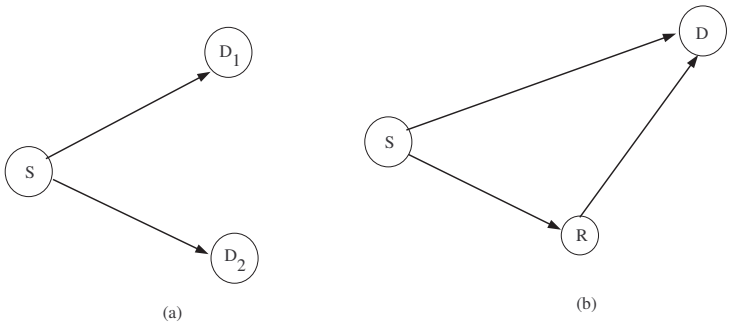


Figure 1.5: (a) A broadcast channel, (b) a relay channel.

In the RC the source S sends a message to the destination D and is helped by the relay node R, which cooperatively transmits the message to the destination together with the transmitter. We will focus on the security issue for these network building blocks in the thesis. We will assume that the the legitimate receiver has an advantage over the eavesdropper, in that the eavesdropper's channel is noisier than the legitimate receiver's. We also assume knowledge of the legitimate receiver's and the eavesdropper's channels, a passive eavesdropper, uniform random bit sequences as sources of randomness, and that authentication has already been performed. We shall see later that the results suggest a multilevel code construction.

In the next section we will summarize the contributions of the thesis and provide an outline of the rest of the thesis.

1.1 Contributions and Thesis Outline

In this thesis we will use the provable, information theoretic approach to the message confidentiality problem and aim to find coding and signal processing schemes for the BC and RC, which are ubiquitous network building blocks in wireless networks today. We also aim to construct practical or structured codes to implement the secure coding in the presence of an eavesdropper, still following the information theoretic approach.

In Chapter 2, we give an overview of some information theoretic and mathematical notions that will be helpful in the understanding of the thesis. In Chapter 3, we review some basic results in information theoretic security, such as Wyner's wiretap channel [115] and Csiszár and Körner's BC with confidential messages [30], the RC with eavesdropper, and the fundamental coding technique for security.

In Chapter 4, we study the BC under security constraints, focusing on 3 or more receivers. In particular, we found the secure coding scheme for the K-receiver degraded BC with private messages sent to each receiver from the transmitter and each to be kept secret from an eavesdropper. This channel had a degradedness condition that meant that receiver Y_1 was able to decode everything receivers Y_2 to Y_K were able to, receiver Y_2 was able to decode everything receivers Y_3 to Y_K were able to, and so on, with the eavesdropper the weakest among the receivers. We derived the secrecy capacity, which is the maximum secure rate, with matching inner and outer bounds, using superposition coding and code partitioning as our secure scheme. Next, we found the secure scheme for a class of the general 3-receiver BC with degraded message sets, which involved rate splitting, double binning, superposition coding and code partitioning. We found an achievable inner bound for the case of 2 degraded message sets, where there is one private and one common message; outer bounds were also found for special cases.

In Chapter 5, we focus on using signal processing techniques to mitigate the effect of the eavesdropper in the RC with eavesdropper setting. We were able to find an optimal solution to the power allocation problem for a bank of relays sending a jamming signal that also interfered with the desired transmission; a distributed implementation was also proposed. In this chapter we were able to prove the fundamental result that relays performing cooperative jamming can improve the secrecy rate under channel conditions that were unfavorable to the legitimate receiver, which meant that the secrecy rate was very low or zero.

In Chapter 6, we turn our attention to constructing information theoretically secure coding schemes for the Gaussian wiretap channel, a degraded wiretap channel. We used nested lattice codes to implement the coset coding, and were able to show a construction that achieved the secrecy rate (secrecy capacity) of the Gaussian wiretap channel. Along the way, we showed that the information leakage to the eavesdropper was small for large block length, thus showing that our construction is information theoretically secure.

In Chapter 7, we give our conclusions and identify directions for future work.

Lastly, the contributions in this thesis resulted in the following publications:

1. L. C. Choo and K. K. Wong, 'The K-receiver broadcast channel with confidential messages', submitted to *IEEE Transactions on Information Theory*, Dec. 2008.

2. G. Zheng, L. C. Choo and K. K. Wong, 'Optimal cooperative jamming to enhance physical layer security using relays', *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1317-1322, Mar. 2011.

3. L. C. Choo and K. K. Wong, 'On the 3-receiver broadcast channel with degraded message sets and confidential messages', to be submitted.

4. L. C. Choo and K. K. Wong, 'Three-receiver broadcast channel with confidential messages', *10th International Symposium on Communication Theory and Applications*, Ambleside, UK, 13–17 July 2009.

5. L. C. Choo and K. K. Wong, 'Physical layer security for a 3-receiver broadcast channel with degraded message sets', *International Conference on Wireless Communications and Signal Processing 2009*, 13–15 Nov., Nanjing, China, 2009.

6. L. C. Choo, C. Ling and K. K. Wong, 'Achievable Rates for Lattice Coded Gaussian Wiretap Channels', Proc. *IEEE International Conference on Communications (ICC 2011)*,Kyoto, Japan, 5–9 June, 2011.

Chapter 2

Information Theory and Mathematical Preliminaries

In this Chapter, we introduce information theoretic notions that will be used in the rest of the thesis. We will state some essential definitions and theorems. In particular, we will define information theoretic notions such as the entropy, mutual information and typical sequences. The theorems will be stated without proof in general; when extensive use of a particular theorem is needed in the thesis, a proof will be given. For a comprehensive treatment of information theory in general, the reader may wish to consult the reference books by Cover and Thomas [26], Csiszár and Körner [29], Kramer, [67], Gallager [49] and Yeung [117].

2.1 Information Theoretic Notions

Besides the entropy and mutual information, we will introduce the notion of typical sequences. Consider an information source $\{X_i, i = 1, .., n\}$, where X_i are i.i.d. $\sim p(x)$, and let the entropy of the generic r.v. X be denoted as $H(X)$. When n is large, the sequence drawn will have sample entropy close to the true entropy, which is called the typical set. In particular, the probability that the sequence drawn occurs is close to $2^{-nH(X)}$ with high probability, and the total number of typical sequences is about $2^{nH(X)}$. Thus the set of all sequences can be divided into the typical set, and the non-typical set. The typical set then determines the behaviour of the large sample; in the case of the information source $\{X_i, i = 1, .., n\}$ the typical set determines the behaviour.

Joint typicality decoding is usually used to prove coding theorems, as an easier

alternative to using maximum likelihood (ML) decoding. In joint typicality decoding, we look for the codeword that is joint typical with the received sequence. By joint typicality discussed later in Section 2.1.2, the received sequence and codeword will be joint typical with high probability, so that the non-typical sequences will not be decoded, and the probability of error will be small. We should remark that joint typicality decoding is suboptimal, but can still achieve all rates below capacity. See Gallager [49] for an analysis of coding theorems using ML decoding.

In the following we make the distinction between two versions of joint typical sequences, namely, letter-typical sequences (or simply typical sequences) and entropy-typical sequences (or weakly typical sequences) [26, Ch. 3]. Letter-typical sequences are the sequences where the relative frequency of each outcome (of the $\{X_i, i = 1, .., n\}$) is close to the corresponding probability. Letter-typicality is restricted to r.v.s with finite alphabets, but can be used to evaluate the joint typicality when one or more of the variables is fixed. Entropy-typicality can be used for discrete as well as continuous r.v.s, but it cannot evaluate the joint typicality when one or more of the variables is fixed. In the thesis we will use letter-typical sequences, which we will denote as *typical sequences*. We note that the class of strong typical sequences (as elaborated by Csiszár and Körner [29] and Yeung [117]) is included in the class of letter-typical sequences.

Logarithms are taken to base 2 or base e. When taken to base 2, the units are in bits; when taken to base e, the units are in natural units (nats). The number of nats is the number of bits multiplied by $\ln 2$. In the thesis, we will assume logarithms are taken to base 2, unless otherwise stated.

2.1.1 Entropy and Mutual Information

Entropy

For X a discrete random variable (r.v.) with probability mass function (p.m.f.) $p(x) = Pr(X = x), x \in \mathcal{X}$, where $\mathcal{X}$ is the alphabet, then the entropy is

$$H(X) = -\sum_{x \in \mathcal{X}} p(x) \log p(x). \tag{2.1}$$

The units are in bits. The conditional entropy of one r.v. Y given another X is the expected value of the entropies of the conditional distributions, averaged over the con-

ditioning r.v.:

$$\begin{aligned} H(Y|X) &= \sum_{x\in\mathcal{X}} p(x)H(Y|X=x) \\ &= -\sum_{x\in\mathcal{X}} p(x)\sum_{y\in\mathcal{Y}} p(x|y)\log p(x|y) \\ &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x|y) \\ &= -E[\log p(Y|X)]. \end{aligned} \tag{2.2}$$

The joint entropy of X and Y with joint distribution $p(x,y)$ is

$$\begin{aligned} H(X,Y) &= -\sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} p(x,y)\log p(x,y) \\ &= -E[\log p(X,Y)]. \end{aligned} \tag{2.3}$$

We have the following properties:

$$0 \le H(X) \le \log|\mathcal{X}|, \tag{2.4}$$

$$0 \le H(Y|X) \le \log|\mathcal{Y}|, \tag{2.5}$$

$$0 \le H(X,Y) \le \log(|\mathcal{X}|.|\mathcal{Y}|). \tag{2.6}$$

We note that $H(Y|X)=0$ if and only if for every $x\in\mathcal{X}$ there is a y so that $p(y|x)=1$. In this situation it is said that X determines Y.

The joint entropy may be expanded using the *chain rule*

$$H(X,Y) = H(X) + H(X|Y). \tag{2.7}$$

In general, we have

$$\begin{aligned} H(X_1,X_2,...,X_n) &= H(X_1) + H(X_2|X_1) + ... + H(X_n|X_1,X_2,...,X_{n-1}) \\ &= \sum_{i=1}^{n} H(X_i|\mathbf{X}^{i-1}). \end{aligned} \tag{2.8}$$

Finally, if $g(.)$ is a function whose domain is the range of X, then we have

$$H(X) \ge H(g(X)), \tag{2.9}$$

that is the entropy of X is greater than or equal to the entropy of a function of X.

Mutual Information

For two r.v.s X, Y with joint p.m.f. $p(x, y)$ and marginal p.m.f.s $p(x), p(y)$, the mutual information $I(X;Y)$ is the relative entropy between the joint distribution and the product distribution $p(x)p(y)$:

$$I(X;Y) = \sum_{x\in\mathcal{X}} \sum_{y\in\mathcal{Y}} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}. \tag{2.10}$$

The mutual information $I(X;Y)$ may also be defined as

$$I(X;Y) = H(X) - H(X|Y). \tag{2.11}$$

We have the following inequalities:

$$I(X;Y) \geq 0, \tag{2.12}$$

$$H(X,Y) \leq H(X) + H(Y), \tag{2.13}$$

$$H(X|Y) \leq H(X), \tag{2.14}$$

with equality if and only if X and Y are statistically independent. The inequality (2.14) means that conditioning cannot increase the entropy. However, $H(X|Y = y)$ may be greater than $H(X)$.

The mutual information may also be expanded using the chain rule:

$$\begin{aligned} I(X_1, X_2, ..., X_n; Y) &= I(X_1;Y) + I(X_2;Y|X_1) + ... + I(X_n;Y|X_1, X_2, ..., X_{n-1}) \\ &= \sum_{i=1}^{n} I(X_i;Y|\mathbf{X}^{i-1}). \end{aligned} \tag{2.15}$$

The conditional mutual information of r.v.s X and Y given Z is

$$\begin{aligned} I(X;Y|Z) &= H(X|Z) - H(X|Y,Z) \\ &= E_{p(x,y,z)} \log \frac{p(X,Y|Z)}{p(X|Z)p(Y|Z)}, \end{aligned} \tag{2.16}$$

where the expectation is taken over the joint p.m.f. $p(x, y, z)$.

If the r.v.s X, Y, Z form a Markov chain $X \to Y \to Z$, we have the *data processing inequality*

$$I(X;Y) \geq I(X;Z). \tag{2.17}$$

This implies that further processing of Y will not increase the information Y carries about X. Furthermore, we also have the inequality

$$I(X;Y|Z) \leq I(X;Y). \tag{2.18}$$

If $Z = g(Y)$, then $I(X;Y) \geq I(X;g(Y))$.

Finally, we have the useful identity commonly referred to as the *Csiszár sum identity* found in Csiszár and Körner [30, Lemma 7]: For random vectors $\mathbf{X}$ and $\mathbf{Y}$ with arbitrary joint distribution,

$$\sum_{i=1}^{n} I(\tilde{\mathbf{X}}^{i+1}; Y_i|\mathbf{Y}^{i-1}) = \sum_{i=1}^{n} I(\mathbf{Y}^{i-1}; X_i|\tilde{\mathbf{X}}^{i+1}). \tag{2.19}$$

A proof may be found in [30].

2.1.2 Letter-typical Sequences

In this subsection we introduce the notion used here of letter-typical sequences (referred to as typical sequences from here), for discrete r.v.s.[1]. Let $N(x|\mathbf{x})$ be defined as the number of occurrences of x over the alphabet $\mathcal{X}$. That is,

$$N(x|\mathbf{x}) \triangleq |\{i : x_i = x\}|, \quad x \in \mathcal{X}. \tag{2.20}$$

For $\epsilon \geq 0$, a sequence $\mathbf{x}$ is typical with respect to $p(x)$ and denoted as $\mathcal{T}_\epsilon^n(P_X)$ if

$$\mathcal{T}_\epsilon^n(P_X) \triangleq \left\{ \left| \frac{1}{n} N(x|\mathbf{x}) - p(x) \right| \leq \epsilon.p(x) \quad \text{for all } x \in \mathcal{X} \right\}. \tag{2.21}$$

For the corresponding conditions for *strong typical sequences*, replace the right hand side of (2.21) with ϵ and add the condition that $N(x|\mathbf{x}) = 0$ if $p(x) = 0$.

Typical sequences have properties captured in the following theorem:

Theorem 1. *Let $\epsilon \geq 0$, $\epsilon \to 0$ for n sufficiently large, with $\mathbf{x} \in \mathcal{T}_\epsilon^n(P_X)$ and $\mathbf{X} \sim p(x)$. We have*

$$2^{-n(1+\epsilon)H(X)} \leq p(\mathbf{x}) \leq 2^{-n(1-\epsilon)H(X)}, \tag{2.22}$$

$$(1-\gamma)2^{n(1-\epsilon)H(X)} \leq |\mathcal{T}_\epsilon^n(P_X)| \leq 2^{n(1+\epsilon)H(X)}, \tag{2.23}$$

$$1 - \gamma \leq \Pr[\mathbf{X} \in \mathcal{T}_\epsilon^n(P_X)] \leq 1, \tag{2.24}$$

where $\gamma \to 0$ as $\epsilon \to 0$ for n sufficiently large. □

We note then that $\Pr[\mathbf{X} \in \mathcal{T}_\epsilon^n(P_X)] \to 1$ for n sufficiently large.

[1] For full details see Kramer [67] or El Gamal and Kim [35].

Jointly Typical Sequences

Let $N(x,y|\mathbf{x},\mathbf{y})$ be the number of times the pair (x,y) occurs in sequence $(x_1,y_1),...,(x_n,y_n)$, that is

$$N(x,y|\mathbf{x},\mathbf{y}) \triangleq |\{i:(x_i,y_i)=(x,y)\}| \quad (x,y)\in\mathcal{X}\times\mathcal{Y}. \tag{2.25}$$

The *jointly typical set* with respect to $p(x,y)$ is

$$\mathcal{T}_\epsilon^n(P_{XY}) = \left\{(\mathbf{x},\mathbf{y}) : \left|\frac{1}{n}N(x,y|\mathbf{x},\mathbf{y}) - p(x,y)\right| \le \epsilon.p(x,y) \quad \text{for all } (x,y)\in\mathcal{X}\times\mathcal{Y}\right\}. \tag{2.26}$$

We should note that $(\mathbf{x},\mathbf{y})\in\mathcal{T}_\epsilon^n(P_{XY})$ means that $\mathbf{x}\in\mathcal{T}_\epsilon^n(P_X)$ and $\mathbf{y}\in\mathcal{T}_\epsilon^n(P_Y)$. For conditional distribution $p(y|x)$ define $p(\mathbf{y}|\mathbf{x})=\prod_{i=1}^n p(y_i|x_i)$, and

$$\mathcal{T}_\epsilon^n(P_{XY}|\mathbf{x}) = \{\mathbf{y}:(\mathbf{x},\mathbf{y})\in\mathcal{T}_\epsilon^n(P_{XY})\}, \tag{2.27}$$

where $\mathcal{T}_\epsilon^n(P_{XY}|\mathbf{x})=\emptyset$ if $\mathbf{x}\notin\mathcal{T}_\epsilon^n(P_X)$. The following theorem generalizes Theorem 1 for conditional typical sequences:

Theorem 2. *Let $\epsilon_1,\epsilon_2\ge 0$, $\epsilon_1<\epsilon_2$, $(\mathbf{x},\mathbf{y})\in\mathcal{T}_\epsilon^n(P_{XY})$ and $(\mathbf{X},\mathbf{Y})\sim p(x,y)$. Then*

$$2^{-n(1+\epsilon_1)H(Y|X)} \le p(\mathbf{y}|\mathbf{x}) \le 2^{-n(1-\epsilon_1)H(Y|X)}, \tag{2.28}$$

$$(1-\gamma)2^{n(1-\epsilon_2)H(Y|X)} \le |\mathcal{T}_{\epsilon_2}^n(P_{XY}|\mathbf{x})| \le 2^{n(1+\epsilon_2)H(Y|X)}, \tag{2.29}$$

$$1-\gamma \le \Pr[\mathbf{Y}\in\mathcal{T}_{\epsilon_2}^n(P_{XY}|\mathbf{x})|\mathbf{X}=\mathbf{x}] \le 1, \tag{2.30}$$

where $\gamma\to 0$ as $\epsilon_1,\epsilon_2\to 0$ for n sufficiently large. □

We now have the following theorem concerning the probability that $\mathbf{Y}$ is jointly typical with respect to $p(x,y)$, given $\mathbf{x}$:

Theorem 3. *For a joint distribution $p(x,y)$ and $\epsilon_1,\epsilon_2\ge 0$, $\epsilon_1<\epsilon_2$, $\mathbf{Y}\sim p(y)$ and $\mathbf{x}\in\mathcal{T}_{\epsilon_1}^n(P_X)$, we have*

$$(1-\gamma)2^{-n[I(X;Y)+2\epsilon_2 H(Y)]} \le \Pr[\mathbf{Y}\in\mathcal{T}_{\epsilon_2}^n(P_{XY}|\mathbf{x})] \le 2^{-n[I(X;Y)-2\epsilon_2 H(Y)]}, \tag{2.31}$$

where $\gamma\to 0$ as $\epsilon_1,\epsilon_2\to 0$ for n sufficiently large.

Proof. For the upper bound, we have

$$\begin{aligned}\Pr[\mathbf{Y}\in\mathcal{T}_{\epsilon_2}^n(P_{XY}|\mathbf{x})] &= \sum_{\mathbf{y}\in\mathcal{T}_{\epsilon_2}^n(P_{XY}|\mathbf{x})} p(\mathbf{y}) \\ &\overset{(a)}{\le} 2^{nH(Y|X)(1+\epsilon_2)}2^{-nH(Y)(1-\epsilon_2)} \\ &\le 2^{-n[I(X;Y)-2\epsilon_2 H(Y)]},\end{aligned} \tag{2.32}$$

where (a) is by using (2.28) and (2.29). The lower bound may be proved similarly. □

A random version of Theorem 3 with regard to the probability of $(\mathbf{X}, \mathbf{Y})$ being joint typical with respect to $p(x, y)$ is stated as

Theorem 4. *For a joint distribution $p(x, y)$ and $\epsilon \geq 0$, $\mathbf{X} \sim p(x)$ and $\mathbf{Y} \sim p(y)$, we have*

$$(1-\gamma)2^{-n[I(X;Y)+3\epsilon H(X,Y)]} \leq \Pr[(\mathbf{X}, \mathbf{Y}) \in \mathcal{T}_\epsilon^n(P_{XY})] \leq 2^{-n[I(X;Y)-3\epsilon H(X,Y)]}, \quad (2.33)$$

where $\gamma \to 0$ as $\epsilon \to 0$ for n sufficiently large.

Proof. For the upper bound, we have

$$\begin{aligned} \Pr[(\mathbf{X}, \mathbf{Y}) \in \mathcal{T}_\epsilon^n(P_{XY})] &= \sum_{(\mathbf{x},\mathbf{y}) \in \mathcal{T}_\epsilon^n(P_{XY})} p(\mathbf{x})p(\mathbf{y}) \\ &\overset{(a)}{\leq} 2^{nH(X,Y)(1+\epsilon)} 2^{-nH(X)(1-\epsilon)} 2^{-nH(Y)(1-\epsilon)} \\ &\leq 2^{-n[I(X;Y)-3\epsilon H(X,Y)]}. \end{aligned} \quad (2.34)$$

The lower bound may be proved similarly. □

If the r.v.s X, Y, Z form the Markov chain $X \to Y \to Z$, we have the *Markov lemma* for the probability that $\mathbf{Z}$ is joint typical with respect to $p(x, y, z)$, given $\mathbf{Y} = \mathbf{y}$:

Lemma 1. *Markov lemma [10]: Let $\epsilon_1, \epsilon_2 \geq 0$, $\epsilon_1 < \epsilon_2$, $(\mathbf{x}, \mathbf{y}) \in \mathcal{T}_{\epsilon_1}^n(P_{XY})$, and $(\mathbf{X}, \mathbf{Y}, \mathbf{Z}) \sim p(x, y, z)$. Then*

$$\Pr[\mathbf{Z} \in \mathcal{T}_{\epsilon_2}^n(P_{XYZ}|\mathbf{x}, \mathbf{y})|\mathbf{Y} = \mathbf{y}] \geq 1 - \gamma, \quad (2.35)$$

where $\gamma \to 0$ as $\epsilon_1, \epsilon_2 \to 0$ for n sufficiently large.

Proof. It can be easily seen that

$$\begin{aligned} \Pr[\mathbf{Z} \in \mathcal{T}_{\epsilon_2}^n(P_{XYZ}|\mathbf{x}, \mathbf{y})|\mathbf{Y} = \mathbf{y}] &\overset{(a)}{=} \Pr[\mathbf{Z} \in \mathcal{T}_{\epsilon_2}^n(P_{XYZ}|\mathbf{x}, \mathbf{y})|\mathbf{X} = \mathbf{x}, \mathbf{Y} = \mathbf{y}] \\ &\overset{(b)}{\geq} 1 - \gamma, \end{aligned} \quad (2.36)$$

where (a) is by the Markov chain condition $X \to Y \to Z$ and (b) is by (2.30). □

Finally, for r.v.s U, X, Y following $U \to X \to Y$, a useful conditional typicality bound concerning the probability of $\mathbf{X}$ being joint typical with respect to $p(u, x, y)$ given $\mathbf{U} = \mathbf{u}$ is

Theorem 5. *For $\epsilon_1, \epsilon_2 \geq 0$, $\epsilon_1 < \epsilon_2$, $X_i \sim p(x_i|u_i)$ for all $i = 1, 2, ..., n$, and $(\mathbf{u}, \mathbf{y}) \in \mathcal{T}^n_{\epsilon_1}(P_{UY})$, we have*

$$(1-\gamma)2^{-n[I(X;Y|U)+2\epsilon_2 H(X|U)]} \leq \Pr[\mathbf{X} \in \mathcal{T}^n_{\epsilon_2}(P_{UXY}|\mathbf{U}=\mathbf{u})] \leq 2^{-n[I(X;Y|U)-2\epsilon_2 H(X|U)]}, \tag{2.37}$$

where $\gamma \to 0$ as $\epsilon_1, \epsilon_2 \to 0$ for n sufficiently large.

Proof. For the upper bound, we have

$$\begin{aligned}\Pr[\mathbf{X} \in \mathcal{T}^n_{\epsilon_2}(P_{UXY}|\mathbf{U}=\mathbf{u})] &= \sum_{\mathbf{x} \in \mathcal{T}^n_{\epsilon_2}(P_{UXY}|\mathbf{u},\mathbf{y})} p(\mathbf{x}|\mathbf{u}) \\ &\overset{(a)}{\leq} 2^{nH(X|U,Y)(1+\epsilon_2)} 2^{-nH(X|U)(1-\epsilon_2)} \\ &\leq 2^{-n[I(X;Y|U)-2\epsilon_2 H(X|U)]},\end{aligned} \tag{2.38}$$

where (a) is by using (2.28) and (2.29). The lower bound may be proved similarly. □

2.1.3 Inequalities

In this section, we state some useful inequalities which we need in the thesis.

We use the version of Chebyshev's inequality stated below.

Lemma 2. *Chebyshev's inequality [55]: Let X be a random variable with finite mean $\mathrm{E}(X)$ and variance $\mathrm{Var}(X)$ and $\nu > 0$. Then*

$$\Pr\left[|X - \mathrm{E}(X)| \geq \nu \mathrm{E}(X)\right] \leq \frac{\mathrm{Var}(X)}{(\nu \mathrm{E}(X))^2}, \tag{2.39}$$

from which

$$\Pr\left[X \leq (1-\nu)\mathrm{E}(X)\right] \leq \frac{\mathrm{Var}(X)}{(\nu \mathrm{E}(X))^2}, \qquad \Pr\left[X \geq (1+\nu)\mathrm{E}(X)\right] \leq \frac{\mathrm{Var}(X)}{(\nu \mathrm{E}(X))^2}. \tag{2.40}$$

□

Fano's inequality provides a lower bound to the error probability P_e in the situation when we know a r.v. X and its estimate $\hat{X}$, with both $X, \hat{X} \in \mathcal{X}$.

Lemma 3. *Fano's inequality ([26, Ch. 2], [67, Appx.]): Let $X, \hat{X} \in \mathcal{X}$ and $P_e = \Pr[\hat{X} \neq X]$. Then we have*

$$H_2(P_e) + P_e \log(|\mathcal{X}| - 1) \geq H(X|\hat{X}), \tag{2.41}$$

where $H_2(p)$ is the binary entropy function $H_2(p) = -p \log p - (1-p)\log(1-p)$, $p \in [0, 1]$. □

The binary entropy function $H_2(p) \leq 1$ for $p \in [0,1]$. Therefore (2.41) is sometimes written as $1 + P_e \log(|\mathcal{X}| - 1) \geq H(X|\hat{X})$.

Proof. The proof (from [67, Appx.]) is given here since we will use some variations of it in the thesis. Define $E = \mathbb{I}(\hat{X} \neq X)$, where $\mathbb{I}(.)$ is the indicator function. So

$$E = \begin{cases} 1, & \hat{X} \neq X \\ 0, & \hat{X} = X. \end{cases} \tag{2.42}$$

Now expand $H(E, X|\hat{X})$ in two ways using the chain rule. Firstly,

$$H(E, X|\hat{X}) = H(X|\hat{X}) + H(E|X, \hat{X}) = H(X|\hat{X}), \tag{2.43}$$

since E is determined by $X, \hat{X}$. Secondly,

$$\begin{aligned} H(E, X|\hat{X}) &= H(E|\hat{X}) + H(X|E, \hat{X}) \\ &= H(E|\hat{X}) + \Pr[E=0]H(X|E=0, \hat{X}) + \Pr[E=1]H(X|E=1, \hat{X}) \\ &\overset{(a)}{=} H(E|\hat{X}) + \Pr[E=1]H(X|E=1, \hat{X}) \\ &\overset{(b)}{\leq} H(E|\hat{X}) + P_e \log(|\mathcal{X}| - 1) \\ &\leq H(E) + P_e \log(|\mathcal{X}| - 1) \\ &\leq H_2(P_e) + P_e \log(|\mathcal{X}| - 1), \end{aligned} \tag{2.44}$$

where (a) is because $H(X|E=0, \hat{X}) = 0$ and (b) is because X takes on at most $|\mathcal{X}|-1$ values, given $E = 1$ and $\hat{X}$. Combining the two equations, we have the lemma. □

2.2 Convex Optimization

In the thesis, we shall need to solve some optimization problems. The standard form of an optimization problem is given as

$$\begin{aligned} &\text{minimize} && f_0(\mathbf{x}) \\ &\text{subject to} && f_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots, m, \\ &&& h_i(\mathbf{x}) = 0, \quad i = 1, \cdots, p, \end{aligned} \tag{2.45}$$

where $\mathbf{x} \in \mathbb{R}^n$. The function $f_0 : \mathbb{R}^n \to \mathbb{R}$ is the objective function, and for $i = 1, \cdots, m$, functions $f_i : \mathbb{R}^n \to \mathbb{R}$ are inequality constraint functions, and functions $h_i : \mathbb{R}^n \to \mathbb{R}$ are equality constraint functions. Denoting the domain of function f_i

as dom(f_i), the domain of the optimization problem in (2.45) is $\mathcal{D} = \bigcap_{i=1}^{m} \mathrm{dom}(f_i) \cap \bigcap_{i=1}^{p} \mathrm{dom}(h_i)$. If the inequality constraint functions are convex, that is, they satisfy

$$f_i(\alpha\mathbf{x} + \beta\mathbf{y}) \leq \alpha f_i(\mathbf{x}) + \beta f_i(\mathbf{y}), \tag{2.46}$$

with $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ and $\alpha, \beta \in \mathbb{R}$, $\alpha, \beta \geq 0$, $\alpha+\beta = 1$, and the equality constraint functions are affine, so that $h_i(\mathbf{x}) = A\mathbf{x} - \mathbf{b}$, $A \in \mathbf{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$, then the optimization problem is called a *convex* optimization problem.

A feasible convex optimization problem has the general form

$$\begin{aligned} &\text{minimize} \quad f_0(\mathbf{x}) \\ &\text{subject to} \quad f_i(\mathbf{x}) \leq 0, \quad i = 1, \cdots, m, \\ &\qquad\qquad A\mathbf{x} = \mathbf{b}, \end{aligned} \tag{2.47}$$

with $f_1(\mathbf{x}), \cdots, f_m(\mathbf{x})$ convex.

Generally, optimization problems may be hard to solve, but the class of convex optimization problems can be solved efficiently. So, when confronted with a general optimization problem, we can recognize or re-formulate it as a convex optimization problem, which can be solved. In this section we give some mathematical background on convex optimization useful to our analysis, which is from [14]. For the full details, please refer to [14].

2.2.1 Affine and Convex Sets and Functions

Affine Sets and Functions

A set $\mathcal{R} \subseteq \mathbb{R}^n$ is affine if the line between any two points $\mathbf{x}, \mathbf{y} \in \mathcal{R}$, lies in $\mathcal{R}$. We have, for $\gamma \in \mathbb{R}$, $0 \leq \gamma \leq 1$, $\gamma\mathbf{x} + (1-\gamma)\mathbf{y} \in \mathcal{R}$. Thus, the affine set $\mathcal{R}$ with $\mathbf{x}_1, \cdots, \mathbf{x}_k \in \mathcal{R}$ contains every affine combination[2] of its points $\sum_{i=1}^{k} \gamma_i \mathbf{x}_i$, where $\sum_{i=1}^{k} \gamma_i = 1$.

The function $f : \mathbb{R}^n \to \mathbb{R}^m$ is affine if it has the form $f(\mathbf{x}) = A\mathbf{x} + \mathbf{b}$, where $A \in \mathbf{R}^{m\times n}$, $\mathbf{b} \in \mathbb{R}^m$. The function f has the form of a sum of a linear function and a constant.

Convex Sets

A set $\mathcal{R} \subseteq \mathbb{R}^n$ is convex if the line segment between any two points $\mathbf{x}, \mathbf{y} \in \mathcal{R}$, lies in $\mathcal{R}$. So, for $\gamma \in \mathbb{R}$, $0 \leq \gamma \leq 1$, $\gamma\mathbf{x} + (1-\gamma)\mathbf{y} \in \mathcal{R}$. Thus an affine set is also convex.

[2] An affine combination is a linear combination where the sum of the coefficients in the linear combination is one.

A point $\mathbf{z}$ is said to be a convex combination of the points $\mathbf{x}_1, \cdots, \mathbf{x}_k$ if it has the form $\mathbf{z} = \sum_{i=1}^{k} \gamma_i \mathbf{x}_i$, where $\sum_{i=1}^{k} \gamma_i = 1$, and $\gamma_i \geq 0$, for $i = 1, \cdots, k$. A set is convex if and only if it contains all convex combinations of its points.

The convex hull of $\mathcal{R}$ is denoted as $\mathrm{conv}(\mathcal{R})$ and is the set of all convex combinations of points in $\mathcal{R}$. The convex hull is expressed as

$$\mathrm{conv}(\mathcal{R}) = \left\{ \sum_{i=1}^{k} \gamma_i \mathbf{x}_i \,\middle|\, \mathbf{x}_i \in \mathcal{R}, \gamma_i \geq 0, i = 1, \cdots, k, \sum_{i=1}^{k} \gamma_i = 1 \right\}. \tag{2.48}$$

The convex hull is the smallest convex set that contains $\mathcal{R}$; the convex hull is always convex.

Cones

A cone is a set $\mathcal{R}$ which has the property that $\gamma \mathbf{x} \in \mathcal{R}$, for every $\mathbf{x} \in \mathcal{R}$ and $\gamma \geq 0$. The set $\mathcal{R}$ is a convex cone if it is both convex and a cone.

A point $\mathbf{z} = \sum_{i=1}^{k} \gamma_i \mathbf{x}_i$, with $\gamma_i \geq 0$, $i = 1, \cdots k$ is a conic (or nonnegative linear) combination of the $\mathbf{x}_1, \cdots, \mathbf{x}_k$. If $\mathbf{x}_i \in \mathcal{R}$, then every conic combination of $\mathbf{x}_i$ is $\in \mathcal{R}$. Also, $\mathcal{R}$ is a convex cone if and only if every conic combination of $\mathbf{x}_i$ is $\in \mathcal{R}$.

Convex Functions

A function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, with domain denoted as dom(f), is convex if dom(f) is a convex set and $\mathbf{y} \in \mathrm{dom}(f)$ for all $\mathbf{x}$, so that

$$f(\gamma \mathbf{x} + (1 - \gamma)\mathbf{x}) \leq \gamma f(\mathbf{x}) + (1 - \gamma) f(\mathbf{y}), \tag{2.49}$$

where $0 \leq \gamma \leq 1$. A function f is strictly convex if strict inequality holds in (2.49) for $\mathbf{x} \neq \mathbf{y}$. A function is f is concave if $-f$ is convex and strictly concave if $-f$ is strictly convex. Affine functions have equality in (2.49) so that affine functions are both convex and concave.

Whether the function f is convex can also be determined by the first and second order conditions.

First order conditions The first order convexity condition can be stated as follows. Let f be differentiable so that its gradient ∇f exists at each point in dom(f). Then f is convex if and only if dom(f) is convex and for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(f)$, the following holds:

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T (\mathbf{y} - \mathbf{x}). \tag{2.50}$$

We also have that f is concave if and only if dom(f) is convex and for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(f)$, the following holds:

$$f(\mathbf{y}) \leq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}). \tag{2.51}$$

Second order conditions Let f be twice differentiable so that its Hessian or second derivative $\nabla^2 f$ exists at each point in dom(f). Then f is convex if and only if dom(f) is convex and $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(f)$, and the Hessian is positive semi-definite

$$\nabla^2 f(\mathbf{x}) \succeq 0. \tag{2.52}$$

For a function on $\mathbb{R}$, this means that $f''(x) \geq 0$, that is, the derivative is nondecreasing.

We also have that f is concave if and only if dom(f) is convex and $\nabla^2 f(\mathbf{x}) \preceq 0$ for all $\mathbf{x} \in \mathrm{dom}(f)$.

Convexity-preserving operation Some operations on functions help to preserve convexity, and so are useful in that we can use them to construct new convex functions. One operation that we make use of is forming a new function from the pointwise maximum of some convex functions. That is, if $f_1, \cdots, f_m$ are convex functions then their pointwise maximum, defined as

$$f(\mathbf{x}) = \max\{f_1(\mathbf{x}), \cdots, f_m(\mathbf{x})\}, \tag{2.53}$$

is also convex. We also note that the pointwise supremum of a family of affine functions makes up a convex function. A proof is found in [14, Section 3.2].

Quasiconvex Functions

Quasiconvex (or unimodal) functions are a generalization of convex functions; they are useful in the sense that a global minimum is guaranteed to exist over any convex set in the function domain. A function $f : \mathbb{R}^n \to \mathbb{R}$ is quasiconvex if its domain and all its sublevel sets, defined as

$$\mathcal{S}_\alpha = \{\mathbf{x} \in \mathrm{dom}(f) | f(\mathbf{x}) \leq \alpha\}, \quad \alpha \in \mathbb{R}$$

are convex. The function f is quasiconcave if $-f$ is quasiconvex, which means that every superlevel set $\{\mathbf{x}|f(\mathbf{x}) \geq \alpha\}$ is convex.

As an illustration, a continuous function f on $\mathbb{R}$ is quasiconvex if and only if at least one of the following conditions holds:

1. f is nondecreasing;

2. f is nonincreasing;

3. there exists $\mathbf{a}, \mathbf{t} \in \mathrm{dom}(f)$ so that for $\mathbf{t} \leq \mathbf{a}$, f is nonincreasing, while for $\mathbf{t} \geq \mathbf{a}$, f is nondecreasing. The point $\mathbf{a}$ can be chosen as any point which is the global minimizer of f.

Quasiconvex functions which are differentiable obey first and second order conditions as stated below.

First order conditions: Let $f : \mathbb{R}^n \to \mathbb{R}$ be differentiable. Function f is quasiconvex if and only if dom(f) is convex and for all $\mathbf{x}, \mathbf{y} \in \mathrm{dom}(f)$,

$$f(\mathbf{y}) \leq f(\mathbf{x}) \Rightarrow \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \leq 0. \tag{2.54}$$

Unlike in the case of convex function f, where $\nabla f(\mathbf{x}) = 0$ means that $\mathbf{x}$ is a global minimizer of f, for quasiconvex f it may turn out that $\mathbf{x}$ is not a global minimizer of f, even if $\nabla f(\mathbf{x}) = 0$. So a better set of conditions to determine quasiconcavity would be the second order conditions.

Second order conditions: Let $f : \mathbb{R}^n \to \mathbb{R}$ be twice differentiable. If f is quasiconvex, then for all $\mathbf{x} \in \mathrm{dom}(f)$, and all $\mathbf{y} \in \mathbb{R}^n$, we have

$$\mathbf{y}^T \nabla f(\mathbf{x}) = 0 \Rightarrow \mathbf{y}^T \nabla^2 f(\mathbf{x}) \mathbf{y} \geq 0. \tag{2.55}$$

For f on $\mathbb{R}$, this reduces to

$$f'(\mathbf{x}) = 0 \Rightarrow f''(\mathbf{x}) \geq 0. \tag{2.56}$$

2.2.2 Lagrange Dual Problem

The Langrangian associated with a feasible optimization problem (2.45) with optimal value $\mathbf{x}^*$ may be expressed as

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f_0(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \mu_i h_i(\mathbf{x}), \tag{2.57}$$

with $\mathrm{dom}(L) = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$, where $\mathcal{D} = \bigcap_{i=1}^{m} \mathrm{dom}(f_i) \cap \bigcap_{i=1}^{p} \mathrm{dom}(h_i)$, λ_i is the Lagrange multiplier associated with inequality constraint $f_i(\mathbf{x}) \leq 0$, μ_i is the Lagrange multiplier associated with equality constraint $h_i(\mathbf{x}) = 0$.

The Lagrangian dual function $G : \mathbb{R}^m \times \mathbb{R}^p \rightarrow \mathbb{R}$ is the minimum value of the Lagrangian over $\mathbf{x}$:

$$G(\boldsymbol{\lambda}, \boldsymbol{\mu}) = \inf_{\mathbf{x} \in \mathcal{D}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}). \tag{2.58}$$

For each pair $(\boldsymbol{\lambda}, \boldsymbol{\mu})$ with $\lambda_i \geq 0$ for all i, the Lagrange dual function gives a lower bound on the optimal value of the original problem (2.45). The problem of finding the best lower bound that can be obtained from the Lagrange dual function leads to the optimization problem, known as the Lagrange dual problem:

$$\begin{aligned} &\text{maximize} \;\; G(\boldsymbol{\lambda}, \boldsymbol{\mu}) \\ &\text{subject to} \;\; \lambda_i \geq 0, \; i = 1, \cdots, m. \end{aligned} \tag{2.59}$$

In this context, the original optimization problem and its feasible set and optimal value are called the primal problem, primal feasible set and primal optimal value. The optimal pair $(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*)$ is known as the dual optimal pair.

The optimal value of the Lagrange dual problem, denoted as $\mathbf{d}^*$, gives the lower bound to the primal optimal value

$$\mathbf{d}^* \leq \mathbf{x}^*.$$

This property is called weak duality and it holds even when the original problem is not convex. If however,

$$\mathbf{d}^* = \mathbf{x}^*,$$

the bound is tight (there is zero duality gap) and we say that strong duality holds. Strong duality does not hold in general, but when the primal problem is convex, we may have strong duality. A useful condition for strong duality is Slater's condition, which is stated as: If the primal problem is convex and there exists a feasible $\mathbf{x}$ in the relative interior of $\mathcal{D}$, satisfying

$$f_i(\mathbf{x}) < 0, \quad i = 1, \cdots, m, \qquad A\mathbf{x} = \mathbf{b}, \tag{2.60}$$

then strong duality holds.

If strong duality holds, it can be shown that [14, Sect. 5.5.2]

$$f_0(\mathbf{x}^*) = G(\boldsymbol{\lambda}^*, \boldsymbol{\mu}^*), \tag{2.61}$$

which gives rise to the complementary slackness condition:

$$\lambda_i^* f_i(\mathbf{x}^*) = 0, \quad i = 1, \cdots, m. \tag{2.62}$$

If the Lagrangian $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ is differentiable and strong duality holds, any pair of primal and dual optimal points then satisfy the Karush-Kuhn-Tucker (KKT) conditions:

$$\begin{aligned} f_i(\mathbf{x}^*) &\leq 0, \quad i = 1, \cdots, m, \\ h_i(\mathbf{x}^*) &= 0, \quad i = 1, \cdots, p, \\ \lambda_i^* &\geq 0, \quad i = 1, \cdots, m, \\ \lambda_i^* f_i(\mathbf{x}^*) &= 0, \quad i = 1, \cdots, m \\ \nabla f_0(\mathbf{x}^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(\mathbf{x}^*) + \sum_{i=1}^{p} \mu_i^* \nabla h_i(\mathbf{x}^*) &= 0. \end{aligned}$$

The KKT conditions are useful in solving optimization problems. When the primal problem is convex, the KKT conditions are satisfied for the primal and dual optimal points with zero duality gap.

Chapter 3

Background on Information Theoretic Security

In this Chapter we will give some background on the information theoretic approach of Wyner [115] and Csiszár and Körner [30] to characterizing the simultaneously secure and reliable information rates for the two fundamental channels, the wiretap channel and the broadcast channel with confidential messages (BCC). We first set the scene with basic definitions for the channel coding problem for the discrete memoryless channel (DMC) that will be the basis for the analysis. Then we will describe the wiretap channel and the BCC, before giving a survey of some of the literature of interest in this field.

3.1 Channel Coding for the Discrete Memoryless Channel

In Figure 3.1 a discrete memoryless channel is shown. The discrete memoryless channel (DMC) is a system made up of the input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ respectively, and probability transition matrix $p(y|x)$, for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, with the probability distribution of the output depending only on the input at that time instant and conditionally independent of previous channel inputs or outputs. It is denoted $(\mathcal{X}, p(y|x), \mathcal{Y})$. The message W is assumed to be randomly and uniformly distributed over the set $\mathcal{W} = \{1, 2, ..., M\}$ and is to be sent over the channel to the receiver. The message $w \in \mathcal{W}$ is mapped by encoder $f : \mathcal{W} \to \mathcal{X}^n$ to the codeword $\mathbf{x} \in \mathcal{X}^n$, with the number of channel uses given by n. The codeword $\mathbf{x}$ is transmitted over the DMC with transition probability $p(\mathbf{y}|\mathbf{x})$ as depicted in the figure, $p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} p(y_i|x_i)$ being the nth

extension of the DMC (n-length sequence version of the DMC with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$, and probability transition matrix $p(y|x)$), or simply the probability transition for n uses of the channel.

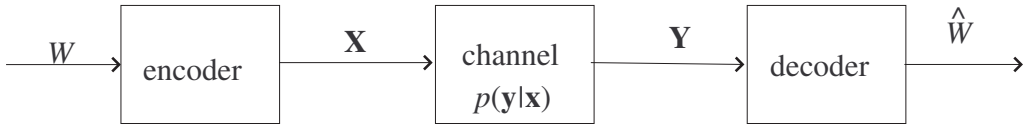


Figure 3.1: A discrete memoryless channel.

At the receiver the received sequence $\mathbf{y} \in \mathcal{Y}^n$ is decoded by the decoder $g : \mathcal{Y}^n \to \mathcal{W}$, mapping the received sequence to $\hat{w} \in \mathcal{W}$, the estimate of the message. Then a $(2^{nR}, n)$-code[1] at rate R for the DMC $(\mathcal{X}, p(y|x), \mathcal{Y})$ is defined as consisting of the three items: an index set $\mathcal{W}$, an encoder f, and the decoder g.

For reliable communications, we define the average probability of error for a $(2^{nR}, n)$-code with W chosen according to the uniform distribution and $\mathbf{X}$ being a deterministic function of W, as

$$P_e^{(n)} = \Pr\left[\hat{W} \neq W\right] = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \Pr\left[\hat{w} \neq w\right]. \tag{3.1}$$

A rate R is achievable if there exists a sequence of $(2^{nR}, n)$-codes such that $P_e^{(n)} \to 0$ as $n \to \infty$. The maximum achievable rate R is the capacity C, which is the famous

$$C = \max_{p(x)} I(X;Y). \tag{3.2}$$

To show an achievable rate R, we proceed in three steps. Firstly, construct a random code by 'choosing' a p.d.f. $p(x)$ and then generating codewords (at rate R) based on this. [2] Secondly, choose an encoding and decoding strategy. Thirdly, show that at the decoder and using the above strategies, the $P_e^{(n)} \to 0$ as $n \to \infty$. Thus there is a sequence of $(2^{nR}, n)$-codes and the rate R is achievable.

The probability of error is calculated over a random code which makes the error probability symmetrical (see (3.1)). The random coding used in the proof also facilitates showing the existence of a good deterministic code. We do not explicitly attempt

[1] Here a $(2^{nR}, n)$-code is simplified notation for $(\lceil 2^{nR} \rceil, n)$-code.

[2] Essentially we just assume that such a p.d.f. exists and proceed with code generation. Of course, codeword generation from this p.d.f. can be as complex or simple as desired for various types of networks.

to find a practical code for the channel. The decoding uses joint typicality where we seek the codeword jointly typical with the received sequence. Joint typicality is suboptimal but is used due to simplicity and by the fact that it still achieves all rates below capacity[3].

Once the achievability proof is shown, we then proceed to show that any sequence of $(2^{nR}, n)$-codes with $P_e^{(n)} \to 0$ as $n \to \infty$ must have $R \leq C$. So no rates greater than C can be achieved. This is known as the converse proof. We note that we essentially show an upper bound to the rates R. In other networks where possibly the achievable rate (region) and the upper bound to the rate (region) do not match, this proof serves as a proof for the outer bound to the rate (region).

The details of the achievability and converse proofs for the DMC may be found in [26], [67] and we will not deal with them here.

3.1.1 Binary Channels

Binary channels are often used in coding theory; these simple channels are used to validate the coding scheme. Two of these are shown in Figure 3.2.

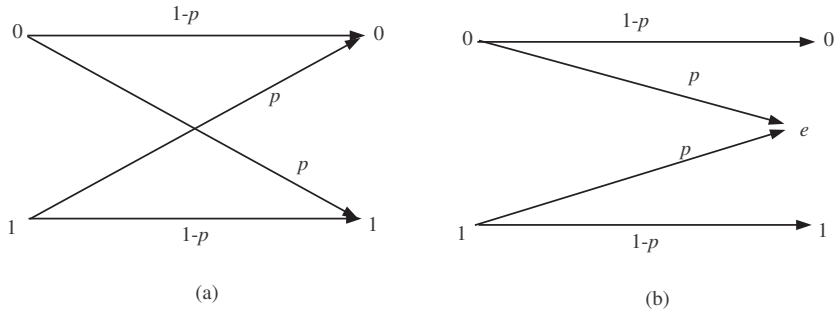


Figure 3.2: Binary channels: (a) binary symmetric channel, (b) binary erasure channel.

The binary symmetric channel (BSC), depicted in Figure 3.2(a), is a binary channel where the input symbols are complemented with probability p. The BSC is so called because all the rows of its probability transition matrix are permutations of each other, and all the columns of its probability transition matrix are permutations of each other. A general symmetric channel has this property as well.

[3]We will depart from using the random coding-joint typicality approach when we consider lattice codes for the wiretap channel in Chapter 6.

The binary erasure channel (BEC), depicted in Figure 3.2(b), is a binary channel where some of the bits are erased with probability p. The node e in the diagram represents erasure, while the receiver will have knowledge of its erased bits.

3.1.2 Maximum Likelihood Decoding for the DMC and the Error Exponent

When we use maximum likelihood (ML) decoding for the DMC instead of joint typical decoding, we should use the results of Gallager [49, Ch. 5] given below to bound the error probability, which is related to the transmission rate. When we want to bound the error probability in terms of the error exponent, we should now take logarithms to base e; this is helpful in emphasizing the exponential dependence of the error probability. The rate is now in natural units and we consider codes with $\left\lceil e^{nR} \right\rceil$ codewords, instead of $\left\lceil 2^{nR} \right\rceil$ codewords.

Now define an (n, R) block code as a code of block length n with $\left\lceil e^{nR} \right\rceil$ codewords. Let a DMC have transition probability matrix $p(y|x)$ for $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, and consider the ensemble of (n, R) block codes where each letter of each codeword is independently selected according to $q(x)$. For each message m, $1 \leq m \leq \left\lceil e^{nR} \right\rceil$, $0 \leq \rho \leq 1$, the ensemble average probability of error with ML decoding and $M - 1 < e^{nR} \leq M$ is given by

$$P_{e,m} \leq \exp\left\{-n\left[E_o(\rho, \mathbf{q}) - \rho R\right]\right\}, \tag{3.3}$$

where

$$E_o(\rho, \mathbf{q}) = -\log \sum_{y \in \mathcal{Y}} \left[\sum_{x \in \mathcal{X}} q(x) p(y|x)^{1/(1+\rho)} \right]^{1+\rho}, \tag{3.4}$$

where the sums are taken over the channel input and output alphabets. The vector $\mathbf{q}$ has elements $q(x), x \in \mathcal{X}$. The average over the messages, for arbitrary message probabilities $p(m)$, is

$$P_e = \sum_{m=1}^{M} p(m) P_{e,m} \leq \exp\left\{-n\left[E_o(\rho, \mathbf{q}) - \rho R\right]\right\}. \tag{3.5}$$

The tightest bound is obtained by maximizing over ρ and $\mathbf{q}$, from which we obtain the random coding error exponent

$$E_r(R) = \max_{0 \leq \rho \leq 1} \max_{\mathbf{q}} \left[E_o(\rho, \mathbf{q}) - \rho R\right]. \tag{3.6}$$

We have, for the $\mathbf{q}$ that maximizes the random coding error exponent,

$$P_{e,m} \leq e^{-nE_r(R)}, \quad 1 \leq m \leq M, \tag{3.7}$$

$$P_e \leq e^{-nE_r(R)}. \tag{3.8}$$

The exponent $E_r(R) > 0$ for all R, $0 \leq R < C$. Furthermore, the exponent is a convex $\cup$, decreasing and positive function of R for $0 \leq R < C$. Thus we can choose codes with error probability increasing exponentially with block length n for rates approaching capacity.

For low rates the random coding error exponent may not be accurate. The error probability should now be bound exponentially in n, by the expurgated error exponent, $E_x(R)$, so that $P_{e,m} \leq e^{-nE_x(R)}$. The expurgated error exponent $E_x(R)$ is derived by expurgating (or removing) poor code words from the ensemble that do not satisfy a given bound. The details may be found in [49, Sect. 5.7]. Our focus will be on the random coding error exponent and rates close to capacity for analysis using ML decoding in the thesis.

3.2 The Wiretap Channel

In this section we briefly look at the discrete memoryless (DM) wiretap channel, shown in Figure 3.3. This channel is the basic channel in physical layer information theoretic security and was originally studied by Wyner [115].

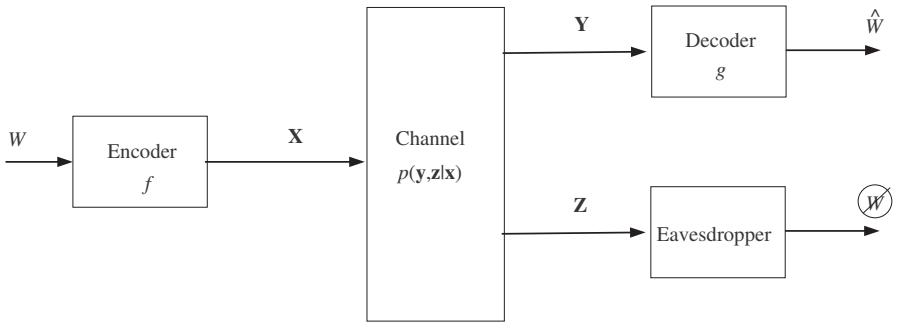


Figure 3.3: The discrete memoryless wiretap channel.

We note that the DM wiretap channel is a 2-receiver DM broadcast channel with the channel as $(\mathcal{X}, p(y, z|x), \mathcal{Y} \times \mathcal{Z})$. The probability distribution for n uses of the

channel is

$$p(\mathbf{y}, \mathbf{z}|\mathbf{x}) = \prod_{i=1}^{n} p(y_i, z_i|x_i).$$

The confidential message W is to be sent to the legitimate receiver Y and kept secret from the eavesdropper Z. A $(2^{nR}, n)$-code for the DM wiretap channel consists of:

- The message set $\{1, ..., 2^{nR}\}$, uniformly and randomly distributed.
- A (stochastic) encoder $f : \mathcal{W} \to \mathcal{X}^n$ which maps message $w \in \{1, ..., 2^{nR}\}$ to codeword $\mathbf{x} \in \mathcal{X}^n$.
- A decoder at the legitimate receiver $g : \mathcal{Y}^n \to \{1, ..., 2^{nR}\}$ maps the received sequence $\mathbf{y} \in \mathcal{Y}^n$ to the estimate of the message $\hat{w} \in \{1, ..., 2^{nR}\}$.

The eavesdropper is assumed to be passive. This channel also models the case of multiple collaborating eavesdroppers, by merging the eavesdroppers into one (a worst-case consideration).

The reliability is still measured by the error probability (3.1). The secrecy level of the confidential message W at the eavesdropper is measured by the equivocation rate

$$R_e^{(n)} = \frac{1}{n} H(W|\mathbf{Z}), \tag{3.9}$$

where $H(W|\mathbf{Z})$ is the conditional entropy of W given $\mathbf{Z}$. The equivocation rate is the uncertainty of the message W at the eavesdropper and the design goal should be to make this as large as possible.

Rate-equivocation region If the information leaked to the eavesdropper, quantified as $\frac{1}{n} I(W; \mathbf{Z})$ does not go to zero for n sufficiently large, we have a rate-equivocation region that is satisfied by the rate-equivocation pair (R, R_e). Now, (R, R_e) is achievable if there exists a sequence of $(2^{nR}, n)$-codes such that $P_e^{(n)} \to 0$ as $n \to \infty$ and the equivocation rate satisfies

$$R_e \leq \lim_{n \to \infty} \inf R_e^{(n)}. \tag{3.10}$$

Alternatively, we say that the rate-equivocation pair (R, R_e) is achievable if there exists a sequence of $(2^{nR}, n)$-codes such that $P_e^{(n)} \leq \eta$ and we have

$$\frac{1}{n} H(W|\mathbf{Z}) \geq R_e - \epsilon_1, \tag{3.11}$$

for any $\eta, \epsilon_1 > 0$ and small for n sufficiently large. The rate-equivocation region (or capacity-equivocation region) $\mathcal{R}$ is the closure of all achievable rate-equivocation pairs (R, R_e).

We should note that there is a stronger notion of secrecy, which makes the equivocation rate satisfy $nR_e \leq \lim_{n\to\infty} \inf R_e^{(n)}$. That is, we consider the equivocation rate for a block of n channel uses instead of just one. For details the reader can consult the work of Maurer and Wolf [86]. This stronger notion of secrecy is commonly called *strong secrecy* and the weaker notion described so far is commonly called *weak secrecy*. Although strong secrecy is generally favored by cryptographers, we shall use the weak secrecy notion throughout the thesis. Using strong secrecy, results from [30, 115] are unchanged, so the coding schemes are justified.

Secrecy capacity region If the information leaked to the eavesdropper goes to zero for n sufficiently large, we have the perfect secrecy condition, which is, for n sufficiently large,

$$\frac{1}{n} I(W; \mathbf{Z}) \to 0 \Rightarrow \frac{1}{n} H(W|\mathbf{Z}) \to \frac{1}{n} H(W). \tag{3.12}$$

So in this case, $R_e = R$ for n sufficiently large and we have the secrecy capacity region that is satisfied by pairs (R, R). The achievability conditions now becomes (R, R) is achievable if there exists a sequence of $(2^{nR}, n)$-codes such that $P_e^{(n)} \leq \eta$ and we have

$$\frac{1}{n} H(W|\mathbf{Z}) \geq R - \epsilon_1, \tag{3.13}$$

for any $\eta, \epsilon_1 > 0$ and small for n sufficiently large. The secrecy capacity is the largest achievable rate with perfect secrecy and is given by:

$$C_S = \max_{(R,R)\in\mathcal{R}} R. \tag{3.14}$$

The original result of Wyner [115] was derived for the case when the channel from legitimate receiver Y to eavesdropper Z is degraded[4], which means that the channel from X to Z is a noisy version of the channel from X to Y. In Theorem 6 below we state the rate equivocation region for the discrete memoryless wiretap channel with general conditions on the channel X to Z. We assume that the channel from X to Y enjoys an advantage over the channel from X to Z. In particular if the channel X to Z is less noisy than the channel X to Y then the equivocation rate $R_e = 0$. This version

[4]See the Appendix A for the ordering of channels.

is a special case of the BC with confidential messages of and is studied by Csiszár and Körner in [29, Problem 3.4.33] and [30]. Liang *et al* [73] provide a helpful elaboration of the method of [30].

Theorem 6. *The rate equivocation region $\mathcal{R}$ for the discrete memoryless wiretap channel is the closure of all rate-tuples (R, R_e) satisfying*

$$R_e \leq R, \tag{3.15}$$

$$0 \leq R \leq I(V;Y), \tag{3.16}$$

$$R_e \leq I(V;Y|U) - I(V;Z|U), \tag{3.17}$$

where the auxiliary random variables U, V satisfy the Markov chain condition $U \rightarrow V \rightarrow X \rightarrow (Y,Z)$ and with ranges $|\mathcal{U}| \leq |\mathcal{X}| + 2$, $|\mathcal{V}| \leq |\mathcal{X}|^2 + 3|\mathcal{X}| + 2$.

The achievability proof uses rate-splitting and code partitioning (also called binning). The rate-splitting is evident in the statement of the theorem. In Theorem 6, V represents the source message which is split into a part that can be decoded by both receiver Y and the eavesdropper Z, represented by U, and another part which is only to be decoded by Y.

Proof. We give an outline of only the code construction in the achievability proof of [29], [30] so that we can have an idea of the achievable coding scheme to be used in wiretap situations. We note that the equivocation calculation and the converse proof is found in [29, Problem 3.4.33] and [73] and we will not repeat it here.

To prove achievability, we proceed in two steps. In the first step, we construct a code and encoding and decoding strategy so that the conditions in Theorem 6 are met. Then we show that the equivocation satisfies (3.11). We will focus on the code construction here. We first note that region $\mathcal{R}$ is convex and use the fact that it is sufficient to prove that (R, R_e) satisfies

$$0 \leq R \leq I(X;Y), \tag{3.18}$$

$$R_e \leq I(X;Y|U) - I(X;Z|U), \tag{3.19}$$

where $U \rightarrow X \rightarrow (Y,Z)$ forms a Markov chain. Then, prefix a DMC with transition probability $p(x|v)$ to the channels $p(y|x)$ and $p(z|x)$ (that is, prefix V to $X \rightarrow (Y,Z)$). This results in channels with transition probabilities $p(x|v)p(y|x)$ and $p(x|v)p(z|x)$,

and (R, R_e) satisfying the conditions in Theorem 6 is contained in $\mathcal{R}$. Since $\mathcal{R}$ is convex, then it is sufficient to show that $(R, R_e) \in \mathcal{R}$ for

$$R_e = \max_U [I(X;Y|U) - I(X;Z|U)] \leq R \leq I(X;Y). \tag{3.20}$$

So fix distribution $p(u)p(x|u)$ for U so that the maximum in (3.20) is achieved. Note that by

$$I(X;Y|U) - I(X;Z|U) = I(X;Y) - I(X;Z) - [I(U;Y) - I(U;Z)] \geq 0$$

this U must satisfy

$$I(U;Y) \leq I(U;Z).$$

To begin the code construction, split the message W into 2 parts, denoted $W_0 \in [1, ..., 2^{nR_0}]$ and $W_1 \in [1, ..., 2^{nR_1}]$ with transmission rates R_0 and R_1 respectively. The receiver decodes W_0, W_1 at rates R_0, R_1 and the eavesdropper decodes W_0 at rate R_0.

The codebook generation is depicted in Figure 3.4:

1. Codebook generation: Generate 2^{nR_0} codewords $\mathbf{u}(j)$ independently and randomly according to $\prod_{i=1}^{n} p(u_i)$. For each $\mathbf{u}(j)$, generate $2^{R_1} = 2^{R_e+R'}$ codewords $\mathbf{x}(j,k,l)$ independently and randomly according to $\prod_{i=1}^{n} p(x_i|u_i)$, where $k = 1, ..., 2^{nR_e}$, $l = 1, ..., 2^{nR'}$. We see from Figure 3.4 that the $\mathbf{x}(j,k,l)$ codewords have been *partitioned* into 2^{nR_e} subcodes (or bins) $\mathcal{C}(1), ..., \mathcal{C}(2^{nR_e})$, each of size $2^{nR'}$. The set of subcodes is known to the encoder, decoder and eavesdropper. The eavesdropper is allowed to decode the transmitted codeword (generated from each subcode) at the capacity (or greater) of the eavesdropper's channel, thus it cannot decode any more information. In this way the subcode index is protected from the eavesdropper and that is where we encode the confidential information.

2. Encoding: Define stochastic encoder $f : \{1, ..., 2^{nR_e}\} \times \{1, ..., 2^{nR'}\}$. To send $k \in \{1, ..., 2^{nR_e}\}$, select an index $l \in \{1, ..., 2^{nR'}\}$ uniformly randomly and and send the codeword $\mathbf{x}(j,k,l)$.

3. Decoding: Using joint typical decoding, we have the following:

(a) The decoder at the receiver tries to find $\hat{j}$ so that $(\mathbf{u}(j), \mathbf{y}) \in \mathcal{T}_\epsilon^n(P_{UY})$. The decoder's error probability is small if $R_0 < I(U;Y)$.

(b) The decoder at the receiver tries to find $(\hat{k}, \hat{l})$ so that $(\mathbf{u}(j), \mathbf{x}(j,k,l), \mathbf{y}) \in \mathcal{T}_\epsilon^n(P_{UXY})$. The decoder's error probability is small if $R_1 = R_e + R' < I(X;Y|U)$.

(c) The eavesdropper tries to find $\hat{j}$ so that $(\mathbf{u}(j), \mathbf{z}) \in \mathcal{T}_\epsilon^n(P_{UZ})$. The decoder's error probability is small if $R_0 < I(U;Z)$.

It may be shown (for the details refer to [29, Problem 3.4.33]) that

$$R' \geq I(X;Z|U). \tag{3.21}$$

Therefore, this condition, together with the successful decoding conditions above, imply that (3.20) is satisfied. We should remark that it is quite common to use the condition that $R' = I(X;Z|U) - \epsilon'$, where ϵ' is small as n gets large, instead.

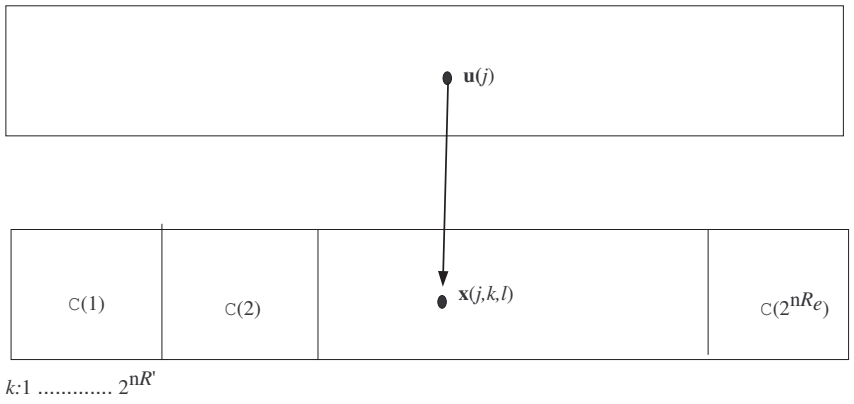


Figure 3.4: Codebook generation for discrete memoryless wiretap channel.

An analysis of the equivocation $\frac{1}{n}H(W|\mathbf{Z})$ along the lines of [29, Problem 3.4.33] or [73, Ch. 2] yields an equivocation rate

$$R_e \leq I(X;Y|U) - I(X;Z|U), \tag{3.22}$$

satisfying (3.11). Combining all of the above and eliminating R', we have that the

region $\mathcal{R}'$, which is the closure of all rate-tuples (R, R_e) satisfying

$$\begin{aligned} R_e \leq & R_1, \\ 0 \leq & R_0 + R_1 = R, \\ 0 \leq & R_0 \leq \min[I(U;Y), I(U;Z)], \\ 0 \leq & R_1 \leq I(X;Y|U), \\ R_e \leq & I(X;Y|U) - I(X;Z|U), \end{aligned} \tag{3.23}$$

where the auxiliary r.v. U satisfies the Markov chain condition $U \rightarrow X \rightarrow (Y, Z)$. Performing channel prefixing as described above, we then have the region $\mathcal{R}''$, which is the closure of all rate-tuples (R, R_e) satisfying

$$\begin{aligned} R_e \leq & R_1, \\ 0 \leq & R_0 + R_1 = R, \\ 0 \leq & R_0 \leq \min[I(U;Y), I(U;Z)], \\ 0 \leq & R_1 \leq I(V;Y|U), \\ R_e \leq & I(V;Y|U) - I(V;Z|U), \end{aligned} \tag{3.24}$$

where the auxiliary r.v.s U, V satisfy the Markov chain condition $U \rightarrow V \rightarrow X \rightarrow (Y, Z)$. Finally, eliminating R_0 and R_1 from the inequalities in (3.24) using Fourier-Motzkin elimination, we can obtain the region $\mathcal{R}$ in Theorem 6. Thus we have shown that $(R, R_e) \in \mathcal{R}$ and the code construction for the DM wiretap channel.

□

Theorem 6 leads to the following Corollary which states the secrecy capacity of the wiretap channel.

Corollary 1. *The secrecy capacity of the wiretap channel is*

$$C_S = \max_{T \rightarrow X \rightarrow (Y,Z)} [I(T;Y) - I(T;Z)] \tag{3.25}$$

where the auxiliary random variable T satisfies the Markov chain condition $T \rightarrow X \rightarrow (Y, Z)$ with range $|\mathcal{T}| \leq |\mathcal{X}| + 1$. □

Proof. We sketch the code construction here. To achieve C_S, we only need for $U =$ constant in Theorem 6. A coding scheme that achieves the secrecy capacity will be as follows:

1. Codebook generation: Define indices $w \in \{1, ..., 2^{nR}\}$, and $w' \in \{1, ..., 2^{nR'}\}$. Independently and randomly generate $\mathbf{t}(k)$, $k \in \{1, ..., 2^{n\tilde{R}}\}$ sequences by $\prod_{i=1}^{n} p(t_i)$. Partition the $\mathbf{t}(k)$ codewords into 2^{nR} subcodes $\mathcal{C}(1), ..., \mathcal{C}(2^{nR})$, with each subcode containing $2^{nR'}$ codewords. Index the set of all the $2^{n\tilde{R}}$ codewords as $\mathbf{t}(w, w')$. The set of subcodes is known to the encoder, decoder and eavesdropper.

2. Encoding: Define stochastic encoder $f : \{1, ..., 2^{nR}\} \times \{1, ..., 2^{nR'}\}$. To send $w \in \{1, ..., 2^{nR}\}$, select an index $w' \in \{1, ..., 2^{nR'}\}$ uniformly randomly and generate $\mathbf{x}$ according to $\prod_{i=1}^{n} p(x_i|t_i)$ and send it.

3. Decoding: Using joint typical decoding, the legitimate decoder declares that $(\hat{w}, \hat{w}')$ is sent if it is the unique index pair such that $(\mathbf{t}(\hat{w}, \hat{w}'), \mathbf{y}) \in \mathcal{T}_{\epsilon}^{n}(P_{TY})$. The legitimate decoder has low probability of error provided $\tilde{R} = R + R' \leq I(T; Y)$.

For each subcode $\mathcal{C}(w)$, the eavesdropper has on average $2^{nR'+\delta} \leq |\mathbf{t}(w, w')| \leq 2^{nR'-\delta}$ sequences that are joint typical with $\mathbf{z}$, where $\delta > 0$ and is small for n sufficiently large. We need to have $R' \geq I(T; Z)$ for the eavesdropper to have almost no information about the message sent[5]. Combining the two conditions $R + R' \leq I(T; Y)$ and $R' \geq I(T; Z)$, we have the statement of the corollary.

□

3.2.1 Gaussian and Multiple-Input Multiple-Output Wiretap Channels

In this section, we will state some results for two important classes of wiretap channels, namely the Gaussian wiretap channel and the Multiple-Input Multiple-Output (MIMO) wiretap channels (or multi-antenna wiretap channels), which we will encounter in Chapters 6 and 5, respectively. We will discuss the input distribution that achieves secrecy capacity.

Gaussian Wiretap Channel

The Gaussian wiretap channel is studied by Leung and Hellman [70] under the same conditions as the original wiretap channel of Wyner. That is, the eavesdropper's channel

[5]Some researchers will use $R' = I(T; Z) - \epsilon'$, $\epsilon' \to 0$ for n sufficiently large, instead.

is a degraded version of the legitimate user's channel. The channel model is

$$\begin{aligned} Y &= X + W, \\ Z &= X + V, \end{aligned} \tag{3.26}$$

for each use of the channel (or the time index), X is the channel input, Y and Z are channel outputs. The noise processes $\{W\}$ and $\{V\}$ are zero-mean i.i.d. proper complex Gaussian with variances σ_M^2 and σ_E^2, respectively. The input sequence $\{X\}$ is subject to the average power constraint

$$\frac{1}{n}\sum_{i=1}^{n} E[X_i^2] \leq P. \tag{3.27}$$

Leung and Hellman [70] proved the following theorem:

Theorem 7. *The secrecy capacity of the Gaussian wiretap channel is*

$$C_S = \frac{1}{2}\log\left(1 + \frac{P}{\sigma_M^2}\right) - \frac{1}{2}\log\left(1 + \frac{P}{\sigma_E^2}\right). \tag{3.28}$$

□

This result applies whether W and V are correlated or not. The achievability proof uses Corollary 1, applying the condition that the channel from X to Z is a degraded version of the channel from X to Y to obtain

$$C_S = \max_{X \to (Y,Z)} I(X;Y) - I(X;Z).$$

Then set $X \sim \mathcal{N}(0, P)$ to obtain (3.28). Thus to achieve the secrecy capacity, we require the input to be $X \sim \mathcal{N}(0, P)$.

Multiple-Input Multiple-Output Wiretap Channels

The MIMO wiretap channel model introduced here has the transmitter, legitimate receiver and eavesdropper all equipped with multiple antennas. This model then subsumes the cases where any of the transmitter, legitimate receiver and eavesdropper are equipped with only a single antenna. The MIMO wiretap channel is not a degraded or ordered channel.

For a multiple antenna wiretap non-fading channel with N_T transmit antennas and N_M and N_E receive antennas at the legitimate recipient and the eavesdropper, the

channel input and output for one channel use is

$$\begin{aligned} \mathbf{Y} &= \mathbf{H}_M \mathbf{X} + \mathbf{W}, \\ \mathbf{Z} &= \mathbf{H}_E \mathbf{X} + \mathbf{V}, \end{aligned} \tag{3.29}$$

where $\mathbf{X}$ is the $N_T \times 1$ transmitted signal vector, $\mathbf{Y}$ is the $N_M \times 1$ received signal vector at the legitimate receiver, $\mathbf{Z}$ is the $N_E \times 1$ received signal vector at the eavesdropper. The channel matrices $\mathbf{H}_M$ and $\mathbf{H}_E$ are fixed $N_M \times N_T$ and $N_E \times N_T$ matrices, the additive noise vectors $\mathbf{W}$, $\mathbf{V}$ are Gaussian vectors with zero mean and identity covariance matrices and are independent from one channel use to another. The channel input may be subject to the average power constraint

$$\frac{1}{n} \sum_{i=1}^{n} E[\mathbf{X}_i^T \mathbf{X}_i] \leq P, \tag{3.30}$$

or a matrix constraint

$$\frac{1}{n} \sum_{i=1}^{n} E[\mathbf{X}_i^T \mathbf{X}_i] \preceq \mathbf{P}. \tag{3.31}$$

Under the average power constraint on the input, in Shafiee *et al* [105] the special case where the transmitter and the legitimate receiver have 2 antennas and the eavesdropper has one antenna was studied and the secrecy capacity derived. Also under the average power constraint, in Khisti and Wornell [62] and Oggier and Hassibi [96], the secrecy capacity is derived for the general case. The secrecy capacity is stated below.

Theorem 8. *The secrecy capacity of the MIMO wiretap channel under the average input power constraint* (3.30) *is*

$$C_S = \max_{\mathbf{K_X} \succeq 0, \mathrm{Tr}(\mathbf{K_X}) \leq P} \frac{1}{2} \log \left| \mathbf{H}_M \mathbf{K_X} \mathbf{H}_M^T + \mathbf{I}_{N_M} \right| - \frac{1}{2} \log \left| \mathbf{H}_E \mathbf{K_X} \mathbf{H}_E^T + \mathbf{I}_{N_E} \right|, \tag{3.32}$$

where $\mathbf{K_X} \succeq 0$ *denotes that the input covariance matrix is positive semi-definite,* $\mathrm{Tr}(.)$ *is the trace operator,* $\mathbf{I}_n$ *is the identity matrix of size* $n \times n$*,* $(.)^T$ *denotes the matrix transpose, and* $|.|$ *denotes the matrix determinant.* □

In Khisti and Wornell [62] and Oggier and Hassibi [96], it was shown that by using Theorem 1, setting (with some abuse of notation) $U = \mathbf{X}$ to be Gaussian with mean zero and covariance matrix $\mathbf{K_X} \succeq 0$, we have the achievable secrecy capacity as in Theorem 8.

The secrecy capacity with input matrix power constraint was evaluated by Bustin *et al* [15], Liu and Shamai [81], and Liu *et al* [77]. It turns out to be the same as stated in Theorem 8, except that $\mathrm{Tr}(\mathbf{K_X}) \leq P$ should be replaced by $\mathrm{Tr}(\mathbf{K_X}) \preceq \mathbf{P}$. In [15, 81], it was shown that using $U = \mathbf{X}$ Gaussian and with covariance matrix satisfying $\mathbf{K_X} \succeq 0$ and $\mathrm{Tr}(\mathbf{K_X}) \preceq \mathbf{P}$, the secrecy capacity as in Theorem 8 is achievable. Finally Liu *et al* [77] showed that we could also set $\mathbf{X} = U + V$, where U and V are independent Gaussian vectors with mean zero and covariance matrices $\mathbf{K_X} - \mathbf{B}$ and $\mathbf{B}$ respectively. Thus $p(\mathbf{X}|U)$ is a prefix channel similar to the discussion in the wiretap channel model of Csiszár and Körner [29, Problem 3.4.33]; the choice of $U = \mathbf{X}$ as previously is thus not a unique one. We can also view V as artificial noise injected to help the legitimate receiver.

3.2.2 Compound Wiretap Channels

In this section we briefly describe the compound wiretap channel which can model multiple eavesdroppers. In the compound wiretap channel, the transmitter sends message W to J legitimate receivers to be kept secret from K eavesdroppers, as depicted in Figure 3.5. This channel can model the case of multiple non-collaborating eavesdroppers, and the general case is studied by Liang *et al* in [72].

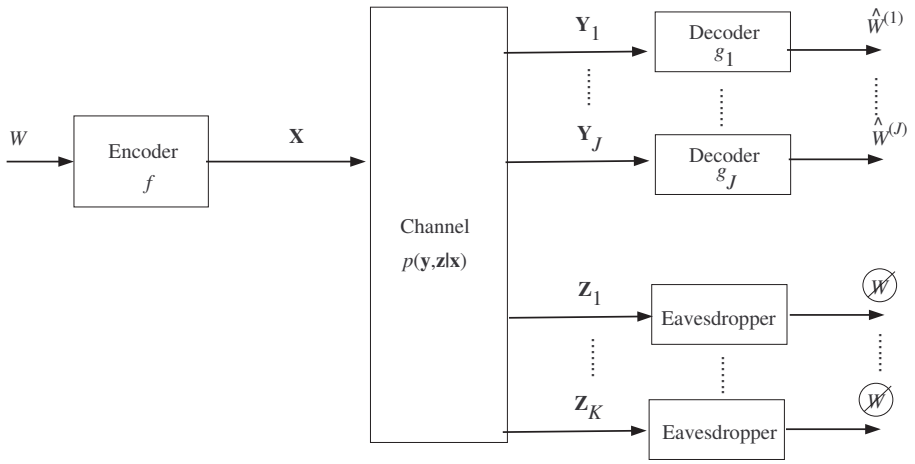


Figure 3.5: The compound wiretap channel.

The probability distribution for n uses of the channel is

$$p(\mathbf{y}_j, \mathbf{z}_k|\mathbf{x}) = \prod_{i=1}^{n} p(y_{j,i}, z_{k,i}|x_i), \quad j = 1, ..., J,\ k = 1, ..., K.$$

This may be re-written as

$$p(\mathbf{y}_j, \mathbf{z}_k|\mathbf{x}) = p(\mathbf{y}_j|\mathbf{x})p(\mathbf{z}_k|\mathbf{x}) = \prod_{i=1}^{n} p(y_{j,i}|x_i)p(z_{k,i}|x_i), \quad j = 1, ..., J,\ k = 1, ..., K, \tag{3.33}$$

since correlation between Y_j and Z_k does not change the secrecy capacity. Then a $(2^{nR}, n)$-code for the compound DM wiretap channel consists of:

- The message set $\{1, ..., 2^{nR}\}$, uniformly and randomly distributed.
- A (stochastic) encoder $f : \mathcal{W} \to \mathcal{X}^n$ which maps message $w \in \{1, ..., 2^{nR}\}$ to codeword $\mathbf{x} \in \mathcal{X}^n$.
- Decoders at the legitimate receiver $g_j : \mathcal{Y}_j^n \to \{1, ..., 2^{nR}\}$ map the received sequence $\mathbf{y}_j \in \mathcal{Y}_j^n$ to the estimate of the message $\hat{w}^{(j)} \in \{1, ..., 2^{nR}\}$, for $j = 1, ..., J$.

The reliability is measured by the average error probability, for $j = 1, ..., J$

$$P_e^{(n)} = \Pr\left[\hat{W}^{(j)} \neq W\right] = \frac{1}{2^{nR}} \sum_{w=1}^{2^{nR}} \Pr\left[\hat{w}^{(j)} \neq w\right], \tag{3.34}$$

while the secrecy level of the confidential message W at the kth eavesdropper, $k = 1, ..., K$, is measured by the equivocation rate

$$R_e^{(n)} = \frac{1}{n} H(W|\mathbf{Z}_k). \tag{3.35}$$

Then rate-equivocation pair (R, R_e) is achievable if there exists a sequence of $(2^{nR}, n)$-codes such that $P_{e,j}^{(n)} \leq \eta$, $j = 1, ..., J$ and we have

$$\frac{1}{n} H(W|\mathbf{Z}_k) \geq R_e - \epsilon_{1,k}, \quad k = 1, ..., K, \tag{3.36}$$

for any $\eta, \epsilon_{1,k} > 0$ and small for n sufficiently large. From Liang *et al* [72], we have the following:

Theorem 9. *The secrecy capacity of the compound wiretap channel is*

$$R = \max \min_{j,k} I(U; Y_j) - I(U; Z_k) \tag{3.37}$$

where U is an auxiliary random variable and the maximum is taken over distributions $p(u)p(x|u)$ that satisfy $U \to X \to (Y_j, Z_k)$, $j = 1, ..., J$, $k = 1, ..., K$. □

We note that the secure coding scheme essentially follows the one in the normal wiretap channel. It is also quite easy to see that this channel can encompass special cases where there are specific conditions on the eavesdroppers' channels, or some subset of the legitimate receivers' channel. Also, note that the legitimate receivers should enjoy an advantage over the eavesdroppers.

3.3 Broadcast Channels

A general 2-receiver broadcast channel (BC) is shown in Figure 3.6. There are two decoders, one for each receiver; messages W_0, W_1, W_2 are statistically independent. The message W_0 is meant for both decoders and is called the *common* or *public* message. The messages W_1, W_2 are called *private* messages. In some of the literature, the common message is omitted.

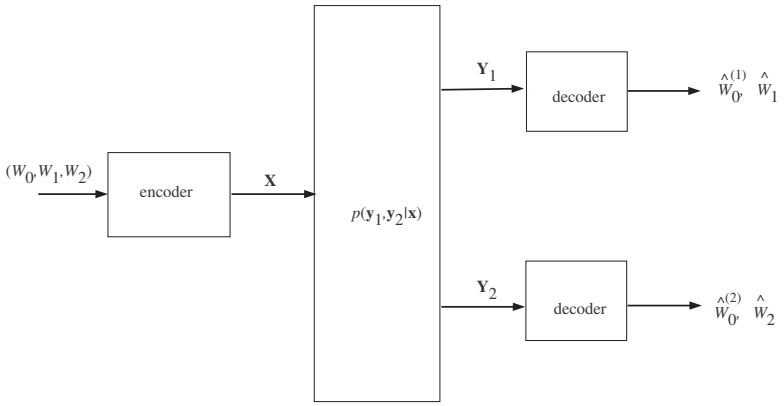


Figure 3.6: General 2-receiver discrete memoryless broadcast channel.

There are two main coding schemes for broadcast channels (BCs) without secrecy: superposition coding (see Bergmans [11], Gallager [50]) and binning, which is also known as Marton's achievability scheme (see Marton [84], El Gamal and van der Meulen [36]; for a discussion see Cover [28]). While it is of much interest to find the capacity region for general BCs with more than 2 receivers, this has been a long standing open problem, despite decades of intense research.

For BCs with security constraints, there are the following main classes:

1. The BC with one confidential message and one common message of Csiszár and Körner [30], a BC with $X \to (Y, Z)$ with Z the eavesdropper. The underlying

coding scheme is superposition coding and the secure scheme is code partitioning. Wyner's wiretap channel is a special case of this.

2. The BC with 2 confidential messages of Liu *et al* [78], with the BC $X \rightarrow (Y_1, Y_2)$ and Y_1 and Y_2 are mutual eavesdroppers. The underlying coding scheme is binning, and the secure scheme is double-binning as proposed in Liu *et al* [78].

3. The degraded BC with confidential messages and external eavesdropper studied in the work of Bagherikaram *et al* [6] (2 confidential messages, 2-receiver BC), our research [20] (K confidential messages, K-receiver BC), and Ekrem and Ulukus [33] (K confidential messages, K-receiver BC). The degradation is such that $X \rightarrow Y_1 \rightarrow Y_2 \rightarrow ... \rightarrow Y_K \rightarrow Z$ forms a Markov chain. The underlying coding scheme is superposition coding and the secure scheme is code partitioning.

4. The general BC with confidential messages and an external eavesdropper as studied in the work of Bagherikaram *et al* [6] (2 confidential messages, 2-receiver BC) and Kobayashi *et al* [65] (K confidential messages, K-receiver BC). The underlying scheme is binning and secure scheme uses double-binning, in the latter case extending Marton's scheme to an achievability scheme for K receivers.

5. The 3-receiver BC with degraded message sets (DMS) with one of the receivers an eavesdropper, studied in our research [22] and by Chia and El Gamal [17, 18]. The 3-receiver BC with DMS can be viewed as a major step towards obtaining the capacity region of the general K-receiver BC, and uses a combination of superposition coding and binning. Thus we have proposed a secure scheme that uses code partitioning and double-binning; we note that in [17, 18] code partitioning only is used for security.

In the next two sections, we will discuss the coding schemes of the first two classes of BCs with confidential messages as the coding schemes provide the basic coding schemes for the other three classes. For the degraded BC and the 3-receiver BC with DMS, we will defer discussion till we present our research in Chapter 4.

3.3.1 The BC with One Confidential Message and One Common Message

This general DM BC[6] with security constraints is shown in Figure 3.7. It is a generalization of the wiretap channel. We assume that the channel from X to Y enjoys an advantage over the channel from X to Z. In particular if the channel X to Z is less noisy than the channel X to Y then the equivocation rate $R_e = 0$. The transmitter sends common message W_0 which is received by both the legitimate receiver Y and the eavesdropper Z, and confidential message W_1 to the legitimate receiver be kept secret from the eavesdropper. A $(2^{nR_0}, 2^{nR_1}, n)$-code for the BC with one confidential message and one common message, consists of the parameters:

$$\begin{aligned}
\mathcal{W}_0 &= \left\{1, \ldots, 2^{nR_0}\right\}, \text{ (common message set)} \\
\mathcal{W}_1 &= \left\{1, \ldots, 2^{nR_1}\right\}, \text{ (private message set)}, \\
f &: \mathcal{W}_0 \times \mathcal{W}_1 \mapsto \mathcal{X}^n, \text{((stochastic) encoding function)}, \\
g_1 &: \mathcal{Y}^n \mapsto \mathcal{W}_0 \times \mathcal{W}_1, \text{ (legitimate user's decoding function)}, \\
g_2 &: \mathcal{Z}^n \mapsto \mathcal{W}_0 \text{ (eavesdropper's decoding function)}.
\end{aligned}$$

We have $g_1(\mathbf{Y}_1) = (\hat{W}_0^{(1)}, \hat{W}_1^{(1)})$, $g_2(\mathbf{Z}) = (\hat{W}_0^{(2)})$, and error probability

$$P_e^{(n)} = \Pr\left[(\hat{W}_0^{(1)}, \hat{W}_0^{(2)}, \hat{W}_1^{(1)}) \neq (W_0, W_0, W_1)\right]. \tag{3.38}$$

The decoders are set up to decode combinations of the messages; in the coding scheme this means that the decoders will decode specific parts of the superposed transmitted codeword. The secrecy level of the message W_1 sent to the legitimate user is defined by the equivocation rate $\frac{1}{n}H(W_1|\mathbf{Y})$. The rate tuple (R_0, R_1, R_{1e}) is said to be achievable if for any $\eta, \epsilon_1 > 0$, there exists a sequence of $(2^{nR_0}, 2^{nR_1}, n)$-codes for which $P_e^{(n)} \leq \eta$ and the equivocation rates R_{1e} satisfies

$$\frac{1}{n}H(W_1|\mathbf{Z}) \geq R_{1e} - \epsilon_1. \tag{3.39}$$

The rate-equivocation region $\mathcal{R}$ is the closure of the set of all achievable (R_0, R_1, R_{1e}). The secrecy capacity region $\mathcal{C}_S$ is the closure of all achievable pairs (R_0, R_1) so that perfect secrecy is achieved ($R_{1e} = R_1$).

[6] A BC is called general if the channel conditions on its receivers are general.

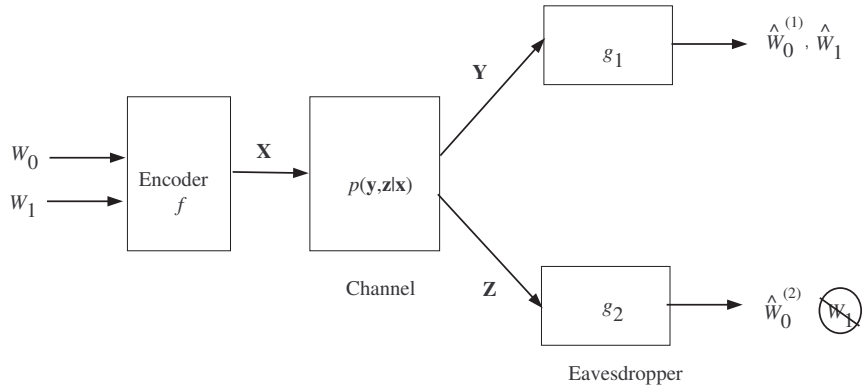


Figure 3.7: Broadcast channel with one confidential and one common message.

Theorem 10. *The rate equivocation region $\mathcal{R}$ for the DM BC with one confidential and one common message is the closure of all rate-tuples (R_0, R_1, R_{1e}) satisfying*

$$R_e \leq R_1, \tag{3.40}$$

$$0 \leq R_0 \leq \min\{I(U;Y), I(U;Z)\} \tag{3.41}$$

$$0 \leq R_0 + R_1 \leq I(V;Y|U) + \min\{I(U;Y), I(U;Z)\} \tag{3.42}$$

$$R_e \leq I(V;Y|U) - I(V;Z|U), \tag{3.43}$$

where the auxiliary random variables U, V satisfy the Markov chain condition $U \to V \to X \to (Y,Z)$ and with ranges $|\mathcal{U}| \leq |\mathcal{X}| + 3$, $|\mathcal{V}| \leq |\mathcal{X}|^2 + 4|\mathcal{X}| + 3$.

Proof. (Outline) We know that the DM wiretap channel is a special case of the BC with one confidential and one common message; specifically, the DM wiretap channel does not have the common message. Thus the coding scheme in the achievability for the BC with one confidential and one common message follows the scheme in Theorem 6. Rate splitting is used on the common message to move some of the rate (denoted as Δ) from R_0 to R_1. We then have the new rates

$$R_0' = R_0 - \Delta, \quad R_1' = R_1 + \Delta, \quad 0 \leq \Delta \leq R_0. \tag{3.44}$$

Now from the coding scheme in Theorem 6, we have (3.24). Substituting the new R_0' and R_1' into (3.24) and using Fourier-Motzkin elimination to eliminate R_0', R_1' and Δ, we can obtain the region in Theorem 10. The equivocation and converse proof and the proof of the ranges of U, V may be found in [30]. □

The following secrecy capacity region was also shown to be achievable in Csiszár and Körner [30].

Corollary 2. *The secrecy capacity region $\mathcal{C}_S$ for the DM BC with one confidential and one common message is the closure of all (R_0, R_1) satisfying*

$$0 \leq R_0 \leq \min\{I(U;Y), I(U;Z)\} \tag{3.45}$$

$$0 \leq R_1 \leq I(V;Y|U) - I(V;Z|U), \tag{3.46}$$

where the auxiliary random variables U, V satisfy the Markov chain condition $U \rightarrow V \rightarrow X \rightarrow (Y,Z)$ and with ranges $|\mathcal{U}| \leq |\mathcal{X}| + 3$, $|\mathcal{V}| \leq |\mathcal{X}|^2 + 4|\mathcal{X}| + 3$. □

3.3.2 BCs with 2 Confidential Messages

This model is studied in Liu *et al* [78] and is shown in Figure 3.8. It is a generalization of the 2-receiver BC with W_1 sent to receiver 1 and W_2 to receiver 2, to the wiretapping case. The transmitter sends no common message W_0, but W_1 is to be kept secret from the receiver 2 and W_2 is to be kept secret from the receiver 1. A $(2^{nR_1}, 2^{nR_2}, n)$-code for the BC with 2 confidential messages, consists of the parameters:

$$\begin{aligned}
\mathcal{W}_1 &= \left\{1, \ldots, 2^{nR_1}\right\}, \text{ (private message set 1)},\\
\mathcal{W}_2 &= \left\{1, \ldots, 2^{nR_2}\right\}, \text{ (private message set 2)},\\
f &: \mathcal{W}_1 \times \mathcal{W}_2 \mapsto \mathcal{X}^n, \text{((stochastic) encoding function)},\\
g_1 &: \mathcal{Y}_1^n \mapsto \mathcal{W}_1, \text{ (receiver 1 decoding function)},\\
g_2 &: \mathcal{Y}_2^n \mapsto \mathcal{W}_2 \text{ (receiver 2 decoding function)}.
\end{aligned}$$

We have $g_1(\mathbf{Y}_1) = (\hat{W}_1)$, $g_2(\mathbf{Y}_2) = (\hat{W}_2)$, and error probability

$$P_e^{(n)} = \Pr\left[(\hat{W}_1, \hat{W}_2) \neq (W_1, W_2)\right]. \tag{3.47}$$

The secrecy level of the message W_1 sent to user 1 is defined by the equivocation rate $\frac{1}{n}H(W_1|\mathbf{Y}_2)$, and that of W_2 sent to user 2 is defined by the equivocation rate $\frac{1}{n}H(W_2|\mathbf{Y}_1)$. The rate tuple $(R_1, R_2, R_{1e}, R_{2e})$ is said to be achievable if for any $\eta, \epsilon_1 > 0, \epsilon_2 > 0$, there exists a sequence of $(2^{nR_1}, 2^{nR_2}, n)$-codes for which $P_e^{(n)} \leq \eta$ and the equivocation rates satisfy

$$\frac{1}{n}H(W_1|\mathbf{Y}_2) \geq R_{1e} - \epsilon_1, \quad \frac{1}{n}H(W_2|\mathbf{Y}_1) \geq R_{2e} - \epsilon_2. \tag{3.48}$$

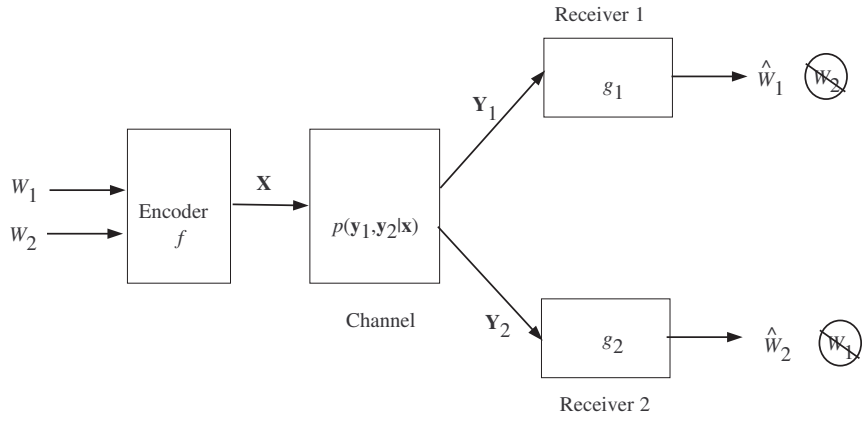


Figure 3.8: Broadcast channel with two confidential messages.

Inner and outer bounds to the secrecy capacity region are derived in [78] and stated below.

Theorem 11. *An inner bound to the secrecy capacity region $\mathcal{C}_S^I$ for the DM BC with two confidential messages is the closure of all rate pairs (R_1, R_2) satisfying*

$$0 \leq R_1 \leq I(V_1;Y_1|U) - I(V_1;Y_2|U,V_2) - I(V_1;V_2|U), \tag{3.49}$$

$$0 \leq R_2 \leq I(V_2;Y_2|U) - I(V_2;Y_1|U,V_1) - I(V_1;V_2|U), \tag{3.50}$$

where the auxiliary random variables V_1, V_2, U satisfy the Markov chain conditions $U \to V_1 \to X \to (Y,Z)$, $U \to V_2 \to X \to (Y,Z)$ and $(U,V_1,V_2) \to X \to (Y,Z)$.

An outer bound $\mathcal{C}_S^O$ is the closure of all rate pairs (R_1, R_2) satisfying

$$0 \leq R_1 \leq \min\{I(V_1;Y_1|U) - I(V_1;Y_2|U), I(V_1;Y_1|V_2;U) - I(V_1;Y_2|V_2,U)\}, \tag{3.51}$$

$$0 \leq R_2 \leq \min\{I(V_2;Y_2|U) - I(V_2;Y_1|U), I(V_2;Y_2|V_1;U) - I(V_2;Y_1|V_1,U)\}, \tag{3.52}$$

where the auxiliary random variables V_1, V_2, U satisfy the same conditions as in the inner bound.

If there exists a distribution that factors as $p(u)p(v_1,v_2|u)p(x|v_1,v_2)p(y_1,y_2|x)$ for which

$$I(V_1;Y_1|U) > I(V_1;Y_2,V_2|U), \quad I(V_2;Y_2|U) > I(V_2;Y_1,V_1|U), \tag{3.53}$$

then both receivers can achieve strictly positive rates while secrecy constraints are satisfied. These conditions can be derived from the inner bound in Theorem 11 by setting $R_1, R_2 > 0$. Also, if one receiver is less noisy than the other, say, X to Y_1 is less noisy than X to Y_2, so that $I(V;Y_1) \geq I(V;Y_2)$ for every $V \to X \to (Y_1, Y_2)$, we have the following secrecy capacity region:

$$\begin{aligned} R_1 &\leq \max_{p(x)} [I(X;Y_1) - I(X;Y_2)] \\ R_2 &= 0. \end{aligned} \tag{3.54}$$

The means that only the user with the better channel can enjoy a non-zero secrecy rate in the less noisy BC with two confidential messages. However, the MIMO Gaussian version of this BC with two confidential messages can have strictly positive rates at both receivers.

Proof. We will outline the coding scheme in the achievability proof, which combines Marton's achievability scheme [84],[36] (which uses Gel'fand Pinsker binning [51]) and random binning to obtain the double binning method. For the proof for the outer bound, which we note does not match the inner bound in general, we refer to Liu *et al* [78].

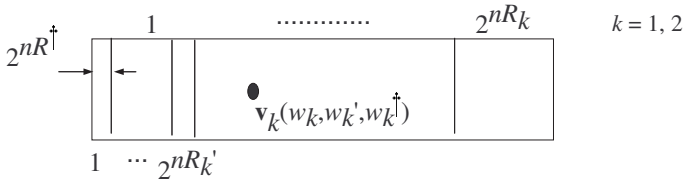


Figure 3.9: Double binning.

The double binning is applied to $\mathbf{v}_k$, $k = 1, 2$ sequences and is shown in Figure 3.9. To begin, fix the distributions $p(u)$, $p(v_1|u)$, $p(v_2|u)$, $p(x|v_1, v_2)$. Let

$$R_1' \triangleq I(V_1;Y_2|V_2,U) - \epsilon_1', \quad R_2' \triangleq I(V_2;Y_1|V_1,U) - \epsilon_1', \quad R^\dagger \triangleq I(V_1;V_2|U) + \epsilon_1', \tag{3.55}$$

where $\epsilon_1' > 0$ and is small as $n \to \infty$.

1. *Codebook generation*: Randomly generate $\mathbf{u}$ typical sequences known to the transmitter and both receivers with probability $p(\mathbf{u}) = \prod_{i=1}^{n} p(u_i)$. For $k = 1, 2$,

then generate $2^{R_k+R'_k+R^\dagger}$ $\mathbf{v}_k$ sequences using $p(\mathbf{v}_k|\mathbf{u}) = \prod_{i=1}^n p(v_{k,i}|u_i)$ and label them as $\mathbf{v}_k(w_k, w'_k, w_k^\dagger)$, where $w_k \in \{1, ..., 2^{nR_k}\}$, $w'_k \in \{1, ..., 2^{nR'_k}\}$, $w_k^\dagger \in \{1, ..., 2^{nR^\dagger}\}$. Thus the code for the kth channel $\mathcal{C}_k = \mathbf{v}_k(w_k, w'_k, w_k^\dagger)$ is partitioned into 2^{nR_k} bins, each of size $2^{nR'_k}$, with each bin further divided into bins of size $2^{nR^\dagger}$.

2. *Encoding*: To send message pair (w_1, w_2), the (stochastic) encoder randomly chooses a sub-bin from within the 2^{nR_k} bins. Then select the unique pair $w_1^\dagger, w_2^\dagger$ so that $\mathbf{v}_1(w_1, w'_1, w_1^\dagger)$ and $\mathbf{v}_2(w_2, w'_2, w_2^\dagger)$ are jointly typical. By the mutual covering lemma [36], this step succeeds with low error probability provided that

$$R^\dagger \geq I(V_1; V_2|U). \tag{3.56}$$

Then generate $\mathbf{x}$ using $p(\mathbf{x}|\mathbf{v}_1, \mathbf{v}_2)$.

3. *Decoding*: The decoders choose the unique w_k so that $(\mathbf{v}_k(w_k, w'_k, w_k^\dagger), \mathbf{y}_k)$ are jointly typical with respect to the distribution $p(v_k, v_k|u)$. The decoding succeeds with low error probability provided that

$$R_1 + R'_1 + R^\dagger \leq I(V_1; Y_1|U), \quad R_2 + R'_2 + R^\dagger \leq I(V_2; Y_2|U). \tag{3.57}$$

Combining the code generation (3.56) and decoding (3.57) conditions with the conditions (3.55), we can get the inner bound in the theorem.

□

We note that Xu *et al* in [116] generalized the BC with 2 confidential messages by the addition of the transmitter sending a common message to both receivers. At the same time, the perfect secrecy condition (required for obtaining secrecy capacity) was removed. The common message has no security constraints. The achievable coding scheme used rate-splitting (for the common message) and double binning. An inner bound to the rate equivocation region was derived, which was shown to reduce to the secrecy capacity inner bound $\mathcal{C}_S^I$ for perfect secrecy and no common message. An outer bound to the rate equivocation region was also derived in [116].

3.4 Relay Channels

The relay channel (RC) is a multiterminal problem where a source node communicates with one or more destination nodes via one or more relays. Coding schemes for the

RC have been well studied by Cover and El Gamal in [27]. For the problem of the RC with secrecy constraints, the coding scheme for achievability is to use the schemes for relaying in [27] together with either code partitioning or double binning for secrecy. For RCs with secrecy constraints, we have four main classes, shown in Figure 3.10.

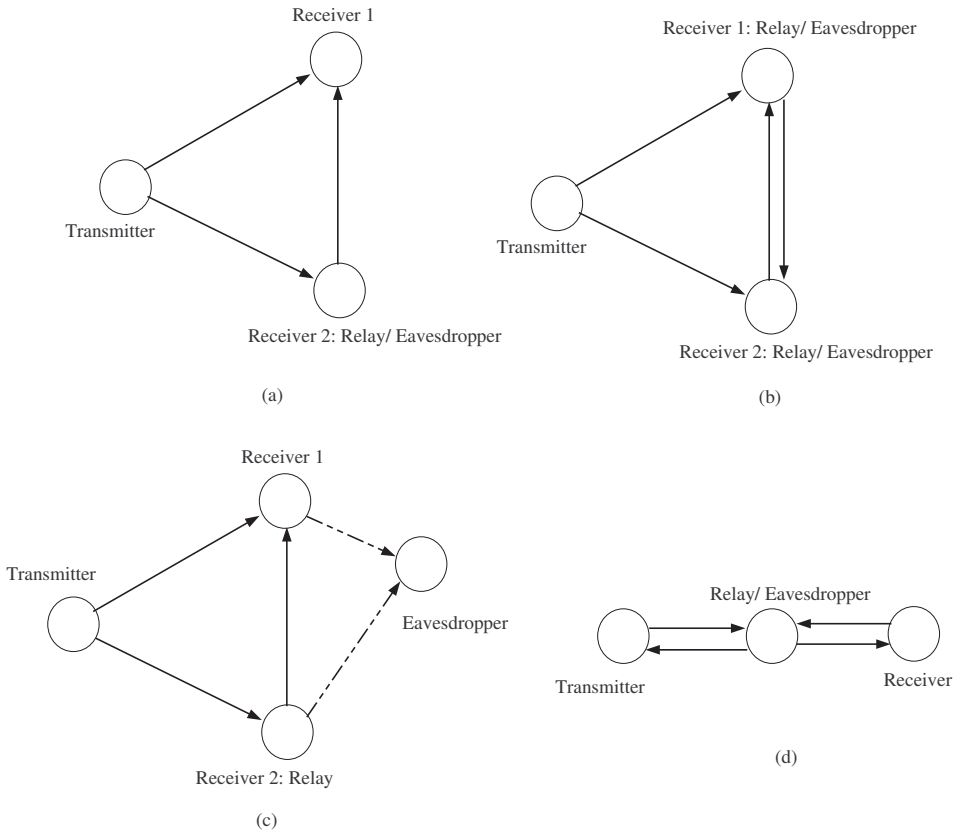


Figure 3.10: Classes of relay channels with confidential messages.

From Figure 3.10, we have the following:

- Class A: shown in Figure 3.10(a), this class has a transmitter sending its message to the receiver while it is helped by a relay that may be an eavesdropper. We can also view this as a BC, specifically a wiretap channel with the transmitter to receiver channel the legitimate channel and transmitter to relay channel the eavesdropper's channel. This is also known as a BC with one-sided cooperation. This model has been studied by Oohama [98]. Then the coding scheme for security will be code partitioning scheme for the wiretap channel and the BC with one confidential message, combined with Cover and El Gamal's schemes for the relaying.

- Class B: shown in Figure 3.10(b), this class has the transmitter sending two messages to the receivers while it is helped by relays at the receivers that may be eavesdroppers. We can also view this as a BC, specifically a 2-user BC with confidential messages with cooperation. This is also known as a BC with 2-sided cooperation. This model has been studied by [34], where the coding scheme for security will be the double binning scheme combined with Cover and El Gamal's schemes for the relaying.

- Class C: shown in Figure 3.10(c), this class has the transmitter sending a message to the receiver while it is helped by a relay and there is an external eavesdropper. This model was studied by Lai and El Gamal [69], and the coding scheme for security is code partitioning combined with relaying schemes. Two other variants of this scheme were studied. In the first variant, Aggarwal *et al* [2] considered splitting the message at the transmitter into two orthogonal components W_{11} and W_{10} and using two orthogonal channels to send W_{11} to the receiver and W_{10} to the relay. The eavesdropper can tap signals on either one or both of the two orthogonal channels. By virtue of the transmitter sending only part of its message to the relay, there is less information for the eavesdropper-relay channel and so the secure region is enhanced. In the second variant in Goel and Negi [52], a bank of relays is used in a MIMO Gaussian channel. The bank of relays act as if they are a combined MIMO relay. A two stage transmission scheme is proposed. In the first stage, the transmitter and receiver both transmit independent artificial noise to the bank of relays. In the second stage the relays send a weighted version of the received signal, while the transmitter transmits the secret message together with a weighted version of its artificial noise, which is to cancel the transmitter artificial noise component at the receiver. At the receiver 1, the known receiver and transmitter artificial noise are both removed. The eavesdropper's channel is then degraded by the artificial noise.

- Class D: shown in Figure 3.10(d), this class has the transmitter sending a message to the receiver which must pass through a relay which is an eavesdropper. This model can include a one-way or two-way operation. In the one-way operation, the transmitter sends the message to the receiver only. In the two way operation,

the transmitter and receiver exchange messages via the relay, using either a half duplex mode or full duplex mode. The two way operation includes the Two-way RC (TWRC) of He and Yener [56], where there are multiple relays through which the transmitter and receiver can communicate. In [56] nested lattice coding is proposed for the secure scheme[7].

Other TWRCs without secrecy can be found in the literature, which include the two-phase channel of Oechtering *et al* [95], which has a multiple access channel (MAC) connected with a BC by a relay; the TWRC of Gunduz *et al* [54]; and the TWRCs of Wilson *et al* [114] and Zhang *et al* [125], where an analogue network coding is used for Gaussian MIMO channels. It is of course of great interest to find out new secure schemes for these TWRCs as these models describe the critical part of a network where one cell (or base station) is connected to the other. In decentralized networks, the connections may have minimal hardware, so it is also important to see how security constraints affect these relays.

We will focus on Class C, that is the RC with external eavesdropper. We see that this configuration is simply a wiretap channel with an extra relay node. Recalling that the secrecy capacity is given by

$$C_S = \max_{T \to X \to (Y,Z)} [I(T;Y) - I(T;Z)],$$

we see that when the legitimate channel is less noisy than the eavesdropper's channel, $I(T;Y) \leq I(T;Z)$ and secrecy capacity goes to zero. By introducing the relay node to help the transmitter, Lai and El Gamal [69] showed that we can achieve positive secrecy capacity even in some scenarios where positive secrecy capacity is not possible without the relay node. So this RC configuration gives us an additional node with which we can enhance the secrecy capacity.

For a more practical implementation, we can use signal processing methods to enhance the secrecy capacity. We will propose such methods in Chapter 5, and we will defer further discussion of the RC with an external eavesdropper till then.

[7] We should remark that the nested lattice coding in [56] is used in conjunction with a jamming signal that turns the wiretap channel into an interference channel with secrecy constraints. However, the secure code was a random code superposed on points from a nested lattice code; the construction is not an explicitly all-lattice wiretap code. So we cannot really view the code in [56] as a 'pure' wiretap code.

3.5 Coding for Wiretap Channels

In this section we look at some practical coding methods to achieve secrecy for the wiretap channel. It should be noted that there are still many open areas for research in practical coding methods for the wiretap channel. Important studies for coding for the wiretap channel are the ones by Ozarow and Wyner [99] and Thangaraj *et al* [110].

The coding strategy by Wyner for the wiretap code is in general a nested code. The nested code structure of Wyner is similar to the well known nested codes of Zamir *et al* [120]. The coding scheme for the wiretap channel uses coset coding (or nested coset coding). The transmitter sends one of M equally likely messages from the secure codebook $\mathcal{C}$ which consists of subcodes $\{\mathcal{C}_1, \ldots, \mathcal{C}_M\}$, while the legitimate receiver can decode the codebook $\mathcal{C}$, but the eavesdropper can decode only within each subcode. The codebook $\mathcal{C}$ is the fine code and the ensemble of subcodes $\{\mathcal{C}_1, \ldots, \mathcal{C}_M\}$ is the coarse code, each of which is a coset of $\mathcal{C}$. To send message $m \in \{1, \ldots, M\}$, the transmitted word is chosen uniformly at random from the subcode $\mathcal{C}_m$; this stochastic encoding is the main source of uncertainty for the wiretapper.

In Thangaraj *et al* [110], the following theorem was shown.

Theorem 12. *If each subcode $\mathcal{C}_m^n$ is from a sequence of codes that achieve capacity over the eavesdropper's channel, then perfect secrecy can be achieved.* □

This theorem forms a key criterion for the design of wiretap codes.

3.5.1 Wiretap Channel Type II

In Ozarow and Wyner [99], the wiretap channel type II was introduced and studied. Here the source input alphabet is $\{0, 1\}$, the legitimate receiver's channel is noiseless and the eavesdropper can choose to see n_w of n symbols of the input sequences. The rate-equivocation region is the set of (R, R_e) so that

$$0 \leq R \leq 1, \quad 0 \leq R_e \leq \min(R, \epsilon), \tag{3.58}$$

where $\epsilon = 1 - n_w/n$ is the fraction of bits that are not observed by the eavesdropper. An achievable scheme for this uses the nested code structure. In a variant of this model in [110] where the eavesdropper can observe approximately n_w of n symbols of the input sequences chosen by the legitimate user, the legitimate user's channel is noiseless and the eavesdropper's channel is a binary erasure channel with erasure probability

$\epsilon = 1 - n_w/n$; this is called the type II binary erasure wiretap channel. A coding scheme also using the nested code structure is given in [110] for this channel, with specific realizations using low-density-parity-check (LDPC) codes. The criterion for the subcode to be capacity achieving on the eavesdropper's channel is for it to contain at least one codeword that matches received sequence $\mathbf{z}$ in the unerased positions, so that the eavesdropper decodes each corresponding message with equal probability with $H(W|\mathbf{Z} = \mathbf{z}) = nR$. Then perfect secrecy is achieved.

In the binary symmetric wiretap channel type II, studied in [110], the legitimate user's channel is noiseless and the eavesdropper's channel is a binary symmetric channel with crossover probability p. A nested code was again proposed, with subcodes using error-detecting codes (for example, Hamming codes and Bose-Chaudhuri-Hocquenghem codes). However, it was noted in [110] that by using error-detecting codes as the subcodes it is difficult to achieve non-zero secrecy capacity.

3.5.2 Non-type-II Wiretap Channels

We look at the more realistic case for channel coding where the legitimate user's channel and the eavesdropper's channel are of the same type. For example, if the legitimate user's channel is BEC, then the eavesdropper's channel is also BEC. For short we will call it the BEC wiretap channel. Now, a final, but important point is that the authors in [110] found that despite using LDPC codes in the coset coding for the BEC wiretap channel, the conditions of Theorem 12 are not met. Thus, unfortunately, the code was not information theoretically secure. This naturally motivates more research into the coding issue.

We will defer further discussion on coding for the wiretap channel until Chapter 6 when we discuss lattice coding for the Gaussian wiretap channel.

Chapter 4

Broadcast Channels with Confidential Messages

In this chapter we present our research on the multiple receiver BC with confidential messages. The work is motivated by the fact that the general multiple receiver BC region is still unknown, and it would be of great interest to be able to derive coding schemes to achieve secrecy rates for the general BC. Our work consists of two parts. The first is on the degraded K-receiver BC with K confidential messages and an external eavesdropper. The second is on the 3-receiver BC with degraded message sets (DMS) and confidential messages, where one of the receivers is an eavesdropper. As noted earlier, the 3-receiver BC with DMS and its achievability scheme represents a major step towards the characterization of the capacity region for the general multiple receiver BC. Thus we choose to study it under secrecy constraints as this will give us insights toward the general problem.

4.1 Introduction

In this section we introduce the various models of the multiple receiver BC with confidential messages. In the most general form, the transmitter sends confidential messages $W_1, ..., W_K$ to receivers $Y_1, ..., Y_K$ respectively, while we want to keep the messages secret from an external eavesdropper. So the model here is a $K + 1$ receiver BC. This is depicted in Figure 4.1. The transmitter may send the common message to all of the receivers, which may include the eavesdropper. This class of BC with confidential messages has two types: depending on whether the receivers have general conditions, or the receivers are degraded, with degradation $X \rightarrow Y_1 \rightarrow ... \rightarrow Y_K$.

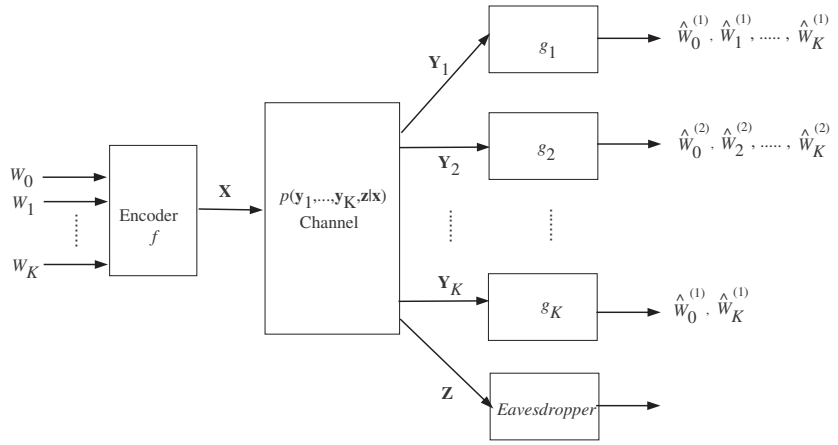


Figure 4.1: General BC with K receivers and confidential messages.

For the degraded BC, we note that the multiple versions of the messages received at say, receiver 1 is due to the degraded nature of the channel. That is, receiver 1 is the strongest receiver and can decode anything the others can decode. This channel has been studied independently in our research [20, 21] (K confidential messages, K-receiver BC), Bagherikaram *et al* [6] (2 confidential messages, 2-receiver BC), and Ekrem and Ulukus [33] (K confidential messages, K-receiver BC, common message to all receivers but not the eavesdropper).

The general 2-receiver BC with confidential messages and external eavesdropper was studied by Bagherikaram *et al* [6] (2 confidential messages, with the common message). A special case of the general K-receiver BC, where each receiver only receives its own message and the common message, was studied by Kobayashi *et al* [65] (K confidential messages, K-receiver BC, common message to all receivers including the eavesdropper). Both [6] and [65] used double binning to prove only inner bounds to the secrecy capacity region. Kobayashi *et al* [65] also extend Marton's scheme to K receivers (but with the limitation that each receiver only receives its own message and the common message). The work in Bagherikaram *et al* [6] is limited to Marton's scheme for 2 receivers, the existing state of the art.

A subclass of the general K-receiver BC is the 3-receiver BC with degraded message sets (DMS). In its general form, a common message W_0 is sent to all three receivers, private message W_1 is sent to receivers 1 and 2, and private message W_2 is sent to receiver 1 only. We can see quite easily then that this is a subclass of the general

K-receiver BC in Figure 4.1. This 3-receiver BC model (and some of its subclasses) was recently introduced in Nair and El Gamal [90, 91]. Our objective is to study this model of the 3-receiver BC with DMS of [90, 91] with *secrecy* constraints. We note that the insights which this model of the 3-receiver BC with degraded message sets might bring are due to it being a more general model than the 2- or 3-receiver degraded BC with secrecy constraints. We will be able to gain some insights on a secure coding scheme for general conditions on more than two receivers.

In particular, we characterize the transmission rates for the 3-receiver BCs with DMS from the model mentioned above where receiver 3 is an eavesdropper, and W_1 is sent to receiver 1. We call this model the 3-receiver BC with 2 DMS. We shall see that our 3-receiver BC with 2 DMS with secrecy constraints is an extension to 3 receivers of the BC with one confidential message and one common message of Csiszár and Körner [30], and a generalization of the 3-receiver degraded BC with secrecy constraints by virtue of the general conditions on the channels. We also note that Chia and El Gamal in [17, 18] have also studied the 3-receiver BC with 2 DMS with receiver 3 being an eavesdropper, but with W_1 sent to receivers 1 and 2, and using a different coding scheme[1].

Lastly we mention that Chia and El Gamal in [17, 18] also studied a 3-receiver BC with a certain degradedness order called the multilevel BC for receivers 2 and 3 being eavesdroppers. Recently this was generalized to a 3-receiver BC with receiver 1 less noisy than receiver 2, which is also less noisy than receiver 3, with receivers 2 and 3 being eavesdroppers by Salehkalaibar and Aref [104]. In both these works, the conditions on the channels were less general than the 3-receiver BC with DMS.

4.2 The K-receiver Degraded BC with Confidential Messages

In this section, we investigate the degraded K-receiver BC with confidential messages sent to each receiver in the presence of an eavesdropper, from which the messages are kept secret. We use the perfect secrecy criteria.

Our results are a generalization of our work for the 3-receiver BC in [21] and

[1] We shall discuss the differences in the coding schemes in Section 4.3.2.

earlier results for 2-receiver BCs in [6]. It is noted that results similar to ours have been established independently in [33], where Ekrem and Ulukus [33] examined the K-receiver degraded BC and one eavesdropper with confidential messages as well as a common message sent to the receivers. However, there are some appreciable differences between our approach and that in [33]. In particular, equivocation calculation and proof of the converse in [33] are accomplished from the perspective of the channel sum rate. In contrast, we provide the error probability analysis and the equivocation calculation with respect to the kth receiver's achievable rate. In our proof of the converse, which we have shown for the kth receiver, we note that our choice of auxiliary random variables is different from that of [6] and [33]. Due to the presence of the wiretapper, it is also different from the choice in Borade *et al.* [13] where the capacity region for the degraded K-receiver BC using superposition coding without confidential messages is studied.

4.2.1 Channel Model

The degraded K-receiver BC with an external eavesdropper is depicted in Figure 4.2. We note that we do not have the common message. The receivers are degraded in that $Y_1 \rightarrow Y_2 \rightarrow ... \rightarrow Y_K \rightarrow Z$ forms a Markov chain.

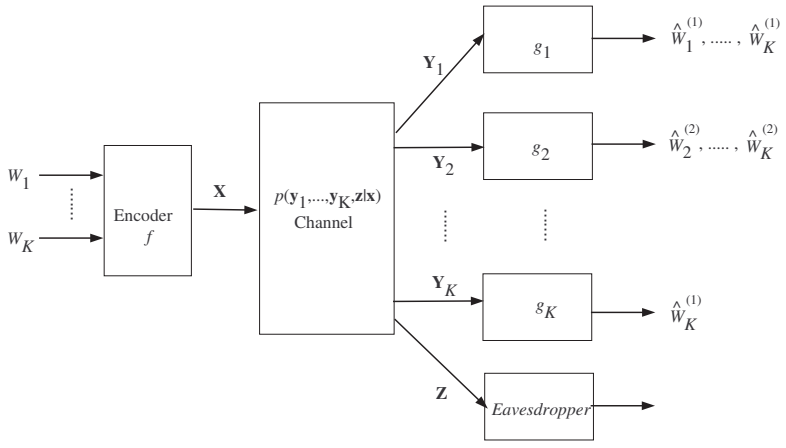


Figure 4.2: Degraded BC with K receivers and confidential messages.

This model consists of a finite input alphabet $\mathcal{X}$ and finite output alphabets $\mathcal{Y}_1, \ldots, \mathcal{Y}_K, \mathcal{Z}$ and has conditional distribution $p(y_1, \ldots, y_K, z|x)$. Thus the discrete memoryless BC with K receivers and an eavesdropper has an input random sequence $\mathbf{X}$, K output random sequences, $\mathbf{Y}_1, \ldots, \mathbf{Y}_K$, at the intended receivers, and an output

random sequence at the eavesdropper $\mathbf{Z}$. Likewise, we have $\mathbf{y}_1 \in \mathcal{Y}_1^n, \ldots, \mathbf{y}_K \in \mathcal{Y}_K^n$ and $\mathbf{z} \in \mathcal{Z}^n$. The conditional distribution for n uses of the channel is

$$p(\mathbf{y}_1, \ldots, \mathbf{y}_K, \mathbf{z}|\mathbf{x}) = \prod_{i=1}^{n} p(y_{1i}, \ldots, y_{Ki}, z_i|x_i). \tag{4.1}$$

The transmitter has to send independent messages $(W_1, \ldots, W_K)$ to the receivers in perfect secrecy. This is done using a $(2^{nR_1}, \ldots, 2^{nR_K}, n)$-code for the BC, which consists of the stochastic encoder

$$f : \left\{1, \ldots, 2^{nR_1}\right\} \times \left\{1, \ldots, 2^{nR_2}\right\} \times \cdots \times \left\{1, \ldots, 2^{nR_K}\right\} \mapsto \mathcal{X}^n, \tag{4.2}$$

and the decoders

$$g_k : \mathcal{Y}_k^n \mapsto \left\{1, \ldots, 2^{nR_k}\right\}, \text{ for } k = 1, 2, \ldots, K. \tag{4.3}$$

The probability of error is defined as the probability that the decoded messages are not equal to the transmitted messages, i.e.,

$$P_e^{(n)} \triangleq \Pr\left[\bigcup_{k=1}^{K} \{g_k(\mathbf{Y}_k) \neq W_k\}\right]. \tag{4.4}$$

The information rate for the kth receiver is $R_k = \frac{1}{n}H(W_k)$. Define $\mathcal{K} \subseteq \{1, ..., K\}$ and $W(\mathcal{K}) \triangleq \{W_k : k \in \mathcal{K}\}$. We now define the following equivocation rates for the K-receiver degraded BC:

$$\begin{cases} \frac{1}{n}H(W_k|\mathbf{Z}), & \text{for } k = 1, \ldots, K, \\ \frac{1}{n}H(W(\mathcal{K})|\mathbf{Z}) & \text{for all } \mathcal{K} \subseteq \{1, ..., K\}. \end{cases} \tag{4.5}$$

For perfect secrecy, we need

$$\begin{cases} \frac{1}{n}H(W_k|\mathbf{Z}) \geq R_k - \eta, & \text{for } k = 1, \ldots, K, \\ \frac{1}{n}H(W(\mathcal{K})|\mathbf{Z}) \geq \sum_{k\in\mathcal{K}} R_k - \eta' & \text{for } k \in \mathcal{K}, \end{cases} \tag{4.6}$$

for arbitrary $\eta, \eta' > 0$.

4.2.2 The Secrecy Capacity Region

The secret rate tuple $(R_1, R_2, \ldots, R_K)$ is achievable if, for any arbitrarily small $\epsilon' > 0$, $\epsilon_k > 0$, $k = 1, \ldots, K$, and $\mathcal{K} \subseteq \{1, \cdots, K\}$, there exist $(2^{nR_1}, \ldots, 2^{nR_K}, n)$-codes for which $P_e^{(n)} \leq \epsilon'$ and

$$\begin{cases} R_{e(k)} \geq R_k - \epsilon_k, & \text{for } k = 1, \ldots, K, \\ \sum_{k\in\mathcal{K}} R_{e(k)} \geq \sum_{k\in\mathcal{K}} R_k - \sum_{k\in\mathcal{K}} \epsilon_k, & \text{for } k \in \mathcal{K}. \end{cases} \tag{4.7}$$

The above equation (4.7) gives the security conditions for the K-receiver BC with an eavesdropper under perfect secrecy requirements in (4.6). We then have the following secrecy capacity region:

Theorem 13. *The secrecy capacity region for the K-receiver degraded BC with an external eavesdropper is the closure of all rate tuples $(R_1, \ldots, R_K)$ satisfying*

$$\begin{aligned} R_1 &\leq I(X;Y_1|U_2) - I(X;Z|U_2), \\ R_k &\leq I(U_k;Y_k|U_{k+1}) - I(U_k;Z|U_{k+1}), \ \textit{for } k = 2, \ldots, K-1, \\ R_K &\leq I(U_K;Y_K) - I(U_K;Z), \end{aligned} \tag{4.8}$$

over all probability distributions of the form $p(u_K)p(u_{K-1}|u_K)\cdots p(u_2|u_3)p(x|u_2)$, so that the auxiliary r.v.s $\{U_k\}_{k=2}^K$ satisfy the Markov chain condition $U_K \to U_{K-1} \to \cdots \to U_2 \to X$.

Proof. The code construction and error analysis is in Section 4.2.3 and equivocation calculation is in Section 4.2.4. We note first that, for the codewords sent to each receiver, code partitioning will give rise, firstly, to an overall rate which includes all the subcodes; and secondly, a rate within each subcode. The conditions on the overall rates for successful decoding at the receivers and the conditions on the rates of the codewords within each subcode satisfying perfect secrecy in the equivocation calculation are then combined to show achievability of Theorem 13. Finally the proof of the converse is found in Section 4.2.5. □

If we use superposition coding with code partitioning to achieve the rates in Theorem 13, then the secrecy capacity region may be interpreted as the capacity region for the K-receiver BC using superposition coding without the eavesdropper, with the rates at each receiver each reduced due to the presence of the eavesdropper. However, we shall see that the choice of auxiliary r.v.s in our proof of converse for the K-receiver BC will be different from that of [13], which is without the secrecy conditions. This is also in contrast to the 2-receiver BC with an eavesdropper in [6], where the same definition for the auxiliary random variables in the converse proof can be used for the scenarios with and without the secrecy conditions.

4.2.3 Code Construction and Error Analysis

Here we employ superposition coding and Wyner's random code partitioning to show the achievable rate tuples $(R_1, \ldots, R_K)$. For brevity, we use $p_{\mathbf{Y}_1|\mathbf{X}}$ to denote the channel from $\mathbf{X}$ to $\mathbf{Y}_1$, similarly for the channels from $\mathbf{X}$ to outputs $\mathbf{Y}_2, \ldots, \mathbf{Y}_K$ and $\mathbf{Z}$, by $p_{\mathbf{Y}_2|\mathbf{X}}, \ldots, p_{\mathbf{Y}_K|\mathbf{X}}$ and $p_{\mathbf{Z}|\mathbf{X}}$, respectively.

The coding strategy is depicted in Fig. 4.3. The message $W_k \in \{1, \ldots, L_k\}$ with $L_k \triangleq 2^{nR_k}$ for $k = 1, \ldots, K$, is sent by a code of length $N_k = L_k L'_k$. This code is partitioned into L_k subcodes each of size L'_k, with $L'_k \triangleq 2^{nR'_k}$ for some R'_k. The R'_k is referred to as 'confusion' rate. Each of the L_k subcodes is a code for the channel $p_{\mathbf{Z}|\mathbf{X}}$, while each of the entire codes of size N_k is a code simultaneously for both the channels $p_{\mathbf{Y}_k|\mathbf{X}}$ and $p_{\mathbf{Z}|\mathbf{X}}$. The codes for simultaneous use for $p_{\mathbf{Y}_k|\mathbf{X}}$ and $p_{\mathbf{Z}|\mathbf{X}}$ have to satisfy the transmission requirements for the BC [11], so that

$$\begin{aligned} &\frac{1}{n}\log N_1 \leq I(X; Y_1|U_2), \\ &\frac{1}{n}\log N_k \leq I(U_k; Y_k|U_{k+1}), \ \text{ for } k = 2, \ldots, K-1, \\ &\frac{1}{n}\log N_K \leq I(U_K; Y_K). \end{aligned} \tag{4.9}$$

Random codebook generation: Suppose that we have the p.d.f.s

$$\begin{cases} p(u_K), \\ p(u_k|u_{k+1}), \ \text{ for } k = 2, \ldots, K-1, \\ p(x|u_2). \end{cases} \tag{4.10}$$

For a given rate tuple $(R_1, \ldots, R_K, R'_1, \ldots, R'_K)$, in order to send message W_K, generate $2^{n(R_K+R'_K)}$ independent codewords $\mathbf{u}_K(w''_K)$, for $w''_K \in \{1, \ldots, 2^{n(R_K+R'_K)}\}$ according to the p.d.f. $p(\mathbf{u}_K) = \prod_{i=1}^{n} p(u_{Ki})$. Then, partition $\mathbf{u}_K(w''_K)$ into $L_K = 2^{nR_K}$ subcodes, $\{\mathcal{C}_i^{(K)}\}_{i=1}^{L_K}$ with $|\mathcal{C}_i^{(K)}| = L'_K = 2^{nR'_K}\ \forall i$.

The message for the kth receiver, for $k = 2, 3, \ldots, K-1$, is sent by generating $2^{n(R_k+R'_k)}$ independent codewords $\mathbf{u}_k(w''_k, \ldots, w''_K)$, for $w''_k \in \{1, \ldots, 2^{n(R_k+R'_k)}\}$ according to the conditional p.d.f.

$$p(\mathbf{u}_k|\mathbf{u}_{k+1}) = \prod_{i=1}^{n} p(u_{ki}|u_{(k+1)i}). \tag{4.11}$$

Then, partition $\mathbf{u}_k(w''_k, \ldots, w''_K)$ into $L_k = 2^{nR_k}$ subcodes, $\{\mathcal{C}_i^{(k)}\}_{i=1}^{L_k}$, with $|\mathcal{C}_i^{(k)}| = L'_k = 2^{nR'_k}\ \forall i$. Finally, to send the message intended for the first receiver, generate

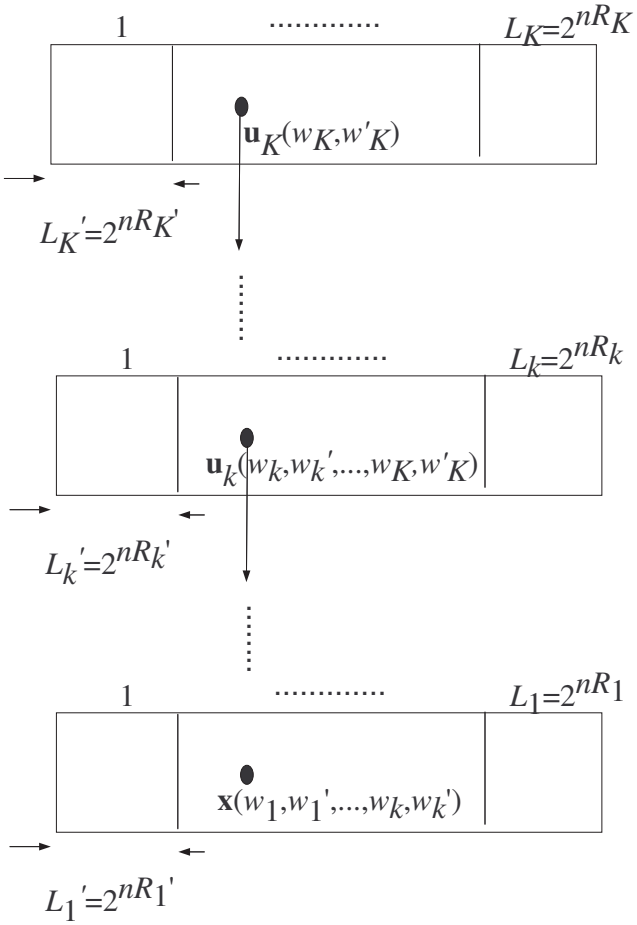


Figure 4.3: Coding for K receiver BC with an eavesdropper.

$2^{n(R_1+R_1')}$ independent codewords $\mathbf{x}(w_1'', \ldots, w_K'')$, for $w_1'' \in \{1, \ldots, 2^{n(R_1+R_1')}\}$ according to the p.d.f. $p(\mathbf{x}|\mathbf{u}_2) = \prod_{i=1}^{n} p(x_i|u_{2i})$. Then, partition $\mathbf{x}(w_1'', \ldots, w_K'')$ into $L_1 = 2^{nR_1}$ subcodes, $\{\mathcal{C}_i^{(1)}\}_{i=1}^{L_1}$, with $|\mathcal{C}_i^{(1)}| = L_1' = 2^{nR_1'}\ \forall i$.

Following this code structure, the codeword indices w_k'' may be expressed as $w_k'' = (w_k, w_k')$, where $w_k \in \{1, \ldots, 2^{nR_k}\}$ is the index of the message transmitted to the kth receiver, and $w_k' \in \{1, \ldots, 2^{nR_k'}\}$ denotes the index of the codeword within the subcodes $\mathcal{C}_i^{(k)}$, selected for transmission along with w_k to ensure secrecy.

Encoding: The encoding is by superposition coding. To send the message $w_K = i_K$, for $1 \leq i_K \leq L_K$, the transmitter chooses one of the $\mathbf{u}_K(w_K'')$ codewords uniformly and randomly from $\{\mathcal{C}_{i_K}^{(K)}\}_{i_K=1}^{L_K}$. Then, to send the message $w_{K-1} = i_{K-1}$, for $1 \leq i_{K-1} \leq L_{K-1}$, the transmitter selects one of the $\mathbf{u}_{K-1}(w_{K-1}'', w_K'')$ uniformly randomly from $\{\mathcal{C}_{i_{K-1}}^{(K-1)}\}_{i_{K-1}=1}^{L_{K-1}}$, given $\mathbf{u}_K(w_K'')$. Sequentially, the transmitter sends the message $w_k =$

i_k, for $1 \leq i_k \leq L_k$ and $k = 2, \ldots, K-2$, to the kth receiver by choosing one of the $\mathbf{u}_k(w_k'', \ldots, w_K'')$ uniformly and randomly from $\{\mathcal{C}_{i_k}^{(k)}\}_{i_k=1}^{L_k}$, given $\mathbf{u}_{k+1}(w_{k+1}'', \ldots, w_K'')$. Lastly, to send $w_1 = i_1$ for $1 \leq i_1 \leq L_1$, given $\mathbf{u}_2(w_2'', \ldots, w_K'')$, the transmitter chooses one of the $\mathbf{x}(w_1'', \ldots, w_K'')$ uniformly randomly from $\{\mathcal{C}_{i_1}^{(1)}\}_{i_1=1}^{L_1}$.

Decoding: For convenience, let $X = U_1$. Then, we have:

1. At receiver K, given that $\mathbf{y}_K$ is received, find a $\hat{w}_K$, such that $(\mathbf{u}_K(\hat{w}_K, w_K'), \mathbf{y}_K) \in \mathcal{T}_\epsilon^n(P_{U_K Y_K})$.

2. At the kth receiver, for $k = 1, \ldots, K-1$, given that $\mathbf{y}_k$ is received, find a $(\hat{w}_k, \ldots, \hat{w}_K)$ such that

$$(\mathbf{u}_K(\hat{w}_K, w_K'), \ldots, \mathbf{u}_k(\hat{w}_k, w_k', \ldots, \hat{w}_K, w_K'), \mathbf{y}_k) \in \mathcal{T}_\epsilon^n(P_{U_K U_{K-1} \cdots U_k Y_k}). \tag{4.12}$$

For each of the above cases, if there is none or more than one possible decoded message, then an error will be declared. Note that w_k' is unimportant for the decoding of w_k at the kth receiver.

Probability of error analysis: Assume without loss of generality that $(w_1, \ldots, w_k) = (1, \ldots, 1)$ is sent and an arbitrary w_k' is sent for the subcodes $\mathcal{C}_{i_k}^{(k)}$ $\forall k$. Let $\epsilon' > 0$ be a small positive value that goes to zero for n sufficiently large.

For Receiver K, the error events are:

(a) $\mathtt{E}_1^{(K)}$: $(w_K, w_K') = (1, 1)$ but $\mathbf{u}_K$ is not jointly typical with $\mathbf{y}_K$. By the properties of typical sequences, $\Pr\left[\mathtt{E}_1^{(K)}\right] \leq \epsilon'$.

(b) $\mathtt{E}_2^{(K)}$: $w_K \neq 1$, w_K' is arbitrary and $\mathbf{u}_K$ is jointly typical with $\mathbf{y}_K$.

The probability of the event $\mathtt{E}_2^{(K)}$ is

$$\begin{aligned} \Pr\left[\mathtt{E}_2^{(K)}\right] &= \Pr\left[\bigcup_{w_K \neq 1} \{(\mathbf{U}_K(w_K, w_K'), \mathbf{y}_K) \in \mathcal{T}_\epsilon^n(P_{U_K Y_K})\}\right] \\ &\leq 2^{n(R_K + R_K' - I(U_K; Y_K) - \delta(\epsilon))}, \end{aligned} \tag{4.13}$$

where $\delta(\epsilon) \to 0$ for n sufficiently large. Thus $P_{e,K} \leq 2\epsilon'$ by the union bound if

$$R_K + R_K' < I(U_K; Y_K). \tag{4.14}$$

The kth receiver, $k = 1, ..., K-1$, should decode $w_k, ..., w_K$. It does so by decoding $\mathbf{u}_K, \mathbf{u}_{K-1}, ..., \mathbf{u}_k$, which it can, due to the degraded nature of the channel. That is, the kth receiver can decode anything that the (k+1)th up to the Kth receiver can. Let us denote the events

$$\mathtt{D_k} : (\mathbf{u}_k, \mathbf{u}_{k+1}, \cdots, \mathbf{u}_K, \mathbf{y}_k) \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} \cdots U_K Y_k})$$
$$\mathtt{D_k}^c : (\mathbf{u}_k, \mathbf{u}_{k+1}, \cdots, \mathbf{u}_K, \mathbf{y}_k) \notin \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} \cdots U_K Y_k}).$$

We have the $K-k+2$ events for the kth receiver:

$\mathtt{E}_1^{(k)} : (w_k, w_k', ..., w_K, w_K') = (1, 1, ..., 1, 1), \mathtt{D_k}^c$ occurred,

$\mathtt{E}_2^{(k)} : w_k \neq 1, w_k'$ arbitrary, $(w_{k+1}, w_{k+1}', ..., w_K, w_K') = (1, 1, ..., 1, 1), \mathtt{D_k}$ occurred,

$\mathtt{E}_3^{(k)} : w_k \neq 1, w_{k+1} \neq 1, w_k', w_{k+1}'$ arbitrary,

$(w_{k+2}, w_{k+2}', ..., w_K, w_K') = (1, 1, ..., 1, 1), \mathtt{D_k}$ occurred

$\vdots$

$\mathtt{E}_{K-k+2}^{(k)} : w_k \neq 1, w_{k+1} \neq 1, \cdots, w_K \neq 1, w_k', w_{k+1}', \cdots, w_K'$ arbitrary, $\mathtt{D_k}$ occurred.

This leads to the following probabilities on the error events for the kth receiver:

$$\Pr\left[\mathtt{E}_1^{(k)}\right] \leq \epsilon',$$

$$\Pr\left[\mathtt{E}_2^{(k)}\right] = \Pr\left[\bigcup_{w_k \neq 1} \left\{(\mathbf{U}_k, \mathbf{u}_{k+1}, \cdots, \mathbf{u}_K, \mathbf{y}_k) \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} \cdots U_K})\right\}\right]$$
$$\overset{(a)}{\leq} 2^{n(R_k + R_k' - I(U_k; Y_k | U_{k+1}, \cdots, U_K) - \delta(\epsilon))} \overset{(b)}{=} 2^{n(R_k + R_k' - I(U_k; Y_k | U_{k+1}) - \delta(\epsilon))},$$

$$\Pr\left[\mathtt{E}_3^{(k)}\right] = \Pr\left[\bigcup_{w_k \neq 1} \bigcup_{w_{k+1} \neq 1} \left\{(\mathbf{U}_k, \mathbf{U}_{k+1}, \mathbf{u}_{k+2}, \cdots, \mathbf{u}_K, \mathbf{y}_k) \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} \cdots U_K})\right\}\right]$$
$$\leq 2^{n(R_k + R_{k+1} + R_k' + R_{k+1}' - I(U_k, U_{k+1}; Y_k | U_{k+2}, \cdots, U_K) - \delta(\epsilon))}$$
$$\overset{(c)}{=} 2^{n(R_k + R_{k+1} + R_k' + R_{k+1}' - I(U_k; Y_k | U_{k+1}) - I(U_{k+1}; Y_k | U_{k+2}) - \delta(\epsilon))},$$
$$\vdots$$
$$\vdots$$

$$\Pr\left[\mathtt{E}_{K-k+2}^{(k)}\right] = \Pr\left[\bigcup_{w_k \neq 1} \cdots \bigcup_{w_K \neq 1} \left\{(\mathbf{U}_k, \mathbf{U}_{k+1}, \cdots, \mathbf{U}_K, \mathbf{y}_k) \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} \cdots U_K})\right\}\right]$$
$$\leq 2^{n\left(\sum_{i=k}^K (R_i + R_i') - I(U_k, \cdots, U_K; Y_k) - \delta(\epsilon)\right)}$$
$$\overset{(d)}{=} 2^{n\left(\sum_{i=k}^K (R_i + R_i') - I(U_k; Y_k | U_{k+1}) - I(U_{k+1}; Y_k | U_{k+2}) - \cdots - I(U_K; Y_k) - \delta(\epsilon)\right)},$$

where (a) uses Theorem 5 and (b),(c) and (d) are by $U_K \to U_{K-1} \to \cdots \to U_1 \to Y_k$. Then the kth receiver has error probability $P_{e,k} \leq (K-k+2)\epsilon'$ if

$$\begin{aligned}
&R_k + R'_k < I(U_k; Y_k|U_{k+1}) \\
&R_k + R_{k+1} + R'_k + R'_{k+1} < I(U_k; Y_k|U_{k+1}) - I(U_{k+1}; Y_k|U_{k+2}) \\
&\vdots \\
&\sum_{i=k}^{K}(R_i + R'_i) < I(U_k; Y_k|U_{k+1}) - I(U_{k+1}; Y_k|U_{k+2}) - \cdots - I(U_K; Y_k). \qquad (4.15)
\end{aligned}$$

Now combining the conditions for low error probability for n sufficiently large for all K receivers and removing the redundant inequalities, we have, for small error probability at all the receivers,

$$\begin{aligned}
R_K + R'_K &\leq I(U_K; Y_K), &(4.16)\\
R_{K-1} + R'_{K-1} &\leq I(U_{K-1}; Y_{K-1}|U_K), &(4.17)\\
&\vdots \\
R_k + R'_k &\leq I(U_k; Y_k|U_{k+1}), &(4.18)\\
&\vdots \\
R_1 + R'_1 &\leq I(U_1; Y_1|U_2) = I(X; Y_1|U_2). &(4.19)
\end{aligned}$$

We have shown the rate conditions for successful decoding at the receivers. We will now show the equivocation calculation.

4.2.4 Equivocation Calculation

Here we show the equivocation calculation. We first note that there is an alternative method to obtaining the equivocation and the sizes of the code partitions, which was presented in our earlier work [20, 21], and is outlined in Appendix B.3. However, the method presented in this section has the advantage of showing how to obtain the code partition size from first principles.

We now show the calculation for the kth receiver, then for the all the receivers $k = 1, \cdots, K$. We still let $X = U_1$ and $U_{K+1} = \emptyset$ for convenience. We shall make use of the relation

$$H(U, V) = H(U) + H(V|U). \qquad (4.20)$$

Let $J_k \in \{1, \cdots, L'_k\}$, $L'_k = 2^{nR'_k}$ for all $k = 1, \ldots, K$. For the kth receiver, $k = 1, \ldots, K$, we have

$$\begin{aligned}
H(W_k|\mathbf{Z}) &\geq H(W_k|\mathbf{Z}, \mathbf{U}_{k+1}) \\
&= H(W_k, J_k|\mathbf{Z}, \mathbf{U}_{k+1}) - H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, W_k) \\
&\overset{(a)}{\geq} H(\mathbf{U}_k|\mathbf{Z}, \mathbf{U}_{k+1}) - H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, W_k) \\
&= H(\mathbf{U}_k, \mathbf{Z}|\mathbf{U}_{k+1}) - H(\mathbf{Z}|\mathbf{U}_{k+1}) - H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, W_k) \\
&= H(\mathbf{U}_k|\mathbf{U}_{k+1}) + H(\mathbf{Z}|\mathbf{U}_k, \mathbf{U}_{k+1}) - H(\mathbf{Z}|\mathbf{U}_{k+1}) - H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, W_k) \\
&= H(\mathbf{U}_k|\mathbf{U}_{k+1}) - I(\mathbf{U}_k; \mathbf{Z}|\mathbf{U}_{k+1}) - H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, W_k), \qquad (4.21)
\end{aligned}$$

where (a) is due to $\mathbf{U}_k$ being a function of (W_k, J_k). We now bound each of the terms in the last line of (4.21). For the first term, given that $\mathbf{U}_{k+1} = \mathbf{u}_{k+1}$, $\mathbf{U}_k$ has $2^{n(R_k+R'_k)}$ possible values with equal probability. As a consequence, we have

$$H(\mathbf{U}_k|\mathbf{U}_{k+1}) = n(R_k + R'_k). \qquad (4.22)$$

For the second term, it can be shown, following the method used in Liu *et al* [78], that

$$I(\mathbf{U}_k; \mathbf{Z}|\mathbf{U}_{k+1}) \leq nI(U_k; Z|U_{k+1}) + n\delta, \qquad (4.23)$$

where $\delta > 0$ and is small for n sufficiently large. To evaluate the last term in the last line of (4.21), we introduce the following lemma, which is proved in Appendix B.1.

Lemma 4. *We have*

$$H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, W_k) \leq n(R'_k - I(U_k; Z|U_{k+1}) + \delta(\epsilon)) + 2, \qquad (4.24)$$

where $\delta(\epsilon) \to 0$ for $\epsilon \to 0$ and n sufficiently large, under the condition that

$$R'_k \geq I(U_k; Z|U_{k+1}). \qquad (4.25)$$

□

Substituting (4.22), (4.23) and (4.24) into the last line of (4.21), we have

$$\frac{1}{n}H(W_k|\mathbf{Z}) \geq R_k - \left(\delta + \delta(\epsilon) + \frac{2}{n}\right) = R_k - \eta, \qquad (4.26)$$

where $\eta \triangleq \delta + \delta(\epsilon) + \frac{2}{n}$, which is > 0 and is small as n is large. We have now shown that the equivocation rate of the kth receiver meets the perfect secrecy requirement (4.6).

We now proceed to show that the sum equivocation rate of the K receivers meets the perfect secrecy requirement (4.6). We have

$$\begin{aligned}
&H(W_1,\cdots,W_K|\mathbf{Z})\\
&= H(W_1,\cdots,W_K,J_1,\cdots,J_K|\mathbf{Z}) - H(J_1,\cdots,J_K|\mathbf{Z},W_1,\cdots,W_K)\\
&= H(W_1,\cdots,W_K,J_1,\cdots,J_K) - I(W_1,\cdots,W_K,J_1,\cdots,J_K;\mathbf{Z})\\
&\quad - H(J_1,\cdots,J_K|\mathbf{Z},W_1,\cdots,W_K)\\
&\geq H(W_1,\cdots,W_K,J_1,\cdots,J_K) - I(W_1,\cdots,W_K,J_1,\cdots,J_K,\mathbf{U}_1,\cdots,\mathbf{U}_K;\mathbf{Z})\\
&\quad - H(J_1,\cdots,J_K|\mathbf{Z},W_1,\cdots,W_K)\\
&= n(R_1+\cdots+R_K+R'_1+\cdots+R'_K) - I(\mathbf{U}_1,\cdots,\mathbf{U}_K;\mathbf{Z})\\
&\quad - H(J_1,\cdots,J_K|\mathbf{Z},W_1,\cdots,W_K),
\end{aligned} \tag{4.27}$$

where the last line above is due to firstly, to all of the $W_1,\cdots,W_K,J_1,\cdots,J_K$ being independent of each other; and secondly, $(W_1,\cdots,W_K,J_1,\cdots,J_K)$ is independent of $\mathbf{Z}$, given $\mathbf{U}_1,\cdots,\mathbf{U}_K$. The second term in the last line of (4.27) can be expressed as

$$I(\mathbf{U}_1,\cdots,\mathbf{U}_K;\mathbf{Z}) = I(\mathbf{U}_1;\mathbf{Z}|\mathbf{U}_2) + \sum_{k=2}^{K-1} I(\mathbf{U}_k;\mathbf{Z}|\mathbf{U}_{k+1}) + I(\mathbf{U}_K;\mathbf{Z}), \tag{4.28}$$

by using the Markov chain condition $\mathbf{U}_K \to \mathbf{U}_{K-1} \to \cdots \to \mathbf{U}_1 \to \mathbf{Z}$. Again, using the the method of [78], we have that

$$I(\mathbf{U}_1,\cdots,\mathbf{U}_K;\mathbf{Z}) \leq nI(U_1;Z|U_2) + \sum_{k=2}^{K-1} nI(U_k;Z|U_{k+1}) + nI(U_K;Z) + nK\delta. \tag{4.29}$$

To evaluate the last term in the last line of (4.27), we need Lemma 5, the K-receiver counterpart of Lemma 4, which is proved in Appendix B.2.

Lemma 5. *We have*

$$H(J_1,\cdots,J_K|\mathbf{Z},W_1,\cdots,W_K) \leq n\left(\sum_{k=1}^{K}[R'_k - I(U_k;Z|U_{k+1})] + \delta(\epsilon)\right) + 2, \tag{4.30}$$

where $\delta(\epsilon)\to 0$ for $\epsilon\to 0$ and n sufficiently large, under the conditions

$$\begin{aligned}
R'_K &\geq I(U_K;Z)\\
R'_{K-1} &\geq I(U_{K-1};Z|U_K)\\
&\vdots\\
R'_1 &\geq I(U_1;Z|U_2) = I(X;Z|U_2).
\end{aligned} \tag{4.31}$$

□

Substituting (4.29) and (4.30) into the last line of (4.27), we have

$$\frac{1}{n}H(W_1,\cdots,W_K|\mathbf{Z}) \geq \sum_{k=1}^{K} R_k - \left(K\delta + \delta(\epsilon) + \frac{2}{n}\right) = \sum_{k=1}^{K} R_k - \eta', \tag{4.32}$$

where $\eta' \triangleq K\delta + \delta(\epsilon) + \frac{2}{n}$, which is > 0 and is small as n is large. We have now shown that the equivocation sum rate of the K receivers meets the perfect secrecy requirement (4.6). Lastly, we note that since the W_k, $k \in \mathcal{K}$ are independent of each other, we have the relation that $H(W(\mathcal{K})|\mathbf{Z}) = \sum_{k\in\mathcal{K}} H(W_k|\mathbf{Z})$. Then we can repeat the proof of (4.21) for each of $H(W_k|\mathbf{Z})$, and by taking the sum, we can see that the sum rate of any subset of the receivers meets the perfect secrecy requirement (4.6)[2].

Thus, the direct part of Theorem 13 is proved.

4.2.5 Proof of Converse

Here, we show the converse proof to Theorem 13. Consider a $(2^{nR_1},\ldots,2^{nR_K},n)$ code with error probability $P_e^{(n)}$ with the code construction so that we have the condition $(W_1\ \cdots\ W_K) \rightarrow \mathbf{X} \rightarrow \mathbf{Y}_1\ \cdots\ \mathbf{Y}_K\mathbf{Z}$. Then, the probability distribution on $\mathcal{W}_1 \times \cdots \times \mathcal{W}_K \times \mathcal{X}^n \times \mathcal{Y}_1^n \times \cdots \times \mathcal{Y}_K^n \times \mathcal{Z}^n$ is given by

$$p(w_1)\ \cdots\ p(w_K)p(\mathbf{x}|w_1,\ldots,w_K)\prod_{i=1}^{n} p(y_{1i},\ldots,y_{Ki},z_i|x_i). \tag{4.33}$$

A state dependency graph for the K-receiver degraded BC and confidential messages is shown in Figure 4.4. The variables $Y_{1,i}, Y_{2,i}, \cdots, Y_{K,i}, Z_i$ follow the degradedness condition $Y_{1,i} \rightarrow Y_{2,i} \rightarrow \cdots \rightarrow Y_{K,i} \rightarrow Z_i$ for $i = 1, \cdots, n$.

In the following, we give the proof for the rate at the kth receiver, while the proof for the Receiver 1 requires a few additional steps. The proof for Receiver K will be shown later.

[2]This observation has also been made in Ekrem and Ulukus [33, Lemma 11, Appendix A], where it is shown that if the sum rate secrecy constraint is satisfied, then the secrecy constraint for any subset of the sum rate is also satisfied. We also note that the proof in [33] is somewhat different from ours, in that we can actually prove that the secrecy constraint for any subset of receivers is satisfied, without assuming that the sum secrecy rate is satisfied.

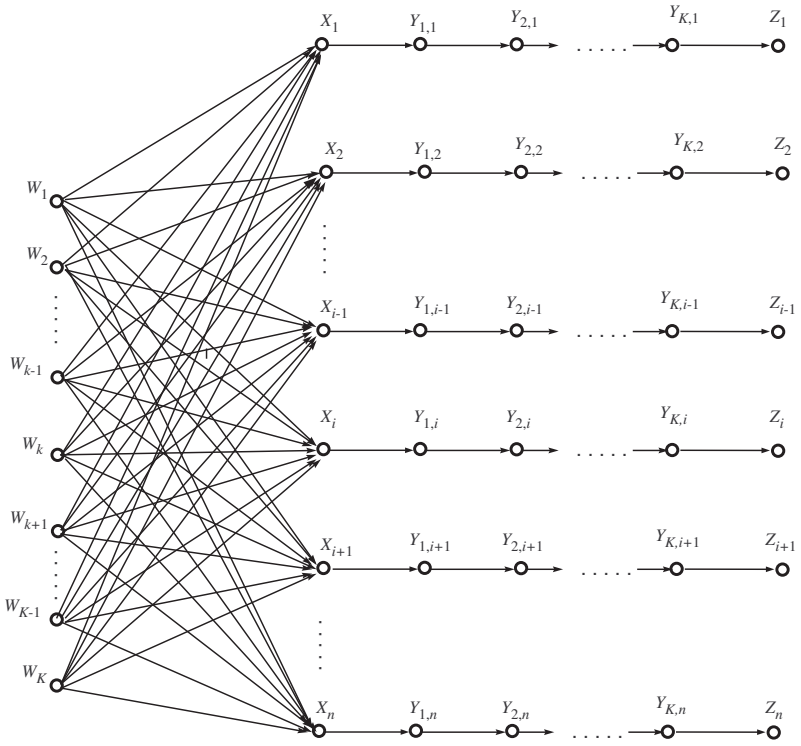


Figure 4.4: State dependency graph for the K-receiver degraded BC and confidential messages.

For $k = 2, \ldots, K-1$, the rate R_k satisfies

$$\begin{aligned}
nR_k &= H(W_k) \le H(W_k|\mathbf{Z}) + n\epsilon_k \qquad \text{by secrecy condition} \\
&\le H(W_k, W_{k+1}, \ldots, W_K|\mathbf{Z}) + n\epsilon_k \\
&= H(W_k|W_{k+1}, \ldots, W_K, \mathbf{Z}) + H(W_{k+1}, \ldots, W_K|\mathbf{Z}) + n\epsilon_k \\
&= H(W_k|W_{k+1}, \ldots, W_K, \mathbf{Z}) - H(W_k|W_{k+1}, \ldots, W_K, \mathbf{Y}_k) \\
&\quad + H(W_k|W_{k+1}, \ldots, W_K, \mathbf{Y}_k) + H(W_{k+1}, \ldots, W_K|\mathbf{Z}) + n\epsilon_k \\
&\overset{(a)}{\le} I(W_k; \mathbf{Y}_k|W_{k+1}, \ldots, W_K) - I(W_k; \mathbf{Z}|W_{k+1}, \ldots, W_K) \\
&\quad + n(\delta''_k + \delta'_{k+1} + \cdots + \delta'_K + \epsilon_k), \qquad (4.34)
\end{aligned}$$

where (a) is by $H(W_{k+1}, \ldots, W_K|\mathbf{Z}) \le H(W_{k+1}|\mathbf{Z}) + \cdots + H(W_K|\mathbf{Z})$, and by Fano's

inequality which gives

$$\begin{cases} H(W_k|\mathbf{Y}_k, W_{k+1}, \ldots, W_K) \leq nR_k P_e^{(n)} + 1 \triangleq n\delta_k'', \\ H(W_{k+1}|\mathbf{Z}) \leq nR_{k+1} P_e^{(n)} + 1 \triangleq n\delta_{k+1}', \\ \vdots \\ H(W_K|\mathbf{Z}) \leq nR_K P_e^{(n)} + 1 \triangleq n\delta_K', \end{cases} \tag{4.35}$$

where $\delta_k'',\ \delta_{k+1}',\ \ldots\ ,\ \delta_K' \to 0$ if $P_e^{(n)} \to 0$.

Expanding the first two terms of the last line of (4.34) by the chain rule gives

$$I(W_k; \mathbf{Y}_k|W_{k+1}, \ldots, W_K) = \sum_{i=1}^{n} I(W_k; Y_{k,i}|W_{k+1}, \ldots, W_K, \mathbf{Y}_k^{i-1}), \tag{4.36}$$

$$I(W_k; \mathbf{Z}|W_{k+1}, \ldots, W_K) = \sum_{i=1}^{n} I(W_k; Z_i|W_{k+1}, \ldots, W_K, \tilde{\mathbf{Z}}^{i+1}). \tag{4.37}$$

From (4.36), by using the identity $I(S_1S_2; T|V) = I(S_1; T|V) + I(S_2; T|S_1V)$, we get

$$\begin{aligned} &I(W_k; Y_{k,i}|W_{k+1}, \ldots, W_K, \mathbf{Y}_k^{i-1}) \\ &\quad = I(W_k, \tilde{\mathbf{Z}}^{i+1}; Y_{k,i}|W_{k+1}, \ldots, W_K, \mathbf{Y}_k^{i-1}) - I(\tilde{\mathbf{Z}}^{i+1}; Y_{k,i}|W_k, W_{k+1}, \ldots, W_K, \mathbf{Y}_1^{i-1}) \\ &\quad = I(W_k; Y_{k,i}|W_{k+1}, \ldots, W_K, \mathbf{Y}_k^{i-1}, \tilde{\mathbf{Z}}^{i+1}) + I(\tilde{\mathbf{Z}}^{i+1}; Y_{k,i}|W_{k+1}, \ldots, W_K, \mathbf{Y}_k^{i-1}) \\ &\qquad - I(\tilde{\mathbf{Z}}^{i+1}; Y_{k,i}|W_k, W_{k+1}, \ldots, W_K, \mathbf{Y}_k^{i-1}). \end{aligned} \tag{4.38}$$

Substituting this into (4.36) we have,

$$\begin{cases} I(W_k; \mathbf{Y}_k|W_{k+1}, \ldots, W_K) = \sum_{i=1}^{n} I(W_k; Y_{k,i}|W_{k+1}, \ldots, W_K, \mathbf{Y}_k^{i-1}, \tilde{\mathbf{Z}}^{i+1}) \\ \qquad\qquad + \Sigma_{k,1} - \Sigma_{k,2} \\ \Sigma_{k,1} = \sum_{i=1}^{n} I(\tilde{\mathbf{Z}}^{i+1}; Y_{k,i}|W_{k+1}, \ldots, W_K, \mathbf{Y}_1^{i-1}), \\ \Sigma_{k,2} = \sum_{i=1}^{n} I(\tilde{\mathbf{Z}}^{i+1}; Y_{k,i}|W_k, W_{k+1}, \ldots, W_K, \mathbf{Y}_k^{i-1}). \end{cases} \tag{4.39}$$

From (4.37), again by using $I(S_1S_2; T|V) = I(S_1; T|V) + I(S_2; T|S_1V)$, and substi-

tuting into (4.36), we get

$$\begin{cases} I(W_k;\mathbf{Z}|W_{k+1},\ldots,W_K) = \sum_{i=1}^{n} I(W_k;Z_i|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},\tilde{\mathbf{Z}}^{i+1}) \\ \qquad\qquad + \Sigma_{k,1}^* - \Sigma_{k,2}^* \\ \Sigma_{k,1}^* = \sum_{i=1}^{n} I(\mathbf{Y}_k^{i-1};Z_i|W_{k+1},\ldots,W_K,\tilde{\mathbf{Z}}^{i+1}), \\ \Sigma_{k,2}^* = \sum_{i=1}^{n} I(\mathbf{Y}_k^{i-1};Z_i|W_k,W_{k+1},\ldots,W_K,\tilde{\mathbf{Z}}^{i+1}). \end{cases} \tag{4.40}$$

It is known by [30, Lemma 7] that $\Sigma_{k,1} = \Sigma_{k,1}^*$ and $\Sigma_{k,2} = \Sigma_{k,2}^*$. Therefore,

$$\begin{aligned} nR_k \le \sum_{i=1}^{n} \Big[& I(W_k;Y_{k,i}|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},\tilde{\mathbf{Z}}^{i+1}) \\ & - I(W_k;Z_i|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},\tilde{\mathbf{Z}}^{i+1})\Big] + n(\delta_k'' + \delta_{k+1}' + \cdots + \delta_K' + \epsilon_k). \end{aligned} \tag{4.41}$$

The terms under the summation are

$$\begin{aligned} & I(W_k;Y_{k,i}|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},\tilde{\mathbf{Z}}^{i+1}) - I(W_k;Z_i|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},\tilde{\mathbf{Z}}^{i+1}) \\ & = H(W_k|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}) - H(W_k|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},Y_{k,i},\tilde{\mathbf{Z}}^{i+1}) \\ & \overset{(a)}{\le} H(W_k|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}) - H(W_k|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},Y_{k,i},Z_i,\tilde{\mathbf{Z}}^{i+1}) \\ & = I(W_k;Y_{k,i}|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}) \\ & = H(Y_{k,i}|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}) - H(Y_{k,i}|W_k,W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}) \\ & = H(Y_{k,i}|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},\mathbf{Y}_{k+1}^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}) \\ & \quad + I(Y_{k,i};\mathbf{Y}_{k+1}^{i-1}|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}) \\ & \quad - H(Y_{k,i}|W_k,W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}) \\ & \overset{(b)}{=} H(Y_{k,i}|W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},\mathbf{Y}_{k+1}^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}) \\ & \quad - H(Y_{k,i}|W_k,W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}) \\ & \overset{(c)}{\le} H(Y_{k,i}|W_{k+1},\ldots,W_K,\mathbf{Y}_{k+1}^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}) \\ & \quad - H(Y_{k,i}|W_k,W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},\mathbf{Y}_{k+1}^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}) \\ & = I(W_k,\mathbf{Y}_k^{i-1};Y_{k,i}|W_{k+1},\ldots,W_K,\mathbf{Y}_{k+1}^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1}), \end{aligned} \tag{4.42}$$

where (a) and (c) are due to conditioning reducing entropy, and (b) is due to the fact that, given $(W_{k+1},\ldots,W_K,\mathbf{Y}_k^{i-1},Z_i,\tilde{\mathbf{Z}}^{i+1})$, $Y_{k,i}$ is independent of $\mathbf{Y}_{k+1}^{i-1}$, which can be

checked by referring to the state dependency graph Figure 4.4, from which we can see that $Y_{1,i}$ only depends on X_i, $Y_{2,i}$ only depends on $Y_{1,i}$, and so on, hence the assertion.

Now, define the random variables

$$\begin{cases} U_{K,i} \triangleq W_K \mathbf{Y}_K^{i-1} \tilde{\mathbf{Z}}^{i+1}, \\ U_{k,i} \triangleq W_k \cdots W_K \mathbf{Y}_k^{i-1} \tilde{\mathbf{Z}}^{i+1} \qquad \text{for } k = 2, \ldots, K-1, \end{cases} \tag{4.43}$$

satisfying the Markov chain

$$U_{K,i} \rightarrow \ \cdots \ \rightarrow U_{2,i} \rightarrow X_i \rightarrow Y_{k,i} \ \cdots \ Y_{K,i} \rightarrow Z_i. \tag{4.44}$$

We note that our choice of auxiliary random variables is different from Bagherikaram *et al.*, which deals with the 2-receiver degraded BC with an external eavesdropper [6], and from [33], which studies the K-receiver degraded BC with a common message and an external eavesdropper. The choice is also different, due to the presence of the eavesdropper, from that of Borade *et al.* in [13] which deals with the K-receiver degraded BC without secrecy conditions. Continuing from (4.42), we have

$$\begin{aligned} & I(W_k, \mathbf{Y}_k^{i-1}; Y_{k,i} | W_{k+1}, \ldots, W_K, \mathbf{Y}_{k+1}^{i-1}, Z_i, \tilde{\mathbf{Z}}^{i+1}) \\ &= I(U_{k,i}; Y_{k,i} | U_{(k+1),i}, Z_i) \\ &= I(U_{k,i}; Y_{k,i}, Z_i | U_{(k+1),i}) - I(U_{k,i}; Z_i | U_{(k+1),i}) \\ &= I(U_{k,i}; Y_{k,i} | U_{(k+1),i}) + I(U_{k,i}; Z_i | U_{(k+1),i}, Y_{k,i}) - I(U_{k,i}; Z_i | U_{(k+1),i}) \\ &\overset{(a)}{=} I(U_{k,i}; Y_{k,i} | U_{(k+1),i}) - I(U_{k,i}; Z_i | U_{(k+1),i}), \end{aligned} \tag{4.45}$$

where (a) is due to $I(U_{k,i}; Z_i | U_{(k+1),i}, Y_{k,i}) = I(U_{k,i}; Z_i | Y_{k,i}) = 0$ since we have $U_{(k+1),i} \rightarrow U_{k,i} \rightarrow Y_{k,i} \rightarrow Z_i$, for $k = 2, \cdots, K-1$. As a result, we have

$$nR_k \leq \sum_{i=1}^{n} \Big[I(U_{k,i}; Y_{k,i} | U_{(k+1),i}) - I(U_{k,i}; Z_i | U_{(k+1),i}) \Big] + n(\delta_k'' + \delta_{k+1}' + \cdots + \delta_K' + \epsilon_k), \tag{4.46}$$

for $k = 2, \cdots, K-1$.

To show the converse for R_1, we follow the same steps as above, but additionally we use (4.42) with $k = 1$ to arrive at the equivalent chain of equalities (4.45) above for

$k = 1$. This results in

$$\begin{aligned}
&I(W_1, \mathbf{Y}_1^{i-1}; Y_{1,i}|W_2, \ldots, W_K, \mathbf{Y}_2^{i-1}, Z_i, \tilde{\mathbf{Z}}^{i+1}) \\
&= I(W_1; Y_{1,i}|W_2, \ldots, W_K, \mathbf{Y}_2^{i-1}, Z_i, \tilde{\mathbf{Z}}^{i+1}) \\
&\quad + I(\mathbf{Y}_1^{i-1}; Y_{1,i}|W_1, W_2, \ldots, W_K, \mathbf{Y}_2^{i-1}, Z_i, \tilde{\mathbf{Z}}^{i+1}) \\
&\overset{(a)}{=} I(W_1; Y_{1,i}|W_2, \ldots, W_K, \mathbf{Y}_2^{i-1}, Z_i, \tilde{\mathbf{Z}}^{i+1}) = I(W_1; Y_{1,i}|U_{2,i}, Z_i) \\
&\overset{(b)}{\leq} I(X_i; Y_{1,i}|U_{2,i}, Z_i) \\
&= I(X_i; Y_{1,i}, Z_i|U_{2,i}) - I(X_i; Z_i|U_{2,i}) \\
&= I(X_i; Y_{1,i}|U_{2,i}) + I(X_i; Z_i|U_{2,i}, Y_{1,i}) - I(X_i; Z_i|U_{2,i}) \\
&\overset{(c)}{=} I(X_i; Y_{1,i}|U_{2,i}) - I(X_i; Z_i|U_{2,i})
\end{aligned} \tag{4.47}$$

where (a) is by $I(\mathbf{Y}_1^{i-1}; Y_{1,i}|W_1, W_2, \ldots, W_K, \mathbf{Y}_2^{i-1}, Z_i, \tilde{\mathbf{Z}}^{i+1}) = 0$ since $\mathbf{Y}_1^{i-1}$ is independent of $Y_{1,i}$, given $(W_1, W_2, \ldots, W_K, \mathbf{Y}_2^{i-1}, Z_i, \tilde{\mathbf{Z}}^{i+1})$, which is checked using Figure 4.4, from which we see that $Y_{1,i}$ depends only on X_i; (b) is by the Markov chain $Y_{1,i} \to X_i \to W_1$; and (c) is by the second term $I(X_i; Z_i|U_{2,i}, Y_{1,i}) = I(X_i; Z_i|Y_{1,i}) = 0$ since $U_{2,i} \to X_i \to Y_{1,i} \to Z_i$. Thus, we have

$$nR_1 \leq \sum_{i=1}^{n} [I(X_i; Y_{1,i}|U_{2,i}) - I(X_i; Z_i|U_{2,i})] + n(\delta_1'' + \delta_2' + \cdots + \delta_K' + \epsilon_1). \tag{4.48}$$

We now obtain the bound for Receiver K:

$$\begin{aligned}
&nR_K = H(W_K) \leq H(W_K|\mathbf{Z}) + n\epsilon_K \\
&= H(W_K|\mathbf{Z}) - H(W_K|\mathbf{Y}_K) + H(W_K|\mathbf{Y}_K) + n\epsilon_K \\
&\leq I(W_K; \mathbf{Y}_K) - I(W_K; \mathbf{Z}) + n(\delta_K' + \epsilon_K) \\
&= \sum_{i=1}^{n} \Big[I(W_K; Y_{K,i}|\mathbf{Y}_K^{i-1}) - I(W_K; Z_i|\tilde{\mathbf{Z}}^{i+1}) \Big] + n(\delta_K' + \epsilon_K) \\
&= \sum_{i=1}^{n} \Big[I(W_K, \tilde{\mathbf{Z}}^{i+1}; Y_{K,i}|\mathbf{Y}_K^{i-1}) - I(\tilde{\mathbf{Z}}^{i+1}; Y_{K,i}|W_K, \mathbf{Y}_K^{i-1}) - I(W_K, \mathbf{Y}_K^{i-1}; Z_i|\tilde{\mathbf{Z}}^{i+1}) \\
&\quad + I(\mathbf{Y}_K^{i-1}; Z_i|W_K, \tilde{\mathbf{Z}}^{i+1}) \Big] + n(\delta_K' + \epsilon_K) \\
&\overset{(a)}{=} \sum_{i=1}^{n} \Big[I(W_K, \tilde{\mathbf{Z}}^{i+1}; Y_{K,i}|\mathbf{Y}_K^{i-1}) - I(W_K, \mathbf{Y}_K^{i-1}; Z_i|\tilde{\mathbf{Z}}^{i+1}) \Big] + n(\delta_K' + \epsilon_K) \\
&= \sum_{i=1}^{n} \Big[I(W_K; Y_{K,i}|\mathbf{Y}_K^{i-1}, \tilde{\mathbf{Z}}^{i+1}) + I(\tilde{\mathbf{Z}}^{i+1}; Y_{K,i}|\mathbf{Y}_K^{i-1}) - I(W_K; Z_i|\mathbf{Y}_K^{i-1}, \tilde{\mathbf{Z}}^{i+1}) \\
&\quad - I(\mathbf{Y}_K^{i-1}; Z_i|\tilde{\mathbf{Z}}^{i+1}) \Big] + n(\delta_K' + \epsilon_K) \\
&\overset{(b)}{=} \sum_{i=1}^{n} \Big[I(W_K; Y_{K,i}|\mathbf{Y}_K^{i-1}, \tilde{\mathbf{Z}}^{i+1}) - I(W_K; Z_i|\mathbf{Y}_K^{i-1}, \tilde{\mathbf{Z}}^{i+1}) \Big] + n(\delta_K' + \epsilon_K)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \Big[I(W_K, \mathbf{Y}_K^{i-1}, \tilde{\mathbf{Z}}^{i+1}; Y_{K,i}) - I(\mathbf{Y}_K^{i-1}, \tilde{\mathbf{Z}}^{i+1}; Y_{K,i}) - I(W_K, \mathbf{Y}_K^{i-1}, \tilde{\mathbf{Z}}^{i+1}; Z_i) \\
&\quad + I(\mathbf{Y}_K^{i-1}, \tilde{\mathbf{Z}}^{i+1}; Z_i) \Big] + n(\delta'_K + \epsilon_K) \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} \Big[I(W_K, \mathbf{Y}_K^{i-1}, \tilde{\mathbf{Z}}^{i+1}; Y_{K,i}) - I(W_K, \mathbf{Y}_K^{i-1}, \tilde{\mathbf{Z}}^{i+1}; Z_i) \Big] + n(\delta'_K + \epsilon_K) \\
&= \sum_{i=1}^{n} \left[I(U_{K,i}; Y_{K,i}) - I(U_{K,i}; Z_i) \right] + n(\delta'_K + \epsilon_K), \qquad (4.49)
\end{aligned}
$$

where (a) and (b) are both due to [30, Lemma 7]; (c) is by the fact that since Z is a degraded version of Y_K, this implies that[3] Y_K is less noisy than Z, which in turn means that $I(\mathbf{Y}_K^{i-1}, \tilde{\mathbf{Z}}^{i+1}; Y_{K,i}) \geq I(\mathbf{Y}_K^{i-1}, \tilde{\mathbf{Z}}^{i+1}; Z_i)$, thus giving the desired inequality.

Now, we introduce the random variable G, which is uniformly distributed among the integers $\{1, 2, \ldots, n\}$ and is independent of all other random variables. Define the following auxiliary random variables

$$U_k = (G, U_{k,G}), \quad k = 2, \cdots, K, \qquad (4.50)$$

$$X = X_G, \qquad (4.51)$$

$$Y_k = Y_{k,G}, \quad k = 1, \cdots, K, \qquad (4.52)$$

$$Z = Z_G. \qquad (4.53)$$

Then (4.46), (4.48), and the last line of (4.49) become

$$R_K \leq I(U_K; Y_K) - I(U_K; Z), \qquad (4.54)$$

$$R_k \leq I(U_k; Y_k | U_{k+1}) - I(U_k; Z | U_{k+1}), \text{ for } k = 2, \ldots, K-1, \qquad (4.55)$$

$$R_1 \leq I(X; Y_1 | U_2) - I(X; Z | U_2), \qquad (4.56)$$

and the converse to Theorem 13 is proved.

4.2.6 Conclusion

We have presented the derivation for the secrecy capacity region for the degraded K-receiver BC with private messages in the presence of an eavesdropper which generalizes previous work [6] which dealt with 2-receiver BCs. In the direct proof we have used superposition coding and code partitioning to show the achievable rate tuples. We have provided error probability analysis and equivocation calculation for the general kth receiver, and for all receivers, from which we can also deduce the result for any

[3] See Appendix A on the ordering of channels.

subset of the receivers. In the converse proof we have used a new definition for the auxiliary random variables which is different from the following cases: the 2-receiver degraded BC with an eavesdropper [6]; the K-receiver degraded BC with common message and an eavesdropper [33]; and the K-receiver BC without security constraints [13].

The results imply a multilevel code construction. Each level will be used to send a message to each of the the respective K receivers. Each level will also be split into 2 sub-levels, where the first sub-level is required for reliability, while the second sub-level is required to generate randomness to confuse the eavesdropper. The multilevel code will have $2K$ levels overall.

4.3 The 3-Receiver Broadcast Channel with DMS and Confidential Messages

In this section, we investigate the 3-receiver BC with 2 DMS with a confidential message sent to one of the receivers in the presence of an eavesdropper, from which the message is to be kept secret. Instead of the perfect secrecy criteria, we derive an achievable equivocation rate, so that the equivocation rate $R_e \leq R$, the transmission rate, for a given message W.

For the 3-receiver BC with DMS without secrecy constraints, the inner capacity bound in [90, 91] is achievable by superposition coding, Marton's achievability technique [84] and indirect decoding, where the receivers decoding the common message only do so via satellite codewords instead of cloud centers[4]. In our secure cod-

[4]The mechanics of indirect decoding is as follows [91]. Suppose that we have 2^{nR_0} i.i.d. sequences (cloud centers) $\mathbf{u}(w_0), w_0 \in \{1, \cdots, 2^{nR_0}\}$. For each $\mathbf{u}(w_0)$, generate 2^{nS_1} sequences (satellite codewords) $\mathbf{v}(w_0, s_1), s_1 \in \{1, \cdots, 2^{nS_1}\}$ in an i.i.d. manner according to $\prod_{i=1}^{n} p(v_i|u_i)$ (superposition coding). Suppose we want to decode which w_0 was sent. Then if receiver Y carries out indirect coding, w_0 is declared to be the unique index such that $\mathbf{v}(w_0, s_1)$ and $\mathbf{y}$ are joint typical *for some* s_1. Receiver Y does so with small error probability provided that $R_0 + S_1 \leq I(V;Y)$. On the other hand, ordinary joint typical decoding requires that $\mathbf{v}$ is decoded first, then $\mathbf{u}$; that is, we obtain s_1 first, then w_0. Ordinary joint decoding requires

$$R_0 + S_1 \leq I(V;Y), \;\; S_1 \leq I(V;Y|U)$$

for small error probability. This indirect decoding is useful in extending coding schemes for the general 2-receiver BCs to 3-receiver BCs.

ing scheme, we shall use a combination of the code partitioning of Wyner [115] and double-binning of Liu *et al.* [78] to show the achievability of an inner bound to the rate-equivocation region for the 3-receiver BC with 2 DMS. Error probability analysis and equivocation calculation for the private messages are provided. Outer bounds to the rate equivocation regions are derived for the following two cases:

1. Receiver 1 is less noisy than receiver 2 and less noisy than the eavesdropper receiver 3;

2. Receivers 1 and 2 are less noisy than the eavesdropper receiver 3.

We see that these conditions are more general than degradedness considered in the previous Section 4.2 [66].

This section is organized as follows. In Section 4.3.1, we describe the model for the 3-receiver BC with DMS. In Section 4.3.2, we establish an achievable inner bound to the rate-equivocation region using our secure coding scheme for the 3-receiver BC with 2 DMS and show error probability analysis and equivocation calculation for the private message. In Section 4.3.3, we derive an outer bound for a subclass of the 3-receiver BC with 2 DMS. Lastly, we give conclusions in Section 4.3.4.

4.3.1 The 3-Receiver BC with DMS

The discrete memoryless BC with 3 receivers has an input random sequence, $\mathbf{X}$, and 3 output random sequences at the receivers, denoted respectively by $\mathbf{Y}_1, \mathbf{Y}_2$ and $\mathbf{Y}_3$, all of length n, with $\mathbf{x} \in \mathcal{X}^n$, $\mathbf{y}_1 \in \mathcal{Y}_1^n$, $\mathbf{y}_2 \in \mathcal{Y}_2^n$, and $\mathbf{y}_3 \in \mathcal{Y}_3^n$. The conditional distribution for n uses of the channel is given by

$$p(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3|\mathbf{x}) = \prod_{i=1}^{n} p(y_{1i}, y_{2i}, y_{3i}|x_i). \tag{4.57}$$

The 3-receiver BC with 2 DMS, is shown in Figure 4.5.

A $(2^{nR_0}, 2^{nR_1}, n)$-code for the 3-receiver BC with 2 DMS consists of the following parameters:

$$\begin{aligned}
\mathcal{W}_0 &= \left\{1, \ldots, 2^{nR_0}\right\}, \text{(common message set)} \\
\mathcal{W}_1 &= \left\{1, \ldots, 2^{nR_1}\right\}, \text{(private message set)}, \\
f &: \mathcal{W}_0 \times \mathcal{W}_1 \mapsto \mathcal{X}^n, \text{((stochastic) encoding function)},
\end{aligned}$$

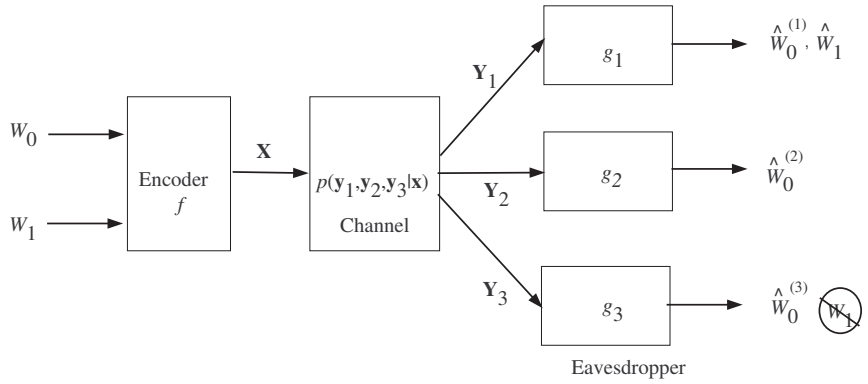


Figure 4.5: The 3-receiver BC with 2 DMS and confidential message.

$$g_1 : \mathcal{Y}_1^n \mapsto \mathcal{W}_0 \times \mathcal{W}_1, \text{(decoding function 1)},$$
$$g_2 : \mathcal{Y}_2^n \mapsto \mathcal{W}_0, \text{(decoding function 2)},$$
$$g_3 : \mathcal{Y}_3^n \mapsto \mathcal{W}_0, \text{(decoding function 3)}.$$

We have $g_1(\mathbf{Y}_1) = (\hat{W}_0^{(1)}, \hat{W}_1^{(1)})$, $g_2(\mathbf{Y}_2) = \hat{W}_0^{(2)}$, and $g_3(\mathbf{Y}_3) = \hat{W}_0^{(3)}$, with the error probability

$$P_e^{(n)} = \Pr\left[(\hat{W}_0^{(1)}, \hat{W}_0^{(2)}, \hat{W}_0^{(3)}, \hat{W}_1^{(1)}) \neq (W_0, W_0, W_0, W_1,)\right]. \tag{4.58}$$

The decoders are set up to decode combinations of the messages; in the coding scheme this means that the decoders will decode specific parts of the transmitted codeword. The private message W_1 is sent to Y_1, with Y_3 the eavesdropper, and the secrecy level of the message sent is $\frac{1}{n}H(W_1|\mathbf{Y}_3)$.

For the 3-receiver BC with 2 DMS, the rate tuple (R_0, R_1, R_{1e}) is said to be achievable if for any $\eta, \epsilon_1 > 0$, there exists a sequence of $(2^{nR_0}, 2^{nR_1}, n)$-codes for which $P_e^{(n)} \leq \eta$ and the equivocation rate R_{1e} satisfies

$$\frac{1}{n}H(W_1|\mathbf{Y}_3) \geq R_{1e} - \epsilon_1. \tag{4.59}$$

Our analysis does not include the case of perfect secrecy (that is $\frac{1}{n}H(W_1|\mathbf{Y}_3) \geq R_1 - \epsilon_1$, the rate region with $R_{1e} = R_1$).

Finally, we remark that the model studied by Chia and El Gamal [17, 18] adds another $\hat{W}_1^{(2)}$ to the output at Y_2, with the appropriate changes in the error probability and decoding function.

4.3.2 Inner Bound to the Rate-equivocation Region for the 3-receiver BC with 2 DMS

In this section we establish an achievable inner bound to the rate-equivocation region for the 3-receiver BC with 2 DMS and a confidential message sent to Y_1.

Theorem 14. *An inner bound to the rate-equivocation region for the 3-receiver BC with 2 DMS and one confidential message is the closure of all rate-tuples* (R_0, R_1, R_{1e}) *satisfying*

$$
\begin{aligned}
&R_0 \leq \min\{I(U_2;Y_2) - I(U_2;Y_3|U_1), I(U_3;Y_3) - I(U_3;Y_3|U_1)\}\\
&2R_0 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3)\\
&R_0 + R_1 \leq \min\{I(X;Y_1), I(U_2;Y_2) + I(X;Y_1|U_2), I(U_3;Y_3) + I(X;Y_1|U_3),\\
&\qquad I(U_2;Y_2) + I(X;Y_1|U_1) - I(U_2;Y_3|U_1),\\
&\qquad I(U_3;Y_3) + I(X;Y_1|U_1) - I(U_3;Y_3|U_1)\}\\
&R_{1e} \leq \min\{R_1, I(X;Y_1|U_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1),\\
&\qquad I(X;Y_1|U_2) + I(X;Y_1|U_3) - 2I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1)\}\\
&R_0 + R_{1e} \leq \min\{I(X;Y_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1),\\
&\qquad I(U_2;Y_2) + I(X;Y_1|U_2) - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1),\\
&\qquad I(U_3;Y_3) + I(X;Y_1|U_3) - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1)\}\\
&2R_0 + R_{1e} \leq I(U_2;Y_2) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3) - I(X;Y_3|U_2,U_3)\\
&\qquad - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1)\\
&R_0 + R_1 + R_{1e} \leq \min\{I(X;Y_1|U_2) - I(U_2;U_3|U_1) + I(X;Y_1|U_3) + I(X;Y_1)\\
&\qquad\qquad - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1),\\
&\qquad\qquad I(U_2;Y_2) + I(X;Y_1|U_1) + I(X;Y_1|U_2) - I(X;Y_3|U_2,U_3)\\
&\qquad\qquad - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1),\\
&\qquad\qquad I(U_2;Y_2) + 2I(X;Y_1|U_2) + I(X;Y_1|U_3) - I(X;Y_3|U_2,U_3)\\
&\qquad\qquad - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1),\\
&\qquad\qquad I(U_3;Y_3) + I(X;Y_1|U_1) + I(X;Y_1|U_3) - I(X;Y_3|U_2,U_3)\\
&\qquad\qquad - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1),\\
&\qquad\qquad I(U_3;Y_3) + I(X;Y_1|U_2) + 2I(X;Y_1|U_3) - I(X;Y_3|U_2,U_3)
\end{aligned}
$$

$$-I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1)\}$$

$$2R_0 + R_1 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3)$$

$$2R_0 + 2R_1 \leq \min\{I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2) + I(X;Y_1|U_3),$$
$$I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_1) + I(X;Y_1|U_2,U_3)\}$$

$$2R_0 + 2R_1 + R_{1e} \leq I(U_2;Y_2) + I(U_3;Y_3) + 2I(X;Y_1|U_2) + 2I(X;Y_1|U_3)$$
$$+ I(X;Y_1|U_2,U_3) - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1),$$

$$2R_0 + 2R_1 + R_{1e} \leq I(U_2;Y_2) + I(U_3;Y_3) + 2I(X;Y_1|U_1) + I(X;Y_1|U_2,U_3)$$
$$- I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1), \tag{4.60}$$

defined over the p.d.f.

$$p(u_1,u_2,u_3,x) = p(u_1)p(u_2,u_3|u_1)p(x|u_2,u_3), \tag{4.61}$$

which is induced by the coding scheme. From the p.d.f. (4.61)*, the auxiliary random variables* U_1*,* U_2 *and* U_3 *satisfy the Markov chain condition*

$$U_1 \to (U_2,U_3) \to X \to (Y_1,Y_2,Y_3). \tag{4.62}$$

□

We now state sufficient conditions for the equivocation rate R_{1e} to be positive: if there exists a distribution $\in p(u_1)p(u_2,u_3|u_1)p(x|u_2,u_3)$ for which

$$I(U_2;Y_1|U_1) > I(U_2;Y_3|U_1),$$
$$I(U_2;Y_1|U_1) > I(U_3;Y_3|U_1),$$
$$I(X;Y_1|U_2,U_3,U_1) > I(X;Y_3|U_2,U_3,U_1),$$
$$I(X;Y_1|U_3) > I(X;Y_3|U_2,U_3,U_1) + I(U_2;Y_3|U_1),$$
$$I(X;Y_1|U_2) > I(X;Y_3|U_2,U_3,U_1) + I(U_3;Y_3|U_1),$$
$$I(X;Y_1|U_1) > I(U_2;U_3|U_1) + I(X;Y_3|U_2,U_3,U_1) + I(U_2,U_3;Y_3|U_1), \tag{4.63}$$

then R_{1e} is positive. (These conditions can be easily derived from the equivocation conditions (4.109) given later.)

From the region specified in Theorem 14, we can make some observations. If we discard the equivocation constraints in the region of Theorem 14, we can obtain the rate

region for the 3 receiver BC with 2 DMS [91, Proposition 5]:

$$\begin{aligned}
R_0 &\leq \min\{I(U_2;Y_2), I(U_3;Y_3)\} \\
2R_0 &\leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) \\
R_0 + R_1 &\leq \min\{I(X;Y_1), I(U_2;Y_2) + I(X;Y_1|U_2), \\
&\qquad I(U_3;Y_3) + I(X;Y_1|U_3)\} \\
2R_0 + R_1 &\leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3) \\
2R_0 + 2R_1 &\leq \min\{I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2) + \\
&\qquad + I(X;Y_1|U_3), I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_1) \\
&\qquad + I(X;Y_1|U_2,U_3)\}
\end{aligned} \tag{4.64}$$

over the p.d.f. $p(u_1,u_2,u_3,x) = p(u_1)p(u_2,u_3|u_1)p(x|u_2,u_3)$. This requires some manipulation and is shown in Appendix C.1.

If we discard the equivocation constraints and impose the condition that Y_1 is less noisy than Y_2 in the region of Theorem 14, we can obtain the rate region for the 3 receiver, 2 DMS for Y_1 less noisy than Y_3 in [91, Proposition 7]:

$$\begin{aligned}
R_0 &\leq \min\{I(U;Y_2), I(V;Y_3)\} \\
R_1 &\leq I(X;Y_1|U) \\
R_0 + R_1 &\leq I(V;Y_3) + I(X;Y_1|V),
\end{aligned} \tag{4.65}$$

over the p.d.f. $p(u,v,x) = p(u)p(v|u)p(x|v)$. This can be shown in Appendix C.2 by setting $U_2 = U_1 = U$ and $U_3 = V$ in Theorem 14.

If we now specialize Theorem 14 to two receivers while keeping the equivocation constraints, by setting $Y_1 = Y_2 = Y$, $Y_3 = Z$, $U_2 = U_1 = U$ and $U_3 = X$ in Theorem 14 and prefixing a DMC with transition probability[5] $p(x|v)$ to the channels $p(y|x)$ and $p(z|x)$ (prefix V to $X \to (Y,Z)$), we can obtain the following rate region

$$\begin{aligned}
R_0 &\leq \min\{I(U;Y), I(U;Z)\} \\
R_0 + R_1 &\leq I(U;Y) + I(V;Y|U) \\
R_0 + R_1 &\leq I(U;Z) + I(V;Y|U) \\
R_e &\leq I(V;Y|U) - I(V;Z|U)
\end{aligned} \tag{4.66}$$

[5]Mentioned earlier in the proof for the DM wiretap channel in Theorem 6.

for the p.d.f. $p(u,v,x) = p(u)p(v|u)p(x|v)p(y,z|x)$, provided that $I(V;Z|U) \leq I(V;Y|U)$. So, the region obtained is the same as the Csiszár and Körner rate-equivocation region for the 2-receiver BC with one common and one confidential message, given previously in Theorem 10. We remark that the re-assignment of r.v.s and the reduction, including channel prefixing, is in several steps, and is given in Appendix C.3.

We now present the proof for Theorem 14.

Proof. (Theorem 14): We use rate splitting, Wyner's code partitioning [115] with the double-binning scheme of [78, 116] to provide secrecy, together with the coding scheme for the 3-receiver BC with DMS in [91].

The scheme of [91] represents W_0 by U_1, then breaks W_1 into 3 parts. The first part of W_1 is combined with U_1 by superposition coding to generate U_2. The second part of W_1 is combined with U_1 by superposition coding to generate U_3. U_2 and U_3 are partitioned into bins and the product bin containing the joint typical pair (achievable by Marton's coding scheme) is combined with the third part of W_1 by superposition coding to obtain X.

At the receivers, Y_1 decodes U_1, U_2, U_3, and X to recover the messages W_0 and W_1 , while Y_2 decodes U_1 indirectly using U_2 to recover message W_0, and Y_3 decodes U_1 indirectly using U_3 to recover W_0. In our secure scheme, the codewords $\mathbf{U}_2$, $\mathbf{U}_3$ are protected by double-binning and codewords $\mathbf{X}$ are protected from receiver Y_3 by code partitioning. This is depicted in Figure 4.6.

We will define the associated variables in Figure 4.6 when we discuss the code generation later. Rate splitting involves splitting the rates of the message to give $R_1 = R_{10} + R_{11}$. We first show an achievable inner bound without rate splitting at rates R_1, then perform rate splitting on the message rate to obtain the final achievable region. Suppose that we have the p.d.f. in (4.61) which induces the Markov chain condition $U_1 \rightarrow (U_2, U_3) \rightarrow X$. From Marton [84], we note that this condition also implies that $U_1 \rightarrow U_2 \rightarrow X$ and $U_1 \rightarrow U_3 \rightarrow X$. The following describes the encoding and decoding processes.

Codebook generation: Recall that we use $\mathcal{T}_\epsilon^n(P_Z)$ to denote the set of jointly typical n-sequences with respect to the p.d.f. $p(z)$. Let $R_1 = L_1 + L_2 + L_3$. The part of W_1

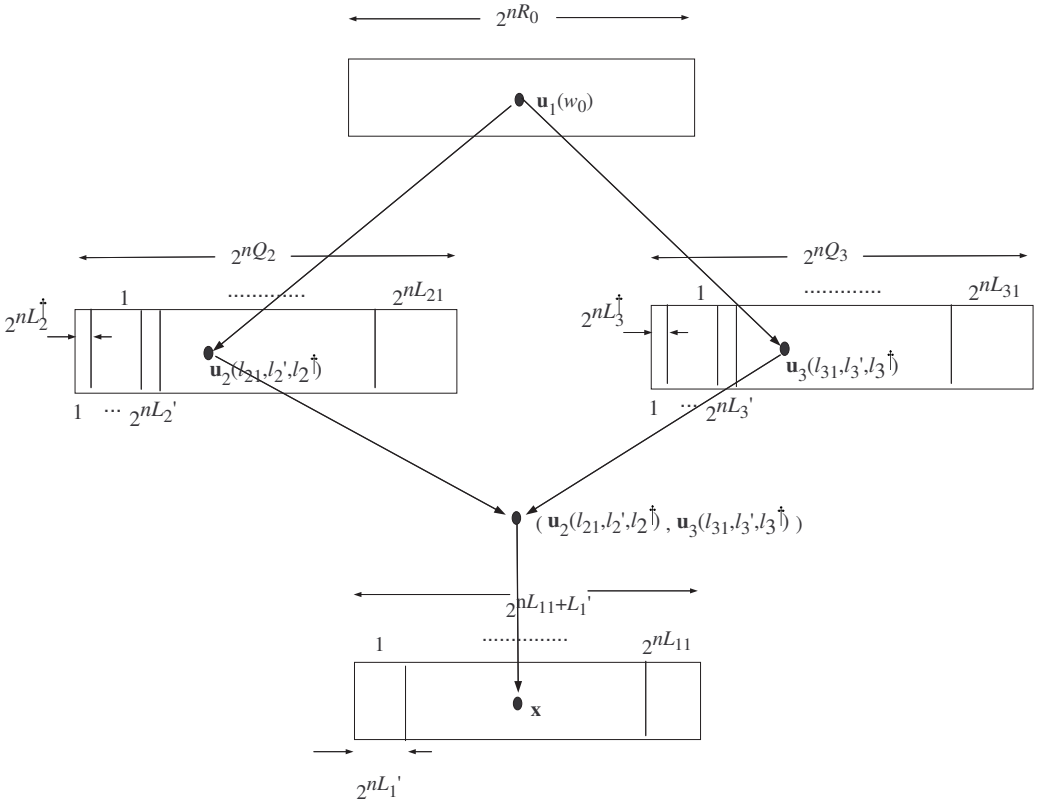


Figure 4.6: Coding for 3-receiver BC with DMS and confidential messages.

indexed by $l_2 \in \{1, ..., 2^{L_2}\}$ is to be combined with U_1 to form U_2, the part indexed by $l_3 \in \{1, ..., 2^{L_3}\}$ is to be combined with U_1 to form U_3, and the part indexed by $l_1 \in \{1, ..., 2^{L_1}\}$ is to be combined with U_2, U_3 to form X.

First of all, generate 2^{nR_0} sequences $\mathbf{U}_1(w_0)$, for $w_0 \in \mathcal{W}_0$, randomly and uniformly from the set of typical $\mathbf{U}_1$ sequences, according to $\prod_{j=1}^n p(u_{1,j})$. For each $\mathbf{U}_1(w_0)$, generate $2^{nQ_2} = 2^{n(L_{21}+L_2'+L_2^\dagger)}$ sequences $\mathbf{U}_2(w_0, q_2)$ randomly and uniformly from the set of conditionally typical $\mathbf{U}_2$ sequences, according to $\prod_{j=1}^n p(u_{2,j}|u_{1,j})$. The $\mathbf{U}_2$ sequences are indexed as $\mathbf{U}_2(l_{21}, l_2', l_2^\dagger)$, where $l_{21} \in \{1, ..., 2^{nL_{21}}\}$, $l_2' \in \{1, ..., 2^{nL_2'}\}$ and $l_2^\dagger \in \{1, ..., 2^{nL_2^\dagger}\}$. Randomly place the $\mathbf{U}_2$ sequences into $2^{nL_{21}}$ bins indexed by l_{21}. The sequences in each bin are placed randomly into $2^{nL_2'}$ sub-bins indexed by l_2'. The $2^{nL_2^\dagger}$ codewords in each sub-bin are indexed by $l_2^\dagger$.

Similarly, for each $\mathbf{U}_1(w_0)$, generate $2^{nQ_3} = 2^{n(L_{31}+L_3'+L_3^\dagger)}$ sequences $\mathbf{U}_3(w_0, q_3)$ randomly and uniformly from the set of conditionally typical $\mathbf{U}_3$ sequences, according to $\prod_{j=1}^n p(u_{3,j}|u_{1,j})$. The $\mathbf{U}_3$ sequences are indexed as $\mathbf{U}_3(l_{31}, l_3', l_3^\dagger)$, where $l_{31} \in \{1, ..., 2^{nL_{31}}\}$, $l_3' \in \{1, ..., 2^{nL_3'}\}$ and $l_3^\dagger \in \{1, ..., 2^{nL_3^\dagger}\}$. Randomly place the

$\mathbf{U}_3$ sequences into $2^{nL_{31}}$ bins indexed by l_{31}. The sequences in each bin are placed randomly into $2^{nL'_3}$ sub-bins indexed by l'_3. The $2^{nL^\dagger_3}$ codewords in each sub-bin are indexed by $l^\dagger_3$.

We now have to map the message indices to bin indices l_{21} and l_{31}. The mapping involves mapping the 2^{nL_2} message part to the $2^{nL_{21}}$ bins for $\mathbf{U}_2$ and mapping the 2^{nL_3} message part to the $2^{nL_{31}}$ bins for $\mathbf{U}_3$. According to [116], if

$$\begin{aligned} L_{21} + L'_2 < L_2 \leq L_{21} + L'_2 + L^\dagger_2, \\ L_{31} + L'_3 < L_3 \leq L_{31} + L'_3 + L^\dagger_3, \end{aligned} \tag{4.67}$$

each sub-bin of $\mathbf{U}_2$ is mapped to at least one of l_2 and each sub-bin of $\mathbf{U}_3$ is mapped to at least one of l_3. We will use these conditions for the subsequent coding process.

After the mapping of the message parts to the bins, the encoder then chooses the joint typical pair $(\mathbf{U}_2(l_{21}, l'_2, l^\dagger_2), \mathbf{U}_3(l_{31}, l'_3, l^\dagger_3))$ satisfying

$$\Big(\mathbf{u}_2(l_{21}, l'_2, l^\dagger_2), \mathbf{u}_3(l_{31}, l'_3, l^\dagger_3), \mathbf{u}_1(w_0)\Big) \in \mathcal{T}^n_\epsilon(P_{U_2U_3U_1}). \tag{4.68}$$

If there is more than one such pair, randomly choose one; if there is no such pair, declare an error.

Given a joint typical pair $(\mathbf{U}_2, \mathbf{U}_3)$, generate $2^{n(L_{11}+L'_1)}$ sequences $\mathbf{X}$ randomly and uniformly from the set of conditionally typical $\mathbf{X}$ sequences, according to $\prod_{j=1}^n p(x_j|u_{2,j}, u_{3,j})$. The $\mathbf{X}$ sequences are indexed as $\mathbf{X}(l_{11}, l'_1, w_0, l_{21}, l'_2, l^\dagger_2, l_{31}, l'_3, l^\dagger_3)$, where $l_{11} \in \{1, ..., 2^{nL_{11}}\}$, $l'_1 \in \{1, ..., 2^{nL'_1}\}$. Randomly place the $\mathbf{X}$ sequences into $2^{nL_{11}}$ bins indexed by l_{11}. The $2^{nL'_1}$ codewords in each bin are indexed by l'_1.

We now map the message part indices l_1 to bin indices l_{11}, by mapping the 2^{nL_1} message part to the $2^{nL_{11}}$ bins. Again, if

$$L_{11} < L_1 \leq L_{11} + L'_1, \tag{4.69}$$

each codeword of $\mathbf{X}$ is mapped to at least one of l_1; this condition is used in the following analysis.

Encoding: To send (w_0, w_1), express w_1 by (l_1, l_2, l_3) and send the codeword $\mathbf{x}(l_{11}, l'_1, w_0, l_{21}, l'_2, l^\dagger_2, l_{31}, l'_3, l^\dagger_3)$.

Decoding: Without loss of generality, assume that the all-ones vector representing $(l_{11}, l'_1, w_0, l_{21}, l'_2, l^\dagger_2, l_{31}, l'_3, l^\dagger_3)$ is sent. The receivers decode as follows:

1. Receiver 1 uses joint typical decoding of $\{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{x}, \mathbf{y}_1\}$ to find the indices $(l_{11}, l'_1, w_0, l_{21}, l'_2, l^\dagger_2, l_{31}, l'_3, l^\dagger_3)$, from which the decoder can calculate the indices (l_2, l_3, l_1).

2. Receiver 2 uses indirect decoding of $\mathbf{u}_2$ [91] to find the index w_0. Once this is known, $\mathbf{u}_1$ is also found.

3. Receiver 3 uses indirect decoding of $\mathbf{u}_3$ to find the index w_0.

Error Analysis: In the encoding process, we need to find the pair $(l^\dagger_2, l^\dagger_3)$ such that (4.68) is satisfied. By the mutual covering lemma [36], if

$$L^\dagger_2 + L^\dagger_3 \geq I(U_2; U_3|U_1) \tag{4.70}$$

is satisfied with high probability, then there will be at least one pair $(l^\dagger_2, l^\dagger_3)$ such that $(\mathbf{u}_2, \mathbf{u}_3, \mathbf{u}_1)$ is jointly typical. We also have $L_{21}+L'_2+L^\dagger_2 = Q_2$ and $L_{31}+L'_3+L^\dagger_3 = Q_3$ from the encoding. Combined with the conditions $L_2 > L_{21} + L'_2$ and $L_3 > L_{31} + L'_3$ from (4.67), we have

$$L^\dagger_2 \leq Q_2 - L_2, \quad L^\dagger_2 \leq Q_3 - L_3. \tag{4.71}$$

Combining (4.70) and (4.71) using Fourier-Motzkin elimination to eliminate $L^\dagger_2$ and $L^\dagger_3$, we get

$$L_2 + L_3 \leq Q_2 + Q_3 - I(U_2; U_3|U_1). \tag{4.72}$$

Also, it is easily seen that $L_2 \leq Q_2$ and $L_3 \leq Q_3$. Putting the last two sets of conditions together, we obtain the following conditions for successful encoding:

$$\begin{aligned} L_2 &\leq Q_2, \\ L_3 &\leq Q_3, \\ L_2 + L_3 &\leq Q_2 + Q_3 - I(U_2; U_3|U_1). \end{aligned} \tag{4.73}$$

At receiver 1, the decoder seeks the indices $(w_0, w_1, l_{21}, l_{31}, l_{11})$ so that

$$\begin{aligned} \Big(\mathbf{u}_1(w_0), \mathbf{u}_2(w_0, l_{21}, l'_2, l^\dagger_2), \mathbf{u}_3(w_0, l_{31}, l'_3, l^\dagger_3), \mathbf{x}(l_{11}, l'_1, w_0, l_{21}, l'_2, l^\dagger_2, l_{31}, l'_3, l^\dagger_3), \mathbf{y}_1\Big) \\ \in \mathcal{T}^n_\epsilon(P_{U_1U_2U_3XY_1}). \end{aligned} \tag{4.74}$$

If there is none or more than one possible codeword, an error is declared. The possible error events are as follows:

(i) $\mathtt{E}_1 : (l_{11}, l'_1, w_0, l_{21}, l'_2, l^\dagger_2, l_{31}, l'_3, l^\dagger_3)$ is equal to the all-ones vector but $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, $\mathbf{x}$ are not jointly typical with $\mathbf{y}_1$. By the properties of typical sequences, $\Pr[\mathtt{E}_1] \leq \epsilon'$, where $\epsilon' \to 0$ for large n.

(ii) $\mathtt{E}_2 : w_0 \neq 1$, with $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, $\mathbf{x}$ jointly typical with $\mathbf{y}_1$. Then, for $\Pr[\mathtt{E}_2] \leq \epsilon'$, we require

$$R_0 + L_{21} + L'_2 + L^\dagger_2 + L_{31} + L'_3 + L^\dagger_3 + L_{11} + L'_1 < I(U_1, U_2, U_3, X; Y_1) = I(X; Y_1) \tag{4.75}$$

since $I(U_1, U_2, U_3; Y_1|X) = 0$ by the Markov chain condition

$$U_1 \to (U_2, U_3) \to X \to Y_1. \tag{4.76}$$

Taking the mapping of messages to indices l_{21}, l_{31}, l_{11} into account, and using the conditions $L_2 \leq L_{21} + L'_2 + L^\dagger_2$, $L_3 \leq L_{31} + L'_3 + L^\dagger_3$ and $L_1 \leq L_{11} + L'_1$, and since $R_1 = L_1 + L_2 + L_3$ we can see from (4.75) that

$$R_0 + R_1 < I(X; Y_1). \tag{4.77}$$

(iii) $\mathtt{E}_3 : w_0 = 1, (l_{21}, l'_2, l^\dagger_2) \neq (1, 1, 1)$ and arbitrary $l_{31}, l'_3, l^\dagger_3, l_{11}, l'_1$, with $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, $\mathbf{x}$ jointly typical with $\mathbf{y}_1$. Then, for $\Pr[\mathtt{E}_3] \leq \epsilon'$, we require

$$\begin{aligned} &L_{21} + L'_2 + L^\dagger_2 + L_{31} + L'_3 + L^\dagger_3 + L_{11} + L'_1 < I(U_2, U_3, X; Y_1|U_1) \\ &= I(X; Y_1|U_1) + I(U_2, U_3; Y_1|X, U_1) = I(X; Y_1|U_1), \end{aligned} \tag{4.78}$$

where the first equality is due to $U_1 \to (U_2, U_3) \to X \to Y_1$. Then, using the conditions to satisfy the mapping of messages to indices l_{21}, l_{31}, l_{11} and $R_1 = L_1 + L_2 + L_3$, we have

$$R_1 < I(X; Y_1|U_1). \tag{4.79}$$

(iv) $\mathtt{E}_4 : w_0 = 1, (l_{21}, l'_2, l^\dagger_2) = (1, 1, 1), (l_{31}, l'_3, l^\dagger_3) \neq (1, 1, 1)$ and arbitrary l_{11}, l'_1, with $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, $\mathbf{x}$ jointly typical with $\mathbf{y}_1$. For $\Pr[\mathtt{E}_4] \leq \epsilon'$, we require

$$\begin{aligned} L_{11} + L'_1 + L_{31} + L'_3 + L^\dagger_3 &< I(U_3, X; Y_1|U_1, U_2) \\ &= I(X; Y_1|U_1, U_2) + I(U_3; Y_1|U_1, U_2, X) \\ &\overset{(a)}{=} I(X; Y_1|U_2) + I(U_3; Y_1|U_2, X) \\ &= I(X; Y_1|U_2) + I(U_2, U_3; Y_1|X) - I(U_2; Y_1|X) \\ &\overset{(b)}{=} I(X; Y_1|U_2), \end{aligned} \tag{4.80}$$

where the first term in (a) is due to $U_1 \to U_2 \to X$ and the second term is due to $U_1 \to (U_2, U_3) \to X \to Y_1$; (b) is due to $U_2 \to X \to Y_1$ and $(U_2, U_3) \to X \to Y_1$. Using the mapping conditions, we have

$$L_3 + L_1 < I(X; Y_1|U_2). \tag{4.81}$$

(v) $\mathtt{E}_5 : w_0 = 1, (l_{21}, l'_2, l_2^\dagger) = (1,1,1), (l_{31}, l'_3, l_3^\dagger) = (1,1,1), (l_{11}, l'_1) \neq (1,1)$ with $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, $\mathbf{x}$ jointly typical with $\mathbf{y}_1$. For $\Pr[\mathtt{E}_5] \leq \epsilon'$, we require

$$L_{11} + L'_1 < I(X; Y_1|U_1, U_2, U_3) = I(X; Y_1|U_2, U_3) \tag{4.82}$$

where the equality is due to $U_1 \to (U_2, U_3) \to X \to Y_1$, from which we obtain, using the mapping conditions

$$L_1 < I(X; Y_1|U_2, U_3). \tag{4.83}$$

(vi) $\mathtt{E}_6 : w_0 = 1, (l_{21}, l'_2, l_2^\dagger) \neq (1,1,1), (l_{31}, l'_3, l_3^\dagger) = (1,1,1)$ and l_{11}, l'_1 arbitrary with $\mathbf{u}_1$, $\mathbf{u}_2$, $\mathbf{u}_3$, $\mathbf{x}$ jointly typical with $\mathbf{y}_1$. Then, to have $\Pr[\mathtt{E}_6] \leq \epsilon'$, we require

$$\begin{aligned} L_{21} + L'_2 + L_2^\dagger + L_{11} + L'_1 &< I(U_2, X; Y_1|U_1, U_3) \\ &= I(X; Y_1|U_1, U_3) + I(U_2; Y_1|U_1, U_3, X) \\ &\overset{(a)}{=} I(X; Y_1|U_3) + I(U_2; Y_1|U_3, X) \\ &= I(X; Y_1|U_3) + I(U_2, U_3; Y_1|X) - I(U_3; Y_1|X) \\ &\overset{(b)}{=} I(X; Y_1|U_3), \end{aligned} \tag{4.84}$$

where the first term of (a) is due to $U_1 \to U_3 \to X$ and the second term of (a) and (b) are due to $(U_2, U_3) \to X \to Y_1$ and $U_3 \to X \to Y_1$, respectively. Under the conditions for mapping indices l_{21}, l_{31}, l_{11} to messages, we then have

$$L_2 + L_1 < I(X; Y_1|U_3). \tag{4.85}$$

Consequently, under the conditions (4.77), (4.79), (4.81), (4.83), (4.85) listed above, the error probability at receiver 1 is less than $\sum_{i=1}^{6} \Pr[\mathtt{E}_i] \leq 6\epsilon'$.

Now, let $q_2 \in \{1, \ldots, 2^{nQ_2}\}$ be the 'super-index' that includes $(l_{21}, l'_2, l_2^\dagger)$. In particular, if $q_2 = 1$, $(l_{21}, l'_2, l_2^\dagger) = (1,1,1)$; if $q_2 \neq 1$, $(l_{21}, l'_2, l_2^\dagger) \neq (1,1,1)$. Assume that $(w_0, q_2) = (1,1)$ is sent to receiver 2. At receiver 2, the decoder finds w_0 by indirect decoding. The error events are:

(i) a) $\mathtt{E}_1' : (w_0, q_2) = (1, 1)$ but $\mathbf{u}_2$ is not jointly typical with $\mathbf{y}_2$ (indirect decoding). In this case, by the properties of typical sequences, we have $\Pr[\mathtt{E}_1'] \leq \epsilon'$.

(ii) $\mathtt{E}_2' : w_0 \neq 1$, q_2 arbitrary and $\mathbf{u}_2$ is jointly typical with $\mathbf{y}_2$ (indirect decoding). This is the same as receiver 2 trying to estimate w_0 such that $(\mathbf{u}_2(w_0, q_2), \mathbf{y}_2) \in T_\epsilon^n(P_{U_2Y_2})$ for any q_2. For $\Pr\{\mathtt{E}_2'\} \leq \epsilon'$, we need

$$R_0 + Q_2 < I(U_2; Y_2). \tag{4.86}$$

Thus the error probability at receiver 2 is less than $\sum_{i=1}^{2} \Pr[\mathtt{E}_i'] \leq 2\epsilon'$.

Now, let $q_3 \in \{1, \ldots, 2^{nQ_3}\}$ be the 'super-index' that includes $(l_{31}, l_3', l_3^\dagger)$. At receiver 3, indirect decoding is used, so that the decoder estimates w_0 such that $(\mathbf{u}_3(w_0, q_3), \mathbf{y}_3) \in T_\epsilon^n(P_{U_3Y_3})$ for any $q_3 \in \{1, \ldots, 2^{nQ_3}\}$. Assuming that $(w_0, q_3) = (1, 1)$ is sent, we require

$$R_0 + Q_3 < I(U_3; Y_3), \tag{4.87}$$

for the error probability at receiver 3 to be small for n sufficiently large.

Equivocation: We now show the bound on the equivocation rate satisfies the security conditions in (4.59), by deriving the bound for $H(W_1|\mathbf{Y}_3)$. A point to note is that, from the coding scheme, W_1 is split into independent parts L_1, L_2, L_3. So, to see what equivocation rates can be achieved for the message parts L_2, L_3 and L_1, we should also derive the bounds for $H(L_2|\mathbf{Y}_3)$, $H(L_3|\mathbf{Y}_3)$ and $H(L_1|\mathbf{Y}_3)$.

Let us denote $\mathcal{I}(L) := (L_2', L_2^\dagger, L_3', L_3^\dagger, L_1')$. For W_1, we have

$$\begin{aligned}
H(W_1|\mathbf{Y}_3) &\geq H(W_1|\mathbf{Y}_3, \mathbf{U}_1) \\
&= H(W_1, \mathcal{I}(L)|\mathbf{Y}_3, \mathbf{U}_1) - H(\mathcal{I}(L)|\mathbf{Y}_3, \mathbf{U}_1, W_1) \\
&\geq H(\mathcal{I}(L)|\mathbf{Y}_3, \mathbf{U}_1) - H(\mathcal{I}(L)|\mathbf{Y}_3, \mathbf{U}_1, W_1) \\
&\overset{(a)}{\geq} H(\mathbf{U}_2, \mathbf{U}_3, \mathbf{X}|\mathbf{Y}_3, \mathbf{U}_1) - H(\mathcal{I}(L)|\mathbf{Y}_3, \mathbf{U}_1, W_1) \\
&= H(\mathbf{U}_2, \mathbf{U}_3, \mathbf{X}, \mathbf{Y}_3|\mathbf{U}_1) - H(\mathbf{Y}_3|\mathbf{U}_1) - H(\mathcal{I}(L)|\mathbf{Y}_3, \mathbf{U}_1, W_1) \\
&= H(\mathbf{U}_2|\mathbf{U}_1) + H(\mathbf{U}_3|\mathbf{U}_2, \mathbf{U}_1) + H(\mathbf{X}|\mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_1) \\
&\quad + H(\mathbf{Y}_3|\mathbf{U}_2, \mathbf{U}_3, \mathbf{X}, \mathbf{U}_1) - H(\mathbf{Y}_3|\mathbf{U}_1) - H(\mathcal{I}(L)|\mathbf{Y}_3, \mathbf{U}_1, W_1) \\
&= H(\mathbf{U}_2|\mathbf{U}_1) + H(\mathbf{U}_3|\mathbf{U}_1) + H(\mathbf{X}|\mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_1) - I(\mathbf{U}_2; \mathbf{U}_3|\mathbf{U}_1) \\
&\quad - I(\mathbf{U}_2, \mathbf{U}_3, \mathbf{X}; \mathbf{Y}_3|\mathbf{U}_1) - H(\mathcal{I}(L)|\mathbf{Y}_3, \mathbf{U}_1, W_1),
\end{aligned} \tag{4.88}$$

where (a) is because $\mathbf{U}_2, \mathbf{U}_3, \mathbf{X}$ is a function of $\mathcal{I}(L)$.

We can now bound each of the terms in (4.88). For the first three terms, by the code generation process,

$$H(\mathbf{U}_2|\mathbf{U}_1) = \log 2^{n(L_{21}+L_2'+L_2^\dagger)} = n(L_{21} + L_2' + L_2^\dagger), \tag{4.89}$$

$$H(\mathbf{U}_3|\mathbf{U}_1) = \log 2^{n(L_{31}+L_3'+L_3^\dagger)} = n(L_{31} + L_3' + L_3^\dagger), \tag{4.90}$$

$$H(\mathbf{X}|\mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_1) = \log 2^{n(L_{11}+L_1')} = n(L_{11} + L_1'). \tag{4.91}$$

For the fourth and fifth terms, using standard methods and following the method in [78], we have

$$I(\mathbf{U}_2; \mathbf{U}_3|\mathbf{U}_1) \leq nI(U_2; U_3|U_1) + n\delta', \tag{4.92}$$

$$I(\mathbf{U}_2, \mathbf{U}_3, \mathbf{X}; \mathbf{Y}_3|, \mathbf{U}_1) \leq nI(U_2, U_3, X; Y_3|U_1) + n\delta', \tag{4.93}$$

where $\delta' > 0$ and is small for n sufficiently large. To evaluate the last term in (4.88), we introduce the following lemma, which is proved in Appendix B.4.

Lemma 6. *We have*

$$\begin{aligned} H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, W_1) \leq n(&L_2' + L_2^\dagger + L_3' + L_3^\dagger + L_1' - I(U_2; U_3|U_1) \\ &- I(U_2, U_3; Y_3|U_1) - I(X; Y_3|U_2, U_3, U_1) + \delta(\epsilon)) + 2, \end{aligned} \tag{4.94}$$

where $\delta(\epsilon) \to 0$ for $\epsilon \to 0$ and n sufficiently large, under the conditions

$$\begin{aligned} L_2' + L_2^\dagger + L_3' + L_3^\dagger + L_1' &\geq I(U_2; U_3|U_1) + I(U_2, U_3; Y_3|U_1) + I(X; Y_3|U_2, U_3, U_1), \\ L_3' + L_3^\dagger + L_1' &\geq I(U_3; Y_3|U_1) + I(X; Y_3|U_2, U_3, U_1) \\ L_2' + L_2^\dagger + L_1' &\geq I(U_2; Y_3|U_1) + I(X; Y_3|U_2, U_3, U_1) \\ L_2' + L_2^\dagger + L_3' + L_3^\dagger &\geq I(U_2; U_3|U_1) + I(U_2, U_3; Y_3|U_1) \\ L_1' &\geq I(X; Y_3|U_2, U_3, U_1) \\ L_3' + L_3^\dagger &\geq I(U_3; Y_3|U_1) \\ L_2' + L_2^\dagger &\geq I(U_2; Y_3|U_1). \end{aligned} \tag{4.95}$$

□

Now, substituting (4.89) - (4.94) into the last line of (4.88), we have

$$H(W_1|\mathbf{Y}_3) \geq n(L_{21} + L_{31} + L_{11} - 2\delta' - \delta(\epsilon)) - 2. \tag{4.96}$$

This means that the equivocation rate

$$R_{1e} \leq L_{21} + L_{31} + L_{11} \tag{4.97}$$

is achievable.

To bound $H(L_2|\mathbf{Y}_3)$, we follow the steps resulting in (4.88), which gives

$$\begin{aligned} H(L_2|\mathbf{Y}_3) &\geq H(L_2|\mathbf{Y}_3, \mathbf{U}_1) \\ &= H(L_2, L_2', L_2^\dagger|\mathbf{Y}_3, \mathbf{U}_1) - H(L_2', L_2^\dagger|\mathbf{Y}_3, \mathbf{U}_1, L_2) \\ &\geq H(L_2', L_2^\dagger|\mathbf{Y}_3, \mathbf{U}_1) - H(L_2', L_2^\dagger|\mathbf{Y}_3, \mathbf{U}_1, L_2) \\ &\geq H(\mathbf{U}_2, |\mathbf{Y}_3, \mathbf{U}_1) - H(L_2', L_2^\dagger|\mathbf{Y}_3, \mathbf{U}_1, L_2) \\ &= H(\mathbf{U}_2, \mathbf{Y}_3|\mathbf{U}_1) - H(\mathbf{Y}_3|\mathbf{U}_1) - H(L_2', L_2^\dagger|\mathbf{Y}_3, \mathbf{U}_1, L_2) \\ &= H(\mathbf{U}_2|\mathbf{U}_1) - I(\mathbf{U}_2; \mathbf{Y}_3|\mathbf{U}_1) - H(L_2', L_2^\dagger|\mathbf{Y}_3, \mathbf{U}_1, L_2). \end{aligned} \tag{4.98}$$

Similarly,

$$H(L_3|\mathbf{Y}_3) \geq H(\mathbf{U}_3|\mathbf{U}_1) - I(\mathbf{U}_3; \mathbf{Y}_3|\mathbf{U}_1) - H(L_3', L_3^\dagger|\mathbf{Y}_3, \mathbf{U}_1, L_3), \tag{4.99}$$

$$H(L_1|\mathbf{Y}_3) \geq H(\mathbf{X}|\mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_1) - I(\mathbf{X}; \mathbf{Y}_3|\mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_1) - H(L_1'|\mathbf{Y}_3, \mathbf{U}_1, L_1). \tag{4.100}$$

Since in (4.98) - (4.100) the terms involving mutual information are

$$I(\mathbf{U}_2; \mathbf{Y}_3|\mathbf{U}_1) \leq nI(U_2; Y_3|U_1) + n\delta', \tag{4.101}$$

$$I(\mathbf{U}_3; \mathbf{Y}_3|\mathbf{U}_1) \leq nI(U_3; Y_3|U_1) + n\delta', \tag{4.102}$$

$$I(\mathbf{X}; \mathbf{Y}_3|\mathbf{U}_2, \mathbf{U}_3, \mathbf{U}_1) \leq nI(X; Y_3|U_2, U_3, U_1) + n\delta', \tag{4.103}$$

it remains to bound the last terms in (4.98) - (4.100). Following the steps similar to obtaining Lemma 6, we can get

$$H(L_2', L_2^\dagger|\mathbf{Y}_3, \mathbf{U}_1, L_2) \leq n(L_2' + L_2^\dagger - I(U_2; Y_3|U_1) + \delta(\epsilon)) + 2, \tag{4.104}$$

$$H(L_3', L_3^\dagger|\mathbf{Y}_3, \mathbf{U}_1, L_3) \leq n(L_3' + L_3^\dagger - I(U_3; Y_3|U_1) + \delta(\epsilon)) + 2, \tag{4.105}$$

$$H(L_1'|\mathbf{Y}_3, \mathbf{U}_1, L_1) \leq n(L_1' - I(X; Y_3|U_2, U_3, U_1) + \delta(\epsilon)) + 2, \tag{4.106}$$

under the conditions

$$\begin{aligned} L_1' &\geq I(X; Y_3|U_2, U_3, U_1), \\ L_2' + L_2^\dagger &\geq I(U_2; Y_3|U_1), \\ L_3' + L_3^\dagger &\geq I(U_3; Y_3|U_1), \end{aligned} \tag{4.107}$$

respectively. The proof is omitted as we can observe that the conditional entropy bounds (4.104) - (4.106) are subsets of the conditional entropy bound in Lemma 6. Thus the proof for the bounds (4.104) - (4.106) will be similar to the proof of Lemma 6, only less elaborate. Finally, putting the results together into (4.98) - (4.100), we see that the following equivocation rates are achievable for the message parts:

$$L_{2e} \leq L_{21}, \quad L_{3e} \leq L_{31}, \quad L_{1e} \leq L_{11}. \tag{4.108}$$

Now we combine the rate constraints (4.75), (4.78), (4.80), (4.82), (4.84), the achievable equivocation rates (4.97) and (4.108), the coding rates $Q_2 = L_{21} + L_2' + L_2^\dagger$, $Q_3 = L_{31} + L_3' + L_3^\dagger$ and the equivocation conditions (4.95) by Fourier-Motzkin elimination (eliminating L_{11}, L_{21}, L_{31}) to yield the following equivocation constraints:

$$\begin{aligned}
L_{2e} &\leq Q_2 - I(U_2; Y_3|U_1), \\
L_{3e} &\leq Q_3 - I(U_3; Y_3|U_1), \\
L_{1e} &\leq I(X; Y_1|U_2, U_3, U_1) - I(X; Y_3|U_2, U_3, U_1), \\
L_{2e} + L_{1e} &\leq I(X; Y_1|U_3) - I(X; Y_3|U_2, U_3, U_1) - I(U_2; Y_3|U_1), \\
L_{3e} + L_{1e} &\leq I(X; Y_1|U_2) - I(X; Y_3|U_2, U_3, U_1) - I(U_3; Y_3|U_1), \\
L_{2e} + L_{3e} + L_{1e} &\leq I(X; Y_1|U_1) - I(U_2; U_3|U_1) - I(X; Y_3|U_2, U_3, U_1) \\
&\quad - I(U_2, U_3; Y_3|U_1), \\
R_0 + L_{2e} + L_{3e} + L_{1e} &\leq I(X; Y_1) - I(U_2; U_3|U_1) - I(X; Y_3|U_2, U_3, U_1) \\
&\quad - I(U_2, U_3; Y_3|U_1).
\end{aligned} \tag{4.109}$$

At this point we can make a simple check: by setting $Y_1 = Y_3$, we see that no secret rate is possible for the message W_1. (There is only some positive rate due to R_0 in the last line of the set of inequalities (4.109).) Now, combining (4.73), (4.77), (4.79), (4.81), (4.83), (4.85), (4.86), (4.87) and the set (4.109) using a Fourier-Motzkin elimination with $R_1 = L_1 + L_2 + L_3$ and $R_{1e} = L_{1e} + L_{2e} + L_{3e}$, we can obtain the inner bound to the rate-equivocation region without rate splitting:

$$\begin{aligned}
R_0 &\leq I(U_3; Y_3) - I(U_3; Y_3|U_1) \\
R_0 &\leq I(U_2; Y_2) - I(U_2; Y_3|U_1) \\
2R_0 &\leq I(U_2; Y_2) - I(U_2; U_3|U_1) + I(U_3; Y_3)
\end{aligned}$$

$$
\begin{aligned}
R_1 &\leq I(X;Y_1|U_1) \\
R_1 &\leq I(X;Y_1|U_2) + I(X;Y_1|U_3) \\
R_{1e} &\leq R_1 \\
R_{1e} &\leq I(X;Y_1|U_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) \\
&\quad - I(U_2,U_3;Y_3|U_1) \\
R_{1e} &\leq I(X;Y_1|U_2) + I(X;Y_1|U_3) - 2I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1) \\
&\quad - I(U_3;Y_3|U_1) \\
R_0 + R_1 &\leq I(X;Y_1) \\
R_0 + R_1 &\leq I(U_2;Y_2) + I(X;Y_1|U_2) \\
R_0 + R_1 &\leq I(U_3;Y_3) + I(X;Y_1|U_3) \\
R_0 + R_{1e} &\leq I(X;Y_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1) \\
R_0 + R_{1e} &\leq I(U_2;Y_2) + I(X;Y_1|U_2) - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1) \\
&\quad - I(U_3;Y_3|U_1) \\
R_0 + R_{1e} &\leq I(U_3;Y_3) + I(X;Y_1|U_3) - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1) \\
&\quad - I(U_3;Y_3|U_1) \\
2R_0 + R_1 &\leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3) \\
2R_0 + R_{1e} &\leq I(U_2;Y_2) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3) - I(X;Y_3|U_2,U_3) \\
&\quad - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1) \\
2R_0 + 2R_1 &\leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2) + I(X;Y_1|U_3)
\end{aligned}
\tag{4.110}
$$

Finally, to apply the rate splitting procedure mentioned earlier, we move some of the rate of R_1 in the set (4.110) without rate splitting to R_0. Substitute $R_{1n} = R_{10} + R_{11}$ and $R_1 = R_{11}$ and $R_0 = R_{0n} - R_{10}$ into (4.110). The rates R_{1n} and R_{0n} are the new rates for R_1 and R_0 in (4.110) achieved using rate splitting. Using $0 \leq R_{10}$ and $0 \leq R_{11}$, we perform a Fourier-Motzkin elimination to remove R_{10} and R_{11} and obtain the region in the theorem, with $R_1 = R_{1n}$, $R_0 = R_{0n}$. The details of the entire Fourier-Motzkin elimination can be found in Appendix B.5.

□

We should remark that it is possible to extend our secure coding scheme to the

3-receiver BC with 3 DMS. In the case without security [91, Sect. VB], the additional message W_2 is encoded into U_2 together with part message L_2, and Y_2 performs joint typical decoding on U_2 instead of indirect decoding to obtain W_2. All the other encoding and decoding functions stay the same, so that Y_1 obtains W_1, W_2, W_0 via X, Y_2 obtains W_2, W_0 via U_2 and Y_3 obtains W_0 indirectly from U_3. We see that since W_2 and L_2 now 'share' U_2, we can still use our secure scheme, but we now make the secure parts R_{2e} and L_{2e} share U_2, instead of only L_{2e} using U_2. Indeed the equivocation calculation still goes through unchanged even by adding the additional message to the encoding of U_2. Applying this idea to obtain an inner bound for the 3-receiver BC with 3 DMS and two confidential messages is now part of our ongoing work; the drawback of course is that the Fourier-Motzkin elimination has now increased many times in complexity.

We have noted that Chia and El Gamal in [17, 18] also studied the 3-receiver BC with DMS and Y_3 an eavesdropper, but with W_1 sent to Y_1 and Y_2. Hence the coding scheme in [17, 18] is somewhat different compared to ours. In the scheme of [17, 18], set $V_0 = U_1$, $V_1 = U_2$, $V_2 = U_3$ and introduce a time sharing r.v. U; the r.v.s satisfy $U \to V_0 \to (V_1, V_2) \to X$. Their inner bound was established by coding W_1 into V_0, then V_1 and V_2 were generated using superposition coding via $p(v_1|v_0)$ and $p(v_2|v_0)$. To encode V_1, V_2, V_0 into codeword X, use Marton's achievability scheme to find the joint typical pair (V_1, V_2) and encode into X. Security is provided by code partitioning for V_1, V_2, V_0. This is in contrast to our scheme where we use double binning for U_2, U_3 and code partitioning for X. We also note that by specializing to 2 receivers, the inner bound in [17, 18] can also reduce to the Csiszár-Körner BC with one common and one confidential message region. However, due to the differences in the coding schemes, it is not easy to say conclusively whether one is better than the other in the general case.

Lastly we remark that the coding scheme for the 3-receiver, 2 eavesdropper BC models in the multilevel BC and its generalization [17, 18, 104] was superposition coding, similar to our K-receiver degraded BC scheme in Section 4.2.

4.3.3 Outer Bounds to the Rate-equivocation of the 3-receiver BC with 2 DMS

In this section, we derive outer bounds to the rate-equivocation of the 3-receiver BC with 2 DMS and one confidential message. We note that the outer bound for the general case where the conditions on all the channels are general, is unknown[6]; it is known only for the case without security [91, Proposition 6]. The main difficulty in deriving the outer bound for the general case lies in using the Csiszár sum lemma [30, Lemma 7](like in Section 4.2.5) which is not easily generalized to three receivers. Methods to derive the outer bounds using alternative methods may be investigated in future work; at the moment we shall consider some subclasses of the general 3-receiver BC where outer bounds can be derived.

We now consider subclasses of the general 3-receiver BC with 2 DMS where we have the following conditions on the receivers:

1. Y_1 is less noisy than Y_2 and Y_3;
2. Y_1 and Y_2 are less noisy than Y_3.

Both these two conditions result in 3-receiver BCs which have more general conditions than the 3-receiver degraded BC, or the 3-receiver multilevel BC [17, 18] and its generalization in [104]. We also note that under these conditions, the inner and outer bounds match. The proof for the outer bound is termed the 'converse proof' in this case. In the following we will state the rate-equivocation regions for the above two types of 3-receiver BCs, outline the achievability proofs and show the converse proofs.

Y_1 is Less Noisy Than Y_2 and Y_3

Corollary 3. *The rate-equivocation region for the 3-receiver BC with 2 DMS and one confidential message where Y_1 is less noisy than Y_2 and Y_3 is the closure of all rate-tuples (R_0, R_1, R_{1e}) satisfying*

$$\begin{aligned} R_0 &\leq \min\{I(U;Y_2), I(U;Y_3)\} \\ R_1 &\leq I(X;Y_1|U) \\ R_{1e} &\leq I(X;Y_1|U) - I(X;Y_3|U), \end{aligned} \tag{4.111}$$

[6] In an earlier version of this work [22], we had derived a general outer bound by insertion of auxiliary r.v.s, which is now been realized to be mistaken. This section thus serves to correct the error.

over the p.d.f. $p(u,x) = p(u)p(x|u)$. □

Proof. We show achievability and the converse.

Achievability: The region in Corollary 3 can be obtained from the region in Theorem 14 by setting $U_2 = U_3 = U_1 = U$ and removing redundancies. A secure coding scheme to achieve the rate-equivocation region needs only to use code partitioning and superposition coding like the Csiszár-Körner scheme [30].

Converse: We now use a $(2^{nR_0}, 2^{nR_1}, n)$-code with error probability $P_e^{(n)}$ and code construction so that we have the Markov chain condition $(W_0, W_1) \to \mathbf{X} \to (\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3)$. Then, the probability distribution on $\mathcal{W}_0 \times \mathcal{W}_1 \times \mathcal{X}^n \times \mathcal{Y}_1^n \times \mathcal{Y}_2^n \times \mathcal{Y}_3^n$ is given by

$$p(w_0)p(w_1)p(\mathbf{x}|w_0,w_1)\prod_{i=1}^{n} p(y_{1i}, y_{2i}, y_{3i}|x_i). \tag{4.112}$$

We first note that from the definition of more capable and less noisy channels (see Appendix A), when Y_1 is less noisy than Y_2 or Y_3, then it also follows that Y_1 is more capable than Y_2 or Y_3.

We now also define the new auxiliary random variable $U_i \triangleq (W_0, \mathbf{Y}_1^{i-1})$ satisfying the condition

$$U_i \to X_i \to (Y_{1,i}, Y_{2,i}, Y_{3,i}), \qquad \forall i. \tag{4.113}$$

To see that the above assertion is true, we refer to the state dependency graph for the 3-receiver BC with 2 DMS in Figure 4.7; we see that the group of states $(Y_{1,i}, Y_{2,i}, Y_{3,i})$ depends on $(W_0, \mathbf{Y}_1^{i-1})$ only through X_i, hence the Markov chain.

To begin, we have, by Fano's inequality

$$\begin{aligned} &H(W_1|\mathbf{Y}_1) \le n\gamma_1, \quad H(W_1|\mathbf{Y}_1, W_0) \le n\gamma_2, \\ &H(W_0|\mathbf{Y}_1) \le n\gamma_3, \quad H(W_0|\mathbf{Y}_2) \le n\gamma_4, \quad H(W_1|\mathbf{Y}_3) \le n\gamma_5, \end{aligned} \tag{4.114}$$

where the $\gamma_i \to 0$ as $n \to \infty$, for $i = 1, \cdots, 5$. Next, we have the equivocation rate

$$\begin{aligned} nR_{1e} &\overset{(a)}{\le} H(W_1|\mathbf{Y}_3) + n\epsilon_1 \\ &\le H(W_0, W_1|\mathbf{Y}_3) + n\epsilon_1 \\ &\le H(W_1|\mathbf{Y}_3, W_0) + H(W_0|\mathbf{Y}_3) + n\epsilon_1 \\ &= H(W_1|\mathbf{Y}_3, W_0) + H(W_0|\mathbf{Y}_3) - H(W_1|\mathbf{Y}_1, W_0) + H(W_1|\mathbf{Y}_1, W_0) + n\epsilon_1 \\ &= I(W_1; \mathbf{Y}_1|W_0) - I(W_1; \mathbf{Y}_3|W_0) + H(W_1|\mathbf{Y}_3, W_0) + H(W_0|\mathbf{Y}_3) + n\epsilon_1 \\ &\overset{(b)}{\le} I(W_1; \mathbf{Y}_1|W_0) - I(W_1; \mathbf{Y}_3|W_0) + n(\epsilon_1 + \gamma_2 + \gamma_5), \end{aligned} \tag{4.115}$$

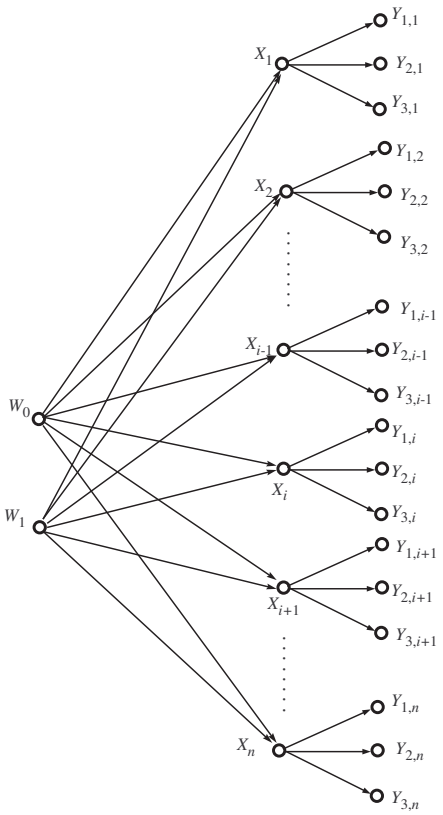


Figure 4.7: State dependency graph for 3-receiver BC with 2 DMS.

where (a) is by the secrecy condition, and (b) is by Fano's inequality. Continuing the chain of inequalities by expanding the mutual informations in (4.115) using the chain rule, we have, defining $\delta' = \epsilon_1 + \gamma_2 + \gamma_5$

$$
\begin{aligned}
nR_{1e} &\leq \sum_{i=1}^{n} \Big[I(W_1; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) - I(W_1; Y_{3,i}|W_0, \tilde{\mathbf{Y}}_3^{i+1}) \Big] + n\delta' \\
&= \sum_{i=1}^{n} \Big[I(W_1, \tilde{\mathbf{Y}}_3^{i+1}; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) - I(\tilde{\mathbf{Y}}_3^{i+1}; Y_{1,i}|W_0, W_1, \mathbf{Y}_1^{i-1}) \\
&\quad - I(W_1, \mathbf{Y}_1^{i-1}; Y_{3,i}|W_0, \tilde{\mathbf{Y}}_3^{i+1}) + I(\mathbf{Y}_1^{i-1}; Y_{3,i}|W_0, W_1, \tilde{\mathbf{Y}}_3^{i+1}) \Big] + n\delta' \\
&\overset{(a)}{=} \sum_{i=1}^{n} \Big[I(W_1, \tilde{\mathbf{Y}}_3^{i+1}; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) - I(W_1, \mathbf{Y}_1^{i-1}; Y_{3,i}|W_0, \tilde{\mathbf{Y}}_3^{i+1}) \Big] + n\delta' \\
&= \sum_{i=1}^{n} \Big[I(W_1; Y_{1,i}|W_0, \tilde{\mathbf{Y}}_3^{i+1}, \mathbf{Y}_1^{i-1}) + I(\tilde{\mathbf{Y}}_3^{i+1}; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) \\
&\quad - I(W_1; Y_{3,i}|W_0, \tilde{\mathbf{Y}}_3^{i+1}, \mathbf{Y}_1^{i-1}) - I(\mathbf{Y}_1^{i-1}; Y_{3,i}|W_0, \tilde{\mathbf{Y}}_3^{i+1}) \Big] + n\delta' \\
&\overset{(b)}{=} \sum_{i=1}^{n} \Big[I(W_1; Y_{1,i}|W_0, \tilde{\mathbf{Y}}_3^{i+1}, \mathbf{Y}_1^{i-1}) - I(W_1; Y_{3,i}|W_0, \tilde{\mathbf{Y}}_3^{i+1}, \mathbf{Y}_1^{i-1}) \Big] + n\delta' \\
&= \sum_{i=1}^{n} \Big[I(W_1, \tilde{\mathbf{Y}}_3^{i+1}; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) - I(\tilde{\mathbf{Y}}_3^{i+1}; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1})
\end{aligned}
$$

$$
\begin{aligned}
&-I(W_1, \tilde{\mathbf{Y}}_3^{i+1}; Y_{3,i}|W_0, \mathbf{Y}_1^{i-1}) + I(\tilde{\mathbf{Y}}_3^{i+1}; Y_{3,i}|W_0, \mathbf{Y}_1^{i-1})\Big] + n\delta' \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} \Big[I(W_1, \tilde{\mathbf{Y}}_3^{i+1}; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) - I(W_1, \tilde{\mathbf{Y}}_3^{i+1}; Y_{3,i}|W_0, \mathbf{Y}_1^{i-1})\Big] + n\delta' \\
&= \sum_{i=1}^{n} \Big[I(X_i, W_1, \tilde{\mathbf{Y}}_3^{i+1}; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) - I(X_i; Y_{1,i}|W_0, W_1, \tilde{\mathbf{Y}}_3^{i+1}\mathbf{Y}_1^{i-1}) \\
&-I(X_i, W_1, \tilde{\mathbf{Y}}_3^{i+1}; Y_{3,i}|W_0, \mathbf{Y}_1^{i-1}) + I(X_i; Y_{1,i}|W_0, W_1, \tilde{\mathbf{Y}}_3^{i+1}\mathbf{Y}_1^{i-1})\Big] + n\delta' \\
&\overset{(d)}{\leq} \sum_{i=1}^{n} \Big[I(X_i, W_1, \tilde{\mathbf{Y}}_3^{i+1}; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) - I(X_i, W_1, \tilde{\mathbf{Y}}_3^{i+1}; Y_{3,i}|W_0, \mathbf{Y}_1^{i-1})\Big] + n\delta' \\
&\overset{(e)}{=} \sum_{i=1}^{n} \Big[I(X_i; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) - I(X_i; Y_{3,i}|W_0, \mathbf{Y}_1^{i-1})\Big] + n\delta' \qquad (4.116)
\end{aligned}
$$

where (a) and (b) are by the Csiszár sum identity [30, Lemma 7], (c) is because $I(\tilde{\mathbf{Y}}_3^{i+1}; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) - I(\tilde{\mathbf{Y}}_3^{i+1}; Y_{3,i}|W_0, \mathbf{Y}_1^{i-1}) \geq 0$ since Y_1 is less noisy than Y_3, (d) is because $I(X_i; Y_{1,i}|W_0, W_1, \tilde{\mathbf{Y}}_3^{i+1}, \mathbf{Y}_1^{i-1}) - I(X_i; Y_{3,i}|W_0, W_1, \tilde{\mathbf{Y}}_3^{i+1}, \mathbf{Y}_1^{i-1}) \geq 0$ since Y_1 is more capable than Y_3, and (e) is because, given $(W_0, \mathbf{Y}_1^{i-1})$, $(W_1, \tilde{\mathbf{Y}}_3^{i+1}) \to X_i \to (Y_{1,i}, Y_{3,i})$ forms a Markov chain. The last assertion may be checked using the state dependency graph in Figure 4.7, where we see that the group $(Y_{1,i}, Y_{3,i})$ depends on $(W_1, \tilde{\mathbf{Y}}_3^{i+1})$ only through X_i, hence the Markov chain.

For rates R_0, we first have the rate for W_0 sent to Y_2:

$$
\begin{aligned}
nR_0 &= H(W_0) = I(W_0; \mathbf{Y}_2) + H(W_0|\mathbf{Y}_2) \\
&\overset{(a)}{\leq} I(W_0; \mathbf{Y}_2) + n\gamma_4 \\
&= \sum_{i=1}^{n} I(W_0; Y_{2,i}|\tilde{\mathbf{Y}}_2^{i+1}) + n\gamma_4 \\
&= \sum_{i=1}^{n} \Big[I(W_0, \mathbf{Y}_1^{i-1}; Y_{2,i}|\tilde{\mathbf{Y}}_2^{i+1}) - I(\mathbf{Y}_1^{i-1}; Y_{2,i}|W_0\tilde{\mathbf{Y}}_2^{i+1})\Big] + n\gamma_4 \\
&\leq \sum_{i=1}^{n} \Big[I(W_0, \tilde{\mathbf{Y}}_2^{i+1}, \mathbf{Y}_1^{i-1}; Y_{2,i}) - I(\mathbf{Y}_1^{i-1}; Y_{2,i}|W_0\tilde{\mathbf{Y}}_2^{i+1})\Big] + n\gamma_4 \\
&\overset{(b)}{=} \sum_{i=1}^{n} \Big[I(W_0, \mathbf{Y}_1^{i-1}; Y_{2,i}) + I(\tilde{\mathbf{Y}}_2^{i+1}; Y_{2,i}|W_0, \mathbf{Y}_1^{i-1}) \\
&\quad -I(\tilde{\mathbf{Y}}_2^{i+1}; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1})\Big] + n\gamma_4 \\
&\overset{(c)}{\leq} \sum_{i=1}^{n} I(W_0, \mathbf{Y}_1^{i-1}; Y_{2,i}) + n\gamma_4, \qquad (4.117)
\end{aligned}
$$

where (a) is by Fano's inequality, (b) is by applying the Csiszár sum identity [30, Lemma 7] on the last term in the brackets, and (c) is because $I(\tilde{\mathbf{Y}}_2^{i+1}; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) - I(\tilde{\mathbf{Y}}_2^{i+1}; Y_{2,i}|W_0, \mathbf{Y}_1^{i-1}) \geq 0$ as Y_1 is less noisy than Y_2. Similarly, we can get for W_0

sent to Y_3,

$$nR_0 \leq \sum_{i=1}^{n} I(W_0, \mathbf{Y}_1^{i-1}; Y_{3,i}) + n\gamma_5. \tag{4.118}$$

For rate R_1, we have

$$\begin{aligned}
nR_1 = H(W_1) &= I(W_1; \mathbf{Y}_1) + H(W_1|\mathbf{Y}_1) \\
&\overset{(a)}{\leq} I(W_1; W_0, \mathbf{Y}_1) + n\gamma_1 \\
&\overset{(b)}{=} I(W_1; \mathbf{Y}_1|W_0) + n\gamma_1 \\
&= \sum_{i=1}^{n} I(W_1; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) + n\gamma_1 \\
&\overset{(c)}{=} \sum_{i=1}^{n} \Big[H(Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) - H(Y_{1,i}|W_0, W_1, \mathbf{Y}_1^{i-1}, X_i)\Big] + n\gamma_1 \\
&= \sum_{i=1}^{n} I(W_1, X_i; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) + n\gamma_1 \\
&\overset{(d)}{=} \sum_{i=1}^{n} I(X_i; Y_{1,i}|W_0, \mathbf{Y}_1^{i-1}) + n\gamma_1,
\end{aligned} \tag{4.119}$$

where (a) is by Fano's inequality, (b) is by independence of W_0 and W_1, (c) is because we have included X_i in the conditioning of the second term in the brackets as X_i is a function of W_1, and (d) is by $W_1 \to X_i \to Y_{1,i}$ forming a Markov chain, given $(W_0, \mathbf{Y}_1^{i-1})$.

To finish off the converse proof, introduce r.v. J, independent of $W_0, W_1, \mathbf{X}, \mathbf{Y}_1$, $\mathbf{Y}_2, \mathbf{Y}_3$, and uniformly distributed over $\{1, \cdots, n\}$. Setting $U \triangleq W_0, Y_1^{J-1}J$, $X \triangleq X_J$, $Y_1 \triangleq Y_{1,J}$, $Y_2 \triangleq Y_{2,J}$ and $Y_3 \triangleq Y_{3,J}$ and substituting into (4.116), (4.117), (4.118), (4.119), we can obtain the bounds in Corollary 3. Lastly, the memoryless character of the channel means that the condition $U \to X \to (Y_1, Y_2, Y_3)$ is met.

□

Y_1 and Y_2 are Less Noisy Than Y_3

Corollary 4. *The rate-equivocation region for the 3-receiver BC with 2 DMS and one confidential message where Y_1 and Y_2 are less noisy than Y_3 is the closure of all rate-tuples (R_0, R_1, R_{1e}) satisfying*

$$\begin{aligned}
R_0 &\leq \min\{I(U; Y_3)\} \\
R_1 &\leq I(X; Y_1|U) \\
R_{1e} &\leq I(X; Y_1|U) - I(X; Y_3|U),
\end{aligned} \tag{4.120}$$

over the p.d.f. $p(u, x) = p(u)p(x|u)$. □

We can see that this region is a subset of the region in Corollary 3. The region in Corollary 4 is also a subset of the region in [18, Proposition 2] under the same conditions; the difference is that the region in Corollary 4 is a result of only one message to send to Y_1, while the region in [18, Proposition 2] is a result of sending one message each to Y_1 and Y_2, resulting in one additional bound each for R_1 and R_{1e}.

Proof. Achievability: The region in Corollary 4 can be obtained from the region in Theorem 14 by setting $U_2 = U_3 = U_1 = U$, removing redundancies, and imposing the condition $I(U; Y_2) \geq I(U; Y_3)$. In fact, the region in Corollary 4 can be obtained from the region in Corollary 3 by using $I(U; Y_2) \geq I(U; Y_3)$. Again, the secure coding scheme to achieve the rate-equivocation region needs only to use code partitioning and superposition coding like the Csiszár-Körner scheme [30].

Converse: The converse proof follows exactly from the proof in Corollary 3.

□

4.3.4 Conclusions

Bounds to the rate-equivocation region for the general 3-receiver BC with DMS, in which receiver 3 is an eavesdropper receiving the common message, are presented. This model is a more general model than the 2-receiver BCs with confidential messages with an external eavesdropper, and 3-receiver degraded BCs with confidential messages. We obtain, with secrecy constraints, a new inner bound to the rate-equivocation region for the 3-receiver BC with 2 DMS. This inner bound region reduces to known bounds for the 3-receiver BC with 2 DMS and no security, the 3-receiver BC with 2 DMS with no security where Y_1 is less noisy than Y_2, and the 2-receiver BC with one common and one confidential message. Outer bounds for special cases of the 3-receiver BC with 2 DMS have been derived, where Y_1 is less noisy than Y_2 and Y_3, and where Y_1 and Y_2 are less noisy than Y_3. These outer bounds are shown to match the inner bounds for the special cases.

We can also see that our secure scheme can be straightforwardly used to provide security for the even more general 3-receiver BC with 3 DMS, which is our ongoing work. Lastly, we mention that deriving the outer bound for the general case was not possible as the conventional method relies on the Csiszár sum lemma which is not

easily generalized to three receivers. Furthermore, our secure scheme suggests that, to design a practical code, we should use multilevel coding in conjunction with a dirty paper coding scheme such as in [40].

Chapter 5

Signal Processing for Enhancing Message Secrecy in Relay Channels

In this chapter we will focus on more practical methods to enhance the secrecy rate, especially when the eavesdropper has more favorable conditions compared to the legitimate receiver. We do this by introducing a relay node into the wiretap channel model. This will then help us to find a distributed signal processing method that can enhance the secrecy rate.

5.1 Introduction

5.1.1 Artificial Noise via Beamforming to Enhance Secrecy

We firstly note that there is an alternative method to enhance the secrecy rate, which is to use multiple antennas at the transmitter and legitimate receiver and combine this with beamforming methods to send useful signal to the legitimate receiver and noise to the eavesdropper, as proposed in Goel and Negi [52]. Specifically, the transmitter sends the message in the direction corresponding to the legitimate receiver's channel, and sends white noise (called artificial noise) in all other directions, disregarding any knowledge of the channel state information (CSI) of the eavesdropper's channel. This masked beamforming approach is in the framework of the wiretap channel.

The artificial noise approach of [52] was also investigated for more elaborate models of the CSI knowledge. In Khisti and Wornell [63] fast Rayleigh channels were considered, where the message is transmitted over a block that is long compared to the coherence time of the channel, while in Mukherjee and Swindlehurst [88], the impact of

imperfections in the CSI of the legitimate and eavesdropper channels are investigated. Another way to improve the secrecy rate was to jointly optimize the beamforming vector and the artificial noise covariance matrix, as in the work of Qin *et al* [102] and Liao *et al* [74], where CSI of the receiver's and eavesdropper's channel was assumed. Lastly, Liu *et al* [75] studied the effects of CSI imperfections on the receiver's channel when using quantized channel feedback and the artificial noise approach of [52].

5.1.2 Cooperative Relaying to Enhance Secrecy

However, an arguably more flexible method is to introduce an additional relay or helping node into the wiretap channel. An advantage of studying such a model compared to the beamforming approach is that cooperative networks can be modeled, where the additional flexibility comes from more options in relaying or jamming methods to enhance the rate at the legitimate receiver. We recall that the wiretap channel is a three-node channel, with a transmitter (source $\mathtt{S}$), legitimate receiver (destination $\mathtt{D}$) and an eavesdropper $\mathtt{E}$[1]. To enhance the secrecy rate then, we introduce an additional node into the wiretap channel to turn the wiretap channel into a relay channel with an external eavesdropper, which is depicted in Figure 5.1. The additional node is, of course, the relay node, which in our work represents a bank of relays $\mathtt{R}_1, \cdots \mathtt{R}_M$. We shall see that we can formulate the secrecy rate in the RC with external eavesdropper as an optimization problem, and then propose a distributed signal processing solution to the optimization problem.

There are three common cooperative strategies that could be used to enhance secrecy. They are Decode and Forward (DF) relaying, Amplify and Forward (AF) relaying, and cooperative jamming (CJ).

DF: In DF, there are two stages in the transmission protocol. In the first stage, the source broadcasts its symbols to the M trusted relays. In the second stage, the source is silent and the M relays decode and re-encode the message and each relay then sends a weighted version of the re-encoded symbol to the destination. The eavesdropper taps the signals in both stages. The destination combines the received signals in both stages by maximum ratio combining (MRC).

[1] In line with standard cooperative communication terminology, we call the transmitter the source and the legitimate receiver the destination.

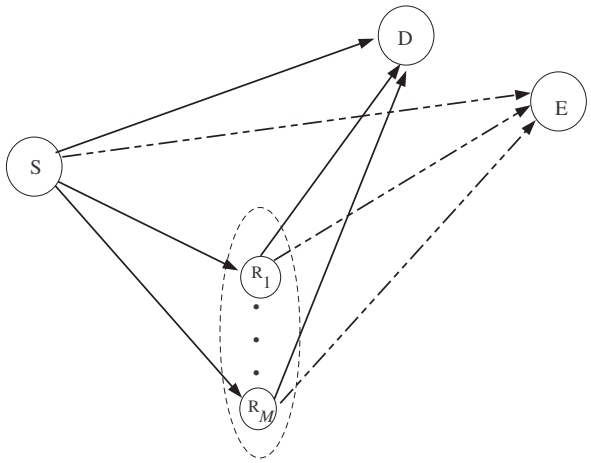


Figure 5.1: System node level model for relay channel with external eavesdropper.

AF: The AF protocol is also a two-stage protocol. In the first stage, the source again broadcasts its symbols to the M trusted relays. In the second stage, each relay sends a weighted version of the noisy signal that it had received from the source in the first stage, to the destination. The eavesdropper taps the signals in both stages, while the destination combines the received signals in both stages by MRC.

CJ: In CJ, there is only one stage. While the source transmits, the relays send a weighted jamming signal which is independent of the source message. We should note that the jamming signal will interfere with the reception of the wanted signal at the destination, while at the same time degrading the eavesdropper's received signal. Thus a careful design of the relaying weights is important.

In the case where there is a direct path between the source and the destination, Dong *et al* [31] and Li *et al* [71] studied the AF and DF variants. Dong *et al* [31] also studied the CJ method. All these studies assumed global CSI (that is the CSI of all the links are known at the source), and solutions found under a total power constraint for the relays. When there is no direct path between the source and the destination, Zhang and Gursoy [123] investigated the AF and DF variants with the eavesdropper's CSI imperfection modeled as the channel realization minus an estimation error.

Of the three cooperative methods, CJ has the potential to be the most cost effective solution. It can also be used in conjunction with either AF or DF. Although CJ was studied in [31] and [71], a limitation of these studies was that a total relay power constraint was imposed and the optimal CJ solution was not known. This motivated our work

which addresses the optimization of collaborative relay weights for CJ in maximizing the secrecy rate with individual relay power constraints. In the following sections, we first give conditions under which positive secrecy rate is possible. Given these conditions, we propose an algorithm to obtain the optimal CJ relay beamforming solution using a combination of convex optimization and a one-dimensional search. The proposed algorithm is also extended to cope with the grouped relays' power constraints. We further develop a distributed implementation which permits each individual relay to derive its own weight based on its local CSI for achieving a near-optimal secrecy rate.

5.1.3 Other Related Work

We note that after the publication of our work [126], some interesting extensions to the CJ scenario have appeared. The work of Fakoorian and Swindlehurst[41] is a generalization of the present CJ scenario to the case of MIMO channels for all of S, D and E in Figure 5.1, with global CSI at S and a total power constraint on the relay; lower bounds to the secrecy rate improvement due to CJ are derived. In another direction, Huang and Swindlehurst [61] and Dong *et al* [32] used a two-stage protocol where there is no direct path in Figure 5.1, but S and even D send jamming signals when they are normally inactive. Unknown CSI for E was considered in [61], while global CSI was assumed in [32].

Evidently the CSI issue is an important one, as the knowledge of CSI at S will help it to design CJ schemes. A way to handle unknown eavesdropper CSI is to use the idea of secrecy outage. That is, we design the system for a certain secrecy rate, say C_s. Then when the actual secrecy rate drops below C_s due to the channel imperfections, then the system is in outage. The probability that this happens is the secrecy outage. The work of Gabry *et al* [48] obtains the secrecy outage probabilities for the DF, AF and CJ protocols in the scenario of Figure 5.1. A two-dimensional model using CJ, where perhaps multiple relays are allowed to shift their position in the plane of S, D and E, is used to obtain regions where the system is in secrecy outage in Vilela *et al* [111].

Another more recent work by Vilela *et al* [112] used CJ, a two-dimensional model and the concept of secure throughput (the probability that a message is successfully received by D, but not by E) to model medium access control layer behaviour with CJ. The channels were either completely known or known statistically. In the latter case,

this just meant that the throughput was reduced.

5.2 System Model And Problem Formulation

Here we give a more detailed description of our RC with external eavesdropper system with CJ. The detailed system model is given in Figure 5.2 below.

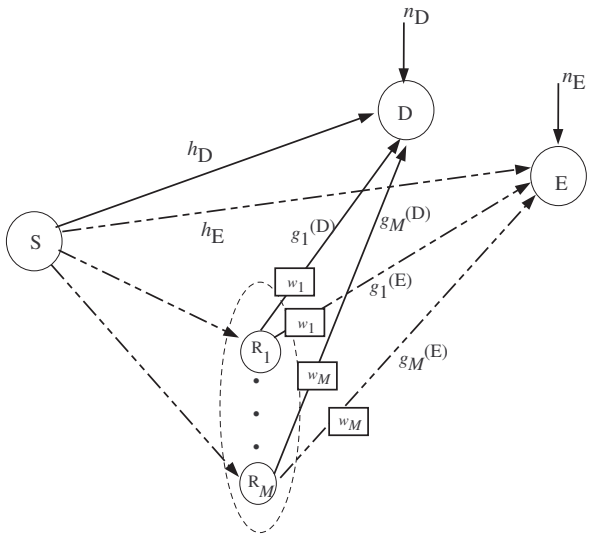


Figure 5.2: System model for relay channel with external eavesdropper and cooperative jamming.

For our system under consideration, we have one source, S, one destination, D, M trusted relays, labeled as $\{\mathtt{R}_1, \ldots, \mathtt{R}_\mathtt{M}\}$ and one eavesdropper, E. All nodes are assumed to have a single antenna. There is a direct link between S and D, and all relays work synchronously in half-duplex mode. The message from S is uniformly distributed over $\mathcal{W} = \{1, \ldots, 2^{nR_0}\}$, for n channel uses and the message rate R_0. The message is mapped to the length-n source codewords x_s^n, and the codewords are transmitted using n time units in a single transmission slot in a time division system.

The source codewords are assumed to be independent zero-mean Gaussian to enable evaluation of the achievable secrecy rate. The reason behind our choice of source codewords is as follows. Our CJ scheme can be viewed alternatively as a multiple-input single-output (MISO) wiretap channel, where the channel from S and relays to D is the MISO main channel, and the channel from S and relays to E is the eavesdropper's MISO channel. Since this MISO wiretap channel is a special case of the MIMO wiretap channel reviewed in Section 3.2.1, the result in Theorem 8 will apply. In particular,

in [62, 96], it was shown that the optimization problem to evaluating the secrecy capacity in Theorem 8 is solved optimally using using an input $\sim \mathcal{N}(0, \mathbf{K})$, where $\mathbf{K}$ is a covariance matrix that satisfies the power constraint for the input. Thus the secrecy capacity of the MIMO wiretap channel (and by extension our MISO wiretap channel) is achieved by the optimal zero mean vector Gaussian input. Thus we assume that our source codewords are zero mean Gaussian.

The channels are assumed to undergo flat fading with CSI perfectly known at S, D and also E. Our assumption of CSI known at E (from the perspective of S) is realistic when the eavesdropper is considered part of the network and its transmissions can be monitored. The assumption is less practical when the eavesdropper is considered as passive, with no information about its CSI. However, we now list some scenarios where our assumption is reasonable. Let the secrecy capacity with S having perfect knowledge of the CSI at E be $C_s(CSI)$. It has been shown for the single antenna fading wiretap channel [12] that the probability that the secrecy rate falls below $C_s(CSI)$, decreases as the SNR increases, if the transmission scheme is designed for perfect knowledge of the CSI at E, while in reality there is imperfect knowledge. Also, a transmission scheme has been proposed in [53] for the single antenna fading wiretap channel with unknown CSI at E. This transmission scheme gives a secrecy rate near to $C_s(CSI)$, if the SNR is large. Our assumption, then, is reasonable for the high SNR case when the eavesdropper is passive and when the eavesdropper can be monitored.

Let h_D denote the channel between S and D, and h_E denote the channel between S and E. In addition, the channel between $\mathtt{R}_m$ and D and that between $\mathtt{R}_m$ and E are denoted, respectively, by $g_m^{(\mathtt{D})}$ and $g_m^{(\mathtt{E})}$. This is depicted in Figure 5.2. For ease of exposition, we define the channel vectors

$$\mathbf{g}_D \triangleq [g_1^{(\mathtt{D})}, g_2^{(\mathtt{D})}, \ldots, g_M^{(\mathtt{D})}]^T, \tag{5.1}$$

$$\mathbf{g}_E \triangleq [g_1^{(\mathtt{E})}, g_2^{(\mathtt{E})}, \ldots, g_M^{(\mathtt{E})}]^T, \tag{5.2}$$

where the superscript $(\cdot)^T$ denotes the transpose operation.

Both S and $\{\mathtt{R}_m\}$ transmit simultaneously to both D and E, while the relays send jamming signals to interfere with E. To be specific, let us focus on a source symbol x_s (with $\mathtt{E}[|x_s|^2] = P_S$) which appears within one time unit of the transmission slot of n time units. In the same time unit, the relays transmit the CJ codewords $\{x_c^{(m)}\}$

that are assumed to be independent zero-mean Gaussian signals, with $\mathtt{E}[|x_c^{(m)}|^2] = 1$. The CJ codeword at $\mathtt{R}_m$ is weighted by w_m, before sending at the same time unit as S transmitting to D via the direct link. We shall define $\mathbf{w} \triangleq [w_1, \ldots, w_M]^T$ as the CJ beamforming vector. Thus, only a 1-stage transmission protocol is considered here as opposed to the conventional AF and DF relay protocols. The received signals at D and E can be, respectively, written as

$$y_D = h_D x_s + \sum_{m=1}^{M} g_m^{(\mathtt{D})} w_m x_c^{(m)} + n_D, \tag{5.3}$$

$$y_E = h_E x_s + \sum_{m=1}^{M} g_m^{(\mathtt{E})} w_m x_c^{(m)} + n_E, \tag{5.4}$$

where n_D and n_E are the zero-mean Gaussian noises at D and E, respectively, with $\mathtt{E}[|n_D|^2] = \sigma_D^2$ and $\mathtt{E}[|n_E|^2] = \sigma_E^2$. The jamming signal introduces interference at both D and E, and it has recently been known that the achievable secrecy rate for this channel can be expressed as

$$R_s = \log_2(1 + \Gamma_D) - \log_2(1 + \Gamma_E), \tag{5.5}$$

where Γ_D and Γ_E are the signal-to-noise ratios (SNRs) at D and E, respectively, and are given by

$$\Gamma_D = \frac{P_S |h_D|^2}{|\mathbf{w}^\dagger \mathbf{g}_D|^2 + \sigma_D^2}, \tag{5.6}$$

$$\Gamma_E = \frac{P_S |h_E|^2}{|\mathbf{w}^\dagger \mathbf{g}_E|^2 + \sigma_E^2}, \tag{5.7}$$

where $(\cdot)^\dagger$ denotes the conjugate transposition. The secrecy rate expression in (5.5) is recognized as the difference of the capacity at D and the capacity at E. This simple form is obtained under the following assumptions [71]: First, the received signals at D and E at time i only depend on the transmitted codewords at the relays at time i (referred to as the memoryless relay channel assumption); secondly, the relays use independent, zero-mean Gaussian codewords, similar to the source, to send the jamming signal. Existing results for the MIMO wiretap channel can be applied to (5.5) under the first assumption.

Our aim is to maximize the secrecy rate via the design of the beamforming vector $\mathbf{w}$. That is,

$$\max_{\mathbf{w}} R \ \text{ s.t. } \ |w_m|^2 \leq p_m \ \forall m, \tag{5.8}$$

where p_m is the transmit power constraint of $\mathtt{R}_m$ and R is defined as

$$R \triangleq \frac{1+\frac{P_s|h_D|^2}{|\mathbf{w}^\dagger \mathbf{g}_D|^2+\sigma_D^2}}{1+\frac{P_s|h_E|^2}{|\mathbf{w}^\dagger \mathbf{g}_E|^2+\sigma_E^2}}. \tag{5.9}$$

5.3 Conditions for Positive Secrecy Rate

The optimization problem given in (5.8) is meaningful only under the conditions that give a positive secrecy rate. Thus we will need to derive these conditions. Given that there exists a $\mathbf{w}$, to obtain a positive secrecy rate we need, from (5.9),

$$\frac{P_s|h_D|^2}{|\mathbf{w}^\dagger \mathbf{g}_D|^2+\sigma_D^2} > \frac{P_s|h_E|^2}{|\mathbf{w}^\dagger \mathbf{g}_E|^2+\sigma_E^2}. \tag{5.10}$$

The equation above can also be expressed as

$$\frac{|\mathbf{w}^\dagger \mathbf{g}_E|^2+\sigma_E^2}{|\mathbf{w}^\dagger \mathbf{g}_D|^2+\sigma_D^2} > \frac{|h_E|^2}{|h_D|^2}. \tag{5.11}$$

The feasibility of (5.11) and thus whether or not positive secrecy rate is possible, can be checked by solving

$$\max_{\{|w_m|^2 \le p_m \forall m\}} \frac{|\mathbf{w}^\dagger \mathbf{g}_E|^2+\sigma_E^2}{|\mathbf{w}^\dagger \mathbf{g}_D|^2+\sigma_D^2} > \frac{|h_E|^2}{|h_D|^2}. \tag{5.12}$$

The optimization problem (5.12) can be re-written as

$$\begin{aligned} &\max_{\{|w_m|^2 \le p_m \forall m\}} t \\ &\text{s.t. } |\mathbf{w}^\dagger \mathbf{g}_E|^2+\sigma_E^2 \ge t(|\mathbf{w}^\dagger \mathbf{g}_D|^2+\sigma_D^2), \end{aligned} \tag{5.13}$$

where $t = \max_{\mathbf{w}} \frac{|\mathbf{w}^\dagger \mathbf{g}_E|^2+\sigma_E^2}{|\mathbf{w}^\dagger \mathbf{g}_D|^2+\sigma_D^2}$. The problem (5.13) can in turn be re-written as

$$\begin{aligned} &\max_{\{|w_m|^2 \le p_m \forall m\}} t \\ &\text{s.t. } \sqrt{t|\mathbf{w}^\dagger \mathbf{g}_D|^2+(t\sigma_D^2-\sigma_E^2)} \le \mathbf{w}^\dagger \mathbf{g}_E. \end{aligned} \tag{5.14}$$

Now, since $\mathbf{w} \succeq \mathbf{0}$, t is lower bounded by $\frac{\sigma_E^2}{\sigma_D^2}$ since when $\mathbf{w} = \mathbf{0}$, $t = \frac{\sigma_E^2}{\sigma_D^2}$. This implies that $t\sigma_D^2 - \sigma_E^2 > 0$, and that all terms within the square root in the constraint of (5.14) are > 0. Given this fact, we can recognize that the optimization problem (5.14) is a second order cone program (SOCP), which can be solved using bisection

search.[2] We remark that a similar technique was used in Zhang and Gursoy [123] for AF relay beamforming but [123] did not show that all terms within the square root are nonnegative to enable the formulation (5.14).

After having now derived the conditions for positive secrecy rate, in the subsequent sections as we describe our method to solve (5.8), we will assume that a positive secrecy rate is achieved.

5.4 Methodology: Fixed $\mathbf{w}^\dagger \mathbf{g}_D$ and One-dimensional Search

The optimization problem (5.8) is challenging because it is non-convex and R is a complicated function of $\mathbf{w}$. Our approach is to first study a sub-problem with $|\mathbf{w}^\dagger \mathbf{g}_D|^2$ fixed and then use a one dimension search to find the solution to (5.8) which is guaranteed to be optimal by its analytical properties. We note that this method is similar in spirit to that in [113], which solves an entirely different problem with a different technique.

5.4.1 Sub-problem with Fixed $|\mathbf{w}^\dagger \mathbf{g}_D|$

We consider the the sub-problem where we fix $|\mathbf{w}^\dagger \mathbf{g}_D|$, which is the interference at D, to a fixed scalar $t \geq 0$. The optimization problem (5.8) is then reduced to

$$\begin{aligned} &\max_{\mathbf{w}} |\mathbf{w}^\dagger \mathbf{g}_E|^2 \\ &\text{s.t. } |\mathbf{w}^\dagger \mathbf{g}_D|^2 = t, \\ &\quad |w_m|^2 \leq p_m \, \forall m. \end{aligned} \tag{5.15}$$

[2] An SOCP has the standard form [14, Sect. 4.4.2]

$$\begin{aligned} &\min_{\mathbf{x}} \mathbf{f}^T \mathbf{x} \\ &\text{s.t. } \|A_i \mathbf{x} + \mathbf{b}_i\| \leq \mathbf{c}_i^T \mathbf{x} + \mathbf{d}_i, \;\; i = 1, \cdots, N, \end{aligned}$$

where $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{f} \in \mathbb{R}^n$, $A_i \in \mathbb{R}^{(n_i - 1) \times n}$, $\mathbf{b}_i \in \mathbb{R}^{n_i - 1}$, $\mathbf{c}_i \in \mathbb{R}^n$, $\mathbf{d}_i \in \mathbb{R}$. The norm in the constraints is the Euclidean norm.

A bisection search for the maximum $t_{\max}$ for a given optimization problem is as follows. We start with an interval $[lb, ub]$ that is known to contain $t_{\max}$. We then solve the optimization problem at the midpoint $t = (lb + ub)/2$ to see whether the optimal value is smaller or larger than t. If the optimal solution is between lb and the midpoint $(lb + ub)/2$, then set $t = lb$, otherwise, set $t = ub$. We then repeat this procedure until the width of the interval is smaller than a determined threshold.

That is, we now maximize the interference to the eavesdropper. Let the optimal solution to the problem (5.15) be $\mathbf{w}_0(t)$ and the corresponding optimal objective value be $f_0(t)$, and let us define $R(t)$, as the original objective (5.9) evaluated at these values:

$$R(t) \triangleq \frac{1+\frac{P_s|h_D|^2}{|\mathbf{w}_0(t)^\dagger \mathbf{g}_D|^2+\sigma_D^2}}{1+\frac{P_s|h_E|^2}{f_0(t)+\sigma_E^2}} = \frac{1+\frac{P_s|h_D|^2}{t+\sigma_D^2}}{1+\frac{P_s|h_E|^2}{f_0(t)+\sigma_E^2}}. \tag{5.16}$$

We can now focus on finding the maximum of $R(t)$ over $t \geq 0$.

However, $R(t)$ is still difficult to evaluate due to the equality constraint in (5.15). To overcome this difficulty, we consider the modified problem, where we replace the equality constraint in (5.15) with an inequality constraint:

$$\begin{aligned} &\max_{\mathbf{w}} |\mathbf{w}^\dagger \mathbf{g}_E|^2 \\ &\text{s.t.} \begin{cases} |\mathbf{w}^\dagger \mathbf{g}_D|^2 \leq t, \\ \quad |w_m|^2 \leq p_m \ \forall m. \end{cases} \end{aligned} \tag{5.17}$$

Now, let the optimal solution to the problem (5.17) be $\mathbf{w}(t)$ and the corresponding optimal objective value be $f(t)$, and let us define $R_1(t)$, as the original objective (5.9) evaluated at these values:

$$R_1(t) \triangleq \frac{1+\frac{P_s|h_D|^2}{t+\sigma_D^2}}{1+\frac{P_s|h_E|^2}{f(t)+\sigma_E^2}}. \tag{5.18}$$

We have the following useful result regarding $R(t)$ and $R_1(t)$.

Theorem 15. *$R_1(t)$ and $R(t)$ have the same maximizer and the same maximum function value.*

Proof. Let the maximizer of $R(t)$ be t^* and the associated beamforming vector solution to (5.15) is $\mathbf{w}_0(t^*)$. Now $\mathbf{w}_0(t^*)$ is also a feasible solution to (5.17), meaning that $f(t^*) \geq f_0(t^*)$ and thus we have,

$$\max_t R_1(t) \geq R_1(t^*) = \frac{1+\frac{P_s|h_D|^2}{t^*+\sigma_D^2}}{1+\frac{P_s|h_E|^2}{f(t^*)+\sigma_E^2}} \geq \frac{1+\frac{P_s|h_D|^2}{t^*+\sigma_D^2}}{1+\frac{P_s|h_E|^2}{f_0(t^*)+\sigma_E^2}} = \max_t R(t). \tag{5.19}$$

On the other hand, suppose that the maximizer of $R(t)$ is t_1 with the corresponding solution to (5.17) as $\mathbf{w}(t_1)$, then we must have $|\mathbf{w}(t_1)^\dagger \mathbf{g}_D|^2 = t_1$.

This can be seen by contradiction as follows. First, let us define $|\mathbf{w}(t_1)^\dagger \mathbf{g}_D|^2 \triangleq t_2$. Now we assume that $t_2 < t_1$, from which we have $f(t_1) = f(t_2)$. But from the

definition of R_1 in (5.18), we have $R_1(t_1) < R_1(t_2)$, which contradicts the assumption that t_1 is the maximizer. Thus, when the maximizer of $R_1(t)$ is t_1, $|\mathbf{w}(t_1)^\dagger \mathbf{g}_D|^2 = t_1$.

The above observation means that, at the optimal value of $R_1(t)$, the first inequality in (5.17) becomes an equality, and this optimal value can also be attained by $R(t)$. Combining this fact and (5.19) completes the proof. □

We can now make use of Theorem 15 to solve the optimization problem (5.17) instead of (5.15), since the solution to (5.17) will be the same as the solution to (5.15). We do this to circumvent the difficulty in solving the optimization problem (5.15), which is due to the equality constraint in (5.15).

To solve the optimization problem (5.17), we need to maximize $R_1(t)$. We first note that in (5.17), $\mathbf{w}^\dagger \mathbf{g}_E$ can be made positive (see [9] for a similar example) without loss of optimality. This step converts (5.17) into the convex optimization problem

$$\begin{aligned} &\max_{\mathbf{w}} \mathbf{w}^\dagger \mathbf{g}_E \\ &\text{s.t.} \quad \begin{cases} |\mathbf{w}^\dagger \mathbf{g}_D|^2 \le t, \\ \quad |w_m|^2 \le p_m \ \forall m, \end{cases} \end{aligned} \tag{5.20}$$

Then (5.20) can be solved optimally, with optimal solution $\mathbf{w}^*(t)$ for a given t and $f(t) \triangleq |\mathbf{w}^*(t)^\dagger \mathbf{g}_E|^2$. That is, $f(t)$ is the optimal value of $R_1(t)$. We now prove a property of $f(t)$ in the following theorem.

Theorem 16. *$f(t)$ is a concave function of t.*

Proof. The Lagrangian of (5.20) is given by

$$\begin{aligned} L(\mathbf{w}, \boldsymbol{\lambda}, \mu) &= -\mathbf{w}^\dagger \mathbf{g}_E + \mu(|\mathbf{w}^\dagger \mathbf{g}_D|^2 - t) + \sum_{m=1}^{M} \lambda_m (|w_m|^2 - p_m) \\ &= -\mathbf{w}^\dagger \mathbf{g}_E + \mathbf{w}^\dagger \left[\mu \mathbf{g}_D \mathbf{g}_D^\dagger + \mathsf{Diag}(\boldsymbol{\lambda})\right] \mathbf{w} - \left(\mu t + \sum_{m=1}^{M} \lambda_m p_m\right) \\ &= \mathbf{w}^\dagger \left(\mu \mathbf{g}_D \mathbf{g}_D^\dagger + \mathsf{Diag}(\boldsymbol{\lambda}) - \frac{\mathbf{g}_E \mathbf{g}_E^\dagger}{\mathbf{w}^\dagger \mathbf{g}_E}\right) \mathbf{w} - \left(\mu t + \sum_{m=1}^{M} \lambda_m p_m\right), \end{aligned} \tag{5.21}$$

where μ and $\boldsymbol{\lambda} \ge \mathbf{0}$ are Lagrange multipliers, and $\mathsf{Diag}(\boldsymbol{\lambda})$ is a square matrix with the diagonal elements being $\lambda_i, i = 1, \cdots, M$. The Lagrange dual objective is

$$G(\boldsymbol{\lambda}, \mu) = \min_{\mathbf{w}} L(\mathbf{w}, \boldsymbol{\lambda}, \mu). \tag{5.22}$$

At the optimal value of the primal variables $\mathbf{w}^*(t)$, the Lagrange dual objective takes on the value

$$G(\boldsymbol{\lambda}, \mu) = \mathbf{w}^*(t)^\dagger \left[\mu \mathbf{g}_D \mathbf{g}_D^\dagger + \mathsf{Diag}(\boldsymbol{\lambda}) - \frac{\mathbf{g}_E \mathbf{g}_E^\dagger}{\mathbf{w}^*(t)^\dagger \mathbf{g}_E} \right] \mathbf{w}^*(t) - \left(\mu t + \sum_{m=1}^{M} \lambda_m p_m \right). \tag{5.23}$$

To ensure that $G(\boldsymbol{\lambda}, \mu)$ is lower bounded, we must have the terms in the square brackets $\mu \mathbf{g}_D \mathbf{g}_D^\dagger + \mathsf{Diag}(\boldsymbol{\lambda}) - \frac{\mathbf{g}_E \mathbf{g}_E^\dagger}{\mathbf{w}^*(t)^\dagger \mathbf{g}_E} \succeq \mathbf{0}$. That is, positive semidefinite. Then, the Lagrange dual problem can be expressed as

$$\begin{aligned} &\min_{\boldsymbol{\lambda} \geq \mathbf{0}, \mu > 0} \mu t + \sum_{m=1}^{M} \lambda_m p_m \\ &\text{s.t. } \mu \mathbf{g}_D \mathbf{g}_D^\dagger + \mathsf{Diag}(\boldsymbol{\lambda}) - \frac{\mathbf{g}_E \mathbf{g}_E^\dagger}{\mathbf{w}^*(t)^\dagger \mathbf{g}_E} \succeq \mathbf{0}. \end{aligned} \tag{5.24}$$

This form of the dual problem is not the usual formulation, but it is useful in deriving the property of $f(t)$. Now the optimization problem (5.24) contains the unknown $\mathbf{w}^*(t)$ in the constraint, which makes (5.24) not solvable. But the solution to problem (5.24), if it exists, gives us the same objective value of the original dual problem (5.22) and thus of the primal problem as well. Thus we proceed to modify (5.24) so that it leads to a solution. We now have, modifying (5.24),

$$\begin{aligned} &\min_{\boldsymbol{\lambda} \geq \mathbf{0}, \mu > 0} \mu t + \sum_{m=1}^{M} \lambda_m p_m \\ &\text{s.t. } \mu \mathbf{g}_D \mathbf{g}_D^\dagger + \mathsf{Diag}(\boldsymbol{\lambda}) - \mathbf{g}_E \mathbf{g}_E^\dagger \succeq \mathbf{0}. \end{aligned} \tag{5.25}$$

We can easily see that the optimal solution pair $(\boldsymbol{\lambda}, \mu)$ to (5.25) is a scaled version (by a factor of $\mathbf{w}^*(t)^\dagger \mathbf{g}_E$) of the optimal solution pair to (5.24). By the convexity of (5.20), strong duality holds and the optimal objective value of (5.24) is $\mathbf{w}^*(t)^\dagger \mathbf{g}_E$, which is the same as (5.20). Multiplied by another $\mathbf{w}^*(t)^\dagger \mathbf{g}_E$, the optimal objective value of (5.25) becomes $|\mathbf{w}^*(t)^\dagger \mathbf{g}_E|^2$ which is exactly $f(t)$ as defined before. It is easily checked that $f(t)$ is a point-wise minimum of a family of affine functions and as a result concave for $t \geq 0$. This completes the proof. □

We have seen that $R_1(t)$ and $R(t)$ have the same maximizer and the same maximum value, and that $f(t)$, which is the optimal value of the optimization problem (5.17) and hence $R_1(t)$, is concave in t. We next examine the analytical properties of $R_1(t)$ to help us design a search method for the optimal solution.

5.4.2 Search for the Optimal Solution

In this section, we study the analytical properties of $R_1(t)$ which will permit us to design an efficient algorithm to find the optimal solution. Our result is summarized in the following theorem.

Theorem 17. *$R_1(t)$ is quasi-concave in t and its maximum can be found via a one-dimensional search.*

Proof. The proof follows from the result in [14] that if $R_1'(t) = 0$ implies that $R_1''(t) < 0$ for any $t \geq 0$, then $R_1(t)$ is quasi-concave in $t \geq 0$. This follows from the properties of quasi-convex functions given in Section 2.2.1. This result was also used in [113].

For convenience, we define $a \triangleq P_s|h_D|^2, b \triangleq P_s|h_E|^2, c \triangleq \sigma_D^2, d \triangleq \sigma_E^2$. Then, we have

$$R_1(t) = \frac{1 + \frac{a}{t+c}}{1 + \frac{b}{f(t)+d}}. \tag{5.26}$$

The first-order derivative of $R_1(t)$ is

$$\begin{aligned} R_1'(t) = &-a\left(t+c\right)^{-2}\left(1+\frac{b}{f(t)+d}\right)^{-1} \\ &+\left(1+\frac{a}{t+c}\right)bf'(t)\left(1+\frac{b}{f(t)+d}\right)^{-2}\left(f(t)+d\right)^{-2}. \end{aligned} \tag{5.27}$$

Setting it $R_1'(t) = 0$, we have

$$a\left(t+c\right)^{-2}\left(1+\frac{b}{f(t)+d}\right)\left(f(t)+d\right)^2 = b\left(1+\frac{a}{t+c}\right)f'(t). \tag{5.28}$$

Now, the second-order derivative of $R_1(t)$ multiplied by $\left(1+\frac{b}{f(t)+d}\right)^3\left(f(t)+d\right)^4$ gives

$$\begin{aligned} &R_1''(t)\left(1+\frac{b}{f(t)+d}\right)^3\left(f(t)+d\right)^4 \\ &= 2a\left(t+c\right)^{-3}\left(1+\frac{b}{f(t)+d}\right)^2\left(f(t)+d\right)^4 \\ &\quad - 2abf'(t)\left(t+c\right)^{-2}\left(1+\frac{b}{f(t)+d}\right)\left(f(t)+d\right)^2 \\ &\quad + 2b^2\left(1+\frac{a}{t+c}\right)\left(f'(t)\right)^2 - 2b\left(1+\frac{a}{t+c}\right)\left(1+\frac{b}{f(t)+d}\right)\left(f(t)+d\right)\left(f'(t)\right)^2 \\ &\quad + \left(1+\frac{a}{t+c}\right)b\left(1+\frac{b}{f(t)+d}\right)\left(f(t)+d\right)^2 f''(t). \end{aligned} \tag{5.29}$$

Combining the third and fourth terms on the RHS of (5.29), we have

$$\begin{aligned}
&R_1''(t)\left(1+\frac{b}{f(t)+d}\right)^3(f(t)+d)^4 \\
&= 2a\,(t+c)^{-3}\left(1+\frac{b}{f(t)+d}\right)^2(f(t)+d)^4 \\
&\quad - 2abf'(t)\,(t+c)^{-2}\left(1+\frac{b}{f(t)+d}\right)(f(t)+d)^2 \\
&\quad - 2b(f'(t))^2\left(1+\frac{a}{t+c}\right)(f(t)+d) \\
&\quad + \left(1+\frac{a}{t+c}\right)b\left(1+\frac{b}{f(t)+d}\right)(f(t)+d)^2\,f''(t). \qquad (5.30)
\end{aligned}$$

Substituting (5.28) into the second term on the RHS of (5.30) and combining with the third term on the RHS of (5.30), we have

$$\begin{aligned}
&R_1''(t)\left(1+\frac{b}{f(t)+d}\right)^3(f(t)+d)^4 \\
&= 2a\,(t+c)^{-3}\left(1+\frac{b}{f(t)+d}\right)^2(f(t)+d)^4 - 2b(f'(t))^2\left(1+\frac{a}{t+c}\right)(f(t)+d+b) \\
&\quad + \left(1+\frac{a}{t+c}\right)b\left(1+\frac{b}{f(t)+d}\right)(f(t)+d)^2\,f''(t). \qquad (5.31)
\end{aligned}$$

Since $f(t)$ is concave, the last term on the RHS of (5.31) is < 0, so that we have

$$\begin{aligned}
&R_1''(t)\left(1+\frac{b}{f(t)+d}\right)^3(f(t)+d)^4 \\
&< 2a\,(t+c)^{-3}\left(1+\frac{b}{f(t)+d}\right)^2(f(t)+d)^4 - 2b(f'(t))^2\left(1+\frac{a}{t+c}\right)(f(t)+d+b). \qquad (5.32)
\end{aligned}$$

Using the square of (5.28) in the first term of the RHS of (5.32), the first term on the RHS of (5.32) becomes

$$2\frac{b^2}{a}(t+c)\left(1+\frac{a}{t+c}\right)^2(f'(t))^2. \qquad (5.33)$$

Substituting back into (5.32), we have

$$\begin{aligned}
&R_1''(t)\left(1+\frac{b}{f(t)+d}\right)^3(f(t)+d)^4 \\
&< 2\frac{b^2}{a}(t+c)\left(1+\frac{a}{t+c}\right)^2(f'(t))^2\left(1-\frac{2b(f'(t))^2\left(1+\frac{a}{t+c}\right)(f(t)+d+b)}{2\frac{b^2}{a}(t+c)\left(1+\frac{a}{t+c}\right)^2(f'(t))^2}\right). \qquad (5.34)
\end{aligned}$$

The ratio in (5.34) may be expressed as

$$\frac{a(f(t)+d+b)}{b(t+c)\left(1+\frac{a}{t+c}\right)} = \frac{a(f(t)+d+b)}{b(a+t+c)} = \frac{a(1+\frac{f(t)+d}{b})}{(a+t+c)} > \frac{a(1+\frac{t+c}{a})}{(a+t+c)} = 1, \quad (5.35)$$

where we have used the assumption that the secrecy rate is positive, that is, $\frac{t+c}{a} < \frac{f(t)+d}{b}$. By substituting (5.35) back into (5.34), we can see that $R_1''(t) < 0$. Thus $R_1(t)$ is a quasi-concave function and its maximum can be efficiently found by a one-dimensional search [14, p101]. □

We can now give the complete algorithm, summarized in Algorithm 1.

5.5 Generalizations of the Method

In this section we shall discuss some possible extensions of our method, to more generalized CJ models.

5.5.1 Generalization to Grouped Relays' Power Constraints

Suppose there are L groups of relays $\{\mathcal{N}_1, \ldots, \mathcal{N}_l, \ldots, \mathcal{N}_L\}$, which form a partition of $\{1, 2, \ldots, M\}$. It is natural to consider grouped relays power constraints, which can be viewed as a generalization of individual relay power constraints, and can also reflect the scenario where each group represents a relay with multiple antennas. In this case, the optimization problem is given by

$$\begin{aligned} &\max_{\mathbf{w}} \mathbf{w}^\dagger \mathbf{g}_E \\ &\text{s.t.} \begin{cases} |\mathbf{w}^\dagger \mathbf{g}_D|^2 \le t, \\ \sum_{m\in\mathcal{N}_l} |w_m|^2 \le p_l \quad \forall l. \end{cases} \end{aligned} \quad (5.36)$$

The only difference to our analysis is that (5.25) becomes

$$\begin{aligned} &\min_{\boldsymbol{\lambda}\succeq\mathbf{0},\mu>0} \mu t + \sum_{l=1}^{L} \lambda_l p_l \\ &\text{s.t. } \mu \mathbf{g}_D \mathbf{g}_D^\dagger + \mathsf{Diag}(\boldsymbol{\lambda}) - \mathbf{g}_E \mathbf{g}_E^\dagger \succeq \mathbf{0}, \end{aligned} \quad (5.37)$$

where $\boldsymbol{\lambda} \triangleq \mathsf{Diag}(\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_L, \ldots, \lambda_L)$. It can be seen that all analysis is still valid, so the proposed algorithm can be easily generalized to handle the grouped relays constraints.

Algorithm 1 Proposed Algorithm

Input: P_s, h_D, h_E, $\mathbf{g}_D$, $\mathbf{g}_E$, σ_D^2 and σ_E^2.
begin
  Use the conditions in Section 5.3 to check whether a positive secrecy rate is possible.
  if a positive secrecy rate is not possible, **then**
    secrecy rate is zero, exit.
  end
  Initialize $t_{\min}$ and $t_{\max}$.
  While $t_{\max} - t_{\min} \geq \epsilon$ where ϵ is a preset threshold,
    $t = \frac{t_{\min}+t_{\max}}{2}$.
    Solve problem (5.17) with the above t and get solution $\mathbf{w}$.
    Solve problem (5.17) with $t + \Delta t$ for very small $\Delta t > 0$.
    Evaluate $R_1'(t)$ using the above two solutions and (5.18).
    if $R_1'(t) > 0$,
      $t_{\min} = t$.
    else
      $t_{\max} = t$.
    end
  end
end
Output: $\mathbf{w}$.

5.5.2 Distributed Implementation

The overall optimization can be performed at either the source S or the destination D, which needs all the necessary system parameters. In practice, S or D may learn $\mathbf{g}_E, \sigma_E^2$ and $\mathbf{g}_D, \sigma_D^2$ from the relays $\{\mathtt{R}_m\}$. After the optimal relay weights $\mathbf{w}_{\mathsf{opt}}$ are obtained, S or D informs each individual relay about its own amplification coefficient, w_m. However, it will require some bandwidth to perform this task. Thus, it would be beneficial if each individual relay can derive its own beamforming weight based on its local CSI. Here, we shall discuss such a distributed implementation algorithm, with the assumption that $\mathtt{R}_m$ knows its local CSI $g_m^{(\mathtt{E})}$ and $g_m^{(\mathtt{D})}$ perfectly.

To facilitate the design, we recall that the dual problem is

$$\min_{\boldsymbol{\lambda}\succeq\mathbf{0},\mu>0} \mu t + \sum_{m=1}^{M} \lambda_m p_m$$
$$\text{s.t. } \mu\mathbf{g}_D\mathbf{g}_D^\dagger + \mathsf{Diag}(\boldsymbol{\lambda}) - \mathbf{g}_E\mathbf{g}_E^\dagger \succeq \mathbf{0}. \tag{5.38}$$

Using one of the conditions for obtaining the optimal value in (5.38), we have

$$\left(\mu\mathbf{g}_D\mathbf{g}_D^\dagger + \mathsf{Diag}(\boldsymbol{\lambda}) - \mathbf{g}_E\mathbf{g}_E^\dagger\right)\mathbf{w} = \mathbf{0}, \tag{5.39}$$

from which we have

$$\mathsf{Diag}(\boldsymbol{\lambda})\mathbf{w} = (\mathbf{g}_E^\dagger\mathbf{w})\mathbf{g}_E - \mu(\mathbf{g}_D^\dagger\mathbf{w})\mathbf{g}_D. \tag{5.40}$$

Since $\mathbf{g}_E^\dagger\mathbf{w} > 0$, we now have

$$\frac{\mathsf{Diag}(\boldsymbol{\lambda})}{(\mathbf{g}_E^\dagger\mathbf{w})}\mathbf{w} = \mathbf{g}_E - \frac{\mu(\mathbf{g}_D^\dagger\mathbf{w})}{(\mathbf{g}_E^\dagger\mathbf{w})}\mathbf{g}_D. \tag{5.41}$$

Therefore, if the coefficient $\lambda_m > 0$, we have

$$w_m = \sqrt{p_m}\frac{g_m^{(\mathtt{E})} - \frac{\mu(\mathbf{g}_D^\dagger\mathbf{w})}{(\mathbf{g}_E^\dagger\mathbf{w})}g_m^{(\mathtt{D})}}{\left|g_m^{(\mathtt{E})} - \frac{\mu(\mathbf{g}_D^\dagger\mathbf{w})}{(\mathbf{g}_E^\dagger\mathbf{w})}g_m^{(\mathtt{D})}\right|}. \tag{5.42}$$

From the complementary slackness conditions of (5.20), we see that $\lambda_m(|w_m|^2 - p_m) = 0$ and $\lambda_m > 0$ implies that $|w_m|^2 = p_m$. So $\mathtt{R}_m$ should use its full transmit power p_m.

From (5.42), we see that for $\lambda_m > 0$, each relay $\mathtt{R}_m$ can learn its own weight based on local CSI $g_m^{(\mathtt{E})}$ and $g_m^{(\mathtt{D})}$ while $\mathtt{S}$ or $\mathtt{D}$ broadcasts the common scalar $\frac{\mu\mathbf{g}_D^\dagger\mathbf{w}}{\mathbf{g}_E^\dagger\mathbf{w}}$ to all relays. In the ideal case where $\lambda_m > 0$ for $m = 1, \cdots, M$, only one positive scalar $\frac{\mu\mathbf{g}_D^\dagger\mathbf{w}}{\mathbf{g}_E^\dagger\mathbf{w}}$ needs to be broadcast, while the relays can find their own optimal weights. However, when $\lambda_m = 0$, the individual relays get no information about their own beamforming weight from the common information, and in this case, (5.42) may not be the optimal solution. However, our simulation results in Section 5.6 actually show that the distributed algorithm can achieve a secrecy rate very close to the optimal one.

5.6 Simulation Results

Computer simulations are performed to evaluate the achievable secrecy rate of the proposed algorithm. For the simulations, we assumed a one-dimensional system model, and place the source, relays, destination and eavesdropper along a line. Furthermore,

the distance between relays is assumed to be short compared to the source-relay, relay-destination, relay-eavesdropper distances. Channels are modeled by a line-of-sight channel model including the path loss, to take into account the effects of distances. Therefore, the channels can be expressed in general as $h = d^{-\frac{c}{2}} e^{j\theta}$, where d is the distance, and c is the path loss exponent chosen as 3.5, and the random phase θ is uniformly distributed over $[0, 2\pi)$. Also, the path loss for nodes to and from the relays can be assumed to be the same since the distance between relays is small. The source-eavesdropper distance is fixed at 50 m. The noise power is 100 dBm while the source power and individual relay power constraints are chosen to be 40 dBm.

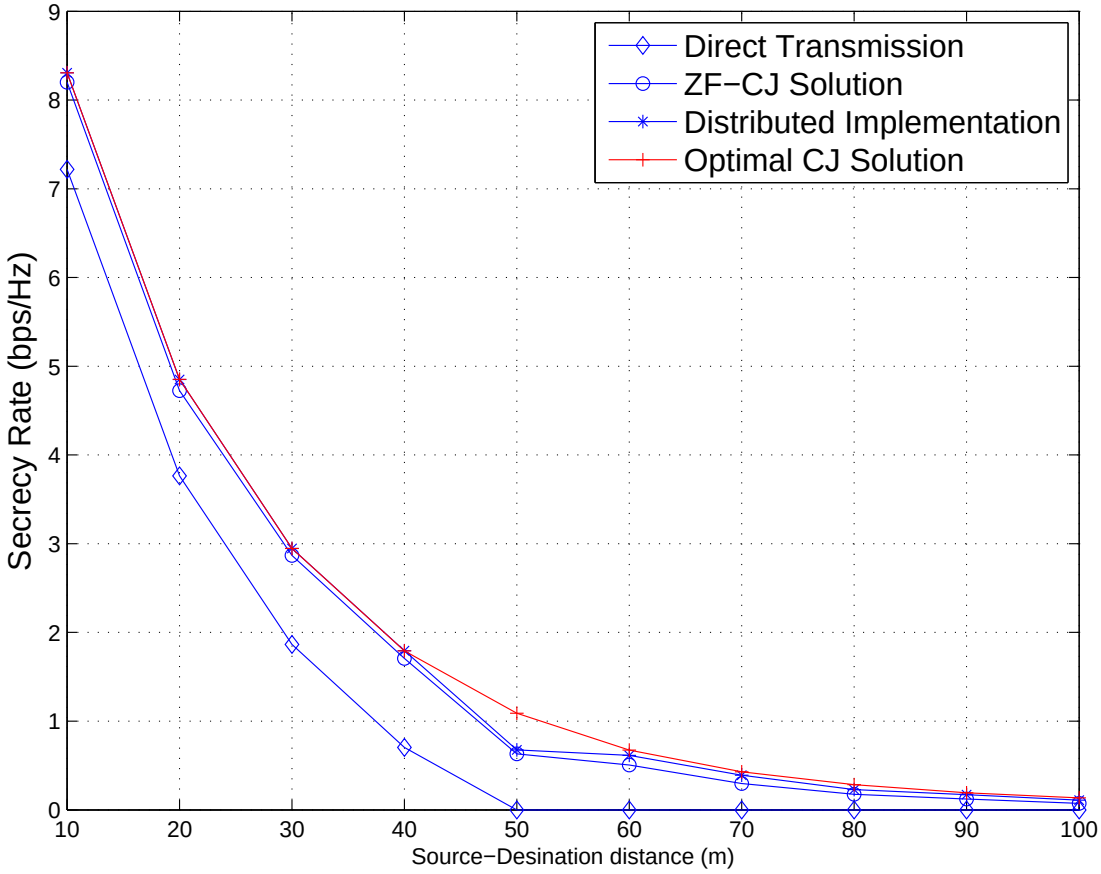


Figure 5.3: Secrecy rate versus source-destination distance.

We first let the source-relays distance be fixed at 25 m. We change the position of the destination so that the source-destination distance varies from 10 m to 100 m. Figure 5.3 shows the secrecy rate comparison between the proposed algorithms and the direct transmission (without the aid of relays). We use 10 relays in the simulations,

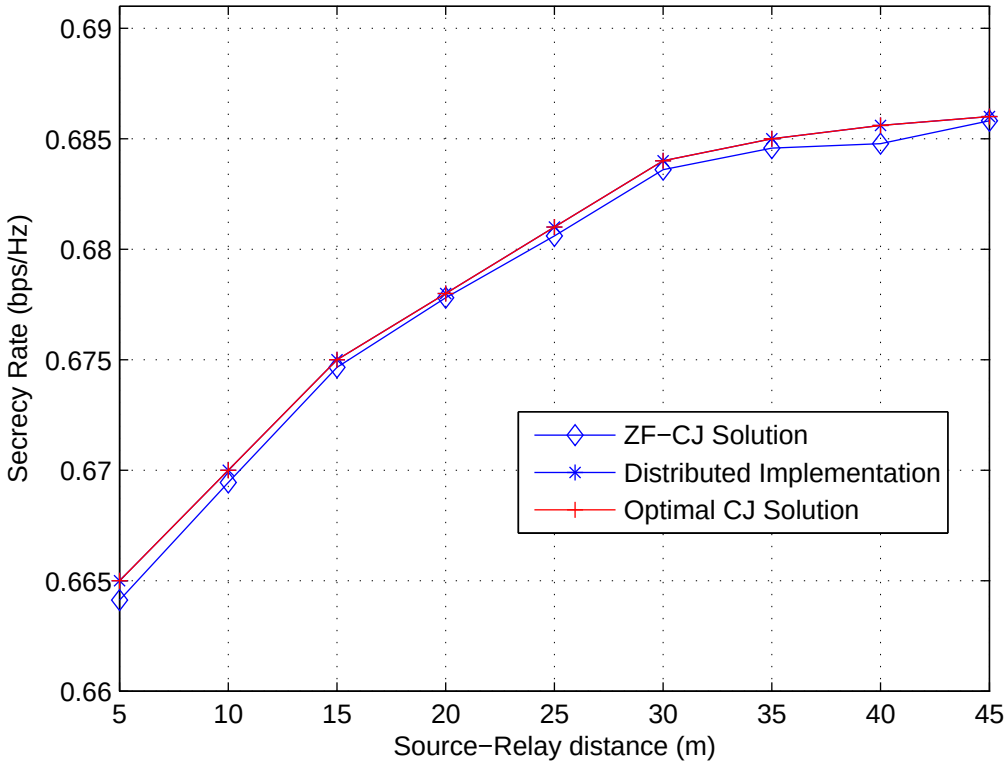


Figure 5.4: a) Secrecy rate versus source-relay distance.

i.e., $M = 10$. The zero-forcing (ZF) solution (labeled as ZF-CJ in the figure) corresponds to forcing the jamming signal at the destination to zero (i.e., $t = 0$ in (5.15)). Note that it is different from that in [31] where only total power constraint was considered. We can see that the secrecy rate is significantly increased using our optimal CJ solution, as compared to the secrecy rate using only direct transmission. We also observe that by using CJ, there is a positive secrecy rate even when the channel for the direct transmission from source to destination is not favorable, when compared to the source to eavesdropper channel. This happens when the source-destination distance is 50 m onwards, corresponding to the eavesdropper masking the direct transmission. The results also illustrate that ZF-CJ can achieve nearly optimal performance in most cases, but when the eavesdropper is close to the destination (for example, when the source-destination distance is around 50 m), a significant gap is observed. This can be explained by the fact that the ZF-CJ solution fails to manage the interference properly at the destination while the optimal solution is able to use a more intelligent strategy

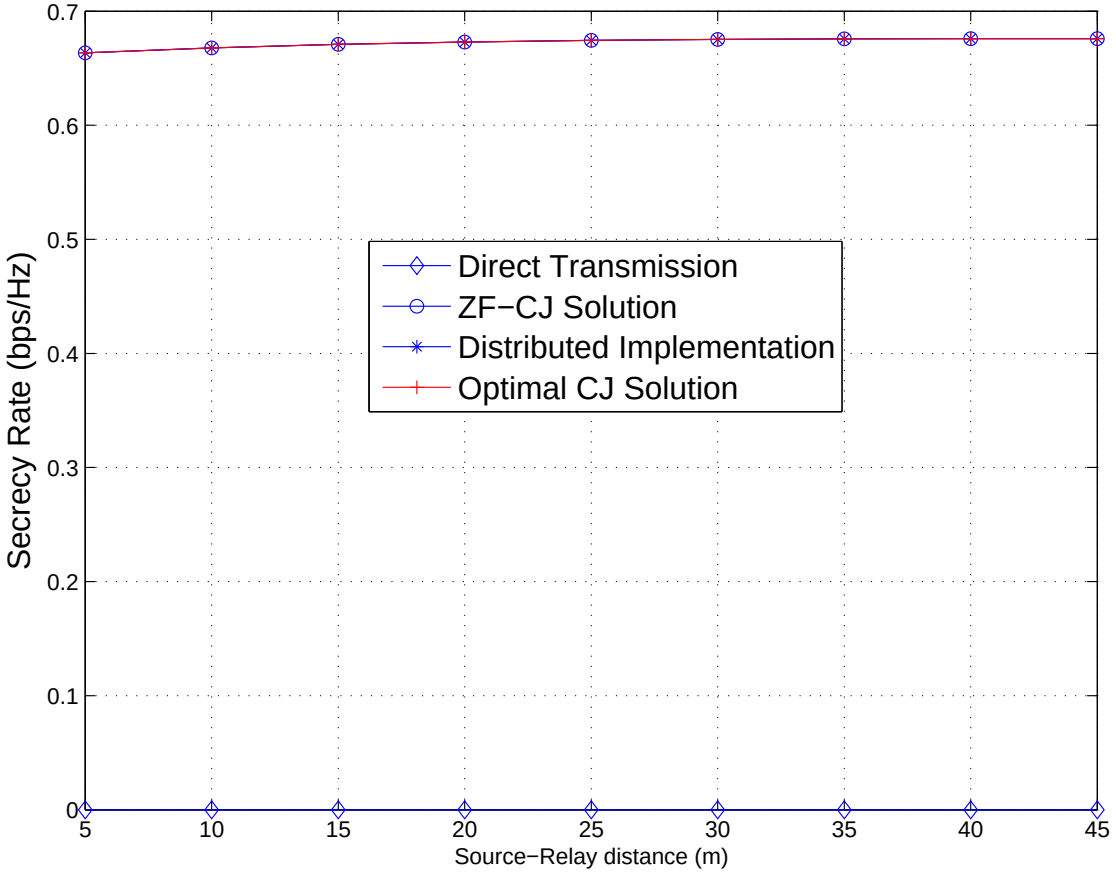


Figure 5.4: b) Secrecy rate versus source-relay distance.

to balance the interference to the eavesdropper and the destination. It can also be seen that the secrecy rate achieved by the distributed algorithm is very close to the optimal rate and this justifies the proposed distributed implementation.

We then fix the source-destination distance at 60 m. The position of the relays is changed so that the source-relays distance varies from 5 m to 45 m. Figures 5.4a and 5.4b show a similar secrecy rate comparison as has already been observed in Figure 5.3 with regard to the source-relays distance. In Figure 5.4a, we see that the optimal solution has the best performance, while the distributed solution is close to optimal, and the ZF-CJ solution is slightly worse than the above two. All the solutions are fairly close to each other. It is observed that the secrecy rate increases as the relays are closer to the eavesdropper. However, the performance gain is not significant, which can be explained by the fact that as the relays are closer to the eavesdropper, they also generate more interference to the destination.

5.7 Conclusions and Discussion

We have studied the CJ protocol via distributed relays to increase the physical layer security. The conditions for positive secrecy rate have been derived and we have shown that the optimal CJ solution can be obtained by a combination of convex optimization and a one-dimensional search. Extension to grouped relays power constraints and a distributed implementation have also been considered. Simulation results have demonstrated the fundamental result that we can enhance security over direct transmission, when channel conditions are favorable to the eavesdropper.

There are two interesting open problems. Suppose there are N eavesdroppers, then the secrecy rate maximization problem becomes

$$\max_{\mathbf{w}} \min_{n} \left\{ \log\left(1+\frac{P_S|h_D|^2}{|\mathbf{w}^\dagger \mathbf{g}_D|^2+\sigma_D^2}\right) - \log\left(1+\frac{P_S|h_{E,n}|^2}{|\mathbf{w}^\dagger \mathbf{g}_{E,n}|^2+\sigma_E^2}\right)\right\}, \quad \text{s.t.}|w_m|^2 \le p_m, \quad \forall m, \tag{5.43}$$

which reminds us of the compound wiretap channel earlier in Section 3.2.2. Hence we may be able to decompose this into N subproblems

$$\max_{\mathbf{w}} \left\{ \log\left(1+\frac{P_S|h_D|^2}{|\mathbf{w}^\dagger \mathbf{g}_D|^2+\sigma_D^2}\right) - \log\left(1+\frac{P_S|h_{E,n}|^2}{|\mathbf{w}^\dagger \mathbf{g}_{E,n}|^2+\sigma_E^2}\right)\right\}, \quad \text{s.t.}|w_m|^2 \le p_m, \quad \forall m, \tag{5.44}$$

for which we can attempt to find the optimal solution; an issue to address would be how to handle the correlations between relays, as the relays will be performing CJ over several eavesdroppers' channels.

We can also consider the relay assignment problem. To illustrate, consider initially a fixed relay assignment and a 2-stage DF protocol with M relays as follows: in stage 1, the source transmits on the direct link to D while M_1 relays send a cooperative jamming signal and $M-M_1$ relays receive the source signal; in stage 2, the source does not transmit while the $M-M_1$ relays transmit to D and the M_1 relays send a cooperative jamming signal. The received signals at D and E are

$$\text{Stage1}: \; y_{D,1} = h_D x_s + \sum_{m=1}^{M_1} g_{D,m} w_m x_c + n_D, \;\; y_{E,1} = h_E x_s + \sum_{m=1}^{M_1} g_{E,m} w_m x_c + n_E,$$

$$\text{Stage2}: \; y_{D,2} = \sum_{m=M_1+1}^{M-M_1} g_{D,m} w_m \frac{x_s}{P_S} + \sum_{m=1}^{M_1} g_{D,m} w_m x_c + n_D,$$

$$y_{E,2} = \sum_{m=M_1+1}^{M-M_1} g_{E,m} w_m \frac{x_s}{P_S} + \sum_{m=1}^{M_1} g_{E,m} w_m x_c + n_E. \tag{5.45}$$

Then the SNR's at D and E are

$$\begin{aligned}
\Gamma_D &= \frac{P_S |h_D|^2}{\sum_{m=1}^{M_1} |g_{D,m} w_m|^2 + \sigma_D^2} + \frac{\sum_{m=M_1+1}^{M-M_1} |g_{D,m} w_m|^2}{\sum_{m=1}^{M_1} |g_{D,m} w_m|^2 + \sigma_D^2}, \\
\Gamma_E &= \frac{P_S |h_E|^2}{\sum_{m=1}^{M_1} |g_{E,m} w_m|^2 + \sigma_E^2} + \frac{\sum_{m=M_1+1}^{M-M_1} |g_{E,m} w_m|^2}{\sum_{m=1}^{M_1} |g_{E,m} w_m|^2 + \sigma_E^2}.
\end{aligned} \tag{5.46}$$

Define $\mathbf{g}_{D,c} = [g_{D,1}, \cdots, g_{D,M_1}]$, $\mathbf{g}_{E,c} = [g_{E,1}, \cdots, g_{E,M_1}]$, $\mathbf{g}_{D,r} = [g_{D,M_1+1}, \cdots, g_{D,M}]$ and $\mathbf{g}_{E,r} = [g_{E,M_1+1}, \cdots, g_{E,M}]$. Then the optimization problem may be written as

$$\begin{aligned}
\max_{\mathbf{w}} \quad & \frac{1}{2} \log \left(1 + \frac{P_S |h_D|^2}{|\mathbf{w}^\dagger \mathbf{g}_{D,c}|^2 + \sigma_D^2} + \frac{|\mathbf{w}^\dagger \mathbf{g}_{D,r}|^2}{|\mathbf{w}^\dagger \mathbf{g}_{D,c}|^2 + \sigma_D^2} \right) \\
& - \frac{1}{2} \log \left(1 + \frac{P_S |h_E|^2}{|\mathbf{w}^\dagger \mathbf{g}_{E,c}|^2 + \sigma_E^2} + \frac{|\mathbf{w}^\dagger \mathbf{g}_{E,r}|^2}{|\mathbf{w}^\dagger \mathbf{g}_{E,c}|^2 + \sigma_E^2} \right).
\end{aligned} \tag{5.47}$$

We can see that this model will allow us to consider dynamic relay assignments. Note that if we consider no CJ and $M - M_1$ relays performing DF, we can easily obtain the optimization problem for DF with $\mathbf{g}_{D,r}, \mathbf{g}_{E,r}$. It is interesting to find the solution to this problem. We could also consider variants where AF is used instead of DF, and various combinations of jamming and relaying for S, D and the relays.

Chapter 6

Lattice Coding for the Gaussian Wiretap Channel

In this chapter, we study structured codes and their construction for the physical layer security problem, focusing on the wiretap channel as it is fundamental and the basic building block in physical layer security. We recall that the coding scheme for the wiretap channel uses coset coding. Here our work, reported in [23], adopts an information-theoretic approach to the lattice-based coset coding problem for the Gaussian wiretap channel.

We begin with some needed lattice definitions, then introduce some useful concepts concerning lattice coding for the Gaussian channel, before presenting our research, where we derive achievable channel rates, equivocation rate, and error probabilities for a nested lattice code. We conclude with a discussion on possible open problems for future research. Lastly, all logarithms in this chapter are to base e (natural logarithms), and we will follow the convention of naming the channel from the transmitter to the legitimate receiver as the *main* channel, and the channel from the transmitter to eavesdropper as the *eavesdropper's* channel.

6.1 Introduction

6.1.1 Channel Model

The Gaussian wiretap channel, studied in [70], has the following input-output relationship for n channel uses:

$$\mathbf{Y} = \mathbf{X} + \mathbf{N} \text{ and } \mathbf{Z} = \mathbf{X} + \mathbf{N}_z, \tag{6.1}$$

where $\mathbf{X}$ denotes the channel input, $\mathbf{Y}$ denotes the legitimate receiver's received signal and $\mathbf{Z}$ denotes the eavesdropper's received signal, $\mathbf{N} \sim \mathcal{N}(0, P_N\mathbf{I}_n)$, $\mathbf{N}_z \sim \mathcal{N}(0, P_{N_z}\mathbf{I}_n)$, with $P_N < P_{N_z}$, the noise is independent over the channel uses, and the channel input is subject to the power constraint $\frac{1}{n}\sum_{i=1}^{n} X_i \leq P_X$. The secrecy rate for this channel is

$$R_e = \tfrac{1}{2}\log(1+\mathrm{SNR}) - \tfrac{1}{2}\log(1+\mathrm{SNR}_z) \triangleq C - C_z, \tag{6.2}$$

where $\mathrm{SNR} = P_X/P_N$, $\mathrm{SNR}_z = P_X/P_{N_z}$, C and C_z are the capacities of the main and eavesdropper's channels, respectively.

The coding scheme to achieve the above secrecy rate is coset coding. We review some related work on the wiretap channel using coset coding and lattice coding in the wiretap scenario in the next section below.

6.1.2 Related Work

We make the distinction that practical codes offer explicit constructions, while structured codes are constructions using a information theoretic point of view, with good properties usually as code dimensions are large. We also note that the wiretap channel type II with BEC or BSC eavesdropper's channel will be called type II-BEC or type II-BSC; if both channels are of the same type, for example BEC or Gaussian, we will call it the BEC or Gaussian wiretap channel.

Coset coding is a form of binning, and we know from Zamir *et al* [120] that nested lattice codes can be used to implement binning.

Practical codes for the wiretap channel with coset encoding have been proposed using LDPC codes by Thangaraj *et al* [110] for the type II-BEC, type II-BSC and combinations of BEC and BSC main and eavesdropper's channels. The more difficult case is the one where the main and eavesdropper's channels are BECs; subsequent work by Rathi *et al* [103], Subramaniam *et al* [107] and Suresh *et al* [108] all further study the BEC wiretap channel, using variations on LDPC codes.

Explicit polar codes were proposed and constructed by Hof and Shamai [60] and Mahdavifar and Vardy [83] for the binary input symmetric channel (BSC wiretap channel).

For the Gaussian wiretap channel, Liu *et al* [76, 80] initially proposed using LDPC codes for the type II Gaussian wiretap channel (noiseless main channel and Gaussian

eavesdropper's channel). Subsequently Klinc *et al* [64] used punctured LDPC codes for the Gaussian wiretap channel, transmitting the message bits over the punctured positions, so forcing the eavesdropper to operate at a bit error rate of > 0.49 and obtaining no information about the message. Then, explicit lattice code constructions were proposed in the work of Oggier, Belfiore and Solé [7, 8, 97] for the Gaussian wiretap channel; here lattice constructions with appropriate parameters where derived for driving the error probability at the eavesdropper to 1.

Lattice codes have been proposed using an information theoretic (non-explicit) point of view in providing security for the Gaussian interference channels in the work of He and Yener [56, 57, 58, 59], and Agrawal and Vishwanath [3].

In this work, we take an information theoretic approach to the lattice-based coset coding problem for the Gaussian wiretap channel. We derive achievable channel rates, equivocation rate, and error probabilities for a nested lattice code. We note that our work is very similar in principle to [7, 8, 97]; however these works used decoding bit error probability as their criteria for secrecy and derived conditions for lattices to meet it, while our focus is on the equivocation rate and capacities. We also note that our work is different from [56, 57, 58, 59] and [3], as these papers consider a jamming signal at the eavesdropper, but the jamming causes no interference at the legitimate receiver. The authors of [3, 58, 56, 57, 59] also did not explicitly construct coarse and fine codes like all the other work using LDPC, polar or lattice codes, including our own.

We also mention that there has also been some work done on concatenating an error correcting code with a wiretap code, for binary channels, which has been reported in Cassuto and Bandic [16].

Finally, we should note that nested lattice constructions were used to provide watermarking security (see [100, 118] and the references within). Also, nested lattice constructions have been used to show achievable rates in Gaussian relay networks [93, 94, 114].

6.2 Lattice Preliminaries

In this section, we introduce notation and definitions for lattices. An extensive treatment for lattices can be found in Zamir [121] and the reference by Conway and Sloane [24].

6.2.1 Lattice Definitions

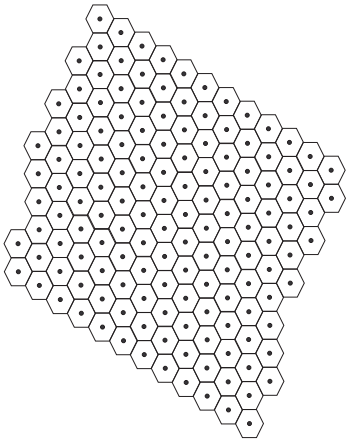


Figure 6.1: Section of a lattice in $\mathbb{R}^2$ with hexagonal Voronoi region.

A lattice Λ is a discrete subgroup of the Euclidean space $\mathbb{R}^n$ with ordinary vector addition. The lattice Λ can be specified in terms of a $n \times n$ real-valued generator matrix $\mathbf{G} = [\mathbf{g}_1|\mathbf{g}_2|\cdots|\mathbf{g}_n]$, for $\mathbf{g}_1, \mathbf{g}_2, \cdots, \mathbf{g}_n \in \mathbb{R}^n$ by

$$\Lambda = \{\boldsymbol{\lambda} = \mathbf{G}\mathbf{x} : \mathbf{x} \in \mathbb{Z}^n\}. \tag{6.3}$$

Alternatively a lattice is generated by taking all integer linear combinations of the basis vectors.

Since a lattice is a subgroup of $\mathbb{R}^n$ under vector addition, we have some useful group properties. A coset of Λ is a translated version of itself. Thus the set $\mathbf{x} + \Lambda$ is a coset of Λ for any $\mathbf{x} \in \mathbb{R}^n$. Let Ω be any fundamental region of Λ. Every $\mathbf{x} \in \mathbb{R}^n$ can be uniquely written as

$$\mathbf{x} = \boldsymbol{\lambda} + \mathbf{e}, \quad \boldsymbol{\lambda} \in \Lambda, \mathbf{e} \in \Omega, \ \ \mathbb{R}^n = \Lambda + \Omega. \tag{6.4}$$

The fundamental region is a building block which is repeated many times to fill the whole lattice space with one lattice point in each block. The fundamental Voronoi region $\mathcal{V}$, with volume V, is the set of minimum Euclidean norm coset representatives of Λ. That is, $\mathcal{V}$ is the set of points in $\mathbb{R}^n$ closest to the zero vector. The volume is defined as the inverse density of the lattice points in space. Then every $\mathbf{x} \in \mathbb{R}^n$ can be uniquely written as

$$\mathbf{x} = \boldsymbol{\lambda} + \mathbf{r}, \quad \boldsymbol{\lambda} \in \Lambda, \mathbf{r} \in \mathcal{V}, \ \ \mathbb{R}^n = \Lambda + \mathcal{V} = \bigcup_{\mathbf{x} \in \mathcal{V}} (\Lambda + \mathbf{x}). \tag{6.5}$$

A section of a lattice $\in \mathbb{R}^2$ is shown in Figure 6.1. The fundamental region of this lattice, each containing a lattice point depicted as a black dot, is marked out as hexagons. We see that the hexagons tile $\mathbb{R}^2$. For an extensive treatment on the group properties of a lattice and coset codes in general, we can refer to the papers by Forney [42],[43], and Forney and Wei [44], [45].

From this point on, we will focus our attention on the fundamental region $\mathcal{V}$ instead of the more general Ω. The quantizer associated with $\mathcal{V}$ is a map that sends a point $\mathbf{x}$ to the nearest lattice point

$$Q_{\mathcal{V}}(\mathbf{x}) = \boldsymbol{\lambda}, \quad \text{if } \mathbf{x} \in \boldsymbol{\lambda} + \mathcal{V}. \tag{6.6}$$

The nearest neighbour quantizer is also stated as the map that sends $\mathbf{x}$ to the nearest lattice point in Euclidean distance

$$Q_{\mathcal{V}}(\mathbf{x}) = \arg\min_{\boldsymbol{\lambda}\in\Lambda} \|\mathbf{x} - \boldsymbol{\lambda}\|, \tag{6.7}$$

where $\|.\|$ denotes the Euclidean norm. The modulo-Λ operation associated with $\mathcal{V}$ is

$$\mathbf{x} \bmod_{\mathcal{V}} \Lambda \triangleq \mathbf{x} \bmod \Lambda = \mathbf{x} - Q_{\mathcal{V}}(\mathbf{x}) \in \mathcal{V}, \ \forall \mathbf{x} \in \mathbb{R}^n. \tag{6.8}$$

The modulo-Λ operation satisfies

$$[\mathbf{x} + \mathbf{y}] \bmod \Lambda = [[\mathbf{x}] \bmod \Lambda + \mathbf{y}\,] \bmod \Lambda \tag{6.9}$$

$$Q_{\mathcal{V}}(\mathbf{x}) \bmod \Lambda = [Q_{\mathcal{V}}(\mathbf{x} \bmod \Lambda)] \bmod \Lambda. \tag{6.10}$$

Let $\mathrm{Ball}(r)$ denote an n-dimensional ball with radius r, with volume $\mathrm{Vol}(\mathrm{Ball}(r))$:

$$\mathrm{Ball}(r) \triangleq \{\mathrm{x} : \ \|\mathrm{x}\| \leq r, \ \mathrm{x} \in \mathbb{R}^n\}. \tag{6.11}$$

The covering radius R_u of a lattice is the smallest real number so that $\mathbb{R}^n \subseteq \Lambda + \mathrm{Ball}\,(R_u)$. The effective radius R_l of a lattice is the real number that satisfies $\mathrm{Vol}(\mathrm{Ball}(r)) = V$, where V is the fundamental volume of the lattice.

The second moment per dimension of Λ associated with $\mathcal{V}$ is the second moment per dimension of a random vector $\mathbf{U}$ that is uniformly distributed over $\mathcal{V}$

$$\sigma^2(\mathcal{V}) = \frac{1}{n} E\,\|\mathbf{U}\|^2 = \frac{1}{n} \frac{\int_{\mathcal{V}} \|\mathbf{x}\|^2 \, d\mathbf{x}}{V}. \tag{6.12}$$

The normalized second moment of Λ with minimized second moment $\sigma^2(\mathcal{V})$ is given by

$$G(\Lambda) \triangleq \frac{\sigma^2(\mathcal{V})}{V^{2/n}} = \frac{1}{n} \frac{\int_{\mathcal{V}} \|\mathbf{x}\|^2 \, d\mathbf{x}}{V^{1+2/n}}. \tag{6.13}$$

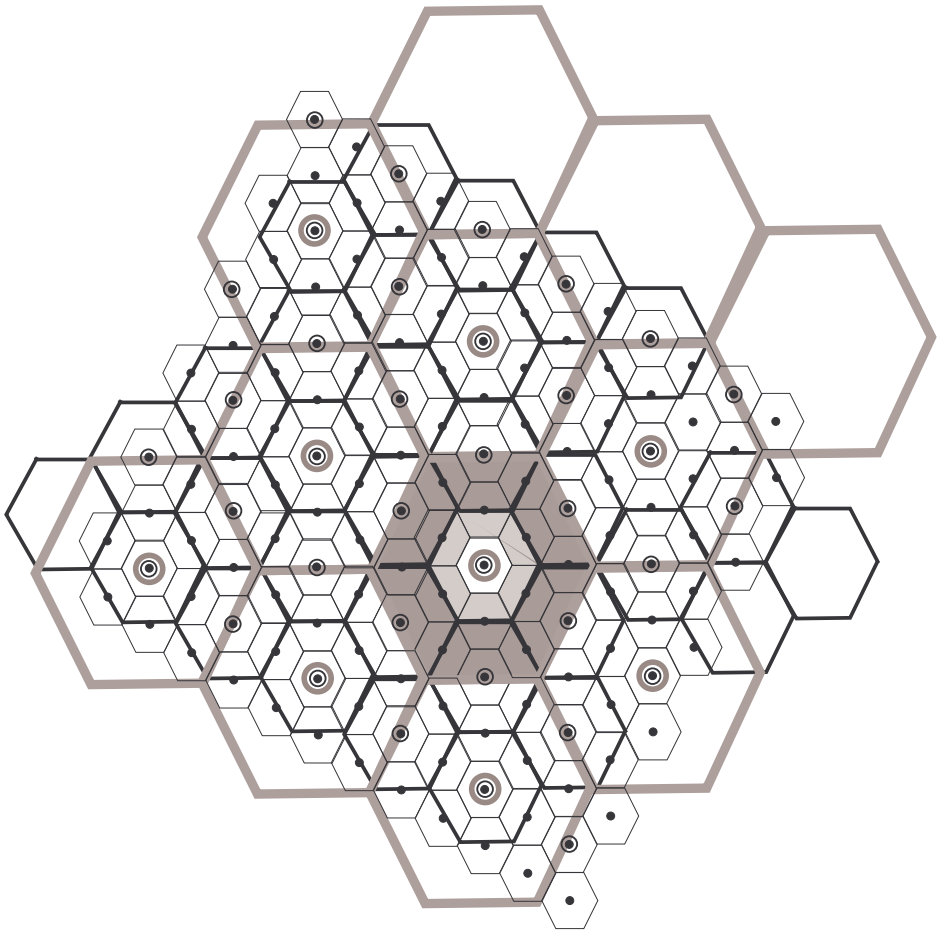


Figure 6.2: Nested lattice chain $\Lambda_3 \subseteq \Lambda_2 \subseteq \Lambda_1$ in $\mathbb{R}^2$, with hexagonal Voronoi region.

Nested lattices: A lattice Λ_2 is nested in another lattice Λ_1 if $\Lambda_2 \subseteq \Lambda_1$. Λ_1 is referred to as the fine lattice and Λ_2 as the coarse lattice. In general, we have a nested lattice chain $\Lambda_L \subseteq \Lambda_{L-1} \subseteq \cdots \subseteq \Lambda_1$. A section of the nested lattice chain $\Lambda_3 \subseteq \Lambda_2 \subseteq \Lambda_1$ is shown in Figure 6.2. The solid black dots are the elements of Λ_1, black circles are elements of Λ_2, and grey circles are elements of Λ_3. For the central point, the Voronoi regions of the lattices are shown. The region $\mathcal{V}_1$ is in white, $\mathcal{V}_2$ is shaded in light grey and includes $\mathcal{V}_1$, and $\mathcal{V}_3$ is in dark grey and includes both $\mathcal{V}_2$ and $\mathcal{V}_1$.

6.2.2 Goodness of Lattices

We use the following definitions of asymptotic goodness of lattices [121]. The existence of a lattice with simultaneous goodness in all the aspects defined below was shown in Erez *et al* [39].

Good for quantization : A sequence of lattices $\Lambda^{(n)} \subset \mathbb{R}^n$ is said to be good for mean-square-error (MSE) quantization if $G(\Lambda^{(n)}) \to \frac{1}{2\pi e}$ as $n \to \infty$.

Rogers-good : Let R_u and R_l be the covering and effective radii of Λ. Then, $1 \leq \left(\frac{R_u}{R_l}\right)^n < cn(\log n)^a$ for constants c, a, which implies that $\frac{1}{n}\log(R_u/R_l)^n \to 0$ as $n \to \infty$.

Poltyrev-good : For the Gaussian channel $\mathbf{Y} = \mathbf{X} + \mathbf{N}_w$, for any $\sigma^2 < \sigma^2(\mathcal{V})$, with $\mathbf{N}_w$ a Gaussian vector with components i.i.d. $\sim \mathcal{N}(0, \sigma^2)$, a sequence of lattices $\Lambda^{(n)}$ is Poltyrev-good, if

$$P_e = \Pr\{\mathbf{N}_w \notin \mathcal{V}\} < e^{-n[E_P(\mu) - o_n(1)]}, \tag{6.14}$$

where $o_n(1) \to 0$ as $n \to \infty$, $\mu = V^{2/n}/(2\pi e\sigma^2)$ is the normalised volume to noise ratio (VNR), and $E_P(\mu)$ is the Poltyrev exponent [38]

$$E_P(\mu) = \begin{cases} E_P^r(\mu) = \frac{1}{2}[(\mu - 1) - \log \mu] & 1 < \mu \leq 2, \\ E_P^r(\mu) = \frac{1}{2}\log\frac{e\mu}{4} & 2 \leq \mu \leq 4, \\ E_P^x(\mu) = \frac{\mu}{8} & \mu \geq 4, \end{cases} \tag{6.15}$$

assuming ML decoding and an unconstrained AWGN channel, that is, no power constraint on the channel input, which is any point of the lattice Λ. $E_P^r(\mu)$ and $E_P^x(\mu)$ are the random coding and expurgated Poltyrev exponents, respectively. They correspond to Gallager's random coding and expurgated error exponents for the DMC reviewed in Section 3.1.2. We shall be interested in the random coding Poltyrev exponent $E_P^r(\mu)$ in particular.

For the power constrained Gaussian channel, the random coding error exponent $E_c^r(\mu, \mathrm{SNR})$ (for Gaussian inputs) is given in [49, Sect. 7.4]. If we compare $E_P^r(\mu)$ with $E_c^r(\mu, \mathrm{SNR})$ (refer to [38], for example), we see that $E_P^r(\mu)$ is a lower bound for $E_c^r(\mu, \mathrm{SNR})$. As the SNR increases, $E_c^r(\mu, \mathrm{SNR}) \to E_P^r(\mu)$. We note that $E_P^r(\mu)$ does not depend on SNR.

For $\mu > 1$, the error probability goes to 0 exponentially in n. Thus $\mu = 1$ has the same significance as 'capacity'. We shall see later in the discussion of lattice coding for Gaussian channels in Section 6.3.1 how μ is related to the channel capacity and transmission rate. If the sequence is simultaneously quantization good as well, $\mu = \sigma^2(\mathcal{V})/\sigma^2$.

AWGN-good : For the same Gaussian channel and ML decoding above, the unnormalized VNR of a sequence of lattices $\Lambda^{(n)}$ is $\mu^*(\Lambda^{(n)}) = V^{2/n}/\sigma^2$. The sequence is AWGN-good if $\lim_{n\to\infty} \mu^*(\Lambda^{(n)}) = 2\pi e$, $0 < P_e < 1$.

6.3 Lattice Coding for Gaussian Channels

We review concepts in lattice coding for Gaussian channels from Erez and Zamir [38] which will be useful for formulating and analyzing our coding scheme for the Gaussian wiretap channel.

6.3.1 Modulo Lattice Additive Noise Channel

To achieve the channel capacity using lattice codes, the additive white Gaussian noise (AWGN) channel is transformed into the modulo lattice additive noise (MLAN) channel. Further useful information on the MLAN channel can be found in Forney *et al* [46]. The following lemma, called the crypto lemma, is useful in the channel transformation and subsequent analysis.

Lemma 7. *"Crypto lemma" (Forney [47], Zamir and Feder [119]): For a fundamental region Ω and any random variable (r.v.) $\mathbf{X} \in \Omega$, which is statistically independent of r.v. $\mathbf{U}$ uniformly distributed over Ω, the sum $\mathbf{Y} = (\mathbf{X} + \mathbf{U}) \bmod_\Omega \Lambda$ is uniformly distributed over Ω and statistically independent of $\mathbf{X}$.* □

Lattice points transmitted over the AWGN channel should be bounded; the bounding region is known as the shaping region. According to Erez and Zamir [38], for the power constrained channel such as the AWGN channel, the best choice for the shaping region is the Voronoi region $\mathcal{V}$. After performing the MLAN channel transformation sketched out below, we can obtain the normal AWGN channel capacity by using an input $\mathbf{c}$ that is uniformly distributed over $\mathcal{V}$ and constrained to be within $\mathcal{V}$. The MLAN channel transformation is carried out with the aid of a "dither" r.v. $\mathbf{U}$ that is also uniformly distributed over the Voronoi region $\mathcal{V}$. The next step in lattice coding is to nest Λ in Λ_1, so that $\Lambda \subseteq \Lambda_1$. Then, we replace $\mathbf{c}$, the input that is uniformly distributed over $\mathcal{V}$, by an element of the fine lattice Λ_1 falling in $\mathcal{V}$, that is the fine lattice with respect to the coarse lattice Λ. It is shown by Erez and Zamir [38] that when we now use points from the fine lattice as codewords, the normal AWGN channel capacity is achieved, in the sense that the error exponent in the probability of error is lower bounded by the

Poltyrev exponent (6.15). In other words, for rates below and approaching the normal AWGN capacity, the error probability goes exponentially to zero with the error exponent lower bounded by the Poltyrev exponent.

We now show the transformation into the MLAN channel. To begin, let us recall that the block of n channel uses for the AWGN channel is $\mathbf{Y} = \mathbf{X} + \mathbf{N}$. The input alphabet is a fundamental region Ω of Λ. Ω is called the shaping region and Λ is called the shaping lattice. Given $\mathbf{c}, \mathbf{U} \in \Omega$, the transmitter output is

$$\mathbf{X} = [\mathbf{c} - \mathbf{U}] \mathrm{mod}_{\Omega} \Lambda. \tag{6.16}$$

At the receiver, the received signal $\mathbf{Y} = \mathbf{X} + \mathbf{N}$ is multiplied by $0 \leq \alpha \leq 1$ and dither $\mathbf{U}$ is added. We have

$$\mathbf{Y}' = [\alpha(\mathbf{X} + \mathbf{N}) + \mathbf{U}] \mathrm{mod}_{\Omega} \Lambda. \tag{6.17}$$

Then we have the following lemma:

Lemma 8. *"Inflated lattice lemma" (Erez, Shamai, Zamir [39]): The channel from* $\mathbf{c}$ *to* $\mathbf{Y}$ *is equivalent in distribution to*

$$\mathbf{Y}' = [\mathbf{c} + \mathbf{N}'] \mathrm{mod}_{\Omega} \Lambda, \tag{6.18}$$

where $\mathbf{N}'$ *has distribution*

$$\mathbf{N}' = [\alpha \mathbf{N} - (1 - \alpha)\mathbf{U}] \mathrm{mod}_{\Omega} \Lambda, \tag{6.19}$$

with r.v. $\mathbf{U} \sim \mathrm{Unif}(\Omega)$ *and statistically independent of* $\mathbf{N}$*, and the term* $-(1-\alpha)\mathbf{U}$ *is called the self-noise.* □

Voronoi shaping region for AWGN channel: According to Erez and Zamir [38], for the power constrained channel such as the AWGN channel, the best choice for shaping region Ω is the Voronoi region $\mathcal{V}$, in which case we now have

$$\mathbf{N}' = [(1 - \alpha)\mathbf{U} + \alpha \mathbf{N}] \mod \Lambda, \tag{6.20}$$

since $\mathcal{V} = -\mathcal{V}$ and we use $\mathrm{mod}\ \Lambda$ to denote $\mathrm{mod}_{\mathcal{V}} \Lambda$. To satisfy the power constraint, scale the lattice so that the second moment per dimension is, from (6.12),

$$\sigma^2(\mathcal{V}) = \frac{1}{n} E \left\| \mathbf{U} \right\|^2 = \frac{1}{n} \frac{\int_{\mathcal{V}} \left\| \mathbf{x} \right\|^2 d\mathbf{x}}{V} = P_X, \tag{6.21}$$

from which, by the crypto lemma, we have the average transmitted power for any $\mathbf{c}$ is

$$\frac{1}{n} E \left\| \mathbf{X} \right\|^2 = \frac{1}{n} E \left\| \mathbf{U} \right\|^2 = P_X. \tag{6.22}$$

Capacity of the MLAN channel using uniform random code over $\mathcal{V}$: If we choose $\mathbf{C} \sim \mathrm{Unif}(\mathcal{V})$, $\alpha = \frac{P_X}{P_X+P_N} = \frac{\mathrm{SNR}}{1+\mathrm{SNR}}$ and a sequence of lattices $\Lambda^{(n)}$ that are "good for quantization" in the sense that $\lim_{n\to\infty} G(\Lambda^{(n)}) \to \frac{1}{2\pi e}$, the capacity of the MLAN channel approaches the capacity of the original AWGN channel C for n sufficiently large, and is given by [38, Thm. 1],

$$\lim_{n\to\infty} \frac{1}{n} I(\mathbf{C}; \mathbf{Y}') = \frac{1}{2}\log(1+\mathrm{SNR}). \tag{6.23}$$

When we use the specified α above, the channel has noise variance given by αP_N and now the normalised VNR, which is the argument of the Poltyrev exponent in (6.15) is given by $\mu = e^{2(C-R)}$. When $\mu = 1$, we have $R = C = \frac{1}{2}\log(1+\mathrm{SNR})$; so μ approaches 1 from above, while $R \to C$.

6.3.2 Nested Lattice Coding for the MLAN Channel

Here we describe the nested lattice coding that can achieve capacity for the MLAN channel, following Erez and Zamir [38]; consequently we conclude that we can achieve the normal AWGN channel capacity using nested lattice codes. The nested lattice structure is as follows. The coarse lattice Λ_2 is nested in the fine lattice Λ_1 if $\Lambda_2 \subseteq \Lambda_1$. To achieve capacity for the MLAN channel, we require that the coarse lattice Λ_2 is chosen so that the average power per dimension is P_X and its normalized second moment approaches that of a sphere, that is, the the coarse lattice Λ_2 satisfies the conditions for lattice Λ in [38, Thm. 1]. The fine lattice Λ_1 is chosen so that it is good for channel coding in the sense that codewords from Λ_1 are uniform over $\mathcal{V}_2$, so that the optimum random coding error exponent for the MLAN channel (Poltyrev exponent) is achieved [38, Sect.s VI-VIII].

The set

$$\mathcal{C} = \{\Lambda_1 \bmod \Lambda_2\} \triangleq \{\Lambda_1 \cap \mathcal{V}_2\} \tag{6.24}$$

contains the coset leaders of Λ_2 relative to Λ_1. Each coset of Λ_2 relative to Λ_1 is given by

$$\Lambda_{\mathbf{c}_m} = \mathbf{c}_m + \Lambda_2, \quad \text{for each } \mathbf{c}_m \in \mathcal{C},\ m \in \{1, \cdots, |\mathcal{C}|\}. \tag{6.25}$$

The set of all cosets is denoted by Λ_1/Λ_2, and we also have

$$\bigcup_{\mathbf{c}_m \in \mathcal{C}} \Lambda_{\mathbf{c}_m} = \Lambda_1. \tag{6.26}$$

The coding rate of the nested lattice code is

$$R = \frac{1}{n}\log|\mathcal{C}| = \frac{1}{n}\log|\Lambda_1/\Lambda_2| = \frac{1}{n}\log\frac{V_2}{V_1}. \tag{6.27}$$

Lastly, nested lattice coding can achieve the normal AWGN channel capacity if we use Euclidean lattice decoding with Euclidean quantization cell $\mathcal{V}_1$, or $Q_{\mathcal{V}_1}(.)$ as the decoder [38, Thm. 3]. In that case the decoding error probability $P_e \to 0$ as $n \to \infty$.

Thus we may write the encoding and decoding process as follows:

1. *Message selection*: Assign a message $\mathbf{c}_m$ to each member of the set of coset leaders $\mathcal{C} = \{\Lambda_1 \mathrm{mod} \Lambda_2\}$.

2. *Encoding*: Let dither $\mathbf{U} \sim \mathrm{Unif}(\mathcal{V}_2)$. The encoder sends

$$\mathbf{X} = [\mathbf{c}_m - \mathbf{U}] \mod \Lambda_2. \tag{6.28}$$

   By the crypto lemma, $\mathbf{X}$ is independent of $\mathbf{c}_m$ and the average transmitted power is P_X.

3. *Decoding*: By the inflated lattice lemma (Lemma 8), the channel from the transmitted codeword $\mathbf{c}_m$ to just before the decoder is $\mathbf{Y}' = [\mathbf{c}_m + \mathbf{N}'] \mathrm{mod} \Lambda_2$, where $\mathbf{N}' = [(1-\alpha)\mathbf{U} + \alpha\mathbf{N}] \mathrm{mod} \Lambda_2$. The decoder is the minimum distance decoder which has decoding metric for each codeword $\mathbf{c}_m$ over its coset $\mathbf{c}_m + \Lambda_2$ as

$$\min_{\boldsymbol{\lambda}_2 \in \Lambda_2} ||\mathbf{y}' - \mathbf{c}_m - \boldsymbol{\lambda}_2||^2. \tag{6.29}$$

   According to [38], this is equivalent to

$$\widehat{\mathbf{c}_m} = Q_{\mathcal{V}_1}(\alpha\mathbf{Y} + \mathbf{U}) \mathrm{mod}\ \Lambda_2 = Q_{\mathcal{V}_1}([\mathbf{c}_m + \mathbf{N}'] \mathrm{mod}\ \Lambda_2) \mathrm{mod}\ \Lambda_2. \tag{6.30}$$

The decoding error probability for any codeword $\mathbf{c}_m$ is then

$$P_e = \Pr\{\mathbf{N}' \notin \mathcal{V}_1\}, \tag{6.31}$$

by the fact that Λ_2 is nested in Λ_1 and the channel is modulo additive [38].

To facilitate a better understanding of the subsequent code construction and analysis for the wiretap channel, we present some more details regarding the lattice decoder (6.30) and P_e in (6.31) in Appendices D.1 and D.2. In particular, we can show that the

lattice decoder, instead of taking the somewhat mysterious form (6.30) which is tied up in quantization theory, actually takes on the much more intuitive form

$$\widehat{\mathbf{c}_m} = \arg \min_{m\in\{1,\ldots,M\}} \left(\min_{\boldsymbol{\lambda}_2\in\Lambda_2} \|\mathbf{y}' - (\mathbf{c}_m + \boldsymbol{\lambda}_2)\| \right), \tag{6.32}$$

which enables us to see immediately that decoding is to the coset specified by $\mathbf{c}_m + \Lambda_2$. Also, we outline essential steps in the analysis of the probability of error in Appendix D.2 that are helpful in our later analysis.

6.4 Nested Lattice Coding for the Gaussian Wiretap Channel

In this section we present our coding scheme for the Gaussian wiretap channel and propose decoding methods.

6.4.1 Coding and Proposed Decoding

Consider the nested lattices $\Lambda_3 \subset \Lambda_2 \subset \Lambda_1$. The encoding follows a 2-level nested coding scheme as follows. Firstly, associate a message $m \in \{1, ..., 2^{nR_e}\}$ with a coset via its coset leader. This is the first level nested lattice code to provide secrecy. Secondly, send a random member of the coset, and constrain this random member to be the set of coset leaders of the shaping lattice for the AWGN channel.

Let us define the following codebooks:

1. The set of coset leaders of Λ_2 relative to Λ_1 is

$$\mathcal{C} = \{\Lambda_1 \bmod \Lambda_2\} \triangleq \{\Lambda_1 \cap \mathcal{V}_2\}. \tag{6.33}$$

2. The set of coset leaders of Λ_3 relative to Λ_2 is

$$\mathcal{C}' = \{\Lambda_2 \bmod \Lambda_3\} \triangleq \{\Lambda_2 \cap \mathcal{V}_3\}. \tag{6.34}$$

Accordingly, we have:

(1) *Message selection and encoding:* Associate each message with a member of the set of coset leaders $\mathcal{C}$. Thus we have $\mathbf{c}_m \in \mathcal{C}$, $\Lambda_{\mathbf{c}_m} = \mathbf{c}_m + \Lambda_2$ is a coset relative to Λ_1 and

$$\Lambda_1 = \bigcup_{\mathbf{c}_m\in\mathcal{C}} \Lambda_{\mathbf{c}_m} = \bigcup_{\mathbf{c}_m\in\mathcal{C}} \mathbf{c}_m + \Lambda_2 = \bigcup_{m=1}^{|\Lambda_1/\Lambda_2|} \mathbf{c}_m + \Lambda_2. \tag{6.35}$$

The order of the partition Λ_1/Λ_2 is $|\Lambda_1/\Lambda_2| = V_2/V_1$, and so $\{\Lambda_1 \cap \mathcal{V}_2\}$ has V_2/V_1 cosets. The rate of the secret message is $R_e = \frac{1}{n}\log|\Lambda_1/\Lambda_2| = \frac{1}{n}\log\frac{V_2}{V_1}$.

We now send a uniformly selected random member of $\Lambda_{\mathbf{c}_m}$. Effectively, we send a uniformly selected random member of Λ_2. Let this random member of of Λ_2 be $\mathbf{a}_l$. This $\mathbf{a}_l$ plays the role of random bits or excess rate that the wiretapper is allowed to decode at its maximum rate, so protecting the actual message carried in the $\mathbf{c}_m$. Then we can write the transmitted point at this stage

$$\mathbf{b}_{m,l} = \mathbf{c}_m + \mathbf{a}_l, \quad \mathbf{b}_{m,l} \in \Lambda_1, \ \ \mathbf{a}_l \in \Lambda_2, \ \ \mathbf{c}_m \in [\Lambda_1/\Lambda_2], \tag{6.36}$$

where $\mathbf{a}_l$ is a uniformly random member of Λ_2. That is, for each $\mathbf{c}_m$, translate it by $\mathbf{a}_l \in \Lambda_2$. Now at this stage, the $\mathbf{a}_l$ and thus the transmitted point is an unbounded member of Λ_2. This is the same as the work of Belfiore and Oggier [7].

To achieve capacity over the AWGN channel, the $\mathbf{b}_{m,l}$ have to be sent using nested lattice coding. Now as we take the $\mathbf{c}_m$ to be an 'indexing' to the particular coset $\mathbf{c}_m + \Lambda_2$, the actual term in $\mathbf{b}_{m,l}$ to 'undergo' nested lattice coding is the $\mathbf{a}_l$. Thus we associate each $\mathbf{a}_l$ with a member of the set of coset leaders $\mathcal{C}'$, from which we have $\mathbf{a}_m \in \mathcal{C}'$, $\Lambda_{\mathbf{a}_l} = \mathbf{a}_l + \Lambda_3$ is a coset relative to Λ_2 and the $\mathbf{a}_l$ are mapped to the $|\Lambda_2/\Lambda_3| = V_3/V_2$ cosets. We have

$$\Lambda_2 = \bigcup_{\mathbf{a}_l \in \mathcal{C}'} \Lambda_{\mathbf{a}_l} = \bigcup_{\mathbf{a}_l \in \mathcal{C}'} \mathbf{a}_l + \Lambda_3 = \bigcup_{l=1}^{|\Lambda_2/\Lambda_3|} \mathbf{a}_l + \Lambda_3, \tag{6.37}$$

$$\Lambda_1 = \bigcup_{m=1}^{|\Lambda_1/\Lambda_2|} \bigcup_{l=1}^{|\Lambda_2/\Lambda_3|} \mathbf{c}_m + \mathbf{a}_l + \Lambda_3. \tag{6.38}$$

The excess rate, or the rate over the eavesdropper's channel is $R' = \frac{1}{n}\log|\Lambda_2/\Lambda_3| = \frac{1}{n}\log\frac{V_3}{V_2}$. The overall rate, over the main channel, is then $R_e + R' = \frac{1}{n}\log\frac{V_3}{V_1}$. A point in Λ_1 may now be written as

$$\boldsymbol{\lambda}_1 = \mathbf{c}_m + \mathbf{a}_l + \boldsymbol{\lambda}_3, \quad \boldsymbol{\lambda}_1 \in \Lambda_1, \ \mathbf{c}_m \in [\Lambda_1/\Lambda_2], \ \mathbf{a}_l \in [\Lambda_2/\Lambda_3], \ \boldsymbol{\lambda}_3 \in \Lambda_3. \tag{6.39}$$

We can also write

$$\boldsymbol{\lambda}_1 = \mathbf{c}_m + \mathbf{a}_l + \boldsymbol{\lambda}_3, \quad \boldsymbol{\lambda}_1 \in \Lambda_1, \ \mathbf{c}_m + \mathbf{a}_l \in [\Lambda_1/\Lambda_3]. \tag{6.40}$$

The partitioning of Λ_1 of may be written as

$$\Lambda_1 = [\Lambda_1/\Lambda_2] + [\Lambda_2/\Lambda_3] + \Lambda_3 = [\Lambda_1/\Lambda_3] + \Lambda_3. \tag{6.41}$$

(2) *Transmission:* For transmission, we add dither, defined as $\mathbf{U} \sim \text{Unif}(\mathcal{V}_3)$, and apply the mod-$\Lambda_3$ map. We note the dither is known to all parties. Then the encoder sends

$$\mathbf{X} = [\boldsymbol{\lambda}_1 - \mathbf{U}] \text{mod } \Lambda_3 = [\mathbf{c}_m + \mathbf{a}_l - \mathbf{U}] \text{mod } \Lambda_3. \tag{6.42}$$

The lattice is scaled so that the second moment of $\mathcal{V}_3$ is $\sigma^2(\mathcal{V}_3) = P_X$ so that by the crypto lemma $\frac{1}{n}E||\mathbf{X}||^2 = P_X$.

(3) *Decoding:* By the inflated lattice lemma and modulo-lattice channel transformation of Erez and Zamir [38], the channel from the transmitted codeword $\mathbf{b}_m$ to just before the decoder for the legitimate receiver is

$$\mathbf{Y}' = [\mathbf{c}_m + \mathbf{a}_l + \mathbf{N}'] \text{mod} \Lambda_3, \tag{6.43}$$

where $\mathbf{N}' = [(1-\alpha)\mathbf{U} + \alpha\mathbf{N}] \text{mod} \Lambda_3$. For the eavesdropper, it is, correspondingly,

$$\mathbf{Y}'_z = [\mathbf{c}_m + \mathbf{a}_l + \mathbf{N}'_z] \text{mod} \Lambda_3, \tag{6.44}$$

where $\mathbf{N}'_z = [(1-\alpha_z)\mathbf{U} + \alpha_z\mathbf{N}_z] \text{mod} \Lambda_3$, $\mathbf{N} \sim \mathcal{N}(0, P_N.\mathbf{I}_n)$, $\mathbf{N}_z \sim \mathcal{N}(0, P_{N_z}.\mathbf{I}_n)$, and $P_N < P_{N_z}$.

In the error probability analysis we will distinguish between the error probability at the legitimate receiver, which has to decode the pair $(\mathbf{c}_m, \mathbf{a}_l)$, and the error probability at the eavesdropper, which has to decode $\mathbf{a}_l$ given $\mathbf{c}_m$. This is the usual way to show achievability of the code rates, see Thangaraj *et al* [110] or Liang *et al* [73]. However, we note that we can also consider the eavesdropper decoding of $(\mathbf{c}_m, \mathbf{a}_l)$, and we will have to show that, for an error probability lower bound, $P_e \to 1$.

(a) *Legitimate receiver*

At the legitimate receiver, there are two possible modes of decoding and determining the decoding error probability. In the first mode, we assume that the receiver decodes $\mathbf{c}_m$, then $\mathbf{a}_l$, given $\mathbf{c}_m$, and so we determine the error probability in decoding $\mathbf{c}_m$, then the error probability in decoding $\mathbf{a}_l$, given $\mathbf{c}_m$. We will denote the first mode the staged decoding mode. In the second mode, we assume that the decoder jointly decodes $(\mathbf{c}_m, \mathbf{a}_l)$ together. Here we determine the error probability in jointly decoding $(\mathbf{c}_m, \mathbf{a}_l)$. We denote the second mode the joint decoding mode.

i. *Staged decoding mode*

We first look at the first mode and note that $\mathcal{V}_1 \subset \mathcal{V}_2 \subset \mathcal{V}_3$. All messages are associated with coset leaders which are contained in $\mathcal{V}_2$, translated by an element of $\mathcal{V}_3$, while each message is contained in $\mathcal{V}_1$. Overall, each message is contained in $\mathcal{V}_3$. Thus the decoding metric at the legitimate receiver is

$$\mu(\mathbf{y}) = \min_{\boldsymbol{\lambda}_3 \in \Lambda_3} ||\mathbf{y} - \boldsymbol{\lambda}_3||^2, \tag{6.45}$$

giving the decoding operation as

$$\widehat{\mathbf{c}_m} = Q_{\mathcal{V}_1}(\mathbf{Y}') \bmod \Lambda_3, \tag{6.46}$$

and the decoding error probability for any codeword $\mathbf{c}_m$ is the probability of error over the coset $\Lambda_2 + \mathbf{c}_m$,

$$P_{e,m} = \Pr\left[\mathbf{N}' \notin \bigcup_{l=1}^{V_3/V_2} \mathcal{V}_1 + \mathbf{a}_l\right]. \tag{6.47}$$

Alternatively, noting that $\mathbf{a}_l \in \Lambda_3 \subset \Lambda_2$, we can also use the decoding metric

$$\mu'(\mathbf{y}) = \min_{\boldsymbol{\lambda}_2 \in \Lambda_2} ||\mathbf{y} - \boldsymbol{\lambda}_2||^2, \tag{6.48}$$

which gives the decoding operation as

$$\begin{aligned}
\widehat{\mathbf{c}_m} &= Q_{\mathcal{V}_1}(\mathbf{Y}') \bmod \Lambda_2 \\
&\overset{(a)}{=} Q_{\mathcal{V}_1}(\mathbf{Y}' \bmod \Lambda_2) \bmod \Lambda_2 \\
&= Q_{\mathcal{V}_1}(([\mathbf{c}_m + \mathbf{a}_l + \mathbf{N}'] \bmod \Lambda_3) \bmod \Lambda_2) \bmod \Lambda_2 \\
&\overset{(b)}{=} Q_{\mathcal{V}_1}([\mathbf{c}_m + \mathbf{a}_l + \mathbf{N}'] \bmod \Lambda_2) \bmod \Lambda_2 \\
&\overset{(c)}{=} Q_{\mathcal{V}_1}([\mathbf{c}_m + \mathbf{N}'] \bmod \Lambda_2) \bmod \Lambda_2 \\
&= Q_{\mathcal{V}_1}(\mathbf{c}_m + \mathbf{N}' \bmod \Lambda_2) \bmod \Lambda_2 \\
&= Q_{\mathcal{V}_1}(\mathbf{c}_m + \mathbf{N}^\dagger) \bmod \Lambda_2
\end{aligned} \tag{6.49}$$

where $\mathbf{N}^\dagger = \mathbf{N}' \bmod \Lambda_2 = [(1-\alpha)\mathbf{U}' + \alpha\mathbf{N}] \bmod \Lambda_2$, where $\mathbf{U}' \sim \mathrm{Unif}(\mathcal{V}_2)$, and (a) follows from the property of the mod-Λ_2 operation,

(b) is due to $\Lambda_3 \subset \Lambda_2$, and (c) is because $\mathbf{a}_l \in \Lambda_3$. The noise is evaluated as

$$\begin{aligned}\mathbf{N}^\dagger = \mathbf{N}' \mathrm{mod} \Lambda_2 &= ([(1-\alpha)\mathbf{U} + \alpha\mathbf{N}] \mathrm{mod} \Lambda_3) \, \mathrm{mod} \Lambda_2 \\ &= [(1-\alpha)\mathbf{U} \mathrm{mod} \Lambda_2 + \alpha\mathbf{N}] \mathrm{mod} \Lambda_2 \\ &= [(1-\alpha)\mathbf{U}' + \alpha\mathbf{N}] \mathrm{mod} \Lambda_2. \end{aligned} \tag{6.50}$$

We need the distribution of $\mathbf{U}' = \mathbf{U} \mathrm{mod} \Lambda_2$, where $\mathbf{U} \sim \mathrm{Unif}(\mathcal{V}_3)$. From the Corollary to Lemma 2 in a recent paper by Zamir [122] we have that $\mathbf{U}' \sim \mathrm{Unif}(\mathcal{V}_2)$, given that $\mathbf{U} \sim \mathrm{Unif}(\mathcal{V}_3)$ and $\Lambda_3 \subset \Lambda_2$.

Then the decoding error probability for any codeword $\mathbf{c}_m$ is now

$$P_{e,m} = \Pr\left[\mathbf{N}^\dagger \notin \mathcal{V}_1\right]. \tag{6.51}$$

A question to answer is whether the two decoding methods using the metrics (6.45) and (6.48) above obtain the same $\mathbf{c}_m$. We see that in (6.45), since we have the chain $\Lambda_3 \subset \Lambda_2 \subset \Lambda_1$, all of $\boldsymbol{\lambda}_3 \in \Lambda_3$ are also $\in \Lambda_2$. Thus it appears that the two decoders using the metrics (6.45) and (6.48) are equivalent and will produce the same $\mathbf{c}_m$.

Next, to obtain $\widehat{\mathbf{a}}_l$, subtract the $\widehat{\mathbf{c}_m}$ from $\mathbf{Y}'$ and use the decoder

$$\begin{aligned}\widehat{\mathbf{a}}_l &= Q_{\mathcal{V}_2}(\mathbf{Y}' - \widehat{\mathbf{c}_m}) \bmod \Lambda_3 \\ &= Q_{\mathcal{V}_2}(\mathbf{a}_l + (\mathbf{c}_m - \widehat{\mathbf{c}_m}) \mathrm{mod} \Lambda_3 + \mathbf{N}') \bmod \Lambda_3, \end{aligned} \tag{6.52}$$

from which the probability of decoding error is

$$P_{e,l} = \Pr\left[((\mathbf{c}_m - \widehat{\mathbf{c}_m}) \mathrm{mod} \Lambda_3 + \mathbf{N}') \notin \mathcal{V}_2\right]. \tag{6.53}$$

By the union bound, the overall probability of decoding error for the staged decoder is

$$P_{e,l,m} \leq P_{e,m} + P_{e,l}. \tag{6.54}$$

If we let the estimated $\widehat{\mathbf{c}_m} \in \Lambda_3$ and by the fact that Λ_3 is centro-symmetric, we have $(\mathbf{c}_m - \widehat{\mathbf{c}_m}) \mathrm{mod} \Lambda_3 = (\mathbf{c}_m + \widehat{\mathbf{c}_m}) \mathrm{mod} \Lambda_3 \in \Lambda_3$. A possible future work is to see what is the distribution of this term and if it can be included into the noise term $\mathbf{N}'$.

ii. *Joint decoding mode*

Here, we assume that the decoder jointly decodes $(\mathbf{c}_m, \mathbf{a}_l)$ together. To determine the decision region, we note that $\mathbf{c}_m \in \mathcal{V}_1$ is translated by $\mathbf{a}_l \in \mathcal{V}_2$. This can be seen as an arbitrary $\mathcal{V}_1$ being moved into the larger region $\mathcal{V}_2$, giving an overall decision region $\mathcal{V}_1 \cap \mathcal{V}_2 = \mathcal{V}_1$. The $\mathbf{c}_m + \mathbf{a}_l$ is enclosed by $\mathcal{V}_3$. The decoder can be expressed as

$$(\widehat{\mathbf{c}_m}, \widehat{\mathbf{a}}_l) = Q_{\mathcal{V}_1 \cap \mathcal{V}_2}(\mathbf{Y}') \bmod \Lambda_3, \tag{6.55}$$

and the joint decoding error probability is then

$$P_{e,l,m} = \Pr\left[\mathbf{N}' \notin \mathcal{V}_1 \cap \mathcal{V}_2\right] = \Pr\left[\mathbf{N}' \notin \mathcal{V}_1\right]. \tag{6.56}$$

(b) *Eavesdropper decoding*

For the eavesdropper, we assume that $\mathbf{c}_m$ is known, and its decoder then attempts to decode $\mathbf{a}_l \in \mathcal{V}_2$. The decoder used is

$$\widehat{\mathbf{a}}_l = Q_{\mathcal{V}_2}(\mathbf{Y}'_z) \bmod \Lambda_3, \tag{6.57}$$

with decoding error probability

$$P_{e,l}^{(z)} = \Pr\left[\mathbf{N}'_z \notin \mathcal{V}_2\right]. \tag{6.58}$$

A couple of alternatives for the eavesdropper are:

i. Staged decoding, with decoders

$$\widehat{\mathbf{c}_m} = Q_{\mathcal{V}_1}(\mathbf{c}_m + \mathbf{N}_z^{\dagger}) \bmod \Lambda_2, \tag{6.59}$$

$$\widehat{\mathbf{a}}_l = Q_{\mathcal{V}_2}(\mathbf{a}_l + (\mathbf{c}_m - \widehat{\mathbf{c}_m}) \bmod \Lambda_3 + \mathbf{N}'_z) \bmod \Lambda_3, \tag{6.60}$$

where $\mathbf{N}_z^{\dagger} = \mathbf{N}'_z \bmod \Lambda_2 = [(1-\alpha)\mathbf{U}' + \alpha \mathbf{N}_z] \bmod \Lambda_2$. The associated error probabilities are

$$P_{e,m}^{(z)} = \Pr\left[\mathbf{N}_z^{\dagger} \notin \mathcal{V}_1\right], \tag{6.61}$$

$$P_{e,l}^{(z)} = \Pr\left[((\mathbf{c}_m - \widehat{\mathbf{c}_m}) \bmod \Lambda_3 + \mathbf{N}'_z) \notin \mathcal{V}_2\right], \tag{6.62}$$

and we should show that $P_{e,m}^{(z)} \to 1$ and $P_{e,l}^{(z)} \to 0$ as $n \to \infty$.

ii. Joint decoding, so that the decoder is

$$(\widehat{\mathbf{c}_m}, \widehat{\mathbf{a}}_l) = Q_{\mathcal{V}_1 \cap \mathcal{V}_2}(\mathbf{Y}'_z) \bmod \Lambda_3, \tag{6.63}$$

and the joint decoding error probability is then

$$P_{e,l,m}^{(z)} = \Pr\left[\mathbf{N}'_z \notin \mathcal{V}_1 \cap \mathcal{V}_2\right] = \Pr\left[\mathbf{N}'_z \notin \mathcal{V}_1\right]. \tag{6.64}$$

We should then show that $P_{e,l,m}^{(z)} \to 1$ as $n \to \infty$.

While the two alternative decoders for the eavesdropper are arguably 'stronger' than the first one in (6.57), we will assume the decoder in (6.57) for our analysis; the other two decoders will be left for future work.

In the next two sections, we show our main result, which is formally stated by the following theorem, where C and C_z denote the capacities of the main and eavesdropper's channels, respectively.

Theorem 18. *For our nested lattice coding scheme described above, $P_{e,l,m}, P_{e,l}^{(z)} \to 0$, as $n \to \infty$ for rates R and R' approaching C and C_z, while the construction achieves C and C_z on the main and eavesdropper's channels, respectively. The equivocation rate $\lim_{n\to\infty} R_e = \frac{1}{n} H(M|\mathbf{Z}) = C - C_z$ satisfies the secrecy rate for the Gaussian wiretap channel and the secrecy constraint $\lim_{n\to\infty} \frac{1}{n} I(M;\mathbf{Z}) = 0$ is achieved.* □

In other words, we shall see that we can meet the important criteria of the code being information theoretically secure, as in Theorem 12.

6.4.2 Rates and Equivocation

Let $\mathbf{C}, \mathbf{A}$ be the r.v.'s uniformly distributed over codebooks $\mathcal{C}$, $\mathcal{C}'$ by construction, of which the realizations are $\mathbf{c}_m$ and $\mathbf{a}_l$, respectively. The equivalent channels are (6.43) and (6.44) with $\mathbf{c}_m$ and $\mathbf{a}_l$ replaced by $\mathbf{C}$ and $\mathbf{A}$. For notational convenience, we write $[\mathbf{C} + \mathbf{A}] \bmod \Lambda_3$ as $\mathbf{C} \oplus \mathbf{A}$.

Channel Rates

The main channel input $\mathbf{C} \oplus \mathbf{A} \in [\Lambda_1/\Lambda_3]$, and has rate

$$R = \frac{1}{n} \log \frac{V_3}{V_1} = \frac{1}{2} \log \frac{V_3^{\frac{2}{n}}}{2\pi e} - \frac{1}{2} \log \frac{V_1^{\frac{2}{n}}}{2\pi e}$$

$$\stackrel{(a)}{=} \frac{1}{2}\log\frac{P_X}{2\pi e G(\Lambda_3)} - \frac{1}{2}\log\frac{V_1^{\frac{2}{n}}}{2\pi e}$$
$$= \frac{1}{2}\log P_X - \frac{1}{2}\log 2\pi e G(\Lambda_3) - \frac{1}{2}\log\frac{V_1^{\frac{2}{n}}}{2\pi e}, \tag{6.65}$$

where (a) is due to the fact that $G(\Lambda_3) = \sigma^2(\mathcal{V}_3)/V_3^{\frac{2}{n}}$ and $\sigma^2(\mathcal{V}_3) = P_X$. Consider the sequence of lattices $\Lambda_3^{(n)}$, good for quantization so that $\lim_{n\to\infty} G(\Lambda_3^{(n)}) = \frac{1}{2\pi e}$, and the AWGN good lattices $\Lambda_1^{(n)}$. As such, from Forney [47, Sect. 2.4], we have $\log\frac{V_1^{\frac{2}{n}}}{2\pi e} \to \log\frac{1}{n}\mathtt{E}\left[\|\mathbf{N}''\|^2\right]$ as $n\to\infty$. Using the minimum MSE (MMSE) scaling $\alpha = \frac{P_X}{P_X+P_N}$, we get

$$\frac{1}{n}\mathtt{E}\left[\|\mathbf{N}''\|^2\right] = \frac{1}{n}\mathtt{E}\left[\|(1-\alpha)\mathbf{U} + \alpha\mathbf{N}\|^2\right] = \alpha P_N. \tag{6.66}$$

Then, from (6.65), we have

$$R = \tfrac{1}{2}\log(1+\mathrm{SNR}) = C, \tag{6.67}$$

as $n\to\infty$. For the eavesdropper's channel, the input $\mathbf{c}_m \oplus \mathbf{A} \in [\Lambda_2/\Lambda_3]$, and using a similar calculation the rate is

$$R' = \tfrac{1}{2}\log(1+\mathrm{SNR}_z) = C_z, \tag{6.68}$$

as $n\to\infty$. This time we use the sequence of AWGN good lattices $\Lambda_2^{(n)}$, and $\alpha_z = \frac{P_X}{P_X+P_{N_z}}$. In summary, we need the sequences $\Lambda_1^{(n)}$, $\Lambda_2^{(n)}$ to be AWGN good, and the sequence $\Lambda_3^{(n)}$ good for quantization.

Calculation of the Equivocation Rate

The equivocation rate satisfies $\lim_{n\to\infty} R_e = \frac{1}{n}H(M|\mathbf{Z})$. The perfect secrecy constraint is given by $\lim_{n\to\infty}\frac{1}{n}I(M;\mathbf{Z}) \to 0$, under which the eavesdropper gets no information about the message.

For the equivocation of the message M,

$$H(M|\mathbf{Z}) = H(M) - I(M;\mathbf{Z}). \tag{6.69}$$

We now use the expansions

$$I(M,\mathbf{X};\mathbf{Z}) = I(M;\mathbf{Z}) + I(\mathbf{X};\mathbf{Z}|M) \tag{6.70}$$
$$= I(\mathbf{X};\mathbf{Z}) + I(M;\mathbf{Z}|\mathbf{X}), \tag{6.71}$$

to give $I(M;\mathbf{Z}) = I(\mathbf{X};\mathbf{Z}) - I(\mathbf{X};\mathbf{Z}|M)$ since $I(M;\mathbf{Z}|\mathbf{X}) = 0$ as $M \to \mathbf{X} \to \mathbf{Z}$ forms a Markov chain. Substituting this into (6.69), we obtain

$$\begin{aligned} H(M|\mathbf{Z}) &= H(M) - I(\mathbf{X};\mathbf{Z}) + I(\mathbf{X};\mathbf{Z}|M) \\ &\overset{(a)}{\geq} H(M) - C_z + I(\mathbf{C},\mathbf{A};\mathbf{Z}|M) \\ &= H(M) - C_z + H(\mathbf{C},\mathbf{A}|M) - H(\mathbf{C},\mathbf{A}|\mathbf{Z},M) \\ &\overset{(b)}{=} \log\tfrac{V_2}{V_1} - C_z + \log\tfrac{V_3}{V_2} - H(\mathbf{C},\mathbf{A}|\mathbf{Z},M), \end{aligned} \tag{6.72}$$

where (a) is by $I(\mathbf{X};\mathbf{Z}) \leq C_z$, since C_z is the maximum possible rate of the wiretapper's channel, and because there is a one-to-one correspondence between $(\mathbf{C},\mathbf{A})$ and $\mathbf{X}$ so that $I(\mathbf{X};\mathbf{Z}|M) = I(\mathbf{C},\mathbf{A};\mathbf{Z}|M)$, and (b) is due to $H(M) = \log(V_2/V_1)$ and $H(\mathbf{C},\mathbf{A}|M) = \log(V_3/V_2)$.

For the last term in (6.72), we now carry out an argument using one similar to Fano's inequality for the pair $(\mathbf{C},\mathbf{A})$ and $\mathbf{Z}$. The last term in (6.72) is the entropy of the codeword conditioned on the coset $\mathbf{C}+\Lambda_2$ and the eavesdropper's observation. This is related to the eavesdropper's decoding error probability $P_{e,l}^{(z)}$. Define the random variable χ as

$$\chi = \begin{cases} 1 & \text{if} \quad \psi(\mathbf{Z}) \neq (\mathbf{C},\mathbf{A}), \\ 0 & \text{if} \quad \psi(\mathbf{Z}) = (\mathbf{C},\mathbf{A}), \end{cases} \tag{6.73}$$

where ψ denotes the eavesdropper's decoding process. We note that we can use $\mathbf{Z}$ instead of $\mathbf{Y}_z'$ in the definition of the decoding process above because the decoding process includes the MLAN transformation. Then we have

$$\begin{aligned} &H(\mathbf{C},\mathbf{A}|M,\mathbf{Z}) \\ &= H(\chi,\mathbf{C},\mathbf{A}|M,\mathbf{Z}) - H(\chi|\mathbf{C},\mathbf{A},M,\mathbf{Z}) \\ &\overset{(a)}{=} H(\chi|M,\mathbf{Z}) + H(\mathbf{C},\mathbf{A}|\chi,M,\mathbf{Z}) \\ &\overset{(b)}{\leq} H(\chi|M) + H(\mathbf{C},\mathbf{A}|\chi,M,\mathbf{Z}) \\ &= \sum_{m=1}^{V_2/V_1} \Pr[M=m]\,[H(\chi|M=m) + \Pr[\chi=0|M=m]H(\mathbf{C},\mathbf{A}|\chi=0,M=m,\mathbf{Z}) \\ &\quad + \Pr[\chi=1|M=m]H(\mathbf{C},\mathbf{A}|\chi=1,M=m,\mathbf{Z})] \\ &\overset{(c)}{\leq} \sum_{m=1}^{V_2/V_1} \Pr[M=m]\,\Big[H_2(P_{e,l}^{(z)}) + (1-P_{e,l}^{(z)})H(\mathbf{C},\mathbf{A}|\chi=0,M=m,\mathbf{Z}) \\ &\quad + P_{e,l}^{(z)}H(\mathbf{C},\mathbf{A}|\chi=1,M=m,\mathbf{Z})\Big] \end{aligned} \tag{6.74}$$

where (a) is because χ is determined by $\mathbf{C}, \mathbf{A}, M, \mathbf{Z}$; (b) is because conditioning reduces entropy; (c) is by letting $P_{e,l}^{(z)} = \Pr[\chi = 1|M = m]$ and since $H(\chi|M = m) \leq H(\chi) \leq H_2(P_{e,l}^{(z)})$. Now we have

$$H(\mathbf{C}, \mathbf{A}|\chi = 0, M = m, \mathbf{Z}) = 0 \tag{6.75}$$

as $\chi = 0$ means that the wiretapper is able to decode the pair $(\mathbf{C}, \mathbf{A})$ with $M = m$ known with probability = 1, thus there is no more uncertainty. Also we have

$$H(\mathbf{C}, \mathbf{A}|\chi = 1, M = m, \mathbf{Z}) \leq \log(V_3/V_2). \tag{6.76}$$

Substituting into (6.74) above, and using the facts that $P_{e,l}^{(z)}$ is independent of the message and that $\Pr[M = m] = \frac{V_1}{V_2}$ (since M is uniformly distributed) we have,

$$H(\mathbf{C}, \mathbf{A}|M, \mathbf{Z}) \leq H_2(P_{e,l}^{(z)}) + P_{e,l}^{(z)} \log(V_3/V_2) \leq n\epsilon, \tag{6.77}$$

where $\epsilon \to 0$ as $n \to \infty$ as long as $P_{e,l}^{(z)} \to 0$ as $n \to \infty$.

Substituting this into (6.72) and dividing by n, we get the equivocation rate

$$R_e = \lim_{n\to\infty} \tfrac{1}{n} H(M|\mathbf{Z}) = \tfrac{1}{n} \log(V_2/V_1) = C - C_z, \tag{6.78}$$

which is the equivocation rate of the Gaussian wiretap channel in [70]. In fact this is the secrecy capacity of the Gaussian wiretap channel. Finally, it is easy to see that

$$\lim_{n\to\infty} \tfrac{1}{n} I(M; \mathbf{Z}) = \lim_{n\to\infty} \left[H(M) - \tfrac{1}{n} H(M|\mathbf{Z}) \right] \tag{6.79}$$

$$= \lim_{n\to\infty} \tfrac{1}{n} \log(V_2/V_1) - (C - C_z) = 0, \tag{6.80}$$

and the secrecy constraint can be achieved.

Calculations in this section have used codewords from codebooks $\mathcal{C}$ and $\mathcal{C}'$. Later we see a nested lattice construction exists with small decoding error probabilities as $n \to \infty$ for the main and eavesdropper's channels at rates $R \to C$ and $R' \to C_z$. We then conclude that, using our coding scheme, we can achieve the capacities for the main and eavesdropper's channel, the secrecy rate of the Gaussian wiretap channel (6.2), and satisfy the security constraint.

6.4.3 Code Construction

The construction follows the method in Nazer and Gastpar [94]. The method uses a coarse lattice and then forms successively fine lattices, taking into account the nesting required.[1].

[1] Here, the coarse lattice to start off the construction is still undetermined explicitly.

We begin with a coarse lattice Λ_3 that is simultaneously covering, quantization, Rogers and Poltyrev good. The existence of such a lattice is shown in [39]. Let Λ_3 have the generator matrix $\mathbf{G}'$, so that $\Lambda_3 = \mathbf{G}'\mathbb{Z}^n$. The fine lattices are constructed in the order of Λ_1 first, then Λ_2.

Let k_1, k_2, n, p be integers such that $k_2 < k_1 \leq n$, and p is a prime. To construct Λ_1, we perform the following:

1. Let $\mathbf{G}_1$ be a $k_1 \times n$ generator matrix with elements $\sim \mathrm{Unif}(0, 1, \ldots, p-1)$, that is, uniform over $\mathbb{Z}_p$.
2. Define discrete codebook $\mathcal{C}_1 = \{\mathbf{x} = \mathbf{y}\mathbf{G}_1 : \mathbf{y} \in \mathbb{Z}_p^{k_1}\}$.
3. Lift $\mathcal{C}_1$ to $\mathbb{R}^n$ to form the lattice $\Lambda_1' = p^{-1}\mathcal{C}_1 + \mathbb{Z}^n$.
4. The fine lattice is given by $\Lambda_1 = \mathbf{G}'\Lambda_1'$.

To construct Λ_2, we do the following:

1. Let $\mathbf{G}_2$ be the $k_2 \times n$ generator matrix which is the first k_2 rows of $\mathbf{G}_1$.
2. Define discrete codebook $\mathcal{C}_2 = \{\mathbf{x} = \mathbf{y}\mathbf{G}_2 : \mathbf{y} \in \mathbb{Z}_p^{k_2}\}$.
3. Lift $\mathcal{C}_2$ to $\mathbb{R}^n$ to form the lattice $\Lambda_2' = p^{-1}\mathcal{C}_2 + \mathbb{Z}^n$.
4. The fine lattice is given by $\Lambda_2 = \mathbf{G}'\Lambda_2'$.

In $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathbf{x} \in \mathbb{Z}_p^n$. By construction, $\mathbb{Z}^n \subset \Lambda_1'$ and $\mathbb{Z}^n \subset \Lambda_2'$. We have $\mathcal{C}_2 \subset \mathcal{C}_1$ since all elements of $\mathcal{C}_2$ can be found in $\mathcal{C}_1$ as $\mathbf{G}_2 \subset \mathbf{G}_1$. This means that we have $\Lambda_3 \subset \Lambda_2 \subset \Lambda_1$. If $\mathbf{G}_1, \mathbf{G}_2$ are of full rank, then the number of fine lattice points in the Voronoi region of the coarse lattice is $|\Lambda_i \cap \mathcal{V}_3| = p^{k_i}$, $i = 1, 2$. The probability that $\mathbf{G}_1, \mathbf{G}_2$ are not of full rank is given by the union bound

$$\begin{aligned}\Pr\left[\bigcup_{i=1}^{2}\{\mathrm{rank}(\mathbf{G}_i) < k_i\}\right] &\leq \sum_{i=1}^{2}\sum_{\mathbf{y}\neq\mathbf{0},\mathbf{y}\in\mathbb{Z}_p^{k_i}} \Pr\left[\mathbf{y}\mathbf{G}_i = \mathbf{0}\right] \\ &\leq p^{-n}(p^{k_1} + p^{k_2} - 2), \end{aligned} \tag{6.81}$$

which $\to 0$ as $n - k_1$ and $n - k_2 \to \infty$. The use of the restriction on n and p, $\frac{n}{p} \to 0$, is seen later in the error probability analysis.

The construction described above gives nested lattices $\Lambda_3 \subset \Lambda_2 \subset \Lambda_1$ all Rogers and Poltyrev good [68]. Furthermore, from Krithivasan and Pradhan [68], the points of the lattices Λ_1, Λ_2 contained in $\mathcal{V}_3$, denoted $\Lambda_i(j)$, $j = 0, 1, \ldots, p^{k_i} - 1$, $i = 1, 2$ satisfy the following properties:

1. $\Lambda_i(j)$ is equally likely to be any of the points in $\{p^{-1}\Lambda_3 \cap \mathcal{V}_3\}$.
2. For any $j \neq k$, $[\Lambda_i(j) - \Lambda_i(k)] \bmod \Lambda_3$ is uniformly distributed over $\{p^{-1}\Lambda_3 \cap \mathcal{V}_3\}$.

6.4.4 Error Analysis

We analyze the error probability and show that the probabilities of error are small for the main and eavesdropper's channels for rates $R \to C$ and $R' \to C_z$, respectively, using Euclidean lattice decoding. The legitimate receiver performs joint decoding of $(\mathbf{c}_m, \mathbf{a}_l)$, while the eavesdropper decodes $\mathbf{a}_l$, given $\mathbf{c}_m$. The lattice Λ_3 is Rogers, Poltyrev, quantization good. Recall that $\alpha = \frac{P_X}{P_X + P_N}$ and $\alpha_z = \frac{P_X}{P_X + P_{N_z}}$. Following [38], we make some necessary definitions.

- Let σ^2 be the second moment of a ball containing $\mathcal{V}_3$, so that $\sigma^2 > P_X$.
- Let $\mathbf{Z}_1 \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_n)$ and $\mathbf{Z} \sim \mathcal{N}(0, \alpha P_N . \mathbf{I}_n)$; for the main channel $\mathbf{Z}_m = \mathcal{N}(0, \alpha P_N . \mathbf{I}_n)$, $\mathbf{Z}_m^* = (1 - \alpha)\mathbf{Z}_1 + \alpha \mathbf{N}$; for the eavesdropper's channel, $\mathbf{Z}_z = \mathcal{N}(0, \alpha_z P_{N_z} . \mathbf{I}_n)$, $\mathbf{Z}_z^* = (1 - \alpha_z)\mathbf{Z}_1 + \alpha_z \mathbf{N}_z$.
- G_n^* denotes the normalized second moment of an n-dimensional sphere, $G_n^* \to \frac{1}{2\pi e}$ as $n \to \infty$.
- Define

$$\epsilon_1(\Lambda_3) \triangleq \log\left(\frac{R_u}{R_l}\right) + \frac{1}{2}\log 2\pi e G_n^* + \frac{1}{n}, \tag{6.82}$$

$$\epsilon_2(\Lambda_3) \triangleq \log\left(\frac{R_u}{R_l}\right) + \frac{1}{2}\log 2\pi e G(\Lambda_3). \tag{6.83}$$

For Λ_3 both Rogers good and quantization good, $\epsilon_1(\Lambda_3), \epsilon_2(\Lambda_3) \to 0$ as $n \to \infty$.

For the main channel, the random coding Poltyrev exponent may be expressed in terms of the channel capacity and transmission rate as

$$E_P^r(\mu) = \max_{0 < \rho \leq 1} \rho \left[\frac{1}{2}\log 2\pi e P_X - h_{\bar{\rho}}(Z_m) - R\right], \tag{6.84}$$

where $\mu = e^{2(C-R)}$, $Z_m \sim \mathcal{N}(0, \alpha P_N)$, $\bar{\rho} = 1/(1+\rho)$ and $h_{\bar{\rho}}(Z_m)$ is the Rényi entropy of order $\bar{\rho}$, defined as

$$h_{\bar{\rho}}(Z_m) \triangleq \frac{\bar{\rho}}{1-\bar{\rho}} \log \left(\int_z f_{Z_m}(z)^{\bar{\rho}} dz \right)^{(1/\bar{\rho})}. \tag{6.85}$$

A similar statement can be made for the random coding error exponent on the eavesdropper's channel, with appropriate changes.

Decoding Error Probability for the Main Channel

We first need to bound the p.d.f. of the unaliased noise $\mathbf{N}''$ by the p.d.f. of the Gaussian $\mathbf{Z}_m^*$. From [38, Lemmas 6 & 11], it is shown that[2]

$$f_{\mathbf{N}''}(\mathbf{x}) \leq e^{\epsilon_1(\Lambda_3)n} f_{\mathbf{Z}_m^*}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{V}_3 \tag{6.86}$$

and $\mathbf{Z}_m^*$ is distributed as $\mathcal{N}(0, P_{Z_m^*}\mathbf{I}_n)$, with

$$\frac{n}{n+2}\alpha P_N \leq P_{Z_m^*} < \left(\frac{R_u}{R_l}\right)^n \alpha P_N. \tag{6.87}$$

The error probability for the pair (m, l) can be bounded as

$$P_{e,l,m} = \Pr\left[\mathbf{N}' \notin \mathcal{V}_1\right] \leq \Pr\left[\mathbf{N}'' \notin \mathcal{V}_1\right] \leq e^{\epsilon_1(\Lambda_3)n} \Pr\left[\mathbf{Z}_m^* \notin \mathcal{V}_1\right]. \tag{6.88}$$

Now, we bound the probability $\Pr\left[\mathbf{Z}_m^* \notin \mathcal{V}_1\right]$ by truncating $\mathbf{Z}_m^*$ to $\mathcal{V}_3$ to give $\mathbf{Z}_{\mathcal{V}_3}^m$. The truncated version is $\mathbf{Z}_{\mathcal{V}_3}^m$ has the distribution

$$f_{\mathbf{Z}_{\mathcal{V}_3}^m}(\mathbf{x}) = \begin{cases} \frac{1}{1-\Pr[\mathbf{Z}_m^* \notin \mathcal{V}_3]} f_{\mathbf{Z}_m^*}(\mathbf{x}) & \mathbf{x} \in \mathcal{V}_3 \\ 0 & \text{otherwise} \end{cases} \tag{6.89}$$

Since $\mathcal{V}_1 \subset \mathcal{V}_3$, we can follow the argument in [38, Eqns. (84)–(88)], to have

$$\Pr\left[\mathbf{Z}_m^* \notin \mathcal{V}_1\right] \leq \Pr\left[\mathbf{Z}_{\mathcal{V}_3}^m \notin \mathcal{V}_1\right] + \Pr\left[\mathbf{Z}_m^* \notin \mathcal{V}_3\right]. \tag{6.90}$$

Next, consider the second term on the RHS of (6.90). If we view Λ_3 as a channel code with respect to the Gaussian $\mathbf{Z}_m^*$, Euclidean decoding is ML for such a channel. So we use (6.14) to bound $\Pr\left[\mathbf{Z}_m^* \notin \mathcal{V}_3\right]$ with the equivalent VNR of Λ_3 viewed as a channel code with respect to $\mathbf{Z}_m^*$ given by

$$\mu = \frac{P_X}{P_{Z_m^*}} \geq 1 + \frac{P_X}{P_N} - o_n(1) = e^{2C} - o_n(1), \tag{6.91}$$

[2]The proofs for [38, Lemmas 6 & 11] go through unchanged because they are derived based on the coarse lattice Λ_3 only.

thus giving

$$\Pr\left[\mathbf{Z}_m^* \notin \mathcal{V}_3\right] \leq e^{-n[E_P(e^{2C})-o_n(1)]}. \tag{6.92}$$

To bound the first term in (6.90), consider the $(\Lambda_3, \mathbf{Z}_{\mathcal{V}_3}^m)$ channel (with generic output $\mathbf{Y}$)

$$\mathbf{Y} = [\mathbf{X} + \mathbf{Z}_{\mathcal{V}_3}^m] \bmod \Lambda_3, \tag{6.93}$$

the modulo additive channel with $\mathbf{Z}_m^*$ restricted to $\mathcal{V}_3$, for which Euclidean decoding is ML with $\mathbf{Z}_{\mathcal{V}_3}^m$ Gaussian in $\mathcal{V}_3$. Then this channel has error probability determined by its error exponent. A random coding error exponent is derived for this channel in [37], given by

$$E_{\Lambda_3}(R; \mathbf{Z}_{\mathcal{V}_3}^m) = \max_{0\leq\rho\leq 1} \rho\left[\frac{1}{n}\log V_3 - \frac{1}{n}h_{\bar{\rho}}(\mathbf{Z}_{\mathcal{V}_3}^m) - R\right], \tag{6.94}$$

where $R = R_e + R'$, $\bar{\rho} \triangleq 1/(1+\rho)$, and the Rényi entropy of order $\bar{\rho}$ is

$$h_{\bar{\rho}}(\mathbf{Z}_{\mathcal{V}_3}^m) \triangleq \frac{\bar{\rho}}{1-\bar{\rho}}\log\left(\int_{\mathbf{x}} f_{\mathbf{Z}_{\mathcal{V}_3}^m}(\mathbf{x})^{\bar{\rho}} d\mathbf{x}\right)^{\frac{1}{\bar{\rho}}}. \tag{6.95}$$

By [38, Eqn. (208)], we have

$$\rho h_{\bar{\rho}}(\mathbf{Z}_{\mathcal{V}_3}^m) \leq \rho h_{\bar{\rho}}(\mathbf{Z}_m^*) \tag{6.96}$$

and therefore

$$\begin{aligned}
&E_{\Lambda_3}(R; \mathbf{Z}_{\mathcal{V}_3}^m) \geq \max_{0\leq\rho\leq 1} \rho\left[\frac{1}{n}\log V_3 - \frac{1}{n}h_{\bar{\rho}}(\mathbf{Z}_m^*) - R\right] - \epsilon_1(\Lambda_3)\\
&= \max_{0\leq\rho\leq 1} \rho\left[\frac{1}{2}\log 2\pi e P_X - h_{\bar{\rho}}(Z_m^*) - R - \frac{1}{2}\log 2\pi e G(\Lambda_3)\right] - \epsilon_1(\Lambda_3)\\
&\overset{(a)}{\geq} \max_{0\leq\rho\leq 1} \rho\left[\frac{1}{2}\log 2\pi e P_X - h_{\bar{\rho}}(Z_m) - R - \log\left(\frac{R_u}{R_l}\right) - \frac{1}{2}\log 2\pi e G(\Lambda_3)\right] - \epsilon_1(\Lambda_3)\\
&= E_P^r\left(e^{2[C-R-\epsilon_2(\Lambda_3)]}\right) - \epsilon_1(\Lambda_3),
\end{aligned} \tag{6.97}$$

by following the steps in [38, Eqns. (126)–(131)], where (a) makes use of (6.87) and the fact that for the Rényi entropy, $h_\beta(aX) = h_\beta(X) + \log a$, and $\epsilon_1(\Lambda_3), \epsilon_2(\Lambda_3)$ are small as n is large.

This shows that the $(\Lambda_3, \mathbf{Z}_{\mathcal{V}_3}^m)$ channel's random coding exponent is asymptotically close to the random coding Poltyrev exponent at $E_P^r(e^{2(C-R)})$ as n is large, assuming the input $\mathbf{c}_m \oplus \mathbf{a}_l$ is randomly uniform over $\mathcal{V}_3$. Next consider the input taken from the random code ensemble taken from a uniform distribution over $\{p^{-1}\Lambda_3 \cap \mathcal{V}_3\}$. Then, the

random coding error exponent for this code ensemble over the $(\Lambda_3, \mathbf{Z}^m_{\mathcal{V}_3})$ channel is, as proved in [38, Appx. C],

$$E_{\Lambda_3}(R; \mathbf{Z}^m_{\mathcal{V}_3}, p) > E_{\Lambda_3}(R; \mathbf{Z}^m_{\mathcal{V}_3}) - o_n(1). \tag{6.98}$$

The proof for [38, Appx. C] is unchanged as it is performed only for $\{p^{-1}\Lambda_3 \cap \mathcal{V}_3\}$, and not our nested lattices. However, we modify it slightly, noticing that [38, Eqn. (224) in Appx. C] is also obtained using our construction with $\frac{n}{p} \to 0$ as n grows. Using this condition, our nested lattices can have $p^{k_i}, i = 1, 2$ points intersecting with $\mathcal{V}_3$.

Now, referring back to the construction, the jointly decoded codeword $\mathbf{c}_m \oplus \mathbf{a}_l$ can be treated as a combined codeword from $\{\Lambda_1 \cap \mathcal{V}_3\}$. From the properties of the construction, codewords from $\{\Lambda_1 \cap \mathcal{V}_3\}$ are uniformly distributed over $\{p^{-1}\Lambda_3 \cap \mathcal{V}_3\}$, and the difference between two codewords $\bmod \Lambda_3$ from $\{\Lambda_1 \cap \mathcal{V}_3\}$ is also uniformly distributed over $\{p^{-1}\Lambda_3 \cap \mathcal{V}_3\}$. Applying the union bound, we have that codewords from $\{\Lambda_1 \cap \mathcal{V}_3\}$ have the same performance as random codewords drawn uniformly from $\{p^{-1}\Lambda_3 \cap \mathcal{V}_3\}$ in terms of error exponent [94]. Thus, codewords from $\{\Lambda_1 \cap \mathcal{V}_3\}$ over the $(\Lambda_3, \mathbf{Z}^m_{\mathcal{V}_3})$ channel have error probability

$$\Pr\left[\mathbf{Z}^m_{\mathcal{V}_3} \notin \mathcal{V}_1\right] \leq e^{-nE_{\Lambda_3}(R; \mathbf{Z}^m_{\mathcal{V}_3}, p)} \leq e^{-n(E_{\Lambda_3}(R; \mathbf{Z}^m_{\mathcal{V}_3}) - o_n(1))}. \tag{6.99}$$

Combining the results in (6.88)–(6.92), (6.97), (6.99) and following [38, Eqns. (95)–(96)], we obtain

$$\begin{aligned} P_{e,l,m} = \Pr\left[\mathbf{N}' \notin \mathcal{V}_1\right] &\leq e^{\epsilon_1(\Lambda_3).n}\left[\Pr\left[\mathbf{Z}^m_{\mathcal{V}_3} \notin \mathcal{V}_1\right] + \Pr\left[\mathbf{Z}^*_m \notin \mathcal{V}_3\right]\right] \\ &\leq e^{\epsilon_1(\Lambda_3).n}\left[e^{-n(E_P^r(e^{2(C-R)}) - o_n(1))} + e^{-n(E_P^r(e^{2C}) - o_n(1))}\right] \\ &\leq e^{-n(E_P^r(e^{2(C-R)}) - o_n(1))}, \end{aligned} \tag{6.100}$$

since as $n \to \infty$, the second term in the second line above is small. At rates R approaching C, the argument in $E_P^r(.)$ approaches 1 from above, so E_P^r is small but as n gets large, $P_{e,l,m} \to 0$.

Decoding Error Probability for the Eavesdropper's Channel

For the eavesdropper's channel, the proof is similar. We provide it here for completeness.

From [38, Lemmas 6, 11],

$$f_{\mathbf{N}''_z}(\mathbf{x}) \leq e^{\epsilon_1(\Lambda_3).n} f_{\mathbf{Z}^*_z}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{V}_3 \tag{6.101}$$

and $\mathbf{Z}_z^*$ is distributed as $\mathcal{N}(0, P_{Z_z^*}.\mathbf{I}^n)$, with

$$\frac{n}{n+2}.\alpha_z P_{N_z} \leq P_{Z_z^*} < \left(\frac{R_u}{R_l}\right)^n .\alpha_z P_{N_z}. \tag{6.102}$$

The error probability for l with m given can be bounded as

$$P_{e,l}^{(z)} = \Pr\left[\mathbf{N}_z' \notin \mathcal{V}_2\right] \leq e^{\epsilon_1(\Lambda_3).n} \Pr\left[\mathbf{Z}_z^* \notin \mathcal{V}_2\right]. \tag{6.103}$$

Truncating $\mathbf{Z}_z^*$ to the Voronoi region $\mathcal{V}_3$, we obtain the distribution

$$f_{\mathbf{Z}_{\mathcal{V}_3}^z}(\mathbf{x}) = \begin{cases} \frac{1}{1-\Pr\{\mathbf{Z}_z^* \notin \mathcal{V}_3\}} f_{\mathbf{Z}_z^*}(\mathbf{x}) & \mathbf{x} \in \mathcal{V}_3 \\ 0 & \text{otherwise} \end{cases} \tag{6.104}$$

Since $\mathcal{V}_2 \subset \mathcal{V}_3$, we can follow the argument in [38, eqn. (84)-(88)] to have

$$\Pr\left[\mathbf{Z}_z^* \notin \mathcal{V}_2\right] \leq \Pr\left[\mathbf{Z}_{\mathcal{V}_3}^z \notin \mathcal{V}_2\right] + \Pr\left[\mathbf{Z}_z^* \notin \mathcal{V}_3\right]. \tag{6.105}$$

The second term in (6.105) can be bound using (6.14) with the equivalent volume to noise ratio of Λ_3 viewed as a channel code with respect to $\mathbf{Z}_z^*$ given by $\mu = \frac{P_X}{P_{Z_z^*}} \geq 1 + \frac{P_X}{P_{N_z}} - o_n(1) = e^{2C_z} - o_n(1)$, so that

$$\Pr\left[\mathbf{Z}_z^* \notin \mathcal{V}_3\right] \leq e^{-n[E_P(e^{2C_z}) - o_n(1)]}. \tag{6.106}$$

For the first term in (6.105), we consider the $(\Lambda_3, \mathbf{Z}_{\mathcal{V}_3}^z)$ channel (with generic output $\mathbf{Y}$)

$$\mathbf{Y} = [\mathbf{X} + \mathbf{Z}_{\mathcal{V}_3}^z] \mod \Lambda_3. \tag{6.107}$$

which has random coding error exponent

$$E_{\Lambda_3}(R'; \mathbf{Z}_{\mathcal{V}_3}^z) = \max_{0 \leq \rho \leq 1} \rho \left[\frac{1}{n} \log V_3 - \frac{1}{n} h_{\bar{\rho}}(\mathbf{Z}_{\mathcal{V}_3}^z) - R'\right], \tag{6.108}$$

assuming the input $\mathbf{a}_l$ is randomly uniform over $\mathcal{V}_3$. Using $\rho h_{\bar{\rho}}(\mathbf{Z}_{\mathcal{V}_3}^z) \leq \rho h_{\bar{\rho}}(\mathbf{Z}_z^*)$, we get

$$\begin{aligned} E_{\Lambda_3}(R'; \mathbf{Z}_{\mathcal{V}_3}^z) &\geq \max_{0 \leq \rho \leq 1} \rho \left[\frac{1}{n} \log V_3 - \frac{1}{n} h_{\bar{\rho}}(\mathbf{Z}_z^*) - R'\right] - \epsilon_1(\Lambda_3) \\ &\overset{(a)}{\geq} E_P^r\left(e^{2[C_z - R' - \epsilon_2(\Lambda_3)]}\right) - \epsilon_1(\Lambda_3), \end{aligned} \tag{6.109}$$

where (a) is by following the steps in [38, eqn.s (126)-(131)], and $\epsilon_1(\Lambda_3), \epsilon_2(\Lambda_3) \to 0$ as $n \to \infty$ for Λ_3 Rogers good. This shows that the $(\Lambda_3, \mathbf{Z}_{\mathcal{V}_3}^z)$ channel has random

coding error exponent asymptotically close to the random coding Poltyrev exponent as n is large.

The random coding error exponent for the input taken from the random code ensemble taken from a uniform distribution over $\{p^{-1}\Lambda_3 \cap \mathcal{V}_3\}$ over the $(\Lambda_3, \mathbf{Z}^z_{\mathcal{V}_3})$ channel is,

$$E_{\Lambda_3}(R'; \mathbf{Z}^z_{\mathcal{V}_3}, p) > E_{\Lambda_3}(R'; \mathbf{Z}^z_{\mathcal{V}_3}) - o_n(1), \tag{6.110}$$

under the condition that $\frac{n}{p} \to 0$ as n grows.

Next, the codeword $\mathbf{a}_l \mod \Lambda_3 \in \{\Lambda_2 \cap \mathcal{V}_3\}$ is uniformly distributed over $\{\Lambda_2 \cap \mathcal{V}_3\}$. From the properties of the construction, we know that codewords from $\{\Lambda_2 \cap \mathcal{V}_3\}$ are uniformly distributed over $\{p^{-1}\Lambda_3 \cap \mathcal{V}_3\}$, and the difference between two codewords $\mod \Lambda_3$ from $\{\Lambda_2 \cap \mathcal{V}_3\}$ is also uniformly distributed over $\{p^{-1}\Lambda_3 \cap \mathcal{V}_3\}$. By the union bound, we have that codewords from $\{\Lambda_2 \cap \mathcal{V}_3\}$ have the same performance as codewords from $\{p^{-1}\Lambda_3 \cap \mathcal{V}_3\}$ [94]. Thus, we have, following the steps in [38, eqn.s (95)-(96)],

$$P^{(z)}_{e,l} = \Pr\{\mathbf{N}'_z \notin \mathcal{V}_2\} \leq e^{-n(E_P(e^{2(C_z - R')}) - o_n(1))}, \tag{6.111}$$

which is small for R' approaching C_z and n large.

Now (6.100), (6.111) show that the decoding error probability at the main and eavesdropper's channels are small for n large and the error exponents achieve the random coding Poltyrev exponent at $E^r_P(e^{2(C-R)})$ and $E^r_P(e^{2(C_z-R')})$, with coding rates R and R' that approach C and C_z, respectively. The constructed nested lattices that achieve the above, $\Lambda_3 \subset \Lambda_2 \subset \Lambda_1$, are all Rogers and Poltyrev good; Λ_3 is also quantization good.

6.5 Conclusion

We showed that using a chain of nested lattices $\Lambda_3 \subset \Lambda_2 \subset \Lambda_1$, lattice coding and decoding can achieve the secrecy rate of the Gaussian wiretap channel; we need the sequence of lattices $\Lambda_1^{(n)}$ and $\Lambda_2^{(n)}$ to be AWGN good, and the sequence $\Lambda_3^{(n)}$ to be good for quantization. We considered a decoder at the legitimate receiver which jointly decoded the transmitted codeword made up of the message bits and random bits, and a lattice construction $\Lambda_3 \subset \Lambda_2 \subset \Lambda_1$ with all lattices Rogers good and Poltyrev good, Λ_3

is also quantization good, and k_1, k_2, p growing appropriately with n. We could then show the achievability of probability of decoding error going to zero at rates approaching the capacities of the main and eavesdropper's channels.

There are some open problems that may be explored in the future. We first notice that in the construction, the coarse lattice to start off the construction Λ_3 was not determined explicitly. It may be interesting to specify Λ_3 explicitly instead of using an existence proof.

At the legitimate receiver, decoding message and random bits separately with a staged decoder, may be considered. This may give us more insight on the requirements of Λ_1 and Λ_2. The difficulty with the staged decoder is that we need to determine the noise distribution of the second stage, given that the first stage occurs some small error.

At the eavesdropper, using a joint decoder for $(\mathbf{c}_m, \mathbf{a}_l)$ may strengthen our argument; while using a staged decoder will perhaps give more insight on the requirements of Λ_1 and Λ_2. Both problems require us to show that the error probability goes to 1 for n large, which requires a lower bound to the error probability. Deriving the lower bound is quite difficult, however. We illustrate the problem encountered by discussing the joint decoder at the eavesdropper below.

We know that $f_{\mathbf{N}''_z}(\mathbf{x}) \leq e^{n.\epsilon(\Lambda_3)} f_{\mathbf{Z}^*_z}(\mathbf{x})$, for $\mathbf{x} \in \mathcal{V}_3$. We can deduce the p.d.f. of $\mathbf{N}'_z$ from the arguments given in Forney *et al* [46]. Firstly, the Λ_3-aliased r.v. $\mathbf{N}'_z \in \mathcal{V}_3$. Secondly, $\mathbf{x}' \in \mathbb{R}^n$ maps to $\mathbf{x} \in \mathcal{V}_3$ if and only if $\mathbf{x}' \in \Lambda_3 + \mathbf{x}$. So, the p.d.f. of $\mathbf{N}'_z$ is

$$f_{\mathbf{N}'_z}(\mathbf{x}) = \sum_{\mathbf{b}\in\Lambda_3} f_{\mathbf{N}''_z}(\mathbf{x}+\mathbf{b}) \leq e^{n.\epsilon(\Lambda_3)} \sum_{\mathbf{b}\in\Lambda_3} f_{\mathbf{Z}^*_z}(\mathbf{x}+\mathbf{b}), \quad \mathbf{x} \in \mathcal{V}_3, \tag{6.112}$$

where the last inequality follows from the bound on the p.d.f. $f_{\mathbf{N}''_z}(\mathbf{x})$ above. The error probability for the eavesdropper for a given pair $\mathbf{c}_m + \mathbf{a}_l$ is

$$P^{(z)}_{e,l,m} = \Pr[\mathbf{N}'_z \notin \mathcal{V}_1(\mathbf{c}_m + \mathbf{a}_l)] = \Pr[\mathbf{N}'_z \notin \mathcal{V}_1], \tag{6.113}$$

where we write the second equality due to the congruency of the Voronoi regions for the lattice. We then have

$$\begin{aligned} P^{(z)}_{e,l,m} &= 1 - \int_{\mathcal{V}_1} f_{\mathbf{N}'_z}(\mathbf{x}) d\mathbf{x} \\ &\geq 1 - e^{n.\epsilon(\Lambda_3)} \sum_{\mathbf{b}\in\Lambda_3} \int_{\mathcal{V}_1} f_{\mathbf{Z}^*_z}(\mathbf{x}+\mathbf{b}) d\mathbf{x} \\ &\overset{(a)}{=} 1 - e^{n.\epsilon(\Lambda_3)} \sum_{\mathbf{b}\in\Lambda_3} \int_{\mathcal{V}_1+\mathbf{b}} f_{\mathbf{Z}^*_z}(\mathbf{u}) d\mathbf{u} \end{aligned}$$

$$\stackrel{(b)}{=} 1 - e^{n.\epsilon(\Lambda_3)} \left[\int_{\mathcal{V}_1} f_{\mathbf{Z}_z^*}(\mathbf{u}) d\mathbf{u} + \sum_{\mathbf{b} \in \Lambda_3 / \{\mathbf{0}\}} \int_{\mathcal{V}_1 + \mathbf{b}} f_{\mathbf{Z}_z^*}(\mathbf{u}) d\mathbf{u} \right] \tag{6.114}$$

where (a) is due to changing the variable of integration to $\mathbf{u} = \mathbf{x} + \mathbf{b}$, and (b) is by writing the sum over Λ_3 as the origin plus the sum over Λ_3 less the origin.

The first integral in (6.114) can be bound using the equivalent sphere argument, as found in Tarokh *et al* [109]. Let $\mathrm{Ball}(r)$ denote a hypersphere in $\mathbb{R}^n$, with radius $r > 0$ and the same volume as $\mathcal{V}_1$, centered at the origin, so that $\mathrm{Ball}(r) \triangleq \{\mathbf{x} \in \mathbb{R}^n, \|\mathbf{x}\| < r\}$. Denote the equivalent sphere of $\mathcal{V}_1$ as $\mathcal{S}_{\mathcal{V}_1}$. Then, using the arguments in Tarokh *et al* [109], we have the following:

$$\int_{\mathcal{V}_1} f_{\mathbf{Z}_z^*}(\mathbf{u}) d\mathbf{u} \leq \int_{\mathcal{S}_{\mathcal{V}_1}} f_{\mathbf{Z}_z^*}(\mathbf{u}) d\mathbf{u} \triangleq \Pr\left[\mathbf{Z}_z^* \in \mathrm{Ball}(r)\right] = \Pr\left[\|\mathbf{Z}_z^*\| \leq r\right]. \tag{6.115}$$

The error probability can now be written as

$$P_{e,l,m}^{(z)} \geq 1 - e^{n.\epsilon(\Lambda_3)} \left[\Pr\left[\|\mathbf{Z}_z^*\| \leq r\right] + \sum_{\mathbf{b} \in \Lambda_3 / \{\mathbf{0}\}} \int_{\mathcal{V}_1 + \mathbf{b}} f_{\mathbf{Z}_z^*}(\mathbf{u}) d\mathbf{u} \right]. \tag{6.116}$$

Unfortunately, the term with the sum cannot be readily evaluated, whether by using an equivalent sphere argument, or otherwise. Another very useful direction for future work will be to see how the error probability at the eavesdropper can be evaluated using this type of argument.

Chapter 7

Conclusions and Future Work

Wireless communications channels today are vulnerable to eavesdropping due to the open nature of the channel, making the characterization of transmission rates for secure and reliable communication for the physical layer an important issue. We use the information theoretic approach to gain fundamental insight into the secure (confidential) codes that give rise to limits on the reliable and secure communication rates between nodes in a network. The challenge is to find coding schemes that have provable security and reliability.

7.1 Summary of Contributions and Insights

In Chapter 4, we have investigated secure coding schemes for the BC, which is an important building block of a network.

- We have characterized the secure and reliable transmission rates for the class of the K-receiver BC with an external eavesdropper, where the receivers and eavesdropper are degraded in the order $X \rightarrow Y_1 \rightarrow .. \rightarrow Y_K \rightarrow Z$. In our coding scheme, we use superposition coding and code partitioning and we have found the secrecy capacity, which is the maximum secure rate achievable.

- For another more general class of BC, which is the 3-receiver BC with DMS, we have found the rate equivocation region with one of the receivers being an eavesdropper. The 3-receiver BC with DMS is an important channel model which gives insights to coding for the general K-receiver BC, whose capacity region is still unknown, even for the case of no security. The 3-receiver BC with DMS and an eavesdropper also generalizes some 2- and 3-receiver BC models with

an eavesdropper. Our secure scheme uses code partitioning and double binning and we have shown that the scheme is secure for the 3-receiver BC with 2 DMS. We can also see that our secure scheme can be straightforwardly used to provide security for the even more general 3-receiver BC with 3 DMS. The BC coding schemes suggest a multilevel code for the K-receiver BC and a multilevel dirty paper code for the 3-receiver BC with DMS.

In Chapter 5, we considered the scenario where the eavesdropper had more favorable channel conditions compared to the legitimate receiver, where it may be possible that the secrecy rate goes to zero. We used a CJ method using a bank of relays, in conjunction with a distributed signal processing method to enhance the secrecy rate, thus proving the fundamental result that we can improve the secrecy rate even if the eavesdropper has a better channel than the legitimate receiver. In addition,

- We derived the conditions for positive secrecy rate and obtained the optimal CJ solution by a combination of convex optimization and a one-dimensional search.
- We also proposed extensions to power constraints for grouped relays and a distributed implementation for the relay power assignment.

In Chapter 6, we considered lattice codes which impose more structure than the random codes, to implement the coset coding for the Gaussian wiretap channel, and adopted an information-theoretic approach to the lattice-based coset coding problem.

- We used a nested lattice chain to perform lattice coding and proposed lattice decoders for the Gaussian wiretap channel; in the thesis we considered a decoder at the legitimate receiver which jointly decoded the transmitted codeword made up of the message bits and random bits and derived the achievable rates and the equivocation rate. We also showed that it is possible to achieve the equivocation rate of the classical Gaussian wiretap channel (the secrecy capacity).
- A construction for the nested lattice chain was proposed and analyzed; a coset code based on this chain was shown to be able to meet the reliability and security criteria. We were able to state the coding requirements on the nested lattices, which is an important step forward in the information-theoretic lattice-based coset coding problem.

7.2 Future Work

We can identify the following directions for future work:

1. We have already seen that the 3-receiver with 3 DMS coding scheme without security essentially 'rides' on the 3-receiver with 2 DMS coding scheme. Thus our secure coding scheme should be able to provide security for the 3-receiver with 3 DMS as well; this is our current work.

2. For the general 3-receiver BC with DMS, no outer bound exists; the challenge is to derive an outer bound by circumventing the use of the Csiszár sum lemma, or to derive an alternative to it. Some progress on this has been made by Nair and Wang [92], but only for the less noisy channel condition. An alternative to the Csiszár sum lemma, or an alternative method to derive the outer bound, for general conditions on 3 receivers, would be a most welcome contribution, as this would open many possibilities for multi-user outer bound derivations. It would also help us to finally be able to quantify the outer bound on the rate-equivocation region for the general 3-receiver BC with DMS.

3. We would like to find secure coding schemes for the two-way relay channel, which models an important part of a network, where the base station has to handle messages from two different cells in a cellular network, for example. This becomes more important as decentralized networks are deployed, as decentralized base 'stations' may be just relays with minimal hardware. A key area where the present work such as in Mukherjee and Swindlehurst [89] concerning an external eavesdropper to a two-way relay network did not address was a information theoretically provable secure coding scheme for both phases of the transmission, that is both MAC (uplink) and BC (downlink) phases. For a parallel problem without security, see Zhang and Gursoy [124]. Thus filling this gap is a very interesting problem by itself. This may involve putting together secure coding schemes for the MAC, BC and relay channels.

4. In the wiretap channel with a bank of relays performing CJ, an open problem is the case for multiple eavesdroppers. A way forward may be treat the channel as a compound wiretap channel [72]. Another line of work may be to split the re-

lays to perform message forwarding and jamming, and find the optimal, dynamic relay assignment. As the CSI issue is an important one, such future work should include unknown CSI models at the eavesdropper.

5. For our lattice coding in the Gaussian wiretap channel, future work will include strengthening the proof for the eavesdropper, such as assuming it performs joint decoding of the message and random bits, and showing that the error probability goes to one. Another possibility would be to use staged decoders for both the legitimate receiver and eavesdropper, where the message and random bits are decoded separately.

6. We know from the coset coding framework of Forney [42, 43] that it is possible to construct the lattice chain $\Lambda_L \subseteq \Lambda_{L-1} \subseteq \cdots \subseteq \Lambda_2 \subseteq \Lambda_1$, so that a multilevel coset code results. Each level or partition can be viewed as a code for a user, and coding and decoding can be carried out independently of other levels [46]. This way of viewing multilevel coding can make it possible to design codes for multiple eavesdroppers, or the compound wiretap channel, by assigning level 1 to the main channel, level 2 to eavesdropper's channel 1, etc. The true advantage in using multilevel coset coding may lie in the extra degree of freedom that it gives to the code designer, so that there is freedom to assign functions other than channel coding to some levels.

7. We know from the random coding results that security is provided for any coding scheme based on binning by using double binning. This was seen in the work on the general BC with DMS, or the BC with 2 receivers and messages to be kept secret from either receiver [78, 116]. Thus multilevel coset coding will be able to provide security using double binning; we need to use, for example, the lattice chain $\Lambda_4 \subseteq \Lambda_3 \subseteq \Lambda_2 \subseteq \Lambda_1$. The partition Λ_4/Λ_3 performs the additional binning. Then, the possible directions would be to provide security for the lattice-based dirty paper coding scheme of Erez and ten Brink [40] and ultimately to form a lattice-based secure coding scheme for either the 2-receiver BC with confidential messages of [78, 116] or the 3-receiver BC with confidential messages and DMS.

8. Lastly, it is also very interesting to consider active eavesdroppers who jam the

network. While some work has been done by Amariucai and Wei[5], designing more practical or structured coding schemes are open problems; a multilevel coding scheme appears to be useful in this possible future work.

To sum up, the information theoretic approach that we have studied in this thesis enables us to provide provable security in network communication scenarios without keys, which is highly beneficial to wireless network design. Coding schemes designed using the information theoretic approach can be further combined with cryptographic schemes to provide robust cross layer security for wireless networks.

Appendix A

On the Ordering of Channels

The definitions given here are from Körner and Marton [66]. Let P_1 and P_2 denote discrete memoryless channels with probability transition matrices $p_1(y|x)$ and $p_2(y|x)$ and the variables $x \in \mathcal{X}$, $y \in \mathcal{Y}$, and $z \in \mathcal{Z}$.

Definition 1 A channel P_2 is the degraded form of P_1 if there exists a probability transition matrix $p_3(z|y)$ such that

$$p_2(z|x) = \sum_{y \in \mathcal{Y}} p_1(y|x) p_3(z|y). \tag{A.1}$$

Definition 2 Channel P_1 is said to be less noisy than P_2 if

$$I(U;Z) \leq I(U;Y) \tag{A.2}$$

for every probability mass function of the form $p(u,x,y,z) = p(u)p(x|u)p(y,z|x)$.

Definition 3 Channel P_1 is said to be more capable than P_2 if

$$I(X;Z) \leq I(X;Y) \tag{A.3}$$

for all probability distributions on $\mathcal{X}$. Körner and Marton [66] also showed that the more capable condition also implies that

$$I(X;Z|U) \leq I(X;Y|U) \tag{A.4}$$

for every probability mass function of the form $p(u,x,y,z) = p(u)p(x|u)p(y,z|x)$.

Lastly, we note that the degraded condition is the strongest, followed by the less noisy condition, followed by the more capable. So if a channel is degraded, for example, it is true that it is also less noisy, but it is not always true the other way around.

Appendix B

Proofs for Chapter 4

B.1 Proof for Lemma 4

We now bound the entropy $H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k)$ for every w_k, following a method in [35]. To begin, we label the $L_k = 2^{nR_k}$ bins from $\mathbf{U}_k$ indexed by w_k corresponding to the message w_k as $\mathcal{B}_k(w_k)$. We then fix $J_k = j_k$ and sequences $(\mathbf{u}_{k+1}, \mathbf{z}) \in \mathcal{T}_\epsilon^n(P_{U_{k+1}Z})$. Let the eavesdropper's estimate of the index j_k be $\widehat{j_k}$, and let us define the set $\mathcal{A}(\widehat{j_k}, j_k)$ as the set of the eavesdropper's estimate $\widehat{j_k}$ of the index j_k that is not equal to the actual transmitted index j_k:

$$\mathcal{A}(\widehat{j_k}, j_k) \triangleq \left\{\widehat{j_k} \in \mathcal{B}_k(w_k) : \left(\mathbf{U}_k(\widehat{j_k}), \mathbf{u}_{k+1}, \mathbf{z}\right) \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z}), \widehat{j_k} \neq j_k\right\}. \quad \text{(B.1)}$$

Let us define the random event

$$\mathcal{E}(\widehat{j_k}, j_k) \triangleq \left\{\left|\mathcal{A}(\widehat{j_k}, j_k)\right| \geq 2\mathrm{E}\left[\left|\mathcal{A}(\widehat{j_k}, j_k)\right|\right]\right\}, \quad \text{(B.2)}$$

where $\mathrm{E}[.]$ denotes the mean. We will now proceed to show that $\Pr\left[\mathcal{E}(\widehat{j_k}, j_k)\right] \to 0$ for n sufficiently large under the appropriate conditions. To do this, we use the version of Chebyshev's inequality stated in Lemma 2. Specifically, we use the second of the inequalities in (2.40) and set $\nu = 1$. That is, if X is a generic r.v. with mean $\mathrm{E}(X)$ and variance $\mathrm{Var}(X)$, then

$$\Pr\{X \geq 2\mathrm{E}(X)\} \leq \frac{\mathrm{Var}(X)}{(\mathrm{E}(X))^2}.$$

So to find the probability of the random event $\mathcal{E}(\widehat{j_k}, j_k)$, we let the generic r.v. in the equation above be $\left|\mathcal{A}(\widehat{j_k}, j_k)\right|$. We now require the mean and variance of $\left|\mathcal{A}(\widehat{j_k}, j_k)\right|$.

For the mean, we have

$$\begin{aligned}\mathrm{E}\left[\left|\mathcal{A}(\widehat{j_k}, j_k)\right|\right] &= \sum_{\widehat{j_k}, \widehat{j_k} \neq j_k} \Pr\left[\left(\mathbf{U}_k(\widehat{j_k}), \mathbf{u}_{k+1}, \mathbf{z}\right) \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z})\right] \\ &= \left(2^{nR'_k} - 1\right) \Pr\left[\left(\mathbf{U}_k(\widehat{j_k}), \mathbf{u}_{k+1}, \mathbf{z}\right) \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z})\right]. \qquad \text{(B.3)}\end{aligned}$$

We have, for $\epsilon \to 0$ for n sufficiently large

$$\begin{aligned}\Pr\left[\left(\mathbf{U}_k(\widehat{j_k}), \mathbf{u}_{k+1}, \mathbf{z}\right) \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z})\right] &\overset{(a)}{=} \sum_{\mathbf{u}_k \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z} | \mathbf{u}_{k+1}, \mathbf{z})} p(\mathbf{u}_k | \mathbf{u}_{k+1}) \\ &\overset{(b)}{\leq} \left|\mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z} | \mathbf{u}_{k+1}, \mathbf{z})\right| . 2^{-nH(U_k|U_{k+1})(1-\epsilon)} \\ &\overset{(c)}{\leq} 2^{nH(U_k|Z,U_{k+1})(1+\epsilon)} . 2^{-nH(U_k|U_{k+1})(1-\epsilon)} \\ &= 2^{-n(I(U_k;Z|U_{k+1}) - \epsilon[H(U_k|U_{k+1}) + H(U_k|Z,U_{k+1})])} \\ &= 2^{-n(I(U_k;Z|U_{k+1}) - \delta(\epsilon))} \qquad \text{(B.4)}\end{aligned}$$

where (a) arises from the code generation process; (b) and (c) are from the properties of typical sequences and the size of the typical set, respectively from (2.22) and (2.23) from Theorem 1; and $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ for n sufficiently large. Similarly using Theorem 1, we have

$$\begin{aligned}\Pr\left[\left(\mathbf{U}_k(\widehat{j_k}), \mathbf{u}_{k+1}, \mathbf{z}\right) \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z})\right] &\geq \left|\mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z} | \mathbf{u}_{k+1}, \mathbf{z})\right| . 2^{-nH(U_k|U_{k+1})(1+\epsilon)} \\ &\geq (1-\epsilon) . 2^{nH(U_k|Z,U_{k+1})(1-\epsilon)} . 2^{-nH(U_k|U_{k+1})(1+\epsilon)} \\ &= (1-\epsilon) . 2^{-n(I(U_k;Z|U_{k+1}) + \delta(\epsilon))}, \qquad \text{(B.5)}\end{aligned}$$

where, as before, $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ for n sufficiently large. Subsituting into (B.3), we can see that, for n sufficiently large,

$$2^{n(R'_k - I((U_k;Z|U_{k+1}) - \delta(\epsilon))} \leq \mathrm{E}\left[\left|\mathcal{A}(\widehat{j_k}, j_k)\right|\right] \leq 2^{n(R'_k - I((U_k;Z|U_{k+1}) + \delta(\epsilon))}. \qquad \text{(B.6)}$$

Next, we have

$$\begin{aligned}\mathrm{E}\left[\left|\mathcal{A}(\widehat{j_k}, j_k)\right|^2\right] &= \sum_{\widehat{j_k}, \widehat{j_k} \neq j_k} \Pr\left[\left(\mathbf{U}_k(\widehat{j_k}), \mathbf{u}_{k+1}, \mathbf{z}\right) \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z})\right] \\ &+ \sum_{\widehat{j_k}, \widehat{j_k} \neq j_k} \sum_{\widehat{\widehat{j_k}} \neq \widehat{j_k}, \widehat{\widehat{j_k}} \neq j_k} \Pr\left[\begin{array}{l}\left(\mathbf{U}_k(\widehat{j_k}), \mathbf{u}_{k+1}, \mathbf{z}\right) \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z}), \\ \left(\mathbf{U}_k(\widehat{\widehat{j_k}}), \mathbf{u}_{k+1}, \mathbf{z}\right) \in \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z})\end{array}\right]. \qquad \text{(B.7)}\end{aligned}$$

Now, if we let

$$p_1 \triangleq \Pr\left[\left(\mathbf{U}_k(\widehat{j_k}), \mathbf{u}_{k+1}, \mathbf{z}\right) \in \mathcal{T}_\epsilon^n(P_{U_kU_{k+1}Z})\right],$$
$$p_2 \triangleq \Pr\left[\left(\mathbf{U}_k(\widehat{j_k}), \mathbf{u}_{k+1}, \mathbf{z}\right) \in \mathcal{T}_\epsilon^n(P_{U_kU_{k+1}Z}), \left(\mathbf{U}_k(\widehat{\widehat{j_k}}), \mathbf{u}_{k+1}, \mathbf{z}\right) \in \mathcal{T}_\epsilon^n(P_{U_kU_{k+1}Z})\right], \tag{B.8}$$

and noting that $p_2 = p_1^2$, we now have

$$\mathrm{E}\left[\left|\mathcal{A}(\widehat{j_k}, j_k)\right|^2\right] \leq 2^{nR'_k} p_1 + 2^{2nR'_k} p_1^2.$$

Thus we have

$$\mathrm{Var}\left[\left|\mathcal{A}(\widehat{j_k}, j_k)\right|\right] = \mathrm{E}\left[\left|\mathcal{A}(\widehat{j_k}, j_k)\right|^2\right] - \mathrm{E}\left[\left|\mathcal{A}(\widehat{j_k}, j_k)\right|\right]^2 \leq 2^{nR'_k} p_1 \tag{B.9}$$

since $\mathrm{E}\left[\left|\mathcal{A}(\widehat{j_k}, j_k)\right|\right]^2 = 2^{2nR'_k} p_1^2$. Finally, as p_1 is already upper bounded in (B.4) and $\mathrm{E}\left[\left|\mathcal{A}(\widehat{j_k}, j_k)\right|\right]$ is lower bounded in (B.6), we have

$$\Pr\left[\mathcal{E}(\widehat{j_k}, j_k)\right] \leq \frac{\mathrm{Var}\left[\left|\mathcal{A}(\widehat{j_k}, j_k)\right|\right]}{\mathrm{E}\left[\left|\mathcal{A}(\widehat{j_k}, j_k)\right|\right]^2} \leq 2^{-n(R'_k - I(U_k; Z|U_{k+1}) - 3\delta(\epsilon))}, \tag{B.10}$$

which becomes small for n sufficiently large provided that $R'_k \geq I(U_k; Z|U_{k+1})$.

Next, for each w_k, we define a random version of $\mathcal{A}(\widehat{j_k}, j_k)$ in (B.1) as

$$\widetilde{\mathcal{A}}(\widehat{j_k}, j_k) \triangleq \left\{\widehat{j_k} \in \mathcal{B}_k(w_k) : \left(\mathbf{U}_k(\widehat{j_k}), \mathbf{u}_{k+1}, \mathbf{Z}\right) \in \mathcal{T}_\epsilon^n(P_{U_kU_{k+1}Z}), \widehat{j_k} \neq J_k\right\}. \tag{B.11}$$

That is, the set of the eavesdropper's estimate of the index j_k that is not equal to the random transmitted index, given *random* eavesdropper received signal. We also define the event

$$\mathcal{E}(w_k) \triangleq \left\{\left|\widetilde{\mathcal{A}}(\widehat{j_k}, j_k)\right| \geq 2\mathrm{E}\left[\left|\widetilde{\mathcal{A}}(\widehat{j_k}, j_k)\right|\right]\right\}, \tag{B.12}$$

and the indicator variables

$$\mathbb{I}(w_k) := 0 \text{ if } \left(\mathbf{U}_k(\widehat{J_k}), \mathbf{u}_{k+1}, \mathbf{Z}\right) \in \mathcal{T}_\epsilon^n(P_{U_kU_{k+1}Z}) \text{ and } \mathcal{E}(w_k)^c \text{ occurs},$$
$$\mathbb{I}(w_k) := 1 \text{ if } \left(\mathbf{U}_k(\widehat{J_k}), \mathbf{u}_{k+1}, \mathbf{Z}\right) \notin \mathcal{T}_\epsilon^n(P_{U_kU_{k+1}Z}) \text{ and } \mathcal{E}(w_k) \text{ occurs}.$$

Let us now find the probability that $\mathbb{I}(w_k) = 1$. We have, by the union bound,

$$\begin{aligned}
&\Pr[\mathbb{I}(w_k) = 1] \leq \Pr\left[\left(\mathbf{U}_k(\widehat{J_k}), \mathbf{u}_{k+1}, \mathbf{Z}\right) \notin \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z})\right] + \Pr[\mathcal{E}(w_k)] \\
&\leq \Pr\left[\left(\mathbf{U}_k(\widehat{J_k}), \mathbf{u}_{k+1}, \mathbf{Z}\right) \notin \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z})\right] + \sum_{\mathbf{z} \in \mathcal{T}_\epsilon^n(P_Z)} p(\mathbf{z}) \Pr\left[\mathcal{E}(w_k) | \mathbf{Z} = \mathbf{z}\right] \\
&\quad + \Pr\left[\mathbf{Z} \notin \mathcal{T}_\epsilon^n(P_Z)\right] \\
&\overset{(a)}{=} \sum_{\mathbf{z} \in \mathcal{T}_\epsilon^n(P_Z)} \sum_{j_k} p(\mathbf{z}) p(j_k|\mathbf{z}) \Pr[\mathcal{E}(w_k) | \mathbf{Z} = \mathbf{z}, J_k = j_k] \\
&= \sum_{\mathbf{z} \in \mathcal{T}_\epsilon^n(P_Z)} \sum_{j_k} p(\mathbf{z}) p(j_k|\mathbf{z}) \Pr[\mathcal{E}(\widehat{j_k}, j_k)], \qquad \text{(B.13)}
\end{aligned}$$

where (a) is by the fact that $\Pr\left[\left(\mathbf{U}_k(\widehat{J_k}), \mathbf{u}_{k+1}, \mathbf{Z}\right) \notin \mathcal{T}_\epsilon^n(P_{U_k U_{k+1} Z})\right] \to 0$ for n sufficiently large by the properties of joint typical sequences, and this implies that $\Pr\left[\mathbf{Z} \notin \mathcal{T}_\epsilon^n(P_Z)\right]$ also goes to 0 for n sufficiently large. Finally, since we know that $\Pr[\mathcal{E}(\widehat{j_k}, j_k)] \to 0$ for n sufficiently large if $R'_k \geq I(U_k; Z|U_{k+1})$, we have $\Pr[\mathbb{I}(w_k) = 1] \to 0$ for n sufficiently large. To bound $H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k)$, we consider the expansion of $H(\mathbb{I}(w_k), J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k)$. We have

$$\begin{aligned}
H(\mathbb{I}(w_k), J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k) &= H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k) + H(\mathbb{I}(w_k)|\mathbf{Z}, \mathbf{U}_{k+1}, J_k, w_k) \\
&= H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k), \qquad \text{(B.14)}
\end{aligned}$$

where the second equality is because $H(\mathbb{I}(w_k)|\mathbf{Z}, \mathbf{U}_{k+1}, J_k, w_k) = 0$ as $\mathbb{I}(w_k)$ is determined by $\mathbf{Z}, J_k, w_k$. So we now have

$$\begin{aligned}
&H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k) = H(\mathbb{I}(w_k), J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k) \\
&= H(\mathbb{I}(w_k)|\mathbf{Z}, \mathbf{U}_{k+1}, w_k) + \Pr[\mathbb{I}(w_k) = 1] H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k, \mathbb{I}(w_k) = 1) \\
&\quad + \Pr[\mathbb{I}(w_k) = 0] H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k, \mathbb{I}(w_k) = 0) \\
&\overset{(a)}{\leq} 1 + \Pr[\mathbb{I}(w_k) = 1] H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k, \mathbb{I}(w_k) = 1) \\
&\quad + (1 - \Pr[\mathbb{I}(w_k) = 1]) H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k, \mathbb{I}(w_k) = 0) \\
&\leq 1 + \Pr[\mathbb{I}(w_k) = 1] H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k, \mathbb{I}(w_k) = 1) \\
&\quad + H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k, \mathbb{I}(w_k) = 0), \qquad \text{(B.15)}
\end{aligned}$$

where (a) is due to $H(\mathbb{I}(w_k)|\mathbf{Z}, \mathbf{U}_{k+1}, w_k) \leq H(\mathbb{I}(w_k)) \leq 1$, since $\mathbb{I}(w_k)$ is binary-valued. In the last line of (B.15), the second term $\to 0$ for n sufficiently large. For the third term, we know that, given $\mathbf{U}_{k+1}, \mathbf{Z}, w_k$ and $\mathbb{I}(w_k) = 0$, J_k takes on

$\leq \log(2\mathrm{E}[|\mathcal{A}(\widehat{j_k}, j_k)|] - 1)$ values. The third term of the last line of (B.15) now becomes

$$\begin{aligned} H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k, \mathbb{I}(w_k) = 0) &\leq \log(2\mathrm{E}[|\mathcal{A}(\widehat{j_k}, j_k)|] - 1) \\ &= \log\left(2^{n(R'_k - I(U_k;Z|U_{k+1}) + \delta(\epsilon))+1} - 1\right) \\ &= n(R'_k - I(U_k; Z|U_{k+1}) + \delta(\epsilon)) + 1 + \log\left(1 - 2^{-n(R'_k - I(U_k;Z|U_{k+1})+\delta(\epsilon))-1}\right) \\ &\overset{(a)}{\leq} n(R'_k - I(U_k; Z|U_{k+1}) + \delta(\epsilon)) + 1 - 2^{-n(R'_k - I(U_k;Z|U_{k+1})+\delta(\epsilon))-1}, \end{aligned} \tag{B.16}$$

where (a) is due to using the relation $\log(x) \leq x - 1$. Then, we have, for n sufficiently large,

$$H(J_k|\mathbf{Z}, \mathbf{U}_{k+1}, w_k) \leq n(R'_k - I(U_k; Z|U_{k+1}) + \delta(\epsilon)) + 2, \tag{B.17}$$

since $R'_k \geq I(U_k; Z|U_{k+1})$. The lemma is then proved by averaging over the w_k.

B.2 Proof for Lemma 5

Here we obtain a bound for the entropy $H(J_1, \cdots, J_K|\mathbf{Z}, w_1, \cdots, w_K)$ for every $w_1, \cdots, w_K$, with a similar, only more elaborate, method as in the proof for Lemma 4 in the previous section.

We label the $L_k = 2^{nR_k}$ bins from $\mathbf{U}_k$ indexed by w_k corresponding to the message w_k as $\mathcal{B}_k(w_k)$, for $k = 1, \cdots, K$. For $k = 1, \cdots, K$, fix $J_k = j_k$ and a sequence $\mathbf{z} \in \mathcal{T}_\epsilon^n(P_Z)$. Now, define the set $\mathcal{A}(\widehat{j_1}, \cdots, \widehat{j_K}, j_1, \cdots, j_K)$ as the set of the eavesdropper's estimate $(\widehat{j_1}, \cdots, \widehat{j_K})$ that is not equal to the actual transmitted indices $(j_1, \cdots, j_K)$:

$$\mathcal{A}(\widehat{j_1}, \cdots, \widehat{j_K}, j_1, \cdots, j_K) \triangleq \left\{ \begin{array}{c} \widehat{j_k} \in \mathcal{B}_k(w_k), k = 1, \cdots, K : \\ \left(\mathbf{U}_1(\widehat{j_1}), \cdots, \mathbf{U}_K(\widehat{j_K}), \mathbf{z}\right) \in \mathcal{T}_\epsilon^n(P_{U_1 \cdots U_K Z}), \\ (\widehat{j_1}, \cdots, \widehat{j_K}) \neq (j_1, \cdots, j_K) \end{array} \right\}. \tag{B.18}$$

We define the random event

$$\begin{aligned} \mathcal{E}(\widehat{j_1}, \cdots, \widehat{j_K}, j_1, \cdots, j_K) \triangleq \Big\{ &\Big|\mathcal{A}(\widehat{j_1}, \cdots, \widehat{j_K}, j_1, \cdots, j_K)\Big| \\ &\geq 2\mathrm{E}\left[\Big|\mathcal{A}(\widehat{j_1}, \cdots, \widehat{j_K}, j_1, \cdots, j_K)\Big|\right]\Big\}, \end{aligned} \tag{B.19}$$

and we aim to show that $\Pr\left[\mathcal{E}(\widehat{j_1}, \cdots, \widehat{j_k}, j_1, \cdots, j_k)\right] \to 0$ for n sufficiently large under the appropriate conditions, using the version of Chebyshev's inequality stated in Lemma 2, as before. We now evaluate the mean and variance of

$\left|\mathcal{A}(\widehat{j_1},\cdots,\widehat{j_K},j_1,\cdots,j_k)\right|$. For the mean, we have

$$\begin{aligned}
&\mathrm{E}\left[\left|\mathcal{A}(\widehat{j_1},\cdots,\widehat{j_K},j_1,\cdots,j_K)\right|\right] \\
&= \sum_{\widehat{j_1},\cdots,\widehat{j_K},(\widehat{j_1},\cdots,\widehat{j_K})\neq j_1,\cdots,j_K} \Pr\left[\left(\mathbf{U}_1(\widehat{j_1}),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{z}\right)\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_K Z})\right] \\
&= \prod_{k=1}^{K}\left(2^{nR'_k}-1\right).\Pr\left[\left(\mathbf{U}_1(\widehat{j_1}),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{z}\right)\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_K Z})\right]. \qquad \text{(B.20)}
\end{aligned}$$

The probability $\Pr\left[\left(\mathbf{U}_1(\widehat{j_1}),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{z}\right)\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_K Z})\right]$ in (B.20) can be upper bounded as, with $\epsilon\to 0$ for n sufficiently large,

$$\begin{aligned}
&\Pr\left[\left(\mathbf{U}_1(\widehat{j_1}),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{z}\right)\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_K Z})\right] \\
&\overset{(a)}{=} \sum_{(\mathbf{u}_1,\cdots,\mathbf{u}_K)\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_K Z}|\mathbf{z})} p(\mathbf{u}_1|\mathbf{u}_2)\cdots p(\mathbf{u}_{K-1}|\mathbf{u}_K)p(\mathbf{u}_K) \\
&\overset{(b)}{\leq} |\mathcal{T}_\epsilon^n(P_{U_1\cdots U_K Z}|\mathbf{z})|\,.2^{-nH(U_1|U_2)(1-\epsilon)}\cdots 2^{-nH(U_{K-1}|U_K)(1-\epsilon)}.2^{-nH(U_K)(1-\epsilon)} \\
&\overset{(c)}{\leq} 2^{nH(U_1,\cdots,U_K|Z)(1+\epsilon)}.2^{-nH(U_1|U_2)(1-\epsilon)}\cdots.2^{-nH(U_{K-1}|U_K)(1-\epsilon)}.2^{-nH(U_K)(1-\epsilon)} \\
&\overset{(d)}{=} 2^{-n\left(\sum_{k=1}^{K} I(U_k;Z|U_{k+1})-\delta(\epsilon)\right)}, \qquad \text{(B.21)}
\end{aligned}$$

where (a) arises from the code generation process; (b) and (c) are from the properties of typical sequences and the size of the typical set, respectively from (2.22) and (2.23) from Theorem 1; (d) is due to the Markov chain $U_K\to\cdots\to U_1\to Z$; and $\delta(\epsilon)\to 0$ as $\epsilon\to 0$ for n sufficiently large. Similarly, the lower bound can be found as

$$\Pr\left[\left(\mathbf{U}_1(\widehat{j_1}),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{z}\right)\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_K Z})\right]\geq(1-\epsilon).2^{-n\left(\sum_{k=1}^{K} I(U_k;Z|U_{k+1})+\delta(\epsilon)\right)}. \qquad \text{(B.22)}$$

This gives us, for n sufficiently large,

$$\begin{aligned}
2^{n\left(\sum_{k=1}^{K}(R'_k-I(U_k;Z|U_{k+1}))-\delta(\epsilon)\right)} &\leq \mathrm{E}\left[\left|\mathcal{A}(\widehat{j_1},\cdots,\widehat{j_K},j_1,\cdots,j_K)\right|\right] \\
&\leq 2^{n\left(\sum_{k=1}^{K}(R'_k-I(U_k;Z|U_{k+1}))+\delta(\epsilon)\right)}. \qquad \text{(B.23)}
\end{aligned}$$

Let us now denote the set $\{1,\cdots,K\}$ as $\mathcal{K}'$. Let the k-subset of $\mathcal{K}'$ be denoted as $\mathcal{P}_k(\mathcal{K}')$, that is, the subset of the set $\mathcal{K}'$ with exactly k elements. For example, 2-subsets of $\{1,2,3\}$ are $\{1,2\}$, $\{1,3\}$ and $\{2,3\}$. We also make the following definitions:

$$\begin{aligned}
&U(\mathcal{S}_k)\triangleq\{U_l : l\in\mathcal{P}_k(\mathcal{K}')\},\;\; U(\mathcal{S}_k^c)\triangleq\{U_l : l\in\mathcal{K}'/\mathcal{P}_k(\mathcal{K}')\}, \\
&\mathbf{U}(\mathcal{S}_k)\triangleq\{U_l : l\in\mathcal{P}_k(\mathcal{K}')\},\;\; \mathbf{u}(\mathcal{S}_k^c)\triangleq\{U_l : l\in\mathcal{K}'/\mathcal{P}_k(\mathcal{K}')\}.
\end{aligned}$$

We also define

$$C\left(\mathcal{P}_k(\mathcal{K}')\right) \triangleq$$
$$\sum_{\substack{\widehat{j}_1,\cdots,\widehat{j_K} \\ \widehat{j}_1\neq j_1,\cdots\widehat{j_K}\neq j_K}} \sum_{\left\{\widehat{\widehat{j}}_k\neq\widehat{j}_k,\widehat{j}_k\neq j_k\right\}_{k\in\mathcal{P}_k(\mathcal{K}')}} \Pr\left[\begin{array}{c}\left(\mathbf{U}_1(\widehat{j}_1),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{z}\right)\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_KZ}),\\ (\mathbf{U}(\mathcal{S}_k),\mathbf{u}(\mathcal{S}_k^c),\mathbf{z})\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_KZ})\end{array}\right]. \tag{B.24}$$

The probability in the equation above can be written as

$$\begin{aligned}
&\Pr\left[\left(\mathbf{U}_1(\widehat{j}_1),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{z}\right)\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_KZ}),(\mathbf{U}(\mathcal{S}_k),\mathbf{u}(\mathcal{S}_k^c),\mathbf{z})\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_KZ})\right]\\
&=\Pr\left[\left(\mathbf{U}_1(\widehat{j}_1),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{z}\right)\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_KZ})\right]\\
&\quad\times\Pr\left[(\mathbf{U}(\mathcal{S}_k),\mathbf{u}(\mathcal{S}_k^c),\mathbf{z})\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_KZ})\right]\\
&\leq\Pr\left[\left(\mathbf{U}_1(\widehat{j}_1),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{z}\right)\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_KZ})\right].2^{-n\left(\sum_{k\in\mathcal{P}_k(\mathcal{K}')}I(U_k;Z|U_{k+1})-\delta(\epsilon)\right)},
\end{aligned} \tag{B.25}$$

where the last equality is by using Theorem 5 and the Markov chain $U_K\rightarrow\cdots\rightarrow U_1\rightarrow Z$. Then $C\left(\mathcal{P}_k(\mathcal{K}')\right)$ can be bounded as

$$\begin{aligned}
C\left(\mathcal{P}_k(\mathcal{K}')\right)\leq&2^{n\left(\sum_{k=1}^K R_k+\sum_{k\in\mathcal{P}_k(\mathcal{K}')}R'_k\right)}\\
&\times 2^{-n\left(\sum_{k=1}^K I(U_k;Z|U_{k+1})+\sum_{k\in\mathcal{P}_k(\mathcal{K}')}I(U_k;Z|U_{k+1})-2\delta(\epsilon)\right)}.
\end{aligned} \tag{B.26}$$

Using (B.24), we can express $\mathrm{E}\left[\left|\mathcal{A}(\widehat{j}_1,\cdots,\widehat{j_K},j_1,\cdots,j_K)\right|^2\right]$ as

$$\begin{aligned}
&\mathrm{E}\left[\left|\mathcal{A}(\widehat{j}_1,\cdots,\widehat{j_K},j_1,\cdots,j_K)\right|^2\right]\\
&=\sum_{\widehat{j}_1,\cdots,\widehat{j_K},(\widehat{j}_1,\cdots,\widehat{j_K})\neq j_1,\cdots,j_K}\Pr\left[\left(\mathbf{U}_1(\widehat{j}_1),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{z}\right)\in\mathcal{T}_\epsilon^n(P_{U_1\cdots U_KZ})\right]\\
&\quad+\sum_{\forall\mathcal{P}_1(\mathcal{K}')}C(\mathcal{P}_1(\mathcal{K}'))+\sum_{\forall\mathcal{P}_2(\mathcal{K}')}C(\mathcal{P}_2(\mathcal{K}'))+\cdots+\sum_{\forall\mathcal{P}_K(\mathcal{K}')}C(\mathcal{P}_K(\mathcal{K}')),
\end{aligned} \tag{B.27}$$

where the sums in the second line in the equation above are taken over all 1-subsets, 2-subsets, and so on till the K-subsets. Let us know define

$$C_{\mathcal{E}}\left(\mathcal{P}_k(\mathcal{K}')\right)\triangleq 2^{-n\left(\sum_{k,k\notin\mathcal{P}_k(\mathcal{K}')}\left[R'_k-I(U_k;Z|U_{k+1})\right]-4\delta(\epsilon)\right)}. \tag{B.28}$$

Using the above, we then have

$$\Pr\left[\mathcal{E}(\widehat{j_1},\cdots,\widehat{j_K},j_1,\cdots,j_K)\right] \le \frac{\mathrm{Var}\left[\left|\mathcal{A}(\widehat{j_1},\cdots,\widehat{j_K},j_1,\cdots,j_K)\right|\right]}{\mathrm{E}\left[\left|\mathcal{A}(\widehat{j_1},\cdots,\widehat{j_K},j_1,\cdots,j_K)\right|\right]^2}$$
$$\le 2^{-n\left(\sum_k [R'_k - I(U_k;Z|U_{k+1})] - 3\delta(\epsilon)\right)}$$
$$+ \sum_{\forall \mathcal{P}_1(\mathcal{K}')} C_{\mathcal{E}}(\mathcal{P}_1(\mathcal{K}')) + \sum_{\forall \mathcal{P}_2(\mathcal{K}')} C_{\mathcal{E}}(\mathcal{P}_2(\mathcal{K}')) + \cdots + \sum_{\forall \mathcal{P}_K(\mathcal{K}')} C_{\mathcal{E}}(\mathcal{P}_K(\mathcal{K}')).$$

Thus for $\Pr\left[\mathcal{E}(\widehat{j_1},\cdots,\widehat{j_K},j_1,\cdots,j_K)\right] \to 0$ as n gets large, we need

$$\begin{aligned} R'_K &\ge I(U_K;Z) \\ R'_{K-1} &\ge I(U_{K-1};Z|U_K) \\ &\vdots \\ R'_1 &\ge I(U_1;Z|U_2) = I(X;Z|U_2), \end{aligned} \tag{B.29}$$

after removing the redundant inequalities. Next, for each $(w_1,\cdots,w_K)$, we define a random version of $\mathcal{A}(\widehat{j_1},\cdots,\widehat{j_K},j_1,\cdots,j_K)$ in (B.18) as

$$\widetilde{\mathcal{A}}(\widehat{j_1},\cdots,\widehat{j_K},j_1,\cdots,j_K) \triangleq \left\{ \begin{array}{c} \widehat{j_k} \in \mathcal{B}_k(w_k), k=1,\cdots,K: \\ \left(\mathbf{U}_1(\widehat{j_1}),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{Z}\right) \in \mathcal{T}_\epsilon^n(P_{U_1\cdots U_K Z}), \\ (\widehat{j_1},\cdots,\widehat{j_K}) \ne (J_1,\cdots,J_K) \end{array} \right\}. \tag{B.30}$$

We also define the event

$$\mathcal{E}(w_1,\cdots,w_K) \triangleq \left\{\left|\widetilde{\mathcal{A}}(\widehat{j_1},\cdots,\widehat{j_K},j_1,\cdots,j_K)\right| \ge 2\mathrm{E}\left[\left|\widetilde{\mathcal{A}}(\widehat{j_1},\cdots,\widehat{j_K},j_1,\cdots,j_K)\right|\right]\right\}, \tag{B.31}$$

and the indicator variables

$$\begin{aligned} \mathbb{I}((w_1,\cdots,w_K) := 0 \text{ if } &\left(\mathbf{U}_1(\widehat{j_1}),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{Z}\right) \in \mathcal{T}_\epsilon^n(P_{U_1\cdots U_K Z}) \\ &\text{and } \mathcal{E}(w_1,\cdots,w_K)^c \text{ occurs}, \\ \mathbb{I}((w_1,\cdots,w_K) := 1 \text{ if } &\left(\mathbf{U}_1(\widehat{j_1}),\cdots,\mathbf{U}_K(\widehat{j_K}),\mathbf{Z}\right) \in \mathcal{T}_\epsilon^n(P_{U_1\cdots U_K Z}) \\ &\text{and } \mathcal{E}(w_1,\cdots,w_K) \text{ occurs}. \end{aligned}$$

The subsequent steps mirror the proof in the single receiver case in Appendix B.1 and we will provide an outline only.

For the probability that $\mathbb{I}((w_1, \cdots, w_K) = 1$, we have

$$\begin{aligned}\Pr[\mathbb{I}(w_1, \cdots, w_K) = 1] \leq & \sum_{\mathbf{z} \in \mathcal{T}_\epsilon^n(P_Z)} \sum_{j_1, \cdots, j_K} p(\mathbf{z}) p(j_1, \cdots, j_K|\mathbf{z}) \times \\ & \Pr[\mathcal{E}(w_1, \cdots, w_K)|\mathbf{Z} = \mathbf{z}, J_1 = j_1, \cdots, J_K = j_K] \\ = & \sum_{\mathbf{z} \in \mathcal{T}_\epsilon^n(P_Z)} \sum_{j_1, \cdots, j_K} p(\mathbf{z}) p(j_1, \cdots, j_K|\mathbf{z}) \Pr[\mathcal{E}(w_1, \cdots, w_K)],\end{aligned} \tag{B.32}$$

which is small for n sufficiently large since $\Pr[\mathcal{E}(w_1, \cdots, w_K)] \to 0$ for n sufficiently large under the conditions (B.29). We then have

$$\begin{aligned}H(J_1, \cdots, J_K|\mathbf{Z}, w_1, \cdots, w_K) \leq & 1 + \Pr[\mathbb{I}(w_1, \cdots, w_K) = 1] \times \\ & H(J_1, \cdots, J_K|\mathbf{Z}, w_1, \cdots, w_K, \mathbb{I}(w_1, \cdots, w_K) = 1) \\ & + H(J_1, \cdots, J_K|\mathbf{Z}, w_1, \cdots, w_K, \mathbb{I}(w_1, \cdots, w_K) = 0).\end{aligned} \tag{B.33}$$

The second term on the RHS of (B.33) is small for n sufficiently large, and the third term is bounded as

$$\begin{aligned}& H(J_1, \cdots, J_K|\mathbf{Z}, w_1, \cdots, w_K, \mathbb{I}(w_1, \cdots, w_K) = 0) \\ & \leq \log(2\mathrm{E}[|\mathcal{A}(\widehat{j_1}, \cdots, \widehat{j_K}, j_1, \cdots, j_K)|] - 1) \\ & = \log\left(2^{n(\sum_{k=1}^K [R'_k - I(U_k; Z|U_{k+1})] + \delta(\epsilon)) + 1} - 1\right) \\ & \leq n\left(\sum_{k=1}^K [R'_k - I(U_k; Z|U_{k+1})] + \delta(\epsilon)\right) + 2 - 2^{-n\left(\sum_{k=1}^K [R'_k - I(U_k; Z|U_{k+1})] + \delta(\epsilon)\right) + 1)},\end{aligned} \tag{B.34}$$

where the last inequality is by using $\log(x) \leq x - 1$. Then, for n sufficiently large,

$$H(J_1, \cdots, J_K|\mathbf{Z}, w_1, \cdots, w_K) \leq n\left(\sum_{k=1}^K [R'_k - I(U_k; Z|U_{k+1})] + \delta(\epsilon)\right) + 2, \tag{B.35}$$

since $\sum_{k=1}^K R'_k \geq \sum_{k=1}^K I(U_k; Z|U_{k+1})$, which can be deduced from the conditions (B.29). The lemma is then proved by averaging over the $(w_1, \cdots, w_K)$.

B.3 Alternative Proofs for K-receiver Degraded BC

B.3.1 Obtaining the Sizes of Subcodes

Here, we follow the approach of Wyner [115], and show how to obtain $\log L'_k$ in the encoding of W_k, for $k = 2, ..., K-1$. Following the same routine, $\log L'_1$ and $\log L'_K$ can be obtained easily, and thus these calculations will be omitted.

To start with, suppose that we have the messages, $w_k = i_k, \ldots, w_K = i_K$. We now define

$$\begin{aligned} q_{i_k}^{(k)} &\triangleq \Pr\left[W_k = i_k | W_{k+1} = i_{k+1}, \ldots, W_K = i_K\right] \\ &= \Pr\left[W_k = i_k | \mathbf{u}_K(i_K, i'_K), \mathbf{u}_{K-1}(i_{K-1}, i'_{K-1}, i_K, i'_K), \ldots, \mathbf{u}_k(i_k, i'_k, \ldots, i_K, i'_K)\right]. \end{aligned} \tag{B.36}$$

The codeword $\mathbf{u}_k(w''_k, \ldots, w''_K)$ is a channel code for $p_{\mathbf{Y}_k|\mathbf{X}}$ and $p_{\mathbf{Z}|\mathbf{X}}$ simultaneously and is comprised of $L_k = 2^{nR_k}$ subcodes $\{C_{i_k}^{(k)}\}_{i_k=1}^{L_k}$. $\mathbf{U}_k$ is an uniformly randomly chosen member of $\{C_{i_k}^{(k)}\}$. Therefore,

$$\Pr\left[\mathbf{U}_k = \mathbf{u}_k(w''_k, \ldots, w''_K) | \mathbf{u}_K(i_K, i'_K), \ldots, \mathbf{u}_{k+1}(i_{k+1}, i'_{k+1}, \ldots, i_K, i'_K)\right] = \frac{q_{i_k}^{(k)}}{L'_k}. \tag{B.37}$$

The codeword $\mathbf{u}_k(w''_k, \ldots, w''_K)$ is a channel code for $p_{\mathbf{Y}_k|\mathbf{X}}$ with prior distribution on codewords given by (B.37). Each of $\mathcal{C}_{i_k}^{(k)}$ is a channel code for the eavesdropper's channel $p_{\mathbf{Z}|\mathbf{X}}$ with L'_k codewords and uniform prior distribution on the codewords. Let $\lambda_{i_k}^{(k)}$ be the error probability for $\mathcal{C}_{i_k}^{(k)}$ with an optimal decoder, when i'_k is chosen as the index for the codeword from $\mathcal{C}_{i_k}^{(k)}$. Then $\bar{\lambda}^{(k)}$ is the average error probability for $\mathcal{C}_{i_k}^{(k)}$ with an optimal decoder, averaged over the probability that $W_k = i_k$ is sent given the previous messages were $W_{k+1} = i_{k+1}, \ldots W_K = i_K$. As a result, we have

$$\begin{cases} \lambda_{i_k}^{(k)} = \Pr\left[\mathbf{X} \neq \mathbf{Z} | W_k = i_k, \mathbf{u}_K(i_K, i'_K), \ldots, \mathbf{u}_{k+1}(i_{k+1}, i'_{k+1}, \ldots, i_K, i'_K)\right], \\ \bar{\lambda}^{(k)} = \sum_{i_k=1}^{L_k} q_{i_k}^{(k)} \lambda_{i_k}^{(k)}. \end{cases} \tag{B.38}$$

By Fano's inequality,

$$\begin{aligned} H(\mathbf{X}|\mathbf{Z}, W_k = i_k, \mathbf{u}_K(i_K, i'_K), \ldots, \mathbf{u}_{k+1}(i_{k+1}, i'_{k+1}, \ldots, i_K, i'_K)) &\leq 1 + \lambda_{i_k}^{(k)} \log L'_k \\ \Rightarrow \quad H(\mathbf{U}_k|\mathbf{Z}, \mathbf{U}_K, \ldots, \mathbf{U}_{k+1}, W_k = i_k) &\leq 1 + \lambda_{i_k}^{(k)} \log L'_k. \end{aligned} \tag{B.39}$$

Since $|\mathcal{C}_{i_k}^{(k)}| = L'_k$ and has probability of error $\lambda_{i_k}^{(k)}$, we have

$$\begin{aligned} & I(\mathbf{U}_k; \mathbf{Z}|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}, W_k = i_k) \\ &= H(\mathbf{U}_k|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}, W_k = i_k) - H(\mathbf{U}_k|\mathbf{Z}, \mathbf{U}_K, \ldots, \mathbf{U}_{k+1}, W_k = i_k) \\ &= \log L'_k - H(\mathbf{U}_k|\mathbf{Z}, \mathbf{U}_K, \ldots, \mathbf{U}_{k+1}, W_k = i_k) \\ \Rightarrow \quad \log L'_k &\leq I(\mathbf{U}_k; \mathbf{Z}|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}, W_k = i_k) + 1 + \lambda_{i_k}^{(k)} \log L'_k. \end{aligned} \tag{B.40}$$

Averaging over i_k using $\{q_{i_k}^{(k)}\}$ gives

$$\begin{aligned}\log L'_k &\leq I(\mathbf{U}_k;\mathbf{Z}|\mathbf{U}_K,\ldots,\mathbf{U}_{k+1},W_k)+1+\bar{\lambda}^{(k)}\log L'_k\\ &\overset{(a)}{\leq} I(\mathbf{U}_k;\mathbf{Z}|\mathbf{U}_K,\ldots,\mathbf{U}_{k+1})+1+\bar{\lambda}^{(k)}\log L'_k\\ &\overset{(b)}{\leq} nI(U_k;Z|U_K,\ldots,U_{k+1})+n\delta+1+\bar{\lambda}^{(k)}\log L'_k,\\ &\overset{(c)}{=} nI(U_k;Z|U_{k+1})+n\delta+1+\bar{\lambda}^{(k)}\log L'_k,\end{aligned} \tag{B.41}$$

where (a) is by $W_k \to (\mathbf{U}_K,\ldots,\mathbf{U}_{k+1}) \to \mathbf{U}_k \to \mathbf{Z}$, (b) results from the fact that (following Liu *et al.* [78])

$$I(\mathbf{U}_k;\mathbf{Z}|\mathbf{U}_K,\ldots,\mathbf{U}_{k+1}) \leq nI(U_k;Z|U_K,\ldots,U_{k+1})+n\delta, \tag{B.42}$$

with $\delta \to 0$ as $n \to \infty$ and (c) is by the Markov chain condition $U_K \to \cdots \to U_{k+1} \to U_k \to Z$ for the degraded BC. Similarly, by substituting $\mathbf{X}$ for $\mathbf{U}_1$ and removing conditioning from (B.36) for $k = K$, we have

$$\begin{cases}\log L_1 \leq nI(X;Z|U_2)+n\delta+1+\bar{\lambda}^{(1)}\log L'_1,\\ \log L_K \leq nI(U_K;Z)+n\delta+1+\bar{\lambda}^{(K)}\log L'_K.\end{cases} \tag{B.43}$$

Based on the above, and since $R'_k = \frac{1}{n}\log L'_k$, we let

$$\begin{cases}R'_1 \triangleq I(X;Z|U_2)-\tau,\\ R'_k \triangleq I(U_k;Z|U_{k+1})-\tau, \quad \text{for } k=2,\ldots,K-1,\\ R'_K \triangleq I(U_K;Z)-\tau,\end{cases} \tag{B.44}$$

where $\tau \to 0$ for sufficiently large n.

We observe that this is a somewhat weaker result than the one presented in Lemma 4. This is because for the method in this section, we can obtain upper bounds on the rates, for example $R'_k \leq I(U_k;Z|U_{k+1})$. However the method used in the proof of Lemma 4 will obtain lower bounds on $R'_k \geq I(U_k;Z|U_{k+1})$, a stronger result.

B.3.2 Equivocation Calculation for the kth Receiver

We only show the calculation for the kth receiver. This method can be extended, only with more elaborate steps, to any combination of receivers which are a subset

of $\{1, \cdots, K\}$. For the kth receiver, we have

$$\begin{aligned}
nR_{e(k)} &= H(W_k|\mathbf{Z}) \\
&\geq H(W_k|\mathbf{Z}, \mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) \qquad \text{since conditioning reduces entropy} \\
&= H(W_k, \mathbf{Z}|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) - H(\mathbf{Z}|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) \\
&= H(W_k, \mathbf{U}_k, \mathbf{Z}|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) - H(\mathbf{U}_k|W_k, \mathbf{Z}, \mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) \\
&\quad - H(\mathbf{Z}|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) \\
&= H(W_k, \mathbf{U}_k|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) + H(\mathbf{Z}|W_k, \mathbf{U}_K, \ldots, \mathbf{U}_{k+1}, \mathbf{U}_k) \\
&\quad - H(\mathbf{Z}|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) - H(\mathbf{U}_k|W_k, \mathbf{Z}, \mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) \\
&\overset{(a)}{\geq} H(\mathbf{U}_k|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) + H(\mathbf{Z}|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}, \mathbf{U}_k) - H(\mathbf{Z}|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) \\
&\quad - H(\mathbf{U}_k|W_k, \mathbf{Z}, \mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) \\
&= H(\mathbf{U}_k|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) - I(\mathbf{U}_k; \mathbf{Z}|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) \\
&\quad - H(\mathbf{U}_k|W_k, \mathbf{Z}, \mathbf{U}_K, \ldots, \mathbf{U}_{k+1}),
\end{aligned} \tag{B.45}$$

where (a) has the second term by the fact that $W_k \to (\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) \to \mathbf{Z}$. We now bound each of the terms in the last line of (B.45). For the first term, given that $\mathbf{U}_K = \mathbf{u}_K$, $\mathbf{U}_{K-1} = \mathbf{u}_{K-1}, \ldots, \mathbf{U}_{k+1} = \mathbf{u}_{k+1}$, $\mathbf{u}_k$ has $2^{n(R_k+R'_k)}$ possible values with equal probability. As a consequence, we have

$$H(\mathbf{U}_k|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) = n(R_k + R'_k). \tag{B.46}$$

For the second term, it can be shown that

$$I(\mathbf{U}_k; \mathbf{Z}|\mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) \leq nI(U_k; Z|U_{k+1}) + n\delta. \tag{B.47}$$

For the last term, we have by Fano's inequality

$$\frac{1}{n}H(\mathbf{U}_k|W_k, \mathbf{Z}, \mathbf{U}_K, \ldots, \mathbf{U}_{k+1}) \leq \frac{1}{n}\left(1 + \bar{\lambda}^{(k)} \log L'_k\right) \triangleq \epsilon'_{k,n} \tag{B.48}$$

where $\epsilon'_{k,n} \to 0$ for n sufficiently large.

To show that $\bar{\lambda}^{(k)} \to 0$ for n sufficiently large so that (B.48) holds, we consider decoding at the eavesdropper and focus on the codebook with rate R'_k to be decoded at the eavesdropper with error probability $\bar{\lambda}^{(k)}$. Let $W_k = i_k$ be fixed. The eavesdropper attempts to decode $\mathbf{u}_k$ given $w_k, \mathbf{u}_K, \ldots, \mathbf{u}_{k+1}$ by finding the estimate for w'_k, $\hat{w}'_k$, so

that

$$(\mathbf{u}_k(w_k, \hat{w}'_k, w_{k+1}, w'_{k+1}, \ldots, w_K, w'_K), \mathbf{z}, \mathbf{u}_{k+1}, \cdots, \mathbf{u}_K) \in \mathcal{T}_\epsilon^n(P_{U_k Z|U_{k+1}\ldots U_K}).$$

where w_k, and all $w_{k+1}, w'_{k+1}, \ldots, w_K, w'_K$ are known. If there is none or more than one possible codeword, an error is declared. Defining the event

$$\mathrm{E}_{i'_k}^{(Z)} \triangleq \left\{(\mathbf{u}_k(i_k, i'_k), \mathbf{z}, \mathbf{u}_{k+1}, \cdots, \mathbf{u}_K) \in \mathcal{T}_\epsilon^n(P_{U_k Z|U_{k+1}\ldots U_K})\right\}, \tag{B.49}$$

and assuming without loss of generality that $w'_k = 1$ is sent, we then have

$$\bar{\lambda}^{(k)} \leq \Pr\left\{\left(\mathrm{E}_1^{(Z)}\right)^{\mathsf{c}}\right\} + \sum_{i'_k \neq 1} \Pr\left\{\left(\mathrm{E}_{i'_k}^{(Z)}\right)\right\} \leq \epsilon + 2^{nR'_k} 2^{-n(I(U_k;Z|U_{k+1},\ldots,U_K)-2\epsilon)}, \tag{B.50}$$

where $\epsilon \to 0$ for n sufficiently large. Since we have chosen from (B.44) that $R'_k = I(U_k; Z|U_{k+1}) - \tau$ which is $= I(U_k; Z|U_{k+1}, \ldots, U_K) - \tau$ by $U_K \to \cdots \to U_{k+1} \to U_k \to Z$, we have $\bar{\lambda}^{(k)} \leq 2\epsilon$, for $\tau > 2\epsilon$. Thus, $\bar{\lambda}^{(k)}$ is small for n sufficiently large and (B.48) holds.

Now substituting (B.46)–(B.48) into the last line of (B.45), we have

$$\begin{aligned} nR_{e(k)} &\geq nR_k + nI(U_k; Z|U_{k+1}) - n\tau - nI(U_k; Z|U_{k+1}) - n\delta - n\epsilon'_{k,n} \\ &= nR_k - n\epsilon_k \end{aligned} \tag{B.51}$$

where $\epsilon_k = \tau + \delta + \epsilon'_{k,n}$. Hence, the security condition in (4.7) is satisfied for the kth receiver. We can see that the procedure can be repeated in a similar way to obtain the equivocation rate of any combination of receivers. However, we again see from (B.50) that the condition for the eavesdropper's error probability to be small is the (weaker) upper bound $R'_k \leq I(U_k; Z|U_{k+1})$.

B.4 Proof of Lemma 6

We bound $H(\mathcal{I}(L)|\mathbf{Y}_3, \mathbf{U}_1, w_1)$ for every w_1, following the method in [35]. To begin, let us label the $2^{nL_{21}}$ bins from $\mathbf{U}_2$ indexed by l_{21} corresponding to the message part L_2 as $\mathcal{B}_2(l_2)$. Similarly, label the $2^{nL_{31}}$ bins from $\mathbf{U}_3$ indexed by l_{31} corresponding to the partial message L_3 as $\mathcal{B}_3(l_3)$, and the $2^{L_{11}}$ bins from $\mathbf{X}$ indexed by l_{11} corresponding to the message part L_1 as $\mathcal{B}_1(l_1)$. Let us fix $L'_2 = l'_2$, $L_2^\dagger = l_2^\dagger$, $L'_3 = l'_3$, $L_3^\dagger = l_3^\dagger$, $L'_1 = l'_1$ and sequences $(\mathbf{u}_1, \mathbf{y}_3) \in \mathcal{T}_\epsilon^n(P_{U_1 Y_3})$, and denote $\iota_l = (l'_2, l_2^\dagger, l'_3, l_3^\dagger, l'_1)$. The typical set with respect to $p(u_1, u_2, u_3, x, y_3)$ which denoted as $\mathcal{T}_\epsilon^n(P_{U_1 U_2 U_3 X Y_3})$ will be abbreviated to $\mathcal{T}_\epsilon^n$ for the rest of this section.

Let the eavesdropper's estimate of the indices be $\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}$, and define the set $\mathcal{A}\left(\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l\right)$ as

$$\mathcal{A}\left(\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l\right) \triangleq \left\{ \begin{array}{c} (\widehat{l'_2}, \widehat{l^\dagger_2}) \in \mathcal{B}_2(l_2), (\widehat{l'_3}, \widehat{l^\dagger_3}) \in \mathcal{B}_3(l_3), \widehat{l'_1} \in \mathcal{B}_1(l_1) : \\ \left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l'_2}, \widehat{l^\dagger_2}), \mathbf{U}_3(\widehat{l'_3}, \widehat{l^\dagger_3}), \mathbf{X}(\widehat{l'_1}), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n, \\ \widehat{l'_2} \neq l'_2, \widehat{l^\dagger_2} \neq l^\dagger_2, \widehat{l'_3} \neq l'_3, \widehat{l^\dagger_3} \neq l^\dagger_3, \widehat{l'_1} \neq l'_1 \end{array} \right\}. \quad \text{(B.52)}$$

That is, the set $\mathcal{A}\left(\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l\right)$ is the set of the eavesdropper's estimate of the indices $\left(l'_2, l^\dagger_2, l'_3, l^\dagger_3, l'_1\right)$ that is not equal to the transmitted indices. Let us define the random event

$$\mathcal{E}\left(\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l\right) \triangleq \left\{ \left| \mathcal{A}\left(\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l\right) \right| \geq 2\mathrm{E}\left[\left| \mathcal{A}(\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l) \right| \right] \right\}. \quad \text{(B.53)}$$

We will now show that $\Pr\left[\mathcal{E}\left(\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l\right)\right] \to 0$ as n is large under appropriate conditions. We use the version of Chebyshev's inequality stated in Lemma 2. We now need the mean and variance of $\left|\mathcal{A}\left(\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l\right)\right|$. For the mean, we have

$$\begin{aligned}
&\mathrm{E}\left[\left|\mathcal{A}(\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l)\right|\right] \\
&= \sum_{\substack{\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1} \\ (\widehat{l'_2}, \widehat{l^\dagger_2}) \neq (l'_2, l^\dagger_2), (\widehat{l'_3}, \widehat{l^\dagger_3}) \neq (l'_3, l^\dagger_3), \widehat{l'_1} \neq l'_1}} \Pr\left[\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l'_2}, \widehat{l^\dagger_2}), \mathbf{U}_3(\widehat{l'_3}, \widehat{l^\dagger_3}), \mathbf{X}(\widehat{l'_1}), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\right] \\
&= \left(2^{n(L'_2 + L^\dagger_2)} - 1\right)\left(2^{n(L'_3 + L^\dagger_3)} - 1\right)\left(2^{nL'_1} - 1\right) \\
&\quad \times \Pr\left[\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l'_2}, \widehat{l^\dagger_2}), \mathbf{U}_3(\widehat{l'_3}, \widehat{l^\dagger_3}), \mathbf{X}(\widehat{l'_1}), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\right]. \quad \text{(B.54)}
\end{aligned}$$

We have, for $\epsilon \to 0$ for n sufficiently large

$$\begin{aligned}
&\Pr\left[\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l'_2}, \widehat{l^\dagger_2}), \mathbf{U}_3(\widehat{l'_3}, \widehat{l^\dagger_3}), \mathbf{X}(\widehat{l'_1}), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\right] \\
&\overset{(a)}{=} \sum_{(\mathbf{u}_2, \mathbf{u}_3, \mathbf{x}) \in \mathcal{T}_\epsilon^n(P_{U_1U_2U_3XY_3}|\mathbf{u}_1, \mathbf{y}_3)} p(\mathbf{u}_2|\mathbf{u}_1) p(\mathbf{u}_3|\mathbf{u}_1) p(\mathbf{x}|\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3) \\
&\overset{(b)}{\leq} \left|\mathcal{T}_\epsilon^n(P_{U_2U_3X}|\mathbf{u}_1, \mathbf{y}_3)\right| . 2^{-nH(U_2|U_1)(1-\epsilon)} . 2^{-nH(U_3|U_1)(1-\epsilon)} . 2^{-nH(X|U_2,U_3,U_1)(1-\epsilon)} \\
&\overset{(c)}{\leq} 2^{nH(U_2,U_3,X|U_1,Y_3)(1+\epsilon)} . 2^{-nH(U_2|U_1)(1-\epsilon)} . 2^{-nH(U_3|U_1)(1-\epsilon)} . 2^{-nH(X|U_2,U_3,U_1)(1-\epsilon)} \\
&= 2^{n(H(U_2|U_1,Y_3) + H(U_3|U_1,U_2,Y_3) + H(X|U_1,U_2,U_3,Y_3) + \epsilon H(U_2,U_3,X|U_1,Y_3))} \\
&\quad \times 2^{-n(H(U_2|U_1) + H(U_3|U_1) + H(X|U_2,U_3,U_1)) + n\epsilon(H(U_2|U_1) + H(U_3|U_1) + H(X|U_2,U_3,U_1))}
\end{aligned}$$

$$
\begin{aligned}
&= 2^{-n(I(U_2;Y_3|U_1)+I(U_3;U_2,Y_3|U_1)+I(X;Y_3|U_2,U_3,U_1))} \\
&\times 2^{n\epsilon(H(U_2,U_3,X|U_1,Y_3)+H(U_2|U_1)+H(U_3|U_1)+H(X|U_2,U_3,U_1))} \\
&= 2^{-n(I(U_2;U_3|U_1)+I(U_2,U_3;Y_3|U_1)+I(X;Y_3|U_2,U_3,U_1)-\delta(\epsilon))} \qquad \text{(B.55)}
\end{aligned}
$$

where (a) arises from the code generation process; (b) and (c) are from the properties of typical sequences and the size of the typical set, respectively from (2.22) and (2.23) from Theorem 1, respectively; and $\delta(\epsilon) \to 0$ as $\epsilon \to 0$ for n sufficiently large. Similarly, from the properties of joint typical sequences and the size of the joint typical set,

$$
\begin{aligned}
&\Pr\left[\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}_2', \widehat{l}_2^{\dagger}), \mathbf{U}_3(\widehat{l}_3', \widehat{l}_3^{\dagger}), \mathbf{X}(\widehat{l}_1'), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\right] \\
&\geq |\mathcal{T}_\epsilon^n(P_{U_2U_3X}|\mathbf{u}_1, \mathbf{y}_3)| \, .2^{-nH(U_2|U_1)(1+\epsilon)}.2^{-nH(U_3|U_1)(1+\epsilon)}.2^{-nH(X|U_2,U_3,U_1)(1+\epsilon)} \\
&\geq (1-\epsilon).2^{nH(U_2,U_3,X|U_1,Y_3)(1-\epsilon)}.2^{-nH(U_2|U_1)(1-\epsilon)}.2^{-nH(U_3|U_1)(1-\epsilon)}.2^{-nH(X|U_2,U_3,U_1)(1-\epsilon)} \\
&= (1-\epsilon).2^{-n(I(U_2;U_3|U_1)+I(U_2,U_3;Y_3|U_1)+I(X;Y_3|U_2,U_3,U_1)+\delta(\epsilon))}. \qquad \text{(B.56)}
\end{aligned}
$$

Substituting into (B.54), we can see that, for n sufficiently large,

$$
\begin{aligned}
&2^{n(L_2'+L_2^{\dagger}+L_3'+L_3^{\dagger}+L_1'-I(U_2;U_3|U_1)-I(U_2,U_3;Y_3|U_1)-I(X;Y_3|U_2,U_3,U_1)-\delta(\epsilon))} \\
&\leq \mathrm{E}\left[\left|\mathcal{A}\left(\widehat{l}_2', \widehat{l}_2^{\dagger}, \widehat{l}_3', \widehat{l}_3^{\dagger}, \widehat{l}_1', \iota_l\right)\right|\right] \\
&\leq 2^{n(L_2'+L_2^{\dagger}+L_3'+L_3^{\dagger}+L_1'-I(U_2;U_3|U_1)-I(U_2,U_3;Y_3|U_1)-I(X;Y_3|U_2,U_3,U_1)+\delta(\epsilon))}. \qquad \text{(B.57)}
\end{aligned}
$$

Let us now define, for ease of presentation, the following probabilities:

$$
\begin{aligned}
p_0 &\triangleq \Pr\left[\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}_2', \widehat{l}_2^{\dagger}), \mathbf{U}_3(\widehat{l}_3', \widehat{l}_3^{\dagger}), \mathbf{X}(\widehat{l}_1'), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\right], && \text{(B.58)} \\
p_2 &\triangleq \Pr\left[\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}_2', \widehat{l}_2^{\dagger}), \mathbf{u}_3, \mathbf{x}, \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\right], && \text{(B.59)} \\
p_3 &\triangleq \Pr\left[\left(\mathbf{u}_1, \mathbf{u}_2, \mathbf{U}_3(\widehat{l}_3', \widehat{l}_3^{\dagger}), \mathbf{x}, \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\right], && \text{(B.60)} \\
p_1 &\triangleq \Pr\left[\left(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{X}(\widehat{l}_1'), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\right], && \text{(B.61)} \\
p_{23} &\triangleq \Pr\left[\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}_2', \widehat{l}_2^{\dagger}), \mathbf{U}_3(\widehat{l}_3', \widehat{l}_3^{\dagger}), \mathbf{x}, \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\right] && \text{(B.62)} \\
p_{21} &\triangleq \Pr\left[\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}_2', \widehat{l}_2^{\dagger}), \mathbf{u}_3, \mathbf{X}(\widehat{l}_1'), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\right] && \text{(B.63)} \\
p_{31} &\triangleq \Pr\left[\left(\mathbf{u}_1, \mathbf{u}_2, \mathbf{U}_3(\widehat{l}_3', \widehat{l}_3^{\dagger}), \mathbf{X}(\widehat{l}_1'), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\right]. && \text{(B.64)}
\end{aligned}
$$

We also define

$$\mathcal{P}_1 \triangleq \Pr\left[\begin{array}{c}\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}'_2, \widehat{l}^{\dagger}_2), \mathbf{U}_3(\widehat{l}'_3, \widehat{l}^{\dagger}_3), \mathbf{X}(\widehat{l}'_1), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n, \\ \left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}'_2, \widehat{l}^{\dagger}_2), \mathbf{u}_3, \mathbf{x}, \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\end{array}\right],$$

$$\mathcal{P}_2 \triangleq \Pr\left[\begin{array}{c}\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}'_2, \widehat{l}^{\dagger}_2), \mathbf{U}_3(\widehat{l}'_3, \widehat{l}^{\dagger}_3), \mathbf{X}(\widehat{l}'_1), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n, \\ \left(\mathbf{u}_1, \mathbf{u}_2, \mathbf{U}_3(\widehat{l}'_3, \widehat{l}^{\dagger}_3), \mathbf{x}, \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\end{array}\right],$$

$$\mathcal{P}_3 \triangleq \Pr\left[\begin{array}{c}\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}'_2, \widehat{l}^{\dagger}_2), \mathbf{U}_3(\widehat{l}'_3, \widehat{l}^{\dagger}_3), \mathbf{X}(\widehat{l}'_1), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n, \\ \left(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3, \mathbf{X}(\widehat{l}'_1), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\end{array}\right],$$

$$\mathcal{P}_4 \triangleq \Pr\left[\begin{array}{c}\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}'_2, \widehat{l}^{\dagger}_2), \mathbf{U}_3(\widehat{l}'_3, \widehat{l}^{\dagger}_3), \mathbf{X}(\widehat{l}'_1), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n, \\ \left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}'_2, \widehat{l}^{\dagger}_2), \mathbf{u}_3, \mathbf{X}(\widehat{l}'_1), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\end{array}\right],$$

$$\mathcal{P}_5 \triangleq \Pr\left[\begin{array}{c}\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}'_2, \widehat{l}^{\dagger}_2), \mathbf{U}_3(\widehat{l}'_3, \widehat{l}^{\dagger}_3), \mathbf{X}(\widehat{l}'_1), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n, \\ \left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}'_2, \widehat{l}^{\dagger}_2), \mathbf{U}_3(\widehat{l}'_3, \widehat{l}^{\dagger}_3), \mathbf{x}, \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\end{array}\right],$$

$$\mathcal{P}_6 \triangleq \Pr\left[\begin{array}{c}\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}'_2, \widehat{l}^{\dagger}_2), \mathbf{U}_3(\widehat{l}'_3, \widehat{l}^{\dagger}_3), \mathbf{X}(\widehat{l}'_1), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n, \\ \left(\mathbf{u}_1, \mathbf{u}_2, \mathbf{U}_3(\widehat{l}'_3, \widehat{l}^{\dagger}_3), \mathbf{X}(\widehat{l}'_1), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\end{array}\right],$$

$$\mathcal{P}_7 \triangleq \Pr\left[\begin{array}{c}\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}'_2, \widehat{l}^{\dagger}_2), \mathbf{U}_3(\widehat{l}'_3, \widehat{l}^{\dagger}_3), \mathbf{X}(\widehat{l}'_1), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n, \\ \left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}'_2, \widehat{l}^{\dagger}_2), \mathbf{U}_3(\widehat{l}'_3, \widehat{l}^{\dagger}_3), \mathbf{X}(\widehat{l}'_1), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\end{array}\right]. \tag{B.65}$$

Next, we have

$$\begin{aligned}
&\mathrm{E}\left[\left|\mathcal{A}(\widehat{l}'_2, \widehat{l}^{\dagger}_2, \widehat{l}'_3, \widehat{l}^{\dagger}_3, \widehat{l}'_1, \iota_l)\right|^2\right] \\
&= \sum_{\substack{\widehat{l}'_2, \widehat{l}^{\dagger}_2, \widehat{l}'_3, \widehat{l}^{\dagger}_3, \widehat{l}'_1 \\ (\widehat{l}'_2, \widehat{l}^{\dagger}_2) \neq (l'_2, l^{\dagger}_2), (\widehat{l}'_3, \widehat{l}^{\dagger}_3) \neq (l'_3, l^{\dagger}_3), \widehat{l}'_1 \neq l'_1}} \Pr\left[\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{l}'_2, \widehat{l}^{\dagger}_2), \mathbf{U}_3(\widehat{l}'_3, \widehat{l}^{\dagger}_3), \mathbf{X}(\widehat{l}'_1), \mathbf{y}_3\right) \in \mathcal{T}_\epsilon^n\right] \\
&+ \sum_{\substack{\widehat{l}'_2, \widehat{l}^{\dagger}_2, \widehat{l}'_3, \widehat{l}^{\dagger}_3, \widehat{l}'_1 \\ (\widehat{l}'_2, \widehat{l}^{\dagger}_2) \neq (l'_2, l^{\dagger}_2), (\widehat{l}'_3, \widehat{l}^{\dagger}_3) \neq (l'_3, l^{\dagger}_3), \widehat{l}'_1 \neq l'_1}} \sum_{\substack{\widehat{\widehat{l}}'_2, \widehat{\widehat{l}}^{\dagger}_2 \\ (\widehat{\widehat{l}}'_2, \widehat{\widehat{l}}^{\dagger}_2) \neq (\widehat{l}'_2, \widehat{l}^{\dagger}_2) \\ (\widehat{\widehat{l}}'_2, \widehat{\widehat{l}}^{\dagger}_2) \neq (l'_2, l^{\dagger}_2)}} \mathcal{P}_1 \\
&+ \sum_{\substack{\widehat{l}'_2, \widehat{l}^{\dagger}_2, \widehat{l}'_3, \widehat{l}^{\dagger}_3, \widehat{l}'_1 \\ (\widehat{l}'_2, \widehat{l}^{\dagger}_2) \neq (l'_2, l^{\dagger}_2), (\widehat{l}'_3, \widehat{l}^{\dagger}_3) \neq (l'_3, l^{\dagger}_3), \widehat{l}'_1 \neq l'_1}} \sum_{\substack{\widehat{\widehat{l}}'_3, \widehat{\widehat{l}}^{\dagger}_3 \\ (\widehat{\widehat{l}}'_3, \widehat{\widehat{l}}^{\dagger}_3) \neq (\widehat{l}'_3, \widehat{l}^{\dagger}_3) \\ (\widehat{\widehat{l}}'_3, \widehat{\widehat{l}}^{\dagger}_3) \neq (l'_3, l^{\dagger}_3)}} \mathcal{P}_2 \\
&+ \sum_{\substack{\widehat{l}'_2, \widehat{l}^{\dagger}_2, \widehat{l}'_3, \widehat{l}^{\dagger}_3, \widehat{l}'_1 \\ (\widehat{l}'_2, \widehat{l}^{\dagger}_2) \neq (l'_2, l^{\dagger}_2), (\widehat{l}'_3, \widehat{l}^{\dagger}_3) \neq (l'_3, l^{\dagger}_3), \widehat{l}'_1 \neq l'_1}} \sum_{\substack{\widehat{\widehat{l}}'_1 \neq \widehat{l}'_1, \widehat{\widehat{l}}'_1 \neq l'_1}} \mathcal{P}_3
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{\substack{\widehat{l'_2},\widehat{l^\dagger_2},\widehat{l'_3},\widehat{l^\dagger_3},\widehat{l'_1} \\ (\widehat{l'_2},\widehat{l^\dagger_2})\neq(l'_2,l^\dagger_2),(\widehat{l'_3},\widehat{l^\dagger_3})\neq(l'_3,l^\dagger_3),\widehat{l'_1}\neq l'_1}} \sum_{\substack{\widehat{\widehat{l'_2}},\widehat{\widehat{l^\dagger_2}})\neq(\widehat{l'_2},\widehat{l^\dagger_2}),(\widehat{\widehat{l'_3}},\widehat{\widehat{l^\dagger_3}})\neq(\widehat{l'_3},\widehat{l^\dagger_3}) \\ (\widehat{\widehat{l'_2}},\widehat{\widehat{l^\dagger_2}})\neq(l'_2,l^\dagger_2),(\widehat{\widehat{l'_3}},\widehat{\widehat{l^\dagger_3}})\neq(l'_3,l^\dagger_3)}} \mathcal{P}_5 \\
&+ \sum_{\substack{\widehat{l'_2},\widehat{l^\dagger_2},\widehat{l'_3},\widehat{l^\dagger_3},\widehat{l'_1} \\ (\widehat{l'_2},\widehat{l^\dagger_2})\neq(l'_2,l^\dagger_2),(\widehat{l'_3},\widehat{l^\dagger_3})\neq(l'_3,l^\dagger_3),\widehat{l'_1}\neq l'_1}} \sum_{\substack{\widehat{\widehat{l'_2}},\widehat{\widehat{l^\dagger_2}})\neq(\widehat{l'_2},\widehat{l^\dagger_2}),\widehat{\widehat{l'_1}}\neq\widehat{l'_1} \\ (\widehat{\widehat{l'_2}},\widehat{\widehat{l^\dagger_2}})\neq(l'_2,l^\dagger_2),\widehat{\widehat{l'_1}}\neq l'_1}} \mathcal{P}_4 \\
&+ \sum_{\substack{\widehat{l'_2},\widehat{l^\dagger_2},\widehat{l'_3},\widehat{l^\dagger_3},\widehat{l'_1} \\ (\widehat{l'_2},\widehat{l^\dagger_2})\neq(l'_2,l^\dagger_2),(\widehat{l'_3},\widehat{l^\dagger_3})\neq(l'_3,l^\dagger_3),\widehat{l'_1}\neq l'_1}} \sum_{\substack{\widehat{\widehat{l'_3}},\widehat{\widehat{l^\dagger_3}})\neq(\widehat{l'_3},\widehat{l^\dagger_3}),\widehat{\widehat{l'_1}}\neq\widehat{l'_1} \\ (\widehat{\widehat{l'_3}},\widehat{\widehat{l^\dagger_3}})\neq(l'_3,l^\dagger_3),\widehat{\widehat{l'_1}}\neq l'_1}} \mathcal{P}_6 \\
&+ \sum_{\substack{\widehat{l'_2},\widehat{l^\dagger_2},\widehat{l'_3},\widehat{l^\dagger_3},\widehat{l'_1} \\ (\widehat{l'_2},\widehat{l^\dagger_2})\neq(l'_2,l^\dagger_2),(\widehat{l'_3},\widehat{l^\dagger_3})\neq(l'_3,l^\dagger_3),\widehat{l'_1}\neq l'_1}} \sum_{\substack{\widehat{\widehat{l'_2}},\widehat{\widehat{l^\dagger_2}})\neq(\widehat{l'_2},\widehat{l^\dagger_2}),(\widehat{\widehat{l'_3}},\widehat{\widehat{l^\dagger_3}})\neq(\widehat{l'_3},\widehat{l^\dagger_3}),\widehat{\widehat{l'_1}}\neq\widehat{l'_1} \\ (\widehat{\widehat{l'_2}},\widehat{\widehat{l^\dagger_2}})\neq(l'_2,l^\dagger_2),(\widehat{\widehat{l'_3}},\widehat{\widehat{l^\dagger_3}})\neq(l'_3,l^\dagger_3),\widehat{\widehat{l'_1}}\neq l'_1}} \mathcal{P}_7. \qquad \text{(B.66)}
\end{aligned}
$$

We have that

$$
\begin{aligned}
\mathcal{P}_1 = p_0 p_2,\ \ \mathcal{P}_2 = p_0 p_3,\ \ \mathcal{P}_3 = p_0 p_1,\ \ \mathcal{P}_4 = p_0 p_{21},\ \ \mathcal{P}_5 = p_0 p_{23}, \\
\mathcal{P}_6 = p_0 p_{31},\ \ \mathcal{P}_7 = p_0^2. \qquad \text{(B.67)}
\end{aligned}
$$

We now need to bound the probabilities in (B.65). The required probability p_0 in (B.58) is already upper bounded in (B.55). In (B.59), we have

$$
\begin{aligned}
p_2 &\leq 2^{n(H(U_2|U_3,U_1,Y_3)-H(U_2|U_1)+\delta(\epsilon))} \\
&= 2^{-n(I(U_2;U_3,X,Y_3|U_1)-\delta(\epsilon))} \leq 2^{-n(I(U_2;U_3|U_1)+I(U_2;Y_3|U_3,U_1)-\delta(\epsilon))}. \qquad \text{(B.68)}
\end{aligned}
$$

In (B.60), we similarly obtain

$$
p_3 \leq 2^{-n(I(U_2;U_3|U_1)+I(U_3;Y_3|U_2,U_1)-\delta(\epsilon))}. \qquad \text{(B.69)}
$$

In (B.61) we have

$$
p_1 \leq 2^{n(H(X|U_1,U_2,U_3,Y_3)-H(X|U_1,U_2,U_3)+\delta(\epsilon))} = 2^{-n(I(X;Y_3|U_2,U_3,U_1)-\delta(\epsilon))}. \qquad \text{(B.70)}
$$

In (B.62) we have

$$
\begin{aligned}
p_{23} &\leq 2^{n(H(U_2,U_3|U_1,X,Y_3)-H(U_2|U_1)-H(U_3|U_1)+\delta(\epsilon))} \\
&= 2^{n(H(U_2,U_3,X,Y_3|U_1)-H(X,Y_3|U_1)-H(U_2|U_1)-H(U_3|U_1)+\delta(\epsilon))} \\
&= 2^{n(H(U_2|U_1)+H(U_3|U_2,U_1)+H(Y_3|U_2,U_3,U_1)+H(X|Y_3,U_2,U_3,U_1))}
\end{aligned}
$$

$$\times 2^{-n(H(Y_3|U_1)+H(X|Y_3,U_1)+H(U_2|U_1)+H(U_3|U_1)-\delta(\epsilon))}$$
$$= 2^{-n(I(U_2;U_3|U_1)+I(U_2,U_3;Y_3|U_1)+I(U_2,U_3;X|Y_3,U_1)-\delta(\epsilon))}$$
$$\leq 2^{-n(I(U_2;U_3|U_1)+I(U_2,U_3;Y_3|U_1)-\delta(\epsilon))}. \tag{B.71}$$

In (B.63) we have

$$p_{21} \leq 2^{n(H(U_2,X|U_1,U_3,Y_3)-H(U_2|U_1)-H(X|U_2,U_3,U_1)+\delta(\epsilon))}$$
$$= 2^{n(H(U_2|U_1,U_3,Y_3)+H(X|U_1,U_2,U_3,Y_3)-H(U_2|U_1)-H(X|U_2,U_3,U_1)+\delta(\epsilon))}$$
$$= 2^{-n(I(U_2;U_3|U_1)+I(U_2;Y_3|U_3,U_1)+I(X;Y_3|U_2,U_3,U_1)-\delta(\epsilon))}. \tag{B.72}$$

In (B.64), we similarly obtain

$$p_{31} \leq 2^{-n(I(U_2;U_3|U_1)+I(U_3;Y_3|U_2,U_1)+I(X;Y_3|U_2,U_3,U_1)-\delta(\epsilon))}. \tag{B.73}$$

Combining (B.55), (B.66), (B.68)-(B.73), we obtain the upper bound for the variance $\mathrm{Var}\left[\left|\mathcal{A}\left(\widehat{l'_2}, \widehat{l^{\dagger}_2}, \widehat{l'_3}, \widehat{l^{\dagger}_3}, \widehat{l'_1}, \iota_l\right)\right|\right]$, which is

$$\begin{aligned}
&\mathrm{Var}\left[\left|\mathcal{A}\left(\widehat{l'_2}, \widehat{l^{\dagger}_2}, \widehat{l'_3}, \widehat{l^{\dagger}_3}, \widehat{l'_1}, \iota_l\right)\right|\right] \\
&= \mathrm{E}\left[\left|\mathcal{A}\left(\widehat{l'_2}, \widehat{l^{\dagger}_2}, \widehat{l'_3}, \widehat{l^{\dagger}_3}, \widehat{l'_1}, \iota_l\right)\right|^2\right] - \mathrm{E}\left[\left|\mathcal{A}\left(\widehat{l'_2}, \widehat{l^{\dagger}_2}, \widehat{l'_3}, \widehat{l^{\dagger}_3}, \widehat{l'_1}, \iota_l\right)\right|\right]^2 \\
&\leq 2^{n(L'_2+L^{\dagger}_2+L'_3+L^{\dagger}_3+L'_1)}.2^{-n(I(U_2;U_3|U_1)+I(U_2,U_3;Y_3|U_1)+I(X;Y_3|U_2,U_3,U_1)-\delta(\epsilon))} \\
&\quad + 2^{n(2L'_2+2L^{\dagger}_2+L'_3+L^{\dagger}_3+L'_1)}.2^{-n(2I(U_2;U_3|U_1)+I(U_2,U_3;Y_3|U_1)+I(U_2;Y_3|U_3,U_1)+I(X;Y_3|U_2,U_3,U_1)-2\delta(\epsilon))} \\
&\quad + 2^{n(L'_2+L^{\dagger}_2+2L'_3+2L^{\dagger}_3+L'_1)}.2^{-n(2I(U_2;U_3|U_1)+I(U_2,U_3;Y_3|U_1)+I(U_3;Y_3|U_2,U_1)+I(X;Y_3|U_2,U_3,U_1)-2\delta(\epsilon))} \\
&\quad + 2^{n(L'_2+L^{\dagger}_2+L'_3+L^{\dagger}_3+2L'_1)}.2^{-n(I(U_2;U_3|U_1)+I(U_2,U_3;Y_3|U_1)+2I(X;Y_3|U_2,U_3,U_1)-2\delta(\epsilon))} \\
&\quad + 2^{n(2L'_2+2L^{\dagger}_2+2L'_3+2L^{\dagger}_3+L'_1)}.2^{-n(2I(U_2;U_3|U_1)+2I(U_2,U_3;Y_3|U_1)+I(X;Y_3|U_2,U_3,U_1)-2\delta(\epsilon))} \\
&\quad + 2^{n(2L'_2+2L^{\dagger}_2+L'_3+L^{\dagger}_3+2L'_1)} \\
&\quad \times 2^{-n(2I(U_2;U_3|U_1)+I(U_2,U_3;Y_3|U_1)+I(U_2;Y_3|U_3,U_1)+2I(X;Y_3|U_2,U_3,U_1)-2\delta(\epsilon))} \\
&\quad + 2^{n(L'_2+L^{\dagger}_2+2L'_3+2L^{\dagger}_3+2L'_1)} \\
&\quad \times 2^{-n(2I(U_2;U_3|U_1)+I(U_2,U_3;Y_3|U_1)+I(U_3;Y_3|U_2,U_1)+2I(X;Y_3|U_2,U_3,U_1)-2\delta(\epsilon))}.
\end{aligned} \tag{B.74}$$

We also know that

$$\Pr\left[\mathcal{E}\left(\widehat{l'_2}, \widehat{l^{\dagger}_2}, \widehat{l'_3}, \widehat{l^{\dagger}_3}, \widehat{l'_1}, \iota_l\right)\right] \leq \frac{\mathrm{Var}\left[\left|\mathcal{A}(\widehat{l'_2}, \widehat{l^{\dagger}_2}, \widehat{l'_3}, \widehat{l^{\dagger}_3}, \widehat{l'_1}, \iota_l)\right|\right]}{\left(\mathrm{E}\left[\left|\mathcal{A}\left(\widehat{l'_2}, \widehat{l^{\dagger}_2}, \widehat{l'_3}, \widehat{l^{\dagger}_3}, \widehat{l'_1}, \iota_l\right)\right|\right]\right)^2}. \tag{B.75}$$

Combining (B.74) and the lower bound in (B.55) using (B.75), we get

$$\begin{aligned}
&\Pr\left[\mathcal{E}\left(\widehat{l'_2},\widehat{l^\dagger_2},\widehat{l'_3},\widehat{l^\dagger_3},\widehat{l'_1},\iota_l\right)\right]\\
&\le 2^{-n(L'_2+L^\dagger_2+L'_3+L^\dagger_3+L'_1-I(U_2;U_3|U_1)-I(U_2,U_3;Y_3|U_1)-I(X;Y_3|U_2,U_3,U_1)-3\delta(\epsilon))}\\
&\quad+2^{-n(L'_3+L^\dagger_3+L'_1-I(U_2,U_3;Y_3|U_1)+I(U_2;Y_3|U_3,U_1)-I(X;Y_3|U_2,U_3,U_1)-4\delta(\epsilon))}\\
&\quad+2^{-n(L'_2+L^\dagger_2+L'_1-I(U_2,U_3;Y_3|U_1)+I(U_3;Y_3|U_2,U_1)-I(X;Y_3|U_2,U_3,U_1)-4\delta(\epsilon))}\\
&\quad+2^{-n(L'_2+L^\dagger_2+L'_3+L^\dagger_3-I(U_2;U_3|U_1)-I(U_2,U_3;Y_3|U_1)-4\delta(\epsilon))}\\
&\quad+2^{-n(L'_1-I(X;Y_3|U_2,U_3,U_1)-4\delta(\epsilon))}\\
&\quad+2^{-n(L'_3+L^\dagger_3-I(U_2,U_3;Y_3|U_1)+I(U_2;Y_3|U_3,U_1)-4\delta(\epsilon))}\\
&\quad+2^{-n(L'_2+L^\dagger_2-I(U_2,U_3;Y_3|U_1)+I(U_3;Y_3|U_2,U_1)-4\delta(\epsilon))}.
\end{aligned}\tag{B.76}$$

Thus $\Pr\left[\mathcal{E}\left(\widehat{l'_2},\widehat{l^\dagger_2},\widehat{l'_3},\widehat{l^\dagger_3},\widehat{l'_1},\iota_l\right)\right]\to 0$ for n sufficiently large under the following conditions:

$$\begin{aligned}
&L'_2+L^\dagger_2+L'_3+L^\dagger_3+L'_1\ge I(U_2;U_3|U_1)+I(U_2,U_3;Y_3|U_1)+I(X;Y_3|U_2,U_3,U_1),\\
&L'_3+L^\dagger_3+L'_1\ge I(U_2,U_3;Y_3|U_1)-I(U_2;Y_3|U_3,U_1)+I(X;Y_3|U_2,U_3,U_1)\\
&\qquad\qquad=I(U_3;Y_3|U_1)+I(X;Y_3|U_2,U_3,U_1),\\
&L'_2+L^\dagger_2+L'_1\ge I(U_2,U_3;Y_3|U_1)-I(U_3;Y_3|U_2,U_1)+I(X;Y_3|U_2,U_3,U_1)\\
&\qquad\qquad=I(U_2;Y_3|U_1)+I(X;Y_3|U_2,U_3,U_1),\\
&L'_2+L^\dagger_2+L'_3+L^\dagger_3\ge I(U_2;U_3|U_1)+I(U_2,U_3;Y_3|U_1),\\
&L'_1\ge I(X;Y_3|U_2,U_3,U_1),\\
&L'_3+L^\dagger_3\ge I(U_2,U_3;Y_3|U_1)-I(U_2;Y_3|U_3,U_1)=I(U_3;Y_3|U_1),\\
&L'_2+L^\dagger_2\ge I(U_2,U_3;Y_3|U_1)-I(U_3;Y_3|U_2,U_1)=I(U_2;Y_3|U_1).
\end{aligned}\tag{B.77}$$

Next, for each w_1, we define a random version of $\mathcal{A}\left(\widehat{l'_2},\widehat{l^\dagger_2},\widehat{l'_3},\widehat{l^\dagger_3},\widehat{l'_1},\iota_l\right)$ in (B.52) as

$$\widetilde{\mathcal{A}}\left(\widehat{l'_2},\widehat{l^\dagger_2},\widehat{l'_3},\widehat{l^\dagger_3},\widehat{l'_1},\iota_l\right)\triangleq\left\{\begin{array}{c}(\widehat{l'_2},\widehat{l^\dagger_2})\in\mathcal{B}_2(w_1,l_2),(\widehat{l'_3},\widehat{l^\dagger_3})\in\mathcal{B}_3(l_3),\widehat{l'_1}\in\mathcal{B}_1(l_1):\\ \left(\mathbf{u}_1,\mathbf{U}_2(\widehat{l'_2},\widehat{l^\dagger_2}),\mathbf{U}_3(\widehat{l'_3},\widehat{l^\dagger_3}),\mathbf{X}(\widehat{l'_1}),\mathbf{Y}_3\right)\in\mathcal{T}^n_\epsilon,\\ \widehat{l'_2}\ne L'_2,\widehat{l^\dagger_2}\ne L^\dagger_2,\widehat{l'_3}\ne L'_3,\widehat{l^\dagger_3}\ne L^\dagger_3,\widehat{l'_1}\ne L'_1\end{array}\right\}.\tag{B.78}$$

That is, the set of the eavesdropper's estimate of the indices $\left(l'_2,l^\dagger_2,l'_3,l^\dagger_3,l'_1\right)$ that is not equal to the random transmitted indices $\left(L'_2,L^\dagger_2,L'_3,L^\dagger_3,L'_1\right)$, given random eavesdrop-

per received signal. We also define

$$\mathcal{E}(w_1) \triangleq \left\{ \left| \widetilde{\mathcal{A}}\left(\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l\right) \right| \geq 2\mathrm{E}\left[\left| \widetilde{\mathcal{A}}\left((\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l\right) \right| \right] \right\}, \tag{B.79}$$

and the indicator variables

$$\mathbb{I}(w_1) := 0 \text{ if } \left(\mathbf{u}_1, \mathbf{U}_2(\widehat{L'_2}, \widehat{L^\dagger_2}), \mathbf{U}_3(\widehat{L'_3}, \widehat{L^\dagger_3}), \mathbf{X}(\widehat{L'_1}), \mathbf{Y}_3\right) \in \mathcal{T}_\epsilon^n \text{ and } \mathcal{E}(w_1)^c \text{ occurs},$$
$$\mathbb{I}(w_1) := 1 \text{ if } \left(\mathbf{u}_1, \mathbf{U}_2(\widehat{L'_2}, \widehat{L^\dagger_2}), \mathbf{U}_3(\widehat{L'_3}, \widehat{L^\dagger_3}), \mathbf{X}(\widehat{L'_1}), \mathbf{Y}_3\right) \notin \mathcal{T}_\epsilon^n \text{ and } \mathcal{E}(w_1) \text{ occurs}.$$

We now want to find the probability that $\mathbb{I}(w_1) = 1$. Using the union bound,

$$\begin{aligned}
&\Pr\left[\mathbb{I}(w_1) = 1\right] \\
&\leq \Pr\left[\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{L'_2}, \widehat{L^\dagger_2}), \mathbf{U}_3(\widehat{L'_3}, \widehat{L^\dagger_3}), \mathbf{X}(\widehat{L'_1}), \mathbf{Y}_3\right) \notin \mathcal{T}_\epsilon^n\right] + \Pr\left[\mathcal{E}(w_1)\right] \\
&\leq \Pr\left[\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{L'_2}, \widehat{L^\dagger_2}), \mathbf{U}_3(\widehat{L'_3}, \widehat{L^\dagger_3}), \mathbf{X}(\widehat{L'_1}), \mathbf{Y}_3\right) \notin \mathcal{T}_\epsilon^n\right] \\
&\quad + \sum_{\mathbf{y}_3 \in \mathcal{T}_\epsilon^n} p(\mathbf{y}_3) \Pr\left[\mathcal{E}(w_1) | \mathbf{Y}_3 = \mathbf{y}_3\right] + \Pr\left[\mathbf{Y}_3 \notin \mathcal{T}_\epsilon^n(P_{Y_3})\right] \\
&\overset{(a)}{=} \sum_{\mathbf{y}_3 \in \mathcal{T}_\epsilon^n} \sum_{\iota_l} p(\mathbf{y}_3) p(\iota_l | \mathbf{y}_3) \Pr\left[\mathcal{E}(w_1) | \mathbf{Y}_3 = \mathbf{y}_3, \mathcal{I}(L) = \iota_l\right] \\
&= \sum_{\mathbf{y}_3 \in \mathcal{T}_\epsilon^n} \sum_{\iota_l} p(\mathbf{y}_3) p(\iota_l | \mathbf{y}_3) \Pr\left[\mathcal{E}\left(\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l\right)\right],
\end{aligned} \tag{B.80}$$

where (a) is by the fact that the first term

$$\Pr\left[\left(\mathbf{u}_1, \mathbf{U}_2(\widehat{L'_2}, \widehat{L^\dagger_2}), \mathbf{U}_3(\widehat{L'_3}, \widehat{L^\dagger_3}), \mathbf{X}(\widehat{L'_1}), \mathbf{Y}_3\right) \notin \mathcal{T}_\epsilon^n\right] \to 0$$

for n sufficiently large by the properties of joint typical sequences, and this implies that the third term also goes to 0 for n sufficiently large. Finally, since we know that $\Pr\left[\{\mathcal{E}\left(\widehat{l'_2}, \widehat{l^\dagger_2}, \widehat{l'_3}, \widehat{l^\dagger_3}, \widehat{l'_1}, \iota_l\right)\right] \to 0$ for n sufficiently large under the conditions of (B.77), we have $\Pr[\mathbb{I}(w_1) = 1] \to 0$ for n sufficiently large.

To bound $H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1)$, we consider the expansion of the entropy $H(\mathbb{I}(w_1), \mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1)$. We have

$$\begin{aligned}
H(\mathbb{I}(w_1), \mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1) &= H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1) + H(\mathbb{I}(w_1)|\mathbf{U}_1, \mathbf{Y}_3, \mathcal{I}(L), w_1) \\
&= H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1),
\end{aligned} \tag{B.81}$$

where the second equality is because $H(\mathbb{I}(w_1)|\mathbf{U}_1, \mathbf{Y}_3, \mathcal{I}(L), w_1) = 0$ as $\mathbb{I}(w_1)$ is de-

termined by $(\mathbf{Y}_3, \mathcal{I}(L), w_1)$. So we now have

$$\begin{aligned}
&H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1) = H(\mathbb{I}(w_1), \mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1) \\
&= H(\mathbb{I}(w_1)|\mathbf{U}_1, \mathbf{Y}_3, w_1) + \Pr\left[\mathbb{I}(w_1) = 1\right] H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1, \mathbb{I}(w_1) = 1) \\
&\quad + \Pr\left[\mathbb{I}(w_1) = 0\right] H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1, \mathbb{I}(w_1) = 0) \\
&\overset{(a)}{\leq} 1 + \Pr\left[\mathbb{I}(w_1) = 1\right] H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1, \mathbb{I}(w_1) = 1) \\
&\quad + (1 - \Pr\left[\mathbb{I}(w_1) = 1\right]) H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1, \mathbb{I}(w_1) = 0) \\
&\leq 1 + \Pr\left[\mathbb{I}(w_1) = 1\right] H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1, \mathbb{I}(w_1) = 1) \\
&\quad + H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1, \mathbb{I}(w_1) = 0), \qquad \text{(B.82)}
\end{aligned}$$

where (a) is due to $H(\mathbb{I}(w_1)|\mathbf{U}_1, \mathbf{Y}_3, w_1) \leq H(\mathbb{I}(w_1)) \leq 1$ as $\mathbb{I}(w_1)$ is binary-valued. In (B.82), the second term $\to 0$ for n sufficiently large. For the third term, we know that, given $(\mathbf{U}_1, \mathbf{Y}_3, w_1)$ and $\mathbb{I}(w_1) = 0$, the number of values $\mathcal{I}(L)$ takes on is upper bounded by $\log\left(2\mathrm{E}\left[\left|\mathcal{A}\left(\widehat{l_2'}, \widehat{l_2^\dagger}, \widehat{l_3'}, \widehat{l_3^\dagger}, \widehat{l_1'}, \iota_l\right)\right|\right] - 1\right)$. The third term of (B.82) now becomes

$$\begin{aligned}
&H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1, \mathbb{I}(w_1) = 0) \leq \log\left(2\mathrm{E}\left[\left|\mathcal{A}\left(\widehat{l_2'}, \widehat{l_2^\dagger}, \widehat{l_3'}, \widehat{l_3^\dagger}, \widehat{l_1'}, \iota_l\right)\right|\right] - 1\right) \\
&= \log\left(2^{n(L_2' + L_2^\dagger + L_3' + L_3^\dagger + L_1' - I(U_2;U_3|U_1) - I(U_2,U_3;Y_3|U_1) - I(X;Y_3|U_2,U_3,U_1) + \delta(\epsilon)) + 1} - 1\right) \\
&= n(L_2' + L_2^\dagger + L_3' + L_3^\dagger + L_1' - I(U_2;U_3|U_1) \\
&\quad - I(U_2,U_3;Y_3|U_1) - I(X;Y_3|U_2,U_3,U_1) + \delta(\epsilon)) \\
&\quad + \log\left(2 - 2^{-n(L_2' + L_2^\dagger + L_3' + L_3^\dagger + L_1' - I(U_2;U_3|U_1) - I(U_2,U_3;Y_3|U_1) - I(X;Y_3|U_2,U_3,U_1) + \delta(\epsilon))}\right) \\
&\overset{(a)}{\leq} n(L_2' + L_2^\dagger + L_3' + L_3^\dagger + L_1' - I(U_2;U_3|U_1) \\
&\quad - I(U_2,U_3;Y_3|U_1) - I(X;Y_3|U_2,U_3,U_1) + \delta(\epsilon)) \\
&\quad + 1 - 2^{-n(L_2' + L_2^\dagger + L_3' + L_3^\dagger + L_1' - I(U_2;U_3|U_1) - I(U_2,U_3;Y_3|U_1) - I(X;Y_3|U_2,U_3,U_1) + \delta(\epsilon))},
\end{aligned}$$

where (a) is due to using the relation $\log(x) \leq x - 1$. Then, for n sufficiently large,

$$\begin{aligned}
H(\mathcal{I}(L)|\mathbf{U}_1, \mathbf{Y}_3, w_1) \leq n(&L_2' + L_2^\dagger + L_3' + L_3^\dagger + L_1' - I(U_2;U_3|U_1) - I(U_2,U_3;Y_3|U_1) \\
&- I(X;Y_3|U_2,U_3,U_1) + \delta(\epsilon)) + 2, \qquad \text{(B.83)}
\end{aligned}$$

since $L_2' + L_2^\dagger + L_3' + L_3^\dagger + L_1' \geq I(U_2;U_3|U_1) + I(U_2,U_3;Y_3|U_1) + I(X;Y_3|U_2,U_3,U_1)$. By averaging over w_1, the lemma is proved.

B.5 Fourier-Motzkin Elimination for Inner Bound in Theorem 14

Our starting point is the set of inequalities, using $R_1 = L_1 + L_2 + L_3$ and $R_{1e} = L_{1e} + L_{2e} + L_{3e}$

$$
\begin{aligned}
&L_{2e} \le Q_2 - I(U_2;Y_3|U_1)\\
&L_{3e} \le Q_3 - I(U_3;Y_3|U_1)\\
&L_{1e} \le I(X;Y_1|U_2,U_3) - I(X;Y_3|U_2,U_3)\\
&L_{2e} + L_{1e} \le I(X;Y_1|U_3) - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1)\\
&L_{3e} + L_{1e} \le I(X;Y_1|U_2) - I(X;Y_3|U_2,U_3) - I(U_3;Y_3|U_1)\\
&R_{1e} \le I(X;Y_1|U_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1)\\
&R_0 + R_{1e} \le I(X;Y_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1)\\
&L_2 \le Q_2, \;\; L_3 \le Q_3\\
&L_2 + L_3 \le Q_2 + Q_3 - I(U_2;U_3|U_1)\\
&R_0 + Q_2 \le I(U_2;Y_2)\\
&R_0 + Q_3 \le I(U_3;Y_3)\\
&R_0 + R_1 \le I(X;Y_1)\\
&R_1 \le I(X;Y_1|U_1)\\
&L_1 + L_3 \le I(X;Y_1|U_2)\\
&L_1 + L_2 \le I(X;Y_1|U_3)\\
&L_1 \le I(X;Y_1|U_2,U_3)\\
&0 \le L_1, \;\; 0 \le L_2, \;\; 0 \le L_3\\
&0 \le L_{1e} \;\; 0 \le L_{2e} \;\; 0 \le L_{3e}.
\end{aligned}
\tag{B.84}
$$

Eliminating Q_2, we have

$$
\begin{aligned}
&R_0 + L_{2e} \le I(U_2;Y_2) - I(U_2;Y_3|U_1)\\
&R_0 + L_2 \le I(U_2;Y_2)\\
&R_0 + L_2 + L_3 \le I(U_2;Y_2) + Q_3 - I(U_2;U_3|U_1)\\
&L_{3e} \le Q_3 - I(U_3;Y_3|U_1)
\end{aligned}
$$

$$
\begin{aligned}
&L_{1e} \le I(X;Y_1|U_2,U_3) - I(X;Y_3|U_2,U_3)\\
&L_{2e} + L_{1e} \le I(X;Y_1|U_3) - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1)\\
&L_{3e} + L_{1e} \le I(X;Y_1|U_2) - I(X;Y_3|U_2,U_3) - I(U_3;Y_3|U_1)\\
&R_{1e} \le I(X;Y_1|U_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1)\\
&R_0 + R_{1e} \le I(X;Y_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1)\\
&L_3 \le Q_3\\
&R_0 + Q_3 \le I(U_3;Y_3)\\
&R_0 + R_1 \le I(X;Y_1)\\
&R_1 \le I(X;Y_1|U_1)\\
&L_1 + L_3 \le I(X;Y_1|U_2)\\
&L_1 + L_2 \le I(X;Y_1|U_3)\\
&L_1 \le I(X;Y_1|U_2,U_3)\\
&0 \le L_1,\ \ 0 \le L_2,\ \ 0 \le L_3\\
&0 \le L_{1e}\ \ 0 \le L_{2e}\ \ 0 \le L_{3e}.
\end{aligned}
\tag{B.85}
$$

Eliminating Q_3, we have

$$
\begin{aligned}
&R_0 + L_{3e} \le I(U_3;Y_3) - I(U_3;Y_3|U_1)\\
&R_0 + L_{2e} \le I(U_2;Y_2) - I(U_2;Y_3|U_1)\\
&R_0 + L_2 \le I(U_2;Y_2)\\
&R_0 + L_3 \le I(U_3;Y_3)\\
&2R_0 + L_2 + L_3 \le I(U_2;Y_2) + I(U_3;Y_3) - I(U_2;U_3|U_1)\\
&L_{1e} \le I(X;Y_1|U_2,U_3) - I(X;Y_3|U_2,U_3)\\
&L_{2e} + L_{1e} \le I(X;Y_1|U_3) - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1)\\
&L_{3e} + L_{1e} \le I(X;Y_1|U_2) - I(X;Y_3|U_2,U_3) - I(U_3;Y_3|U_1)\\
&R_{1e} \le I(X;Y_1|U_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1)\\
&R_0 + R_{1e} \le I(X;Y_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1)\\
&R_0 + R_1 \le I(X;Y_1)\\
&R_1 \le I(X;Y_1|U_1)\\
&L_1 + L_3 \le I(X;Y_1|U_2)
\end{aligned}
$$

$$
\begin{aligned}
&L_1 + L_2 \le I(X;Y_1|U_3)\\
&L_1 \le I(X;Y_1|U_2,U_3)\\
&0 \le L_1,\ \ 0 \le L_2,\ \ 0 \le L_3\\
&0 \le L_{1e}\ \ 0 \le L_{2e}\ \ 0 \le L_{3e}.
\end{aligned} \tag{B.86}
$$

Now substitute $L_{1e} = R_{1e} - L_{2e} - L_{3e}$ and $L_1 = R_1 - L_2 - L_3$ into (B.86) and we get

$$
\begin{aligned}
&R_0 + L_{3e} \le I(U_3;Y_3) - I(U_3;Y_3|U_1)\\
&R_0 + L_{2e} \le I(U_2;Y_2) - I(U_2;Y_3|U_1)\\
&R_0 + L_2 \le I(U_2;Y_2)\\
&R_0 + L_3 \le I(U_3;Y_3)\\
&2R_0 + L_2 + L_3 \le I(U_2;Y_2) + I(U_3;Y_3) - I(U_2;U_3|U_1)\\
&R_{1e} \le L_{2e} + L_{3e} + I(X;Y_1|U_2,U_3) - I(X;Y_3|U_2,U_3)\\
&R_{1e} \le L_{3e} + I(X;Y_1|U_3) - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1)\\
&R_{1e} \le L_{2e} + I(X;Y_1|U_2) - I(X;Y_3|U_2,U_3) - I(U_3;Y_3|U_1)\\
&R_{1e} \le I(X;Y_1|U_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1)\\
&R_0 + R_{1e} \le I(X;Y_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1)\\
&R_0 + R_1 \le I(X;Y_1)\\
&R_1 \le I(X;Y_1|U_1)\\
&R_1 \le L_2 + I(X;Y_1|U_2)\\
&R_1 \le L_3 + I(X;Y_1|U_3)\\
&R_1 \le L_2 + L_3 + I(X;Y_1|U_2,U_3)\\
&L_2 + L_3 \le R_1,\ \ 0 \le L_2,\ \ 0 \le L_3\\
&L_{2e} + L_{3e} \le R_{1e}\ \ 0 \le L_{2e}\ \ 0 \le L_{3e}.
\end{aligned} \tag{B.87}
$$

Eliminating L_2 from (B.87), we obtain, after removing redundant inequalities,

$$
\begin{aligned}
&0 \le L_3\\
&L_3 + 2R_0 + R_1 \le I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2)\\
&L_3 + R_0 \le I(U_3;Y_3)\\
&L_3 \le I(X;Y_1|U_2)
\end{aligned}
$$

$$
\begin{aligned}
&R_0 + R_1 \leq I(U_2;Y_2) + I(X;Y_1|U_2,U_3) + L_3 \\
&L_3 + 2R_0 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) \\
&L_3 \leq R_1 \\
&0 \leq L_{2e} \\
&0 \leq L_{3e} \\
&L_{2e} + R_0 \leq I(U_2;Y_2) - I(U_2;Y_3|U_1) \\
&L_{3e} + R_0 \leq I(U_3;Y_3) - I(U_3;Y_3|U_1) \\
&L_{2e} + L_{3e} \leq R_{1e} \\
&R_{1e} \leq I(X;Y_1|U_2,U_3) + L_{2e} + L_{3e} - I(X;Y_3|U_2,U_3) \\
&R_{1e} \leq I(X;Y_1|U_2) + L_{2e} - I(X;Y_3|U_2,U_3) - I(U_3;Y_3|U_1) \\
&R_{1e} \leq I(X;Y_1|U_3) + L_{3e} - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1) \\
&R_0 + R_1 \leq I(U_2;Y_2) + I(X;Y_1|U_2) \\
&R_0 + R_1 \leq I(X;Y_1) \\
&R_1 \leq I(X;Y_1|U_1) \\
&R_1 \leq I(X;Y_1|U_3) + L_3 \\
&R_0 + R_{1e} \leq I(X;Y_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1) \\
&R_{1e} \leq I(X;Y_1|U_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1) \\
&R_{1e} \leq R_1 \\
&2R_0 + R_1 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3) \\
&R_0 \leq I(U_2;Y_2)
\end{aligned}
\tag{B.88}
$$

We now eliminate L_3 from (B.88). After removing redundant inequalities, we have

$$
\begin{aligned}
&R_0 + R_1 \leq I(X;Y_1) \\
&R_1 \leq I(X;Y_1|U_1) \\
&R_{1e} \leq I(X;Y_1|U_2,U_3) + L_{2e} + L_{3e} - I(X;Y_3|U_2,U_3) \\
&R_0 + R_{1e} \leq I(X;Y_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1) \\
&R_{1e} \leq I(X;Y_1|U_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1) \\
&R_{1e} \leq I(X;Y_1|U_2) + L_{2e} - I(X;Y_3|U_2,U_3) - I(U_3;Y_3|U_1) \\
&R_{1e} \leq I(X;Y_1|U_3) + L_{3e} - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1)
\end{aligned}
$$

$$L_{2e} + R_0 \leq I(U_2;Y_2) - I(U_2;Y_3|U_1)$$

$$L_{3e} + R_0 \leq I(U_3;Y_3) - I(U_3;Y_3|U_1)$$

$$L_{2e} + L_{3e} \leq R_{1e}$$

$$0 \leq L_{2e}$$

$$0 \leq L_{3e}$$

$$R_{1e} \leq R_1$$

$$R_0 + R_1 \leq I(U_2;Y_2) + I(X;Y_1|U_2)$$

$$2R_0 + R_1 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3)$$

$$R_0 \leq I(U_2;Y_2)$$

$$R_0 + R_1 \leq I(U_3;Y_3) + I(X;Y_1|U_3)$$

$$2R_0 + 2R_1 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2) + I(X;Y_1|U_3)$$

$$R_1 \leq I(X;Y_1|U_2) + I(X;Y_1|U_3)$$

$$R_0 \leq I(U_3;Y_3)$$

$$2R_0 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3). \tag{B.89}$$

Eliminating L_{2e} from the inequalities in (B.89) and removing redundancies, we have

$$R_0 + R_1 \leq I(X;Y_1)$$

$$R_1 \leq I(X;Y_1|U_1)$$

$$R_0 + R_{1e} \leq I(X;Y_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1)$$

$$R_{1e} \leq I(X;Y_1|U_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1)$$

$$R_{1e} \leq I(X;Y_1|U_3) + L_{3e} - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1)$$

$$L_{3e} + R_0 \leq I(U_3;Y_3) - I(U_3;Y_3|U_1)$$

$$0 \leq L_{3e}$$

$$R_{1e} \leq R_1$$

$$R_0 + R_1 \leq I(U_2;Y_2) + I(X;Y_1|U_2)$$

$$2R_0 + R_1 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3)$$

$$R_0 + R_1 \leq I(U_3;Y_3) + I(X;Y_1|U_3)$$

$$2R_0 + 2R_1 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2) + I(X;Y_1|U_3)$$

$$R_1 \leq I(X;Y_1|U_2) + I(X;Y_1|U_3)$$

$$
\begin{aligned}
&R_0 \leq I(U_3;Y_3)\\
&2R_0 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3)\\
&R_0 + R_{1e} \leq I(U_2;Y_2) + I(X;Y_1|U_2,U_3) + L_{3e} - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1)\\
&R_0 + R_{1e} \leq I(U_2;Y_2) + I(X;Y_1|U_2) - I(X;Y_3|U_2,U_3)\\
&\qquad\qquad - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1)\\
&L_{3e} \leq I(X;Y_1|U_2) - I(X;Y_3|U_2,U_3) - I(U_3;Y_3|U_1)\\
&R_0 \leq I(U_2;Y_2) - I(U_2;Y_3|U_1)\\
&L_{3e} \leq R_{1e}
\end{aligned}
\tag{B.90}
$$

along with the condition $I(X;Y_1|U_2,U_3) - I(X;Y_3|U_2,U_3) \geq 0$. Next, eliminating L_{3e} from the inequalities in (B.90) and removing redundancies with the help of $I(X;Y_1|U_2,U_3) - I(X;Y_3|U_2,U_3) \geq 0$, we have

$$
\begin{aligned}
&R_0 + R_1 \leq I(X;Y_1)\\
&R_1 \leq I(X;Y_1|U_1)\\
&R_0 + R_{1e} \leq I(X;Y_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1)\\
&R_{1e} \leq I(X;Y_1|U_1) - I(U_2;U_3|U_1) - I(X;Y_3|U_2,U_3) - I(U_2,U_3;Y_3|U_1)\\
&R_{1e} \leq R_1\\
&R_0 + R_1 \leq I(U_2;Y_2) + I(X;Y_1|U_2)\\
&2R_0 + R_1 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3)\\
&R_0 + R_1 \leq I(U_3;Y_3) + I(X;Y_1|U_3)\\
&2R_0 + 2R_1 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2)\\
&\qquad\qquad + I(X;Y_1|U_3)\\
&R_1 \leq I(X;Y_1|U_2) + I(X;Y_1|U_3)\\
&2R_0 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3)\\
&R_0 + R_{1e} \leq I(U_2;Y_2) + I(X;Y_1|U_2) - I(X;Y_3|U_2,U_3)\\
&\qquad\qquad - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1)\\
&R_0 \leq I(U_2;Y_2) - I(U_2;Y_3|U_1)\\
&R_0 + R_{1e} \leq I(U_3;Y_3) + I(X;Y_1|U_3) - I(X;Y_3|U_2,U_3)\\
&\qquad\qquad - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1)
\end{aligned}
$$

$$
\begin{aligned}
R_{1e} &\leq I(X;Y_1|U_2) + I(X;Y_1|U_3) - 2I(X;Y_3|U_2,U_3) \\
&\quad - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1) \\
R_0 &\leq I(U_3;Y_3) - I(U_3;Y_3|U_1) \\
2R_0 + R_{1e} &\leq I(U_2;Y_2) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3) - I(X;Y_3|U_2,U_3) \\
&\quad - I(U_2;Y_3|U_1) - I(U_3;Y_3|U_1). \qquad \text{(B.91)}
\end{aligned}
$$

We also have the conditions $I(X;Y_1|U_3) - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1) \geq 0$, $I(X;Y_1|U_2) - I(X;Y_3|U_2,U_3) - I(U_3;Y_3|U_1) \geq 0$.

Next, to perform the rate splitting operation mentioned in the proof of Theorem 14, we make the substitutions $R_0 = R_{0n} + R_{1n} - R_{11}, R_1 = R_{11}$ into (B.91), and add the conditions $R_{11} \leq R_{1n}, 0 \leq R_{11}$. Then, re-assign the new $R_0 = R_{0n}$, $R_1 = R_{1n}$ and eliminate R_{11}. We end up with the set of inequalities of Theorem 14, after removing redundancies with the help of the conditions

$$
\begin{aligned}
&I(X;Y_1|U_3) - I(X;Y_3|U_2,U_3) - I(U_2;Y_3|U_1) \geq 0, \\
&I(X;Y_1|U_2) - I(X;Y_3|U_2,U_3) - I(U_3;Y_3|U_1) \geq 0, \\
&I(X;Y_1|U_1) \geq I(X;Y_1|U_2,U_3). \qquad \text{(B.92)}
\end{aligned}
$$

The last condition can be easily seen by the fact that

$$
\begin{aligned}
I(X;Y_1|U_1) &= I(U_2,U_3;Y_1|U_1) + I(X;Y_1|U_2,U_3,U_1) \\
&= I(U_2,U_3;Y_1|U_1) + I(X;Y_1|U_2,U_3), \qquad \text{(B.93)}
\end{aligned}
$$

given the Markov chain $U_1 \to (U_2,U_3) \to X \to Y_1$.

Appendix C

Reduction of Rate Region in Theorem 14 to Special Cases

C.1 Reduction to General 3-receiver 2 DMS Inner Bound

From Theorem 14, removing the equivocation constraints, we have the following region over $p(u_1)p(u_2,u_3|u_1)p(x|u_2,u_3)$,

$$R_0 \leq \min\{I(U_2;Y_2), I(U_3;Y_3)\}$$

$$R_0 \leq I(X;Y_1) - I(U_2;U_3|U_1) \quad \text{(i)}$$

$$R_0 \leq I(U_2;Y_2) + I(X;Y_1|U_2) \quad (*)$$

$$R_0 \leq I(U_3;Y_3) + I(X;Y_1|U_3) \quad (*)$$

$$2R_0 \leq I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3)$$

$$2R_0 \leq I(U_2;Y_2) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3) \quad (*)$$

$$R_0 + R_1 \leq I(X;Y_1)$$

$$R_0 + R_1 \leq I(U_2;Y_2) + I(X;Y_1|U_2)$$

$$R_0 + R_1 \leq I(U_3;Y_3) + I(X;Y_1|U_3)$$

$$R_0 + R_1 \leq I(U_2;Y_2) + I(X;Y_1|U_1) \quad (*)$$

$$R_0 + R_1 \leq I(U_3;Y_3) + I(X;Y_1|U_1) \quad (*)$$

$$R_0 + R_1 \leq I(X;Y_1|U_2) - I(U_2;U_3|U_1) + I(X;Y_1|U_3) + I(X;Y_1) \quad \text{(ii)}$$

$$R_0 + R_1 \leq I(U_2;Y_2) + I(X;Y_1|U_1) + I(X;Y_1|U_2) \quad (*)$$

$$
\begin{aligned}
&R_0 + R_1 \le I(U_2;Y_2) + 2I(X;Y_1|U_2) + I(X;Y_1|U_3) \quad (*)\\
&R_0 + R_1 \le I(U_3;Y_3) + I(X;Y_1|U_1) + I(X;Y_1|U_3) \quad (*)\\
&R_0 + R_1 \le I(U_3;Y_3) + I(X;Y_1|U_2) + 2I(X;Y_1|U_3) \quad (*)\\
&2R_0 + R_1 \le I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2,U_3)\\
&2R_0 + 2R_1 \le I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_2) + I(X;Y_1|U_3)\\
&2R_0 + 2R_1 \le I(U_2;Y_2) - I(U_2;U_3|U_1) + I(U_3;Y_3) + I(X;Y_1|U_1) + I(X;Y_1|U_2,U_3)\\
&2R_0 + 2R_1 \le I(U_2;Y_2) + I(U_3;Y_3) + 2I(X;Y_1|U_2)\\
&\qquad\quad + 2I(X;Y_1|U_3) + I(X;Y_1|U_2,U_3) \quad (*)\\
&2R_0 + 2R_1 \le I(U_2;Y_2) + I(U_3;Y_3) + 2I(X;Y_1|U_1) + I(X;Y_1|U_2,U_3) \quad (*)
\end{aligned} \tag{C.1}
$$

The starred inequalities are redundant, while the inequalities marked (i) and (ii) can be shown to be redundant as follows. For the inequality marked (i), we have

$$
\begin{aligned}
R_0 &\le I(X;Y_1) - I(U_2;U_3|U_1)\\
&\overset{(a)}{=} I(U_1;Y_1) + I(X;Y_1|U_1) - I(U_2;U_3,Y_1|U_1) + I(U_2;Y_1|U_3,U_1)\\
&\overset{(b)}{=} I(U_1;Y_1) + I(X;Y_1|U_2) + I(U_2;Y_1|U_1) - I(U_2;Y_1|U_1)\\
&\quad - I(U_2;U_3|Y_1,U_1) + I(U_2;Y_1|U_3,U_1)\\
&= I(U_1;Y_1) + I(X;Y_1|U_2) - H(U_2|Y_1,U_1) + H(U_2|U_3,U_1)\\
&\le I(U_1;Y_1) + I(X;Y_1|U_2) - H(U_2|Y_1,U_1) + H(U_2|U_1)\\
&= I(U_1;Y_1) + I(X;Y_1|U_2) + I(U_2;Y_1|U_1)\\
&\overset{(c)}{=} I(U_1;Y_1) + I(X;Y_1|U_1) = I(X;Y_1),
\end{aligned} \tag{C.2}
$$

which is redundant. In the above, (a) is due to $U_1 \to X \to Y_1$ forming a Markov chain so that $I(X;Y_1) = I(U;Y_1) + I(X;Y_1|U_1)$, (b) is due to $U_1 \to U_2 \to X \to Y_1$ so that

$$
\begin{aligned}
I(X;Y_1|U_1) &= I(X;Y_1|U_2,U_1) + I(U_2;Y_1|U_1)\\
&= I(X;Y_1|U_2) + I(U_2;Y_1|U_1).
\end{aligned} \tag{C.3}
$$

For the inequality marked (ii), we have

$$
\begin{aligned}
R_0 + R_1 &\le I(X;Y_1|U_2) - I(U_2;U_3|U_1) + I(X;Y_1|U_3) + I(X;Y_1)\\
&\overset{(a)}{\le} I(X;Y_1) + I(X;Y_1|U_2) + I(X;Y_1|U_3)
\end{aligned} \tag{C.4}
$$

which is redundant; for (a) we used the result of (C.2). After removing all the redundant inequalities from (C.1), we can obtain the rate region for the 3 receiver BC with 2 DMS [91, Proposition 5] over the p.d.f. $p(u_1)p(u_2,u_3|u_1)p(x|u_2,u_3)$.

C.2 Reduction to 3-receiver 2 DMS Region with Y_1 Less Noisy Than Y_2

In Theorem 14, turning off all equivocation constraints and setting $U_2 = U_1 = U$ and $U_3 = V$, we have

$$
\begin{aligned}
R_0 &\le \min\{I(U;Y_2), I(V;Y_3)\} \\
R_0 + R_1 &\le I(X;Y_1) \\
R_0 + R_1 &\le I(U;Y_2) + I(X;Y_1|U) \\
R_0 + R_1 &\le I(V;Y_3) + I(X;Y_1|V) \\
R_0 &\le I(X;Y_1) \quad (*) \\
2R_0 &\le I(U;Y_2) + I(V;Y_3) + I(X;Y_1|V) \quad (*) \\
R_0 + R_1 &\le I(X;Y_1|U) + I(X;Y_1|V) + I(X;Y_1) \quad (*) \\
R_0 + R_1 &\le I(U;Y_2) + 2I(X;Y_1|U) \quad (*) \\
R_0 + R_1 &\le I(U;Y_2) + 2I(X;Y_1|U) + I(X;Y_1|V) \quad (*) \\
R_0 + R_1 &\le I(V;Y_3) + I(X;Y_1|U) + I(X;Y_1|V) \quad (*) \\
R_0 + R_1 &\le I(V;Y_3) + I(X;Y_1|U) + 2I(X;Y_1|V) \quad (*) \\
2R_0 + R_1 &\le I(U;Y_2) + I(V;Y_3) + I(X;Y_1|V) \quad (*) \\
2R_0 + 2R_1 &\le I(U;Y_2) + I(V;Y_3) + I(X;Y_1|U) + I(X;Y_1|V) \quad (*) \\
2R_0 + 2R_1 &\le I(U;Y_2) + I(V;Y_3) + 2I(X;Y_1|U) + 3I(X;Y_1|V) \quad (*) \\
2R_0 + 2R_1 &\le I(U;Y_2) + I(V;Y_3) + 2I(X;Y_1|U) + I(X;Y_1|V). \quad (*)
\end{aligned} \tag{C.5}
$$

The starred inequalities are redundant after applying the condition that Y_1 less noisy than Y_2 so that we have $I(U;Y_2) \le I(U;Y_1)$. Then we can easily see that we can get the region of [91, Proposition 7], using $I(X;Y_1) = I(U;Y_1) + I(X;Y_1|U)$, since $U \to X \to Y_1$ forms a Markov chain.

C.3 Reduction to Region of BC with One Common and One Confidential Message

We start off by setting $Y_1 = Y_2 = Y$, $Y_3 = Z$, $U_2 = U_1 = U$ and $U_3 = T$ in the region of Theorem 14, to obtain the following region over the p.d.f. $p(u)p(t|u)p(x|t)$:

$$R_0 \leq I(U;Y)$$

$$R_0 \leq I(T;Z) - I(T;Z|U) = I(U;Z)$$

$$2R_0 \leq I(U;Y) + I(T;Z)$$

$$R_0 + R_1 \leq I(X;Y)$$

$$R_0 + R_1 \leq I(U;Y) + I(X;Y|U) = I(X;Y) \quad (*)$$

$$R_0 + R_1 \leq I(T;Z) + I(X;Y|T)$$

$$R_0 + R_1 \leq I(U;Y) + I(X;Y|U) \quad (*)$$

$$R_0 + R_1 \leq I(T;Z) + I(X;Y|U) - I(T;Z|U) = I(U;Z) + I(X;Y|U)$$

$$R_{1e} \leq R_1$$

$$R_{1e} \leq I(X;Y|U) - I(X;Z|T) - I(T;Z|U) = I(X;Y|U) - I(X;Z|U)$$

$$R_{1e} \leq I(X;Y|U) + I(X;Y|T) - 2I(X;Y|T) - I(T;Z|U)$$

$$= I(X;Y|U) + I(X;Y|T) - I(X;Y|U) - I(X;Z|T)$$

$$R_0 + R_{1e} \leq I(X;Y) - I(X;Z|T) - I(T;Z|U)$$

$$R_0 + R_{1e} \leq I(U;Y) + I(X;Y|U) - I(X;Z|T) - I(T;Z|U)$$

$$R_0 + R_{1e} \leq I(T;Z) + I(X;Y|T) - I(X;Z|T) - I(T;Z|U)$$

$$2R_0 + R_{1e} \leq I(U;Y) + I(T;Z) + I(X;Y|T) - I(X;Z|T) - I(T;Z|U)$$

$$R_0 + R_1 + R_{1e} \leq I(X;Y|U) + I(X;Y|T) + I(X;Y) - I(X;Y|T) - I(T;Z|U)$$

$$R_0 + R_1 + R_{1e} \leq I(U;Y) + 2I(X;Y|U) - I(X;Z|T) - I(T;Z|U)$$

$$R_0 + R_1 + R_{1e} \leq I(U;Y) + 2I(X;Y|U) + I(X;Y|T) - I(X;Z|T) - I(T;Z|U)$$

$$R_0 + R_1 + R_{1e} \leq I(T;Z) + I(X;Y|U) + I(X;Y|T) - I(X;Z|T) - I(T;Z|U)$$

$$R_0 + R_1 + R_{1e} \leq I(T;Z) + I(X;Y|U) + 2I(X;Y_1|T) - I(X;Z|T) - I(T;Z|U)$$

$$2R_0 + R_1 \leq I(U;Y) + I(T;Z) + I(X;Y|T)$$

$$2R_0 + 2R_1 \leq I(U;Y) + I(T;Z) + I(X;Y|U) + I(X;Y|T)$$

$$2R_0 + 2R_1 + R_{1e} \leq I(U;Y) + I(T;Z) + 2I(X;Y|U) + 3I(X;Y|T)$$

$$
\begin{aligned}
&\quad - I(X;Z|T) - I(T;Z|U) \\
2R_0 + 2R_1 + R_{1e} &\leq I(U;Y) + I(T;Z) + 2I(X;Y|U) + I(X;Y|T) \\
&\quad - I(X;Z|T) - I(T;Z|U) \qquad \text{(C.6)}
\end{aligned}
$$

where we made use of the fact that, for generic r.v.s U, V and Y, we have $I(V;Y) = I(V;Y|U) + I(U;Y)$ if the Markov chain $U \rightarrow V \rightarrow Y$ is satisfied. The starred inequalities are redundant. Now, for the region in (C.6), set $T = X$ to obtain the following region, which is over the p.d.f. $p(u)p(x|u)$

$$
\begin{aligned}
&R_0 \leq I(U;Y) \\
&R_0 \leq I(U;Z) \\
&2R_0 \leq I(U;Y) + I(X;Z) \quad (*) \\
&R_0 + R_1 \leq I(X;Y) \\
&R_0 + R_1 \leq I(X;Z) = I(U;Z) + I(X;Z|U) \\
&R_0 + R_1 \leq I(U;Z) + I(X;Y|U) \\
&R_{1e} \leq R_1 \\
&R_{1e} \leq I(X;Y|U) - I(X;Z|U) \\
&R_0 + R_{1e} \leq I(X;Y) - I(X;Z|U) = I(U;Y) + I(X;Y|U) - I(X;Z|U) \quad (*) \\
&R_0 + R_{1e} \leq I(U;Y) + I(X;Y|U) - I(X;Z|U) \quad (*) \\
&R_0 + R_{1e} \leq I(X;Z) - I(X;Z|U) = I(U;Z) + I(X;Z|U) - I(X;Z|U) \\
&2R_0 + R_{1e} \leq I(U;Y) + I(X;Z) - I(X;Z|U) \quad (*) \\
&R_0 + R_1 + R_{1e} \leq I(X;Y|U) + I(X;Y) - I(X;Z|U) \quad (*) \\
&R_0 + R_1 + R_{1e} \leq I(U;Y) + 2I(X;Y|U) - I(X;Z|U) \quad (*) \\
&R_0 + R_1 + R_{1e} \leq I(X;Z) + I(X;Y|U) - I(X;Z|U) \quad (*) \\
&R_0 + R_1 + R_{1e} \leq I(X;Z) + I(X;Y|U) - I(X;Z|U) \quad (*) \\
&2R_0 + R_1 \leq I(U;Y) + I(X;Z) \quad (*) \\
&2R_0 + 2R_1 \leq I(U;Y) + I(X;Z) + I(X;Y|U) \quad (*) \\
&2R_0 + 2R_1 + R_{1e} \leq I(U;Y) + I(X;Z) + 2I(X;Y|U) - I(X;Z|U). \quad (*) \qquad \text{(C.7)}
\end{aligned}
$$

After removing the redundant starred inequalities, we have, over the p.d.f. $p(u)p(x|u)$,

$$R_0 \leq I(U;Y)$$

$$
\begin{aligned}
&R_0 \leq I(U;Z) \\
&R_0 + R_1 \leq I(X;Y) = I(U;Y) + I(X;Y|U) \\
&R_0 + R_1 \leq I(U;Z) + I(X;Z|U) \leq I(U;Z) + I(X;Y|U) \quad (*) \\
&R_0 + R_1 \leq I(U;Z) + I(X;Y|U) \\
&R_{1e} \leq R_1 \\
&R_{1e} \leq I(X;Y|U) - I(X;Z|U) \\
&R_0 + R_{1e} \leq I(U;Z) + I(X;Z|U) - I(X;Z|U) \\
&\qquad \leq I(U;Z) + I(X;Y|U) - I(X;Z|U) \quad (*).
\end{aligned} \tag{C.8}
$$

The starred inequalities are redundant provided that $I(X;Y|U) \geq I(X;Z|U)$; this means that we can obtain the following region, over the p.d.f. $p(u)p(x|u)$,

$$
\begin{aligned}
&R_0 \leq I(U;Y) \\
&R_0 \leq I(U;Z) \\
&R_0 + R_1 \leq I(U;Y) + I(X;Y|U) \\
&R_0 + R_1 \leq I(U;Z) + I(X;Y|U) \\
&R_{1e} \leq R_1 \\
&R_{1e} \leq I(X;Y|U) - I(X;Z|U).
\end{aligned} \tag{C.9}
$$

This region is the same as the one in [30, Lemmas 2,3] for the BC with one common and one confidential message. Now prefix a DMC with transition probability $p(x|v)$ to the channels $p(y|x)$ and $p(z|x)$ (that is, prefix V to $X \rightarrow (Y,Z)$), as in [30, Lemma 4] resulting in channels with transition probabilities $p(x|v)p(y|x)$ and $p(x|v)p(z|x)$ and the rate-equivocation region over the p.d.f. $p(u)p(v|u)p(x|v)p(y,z|x)$,

$$
\begin{aligned}
&R_0 \leq I(U;Y) \\
&R_0 \leq I(U;Z) \\
&R_0 + R_1 \leq I(U;Y) + I(V;Y|U) \\
&R_0 + R_1 \leq I(U;Z) + I(V;Y|U) \\
&R_{1e} \leq R_1 \\
&R_{1e} \leq I(V;Y|U) - I(V;Z|U),
\end{aligned} \tag{C.10}
$$

under the condition that $I(V;Y|U) \geq I(V;Z|U)$. The region in (C.10) is the rate-equivocation region for the BC with one common and one confidential message of Csiszár-Körner [30, Lemma 4] given in Theorem 10. Thus we have shown that our inner bound in Theorem 14 can reduce to the Csiszár-Körner region for the 2-receiver BC.

Appendix D

Lattice Decoding for AWGN Channel

D.1 Lattice Decoder

Here, we show that the lattice decoder used by Erez and Zamir in [38] is the same as decoding to the nearest coset $\mathbf{c}_m + \Lambda_2$. The decoding operation as given in [38] is

$$\widehat{\mathbf{c}_m} = Q_{\mathcal{V}_1}(\mathbf{Y}')\mathrm{mod}\Lambda. \tag{D.1}$$

Let us recall that the quantizer $Q_{\mathcal{V}_1}(.)$ is given by

$$Q_{\mathcal{V}_1}(\mathbf{x}) = \arg\min_{\boldsymbol{\lambda}_1 \in \Lambda_1} \|\mathbf{x} - \boldsymbol{\lambda}_1\|. \tag{D.2}$$

By a intuitive reasoning, the decoding operation in Erez-Zamir [38] should be decoding to the coset specified by $\mathbf{c}_m$, under correct decoding. To see the form of the decoder (in terms of the norm, like the metric) and to show decoding to the coset specified by $\mathbf{c}_m$ is true, we need some information from the paper by Liu *et al* [79].

In [79, Sect. 3,5], it is stated that the decoding operation in (D.1) is equivalent to

$$\widehat{\mathbf{c}_m} = \arg\min_{m \in \{1,\ldots,M\}} \left(\min_{\boldsymbol{\lambda}_2 \in \Lambda_2} \|\mathbf{y}' - (\mathbf{c}_m + \boldsymbol{\lambda}_2)\| \right), \tag{D.3}$$

with the nesting and the nested code specified as in Erez and Zamir [38]. To show that this is true, we need the following two properties of the L_2 norm from [25, Chap. 3]:

- P1: $\|\mathbf{v} - \mathbf{w}\| = \|\mathbf{w} - \mathbf{v}\|$ for all $\mathbf{v}, \mathbf{w}$ in $\mathbb{R}^n$.
- P2: For any $\mathbf{v}, \mathbf{w}$ in $\mathbb{R}^n$, then $|\|\mathbf{v}\| - \|\mathbf{w}\|| \leq \|\mathbf{v} - \mathbf{w}\|$.

Now, from (D.1) we have

$$\begin{aligned}
\widehat{\mathbf{c}_m} &= Q_{\mathcal{V}_1}(\mathbf{y}')\mathrm{mod}\Lambda_2 \\
&= Q_{\mathcal{V}_1}(\mathbf{y}') - Q_{\mathcal{V}_2}\left(Q_{\mathcal{V}_1}(\mathbf{y}')\right) \\
&= \arg\min_{\mathbf{c}_m\in\Lambda_1} \|\mathbf{y}' - \mathbf{c}_m\| - \arg\min_{\boldsymbol{\lambda}_2\in\Lambda_2} \left\| \arg\min_{\mathbf{c}_m\in\Lambda_1} \|\mathbf{y}' - \mathbf{c}_m\| - \boldsymbol{\lambda}_2 \right\| \\
&\overset{(a)}{=} \arg\min_{\mathbf{c}_m\in\Lambda_1} \|\mathbf{y}' - \mathbf{c}_m\| - \arg\min_{\boldsymbol{\lambda}_2\in\Lambda_2} \left\| \boldsymbol{\lambda}_2 - \arg\min_{\mathbf{c}_m\in\Lambda_1} \|\mathbf{y}' - \mathbf{c}_m\| \right\| \\
&\overset{(b)}{=} \arg\min_{\mathbf{c}_m\in\Lambda_1} \|\mathbf{y}' - \mathbf{c}_m\| - \arg\min_{\boldsymbol{\lambda}_2\in\Lambda_2} \|\boldsymbol{\lambda}_2\| \\
&= \arg\min_{\mathbf{c}_m\in\Lambda_1} \|\mathbf{y}' - \mathbf{c}_m\| - \arg\min_{\boldsymbol{\lambda}_2\in\Lambda_2} \|\mathbf{y}' - (\mathbf{y}' - \boldsymbol{\lambda}_2)\| \\
&\overset{(c)}{\leq} \arg\min_{\mathbf{c}_m\in\Lambda_1} \arg\min_{\boldsymbol{\lambda}_2\in\Lambda_2} \|\mathbf{c}_m - (\mathbf{y}' - \boldsymbol{\lambda}_2)\| = \arg\min_{\mathbf{c}_m\in\Lambda_1} \arg\min_{\boldsymbol{\lambda}_2\in\Lambda_2} \|\mathbf{c}_m + \boldsymbol{\lambda}_2 - \mathbf{y}'\| \\
&= \arg\min_{\mathbf{c}_m\in\Lambda_1} \arg\min_{\boldsymbol{\lambda}_2\in\Lambda_2} \|\mathbf{y}' - \mathbf{c}_m - \boldsymbol{\lambda}_2\|, \qquad \text{(D.4)}
\end{aligned}$$

where (a) is due to the property P1 of the L_2 norm given above; (b) is due to the following argument, similar to argument in Conway and Sloane [24, Sect. 2.1, p 41-42]: Define $\mathbf{d} \triangleq \arg\min_{\mathbf{c}_m\in\Lambda_1} \|\mathbf{y}' - \mathbf{c}_m\|$. Now $\mathbf{d} \in \Lambda_1$ implies that $\mathbf{d} \in \Lambda_2$ since $\Lambda_2 \subset \Lambda_1$, giving rise to $\min\{\|\boldsymbol{\lambda}_2 - \mathbf{d}\| : \boldsymbol{\lambda}_2, \mathbf{d} \in \Lambda_2, \boldsymbol{\lambda}_2 \neq \mathbf{d}\} = \min\{\|\boldsymbol{\lambda}_2\| : \boldsymbol{\lambda}_2 \in \Lambda_2, \boldsymbol{\lambda}_2 \neq \mathbf{0}\}$. Lastly (c) is due to property P2 of the L_2 norm given above. Since we only need $\mathbf{c}_m \in \Lambda_1$ and $\boldsymbol{\lambda}_2 \in \Lambda_2$ that minimise the objective, the inequality above is not important to the outcome of finding the $\mathbf{c}_m$ and $\boldsymbol{\lambda}_2$, and we have

$$\widehat{\mathbf{c}_m} = \arg\min_{\mathbf{c}_m\in\Lambda_1} \arg\min_{\boldsymbol{\lambda}_2\in\Lambda_2} \left(\|\mathbf{y}' - \mathbf{c}_m - \boldsymbol{\lambda}_2\|\right), \qquad \text{(D.5)}$$

from which we can obtain (D.3). Thus we can show that the Erez-Zamir lattice decoder is the same as the one specified by the more intuitive form of (D.3).

D.2 Error Probability for the Lattice Decoder

Here we give more details on the error probability for the Erez-Zamir decoder for the AWGN channel. To begin, we make the distinction between ML decoding and lattice decoding. In ML decoding, we attempt to find the lattice point inside the sphere closest to the received signal. In doing so the decision region is not the fundamental region of the lattice and so the lattice structure and symmetry is lost. In lattice decoding, the structure and symmetry of the underlying lattice is exploited in decoding.

In summary, the mechanics of obtaining an upper bound on P_e is as follows. The P_e with (aliased) noise induced by the MLAN channel $\mathbf{N}'$ is bounded in terms of the

P_e with unaliased noise $\mathbf{N}'' = [(1-\alpha)\mathbf{U} + \alpha\mathbf{N}]$. This is in turn bounded in terms of P_e with a Gaussian noise $\mathbf{Z}^*$. Then, by truncating $\mathbf{Z}^*$ to Voronoi region $\mathcal{V}_2$ to get $\mathbf{Z}_{\mathcal{V}_2}$, the error probability is finally bound in terms of P_e with $\mathbf{Z}_{\mathcal{V}_2}$. Since Euclidean lattice decoding has the same performance as ML decoding in the Voronoi region (see the discussion in [38, Sect. VIII]), we can use Gallager's error exponent method (see Sect. 3.1.2) as in Poltyrev [101] to show that $P_e \to 0$ for rates approaching the normal AWGN channel capacity. We note that the work of Poltyrev [101] considered a lattice without a bounding region and ML decoding.

Now, the error probability for the Erez-Zamir lattice decoder for a given transmitted $\mathbf{c}_m$ is given by

$$P_e = \Pr[\mathbf{N}' \notin \mathcal{V}_1(\mathbf{c}_m)], \tag{D.6}$$

where $\mathcal{V}_1(\mathbf{c}_m)$ is the Voronoi region associated with $\mathbf{c}_m$. Now $\mathbf{N}' = \mathbf{N}'' \mathrm{mod} \Lambda_2$. We want the error probability of the unaliased noise $\mathbf{N}''$. Following [46, Sect. III] and [79, Sect. 3B], we have (by the modulo-Λ_2 operation)

$$\{\mathbf{N}' \in \mathcal{V}_1(\mathbf{c}_m)\} = \bigcup_{\boldsymbol{\lambda}_2 \in \Lambda_2} \{\mathbf{N}'' \in \mathcal{V}_1(\mathbf{c}_m + \boldsymbol{\lambda}_2)\}. \tag{D.7}$$

Conversely, we have

$$\{\mathbf{N}' \notin \mathcal{V}_1(\mathbf{c}_m)\} = \bigcup_{\boldsymbol{\lambda}_2 \in \Lambda_2} \{\mathbf{N}'' \notin \mathcal{V}_1(\mathbf{c}_m + \boldsymbol{\lambda}_2)\}. \tag{D.8}$$

Therefore, the probability of error for the Erez-Zamir decoder may be expressed in terms of the unaliased noise $\mathbf{N}''$ as

$$P_e = \Pr\left[\bigcup_{\boldsymbol{\lambda}_2 \in \Lambda_2} \{\mathbf{N}'' \notin \mathcal{V}_1(\mathbf{c}_m + \boldsymbol{\lambda}_2)\}\right]. \tag{D.9}$$

It is then the union of events that the unaliased noise does not fall into the regions, each defined by the Voronoi region associated with $\mathbf{c}_m$, translated by an element of the coarse lattice Λ_2. We see that the effect of the operation $\mathbf{N}' = \mathbf{N}'' \mathrm{mod} \Lambda_2$, where $\Lambda_2 \subset \Lambda_1$, is to create an infinite constellation (IC) with points spaced $|\Lambda_1/\Lambda_2|$ times further apart than in Λ_1, where the points $\mathbf{c}_m$ with decision regions $\mathcal{V}_1(\mathbf{c}_m)$ lie.

Since all Voronoi regions $\mathcal{V}_1$ are congruent, and the decision region now extends to the IC with decision regions $\mathcal{V}_1(\mathbf{c}_m + \boldsymbol{\lambda}_2)$ for all $\boldsymbol{\lambda}_2 \in \Lambda_2$, we may find upper and lower bounds to P_e by assuming that the $\mathbf{c}_m$ is taken from an IC. According to Poltyrev

[101, Sect. IV], for a lattice (which is a linear IC), by the congruency of the Voronoi cells of all lattice points, the conditional probability of error for a given lattice point does not depend on the lattice point and coincides with the average probability of error over the lattice; and the distance distribution also does not depend on the lattice point. So, for a lattice, it is sufficient to calculate the error probability associated with $\mathcal{V}_1(\mathbf{0})$. Thus we may bound P_e for the IC "induced" by the operation $\mathbf{N}' = \mathbf{N}'' \mathrm{mod} \Lambda_2$ by the error probability associated with $\mathcal{V}_1(\mathbf{0}) = \mathcal{V}_1$.

D.2.1 Upper bound to P_e

The upper bound

$$P_e \leq \Pr\left[\mathbf{N}'' \notin \mathcal{V}_1(\mathbf{c}_m)\right] = \Pr\left[\mathbf{N}'' \notin \mathcal{V}_1\right]. \tag{D.10}$$

can be found by using the argument of Poltyrev above, or as in [79, Sect. 3B, proof of Lemma 2]. In fact the proof of [79, Lemma 2] is valid for a probability of a union of events and validates our assumption that the error probability should take the form (D.9).

To use the results of Poltyrev [101] (which deals with ML decoding in AWGN), the distribution of the noise $\mathbf{N}''$ has to be upper bounded by the distribution of a Gaussian noise $\mathbf{Z}^* \sim \mathcal{N}(0, P_{Z^*}.\mathbf{I}_n)$, so that

$$f_{\mathbf{N}''}(\mathbf{x}) \leq e^{n.\epsilon_1(\Lambda_2)} f_{\mathbf{Z}^*}(\mathbf{x}), \quad \mathbf{x} \in \mathcal{V}_2, \tag{D.11}$$

where P_{Z^*} is defined in [38, Eqn. (81)], but approaches $\frac{P_X P_N}{P_X + P_N}$ as n gets large; $\epsilon_1(\Lambda_2)$ is a function of parameters of the lattice Λ_2 defined in [38, Eqn. (67)] and is independent of $\mathbf{x}$ and is small as n gets large. The probability $\Pr\left[\mathbf{N}'' \notin \mathcal{V}_1\right]$ is then upper bounded by

$$\Pr\left[\mathbf{N}'' \notin \mathcal{V}_1\right] \leq e^{n.\epsilon_1(\Lambda_2)} \Pr\left[\mathbf{Z}^* \notin \mathcal{V}_1\right]. \tag{D.12}$$

Next, following the argument in [38, Sect. VIII], the probability $\Pr\left[\mathbf{Z}^* \notin \mathcal{V}_1\right]$ is bound in terms of probabilities that allow for the use of the ML decoding, truncating $\mathbf{Z}^*$ to $\mathcal{V}_2$. We have

$$\begin{aligned} \Pr\left[\mathbf{Z}^* \notin \mathcal{V}_1\right] &= \Pr\left[\mathbf{Z}^* \notin \mathcal{V}_2\right] + \Pr\left[\mathbf{Z}_{\mathcal{V}_2} \notin \mathcal{V}_1\right] - \Pr\left[\mathbf{Z}^* \notin \mathcal{V}_2\right].\Pr\left[\mathbf{Z}_{\mathcal{V}_2} \notin \mathcal{V}_1\right] \\ &\leq \Pr\left[\mathbf{Z}^* \notin \mathcal{V}_2\right] + \Pr\left[\mathbf{Z}_{\mathcal{V}_2} \notin \mathcal{V}_1\right], \end{aligned} \tag{D.13}$$

where $\mathbf{Z}_{\mathcal{V}_2}$ is $\mathbf{Z}^*$ truncated to $\mathcal{V}_2$, with p.d.f.

$$f_{\mathbf{Z}_{\mathcal{V}_2}}(\mathbf{x}) = \begin{cases} \frac{1}{1-\Pr[\mathbf{Z}^* \notin \mathcal{V}_2]} f_{\mathbf{Z}^*}(\mathbf{x}) & \mathbf{x} \in \mathcal{V}_2 \\ 0 & \text{otherwise} \end{cases} \tag{D.14}$$

In the last line of (D.13), the first term on the RHS is bound by viewing Λ_2 as a channel code with noise $\mathbf{Z}^*$, so the probability can be bound using (6.14) as if the channel inputs were points from the unconstrained lattice Λ_2. To bound the second term on the RHS, we first recognize that we can express this error probability as the error probability in an MLAN channel (with generic output $\mathbf{Y}$)

$$\mathbf{Y} = [\mathbf{X} + \mathbf{Z}_{\mathcal{V}_2}] \text{mod} \Lambda_2 \tag{D.15}$$

That is, an MLAN channel with input defined on $\mathcal{V}_2$ and the noise truncated to $\mathcal{V}_2$. The error probability of the MLAN channel (D.15) is dictated by its error exponent, assuming a uniform input over $\mathcal{V}_2$. The error exponent of the MLAN channel (D.15) can be found from Gallager's error exponent using the steps in [37], and then related to the lattice Λ_2 using steps in [38, Appendices B,C]. Now since both terms on the RHS of the inequality (D.13) above can be bounded, the evaluation of P_e can now proceed as in [38, Sect. VIII].

At this point we defer further discussion of the finer details of the bounding of the P_e until we analyze our wiretap coding scheme.

Bibliography

[1] R. Ahlswede and I. Csiszár, 'Common randomness in information theory and cryptography- part I. Secret sharing', *IEEE Trans. Inf. Theory,* vol. 39, no. 7, p 1121-1132, July 1993.

[2] V. Aggarwal, L. Shankar, A.R. Calderbank and H.V. Poor, 'Secrecy capacity of a class of orthogonal relay eavesdropper channels', *EURASIP J. Wireless Commun. and Networking, special issue on physical layer security* doi:10.1155/2009/494696, June 2009.

[3] S. Agrawal and S. Vishwanath, 'On the secrecy rate of interference networks using structured codes', *IEEE Symp. Inf. Theory 2009 (ISIT 2009)*, Seoul, Korea, 28 June–3 July, 2009.

[4] R. Ashwell, 'The wireless gateways to cybercrime', *The Guardian,* 22 May, 2008.

[5] G. T. Amariucai and S. Wei, 'Active eavesdropping in fast fading channels: A block-markov Wyner secrecy encoding scheme', *IEEE Symp. Inf. Theory 2010 (ISIT 2010)*, Austin, Texas, 13–18 June 2010.

[6] G. Bagherikaram, A. S. Motahari, and A. K. Khandani, 'Secrecy rate region of the broadcast channel', *submitted to IEEE Trans. Inf. Theory,* July 2008.

[7] J.-C. Belfiore and F. Oggier, 'Secrecy gain: a wiretap lattice code design', *Proc. Int. Symp. Inf. Theory and its Applications 2010*, Taiwan, Oct. 2010.

[8] J.-C. Belfiore and P. Solé, 'Unimodular lattices for the Gaussian wiretap channel', *IEEE Inf. Theory Wkshp 2010 (ITW 2010)*, Dublin, 30 Aug.–3 Sept. 2010.

[9] M. Bengtsson and B. Ottersten, 'Optimal and suboptimal transmit beamforming,' in *Handbook of Antennas in Wireless Commun.*, L. C. Godara, Ed., CRC Press, Boca Raton, USA, Aug. 2001.

[10] T. Berger, 'Multidimensional source coding', in *The Information Theory approach to communications,* G. Longo (ed.), Springer-Verlag, Berlin, pp. 171–231, 1978.

[11] P.P. Bergmans, 'Random coding theorem for broadcast channels with degraded components', *IEEE Trans. Inf. Theory,* vol. 19, no. 2, p 197-207, March 1973.

[12] M. Bloch, J. Barros, M.R.D. Rodrigues and S.W. McLaughlin, 'Wireless information-theoretic security', *IEEE Trans. Inf. Theory,* vol. 54, no. 6, p 2515-2534, June 2008.

[13] S. Borade, L. Zheng and M. Trott, 'Multilevel broadcast networks', *IEEE Symp. Inf. Theory 2007 (ISIT 2007)*, Nice, France, pp. 1151–1155, 2007.

[14] S. Boyd and L. Vandenberghe, 'Convex Optimization', *Cambridge Univ. Press*, 2004.

[15] R. Bustin, R. Liu, H.V. Poor and S. Shamai, 'An MMSE approach to the secrecy capacity of the MIMO Gaussian wiretap channel', *EURASIP J. Wireless Commun. and Networking, special issue on physical layer security,* doi:10.1155/2009/370970, June 2009.

[16] Y. Cassuto and Z. Bandic, 'Low-Complexity Wire-Tap Codes with Security and Error-Correction Guarantees', *IEEE Inf. Theory Wkshp 2010 (ITW 2010)*, Dublin, 30th Aug.–3 Sept. 2010.

[17] Y.-K. Chia and A. El Gamal, '3-receiver broadcast channels with common and confidential messages', *IEEE Symp. Inf. Theory (ISIT 2009),* Seoul, Korea, 28 June–3 July, 2009.

[18] Y.-K. Chia and A. El Gamal, '3-receiver broadcast channels with common and confidential messages' *submitted to IEEE Trans. Inf. Theory,* Oct. 2009. [Online] Available: http://arxiv.org/abs/0910.1407v4

[19] K. Cho, D. Yoon, 'On the general BER expression of one- and two-dimensional amplitude modulations', *IEEE Trans. Commun.,* vol. 50, no. 7, p 1074-1080, July 2002.

[20] L. C. Choo and K. K. Wong, 'The K-receiver broadcast channel with confidential messages', *submitted to IEEE Trans. Inf. Theory,* Dec. 2008.

[21] L. C. Choo and K. K. Wong, 'Three-receiver broadcast channel with confidential messages', *10th Int. Symp. Commun. Theory and Applications,* Ambleside, UK, 13–17 July 2009.

[22] L. C. Choo and K. K. Wong, 'Physical layer security for a 3-receiver broadcast channel with degraded message sets', *Int. Conf. Wireless Commun. and Signal Process. 2009 (WCSP 2009),* Nov. 13–15, Nanjing, China, 2009.

[23] L.C. Choo, C. Ling and K.K. Wong, 'Achievable Rates for Lattice Coded Gaussian Wiretap Channels', *IEEE Int. Conf. Commun. (ICC 2011),* Kyoto, Japan, 5–9 June, 2011.

[24] J.H. Conway and N.H.A. Sloane, 'Sphere packings, lattices and groups', *Springer,* 1988.

[25] L.J. Corwin and R.H. Szczarba, 'Calculus in vector spaces', *Marcel-Dekker,* 1994.

[26] T. Cover and J. Thomas, 'Elements of information theory', *Wiley,* 2006.

[27] T.M. Cover and A.A. El Gamal, 'Capacity theorems for the relay channel', *IEEE Trans. Inf. Theory,* vol. 25, No. 5, p572-584, Sept. 1979.

[28] T.M. Cover, 'Comments on broadcast channels', *IEEE Trans. Inf. Theory,* vol. 44, No. 6, p 2524-2530, Oct. 1998.

[29] I. Csiszár and J. Körner, 'Information theory: coding theorems for discrete memoryless systems', *Academic Press,* 1981.

[30] I. Csiszár and J. Körner, 'Broadcast channels with confidential messages', *IEEE Trans. Inf. Theory,* vol. 24, no. 3, p 339-348, May 1978.

[31] L. Dong, Z. Han, A. Petropulu and H. V. Poor, 'Improving wireless physical layer security via cooperating relays,' *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1875–1888, March 2010.

[32] L. Dong, H. Yousefi'zadeh and J. Jarakhani, 'Cooperative jamming and power allocation for wireless relay networks in presence of eavesdropper', *IEEE Int. Conf. Commun. (ICC 2011),* Kyoto, Japan, 5–9 June, 2011.

[33] E. Ekrem and S. Ulukus, 'Secrecy capacity of a class of broadcast channels with an eavesdropper', *EURASIP J. Wireless Commun. and Networking, special issue on wireless physical layer security,* June 2009.

[34] E. Ekrem and S. Ulukus, 'Secrecy in cooperative relay broadcast channels', *submitted to IEEE Trans. Inf. Theory,* Nov. 2008.

[35] A. El Gamal and Y.H. Kim, 'Lecture notes on network Information Theory,' June 2010. [Online] Available: http://arxiv.org/abs/1001.3404v4

[36] A.A. El Gamal and E. van der Meulen, 'A proof of Marton's coding theorem for the discrete memoryless broadcast channel', *IEEE Trans. Inf. Theory,* vol. 27, no. 1, p 120-122, Jan. 1981.

[37] U. Erez and R. Zamir, 'Error exponents of modulo additive noise channels with side information at the transmitter', *IEEE Trans. Inf. Theory,* vol. 47, no. 1, pp. 210-218, Jan. 2001.

[38] U. Erez and R. Zamir, 'Achieving $\frac{1}{2}\log(1+\mathrm{SNR})$ on the AWGN channel with lattice encoding and decoding,' *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2293–2314, Oct. 2004.

[39] U. Erez, R. Zamir and S. Litsyn, 'Lattices which are good for (almost) everything,' *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3401–3416, Oct. 2005.

[40] U. Erez and S. ten Brink, 'A close-to-capacity dirty paper coding scheme,' *IEEE Trans. Inf. Theory*, vol. 51, no. 10, pp. 3417–3432, Oct. 2005.

[41] S.A. Fakoorian and A.L. Swindelhurst, 'Solutions for the MIMO Gaussian wiretap channel with a cooperative jammer', *IEEE Trans. Signal Process.,* vol. 59, no. 10, pp. 5013-5022, Oct. 2011.

[42] G. Forney, 'Coset codes- Part I: Introduction and geometrical classification', *IEEE Trans. Inf. Theory,* vol. 34, no. 5, pp. 1123-1151, Sept. 1988.

[43] G. Forney, 'Coset codes- Part II: Binary lattices and related codes', *IEEE Trans. Inf. Theory,* vol. 34, no. 5, pp. 1152-1187, Sept. 1988.

[44] G. Forney and L.F. Wei, 'Multidimensional constellations- part I. Introduction, figures of merit, and generalized cross-constellations', *IEEE J. Sel. Areas Commun.,* vol. 7, no. 7, pp. 877-892, Aug. 1989.

[45] G. Forney and L.F. Wei, 'Multidimensional constellations- part II. Voronoi constellations', *IEEE J. Sel. Areas Commun.,* vol. 7, no. 7, pp. 941-958, Aug. 1989.

[46] G.D. Forney, M.D. Trott and S.-Y. Chung, 'Sphere-bound-achieving coset codes and multilevel coset codes', *IEEE Trans. Inf. Theory,* vol. 46, no. 3, pp. 829-850, May 2000.

[47] G.D. Forney, 'On the role of MMSE estimation in approaching the information-theoretic limits of Gaussian channels: Shannon meets Wiener', *Proc. 41st Annual Allerton Conf. on Commun., Control and Computing,* Monticello, Ill., USA, Sept. 2003.

[48] F. Gabry, R. Thobaden and M. Skoglund, 'Outage performances for Amplify-and-Forward, Decode-and-Forward and Cooperative Jamming strategies for the wiretap channel', *IEEE Wireless Commun. and Networking Conf. (WCNC 2011),* 2011.

[49] R.G. Gallager, 'Information Theory and reliable communication', *Wiley,* 1968.

[50] R.G. Gallager, 'Capacity and coding for degraded broadcast channels', *Problemy Peredachi Informatsii,* vol. 10, no. 3, p 3-14, July-Sept. 1974.

[51] S.I. Gel'fand and M.S. Pinsker, 'Coding for channel with random parameters', *Problemy Peredachi Informatsii,* vol. 9, no. 1, p 19-31, 1980.

[52] S. Goel and R. Negi, 'Guaranteeing secrecy using artificial noise', *IEEE Trans. Wireless Commun.,* vol. 7, no. 6, p.2180, June 2008.

[53] P.K. Gopala, L. Lai and H. El Gamal, 'On the secrecy capacity of fading channels', *IEEE Trans. Inf. Theory,* vol. 54, no. 10, p 4687-4698, Oct. 2008.

[54] D. Gunduz, E. Tuncel and J. Nayak, 'Rate regions for the separated two-way relay channel', *Proc. 46th Annual Allerton Conf. on Commun., Control and Computing,* Monticello, Ill., USA, Sept. 2008.

[55] G.H. Hardy, J.E. Littlewood and G. Polya, 'Inequalities,' *2nd ed., Cambridge University Press*, 1952.

[56] X. He and A. Yener, 'Providing secrecy with lattice codes', *Proc. 46th Annual Allerton Conf. on Commun., Control and Computing,* Monticello, Ill., USA, Sept. 2008.

[57] X. He and A. Yener, 'Providing secrecy with structured codes: tools and applications to two-user Gaussian channels', *submitted to IEEE Trans. Inf. Theory,* July 2009.

[58] X. He and A. Yener, 'K-user interference channels: achievable secrecy rates and degrees of freedom', *IEEE Inf. Theory Workshop (ITW 2008),* June 2008.

[59] X. He and A. Yener, 'The Gaussian many-to-one interference channel with confidential messages', *IEEE Trans. Inf. Theory,* vol. 57, no. 5, p 2730–2745, May 2011.

[60] E. Hof and S. Shamai, 'Secrecy-achieving polar coding', *IEEE Inf. Theory Wkshp 2010 (ITW 2010)*, Dublin, 30th Aug.-3 Sept. 2010.

[61] J. Huang and A.L. Swindelhurst, 'Cooperative jamming for secure communications in MIMO relay networks', *IEEE Trans. Signal Process.,* vol. 59, no. 10, pp. 4871-4884, Oct. 2011.

[62] A. Khisti and G. W. Wornell, 'The MIMOME channel', *Proc. 45th Annual Allerton Conf. on Commun., Control and Computing,* Monticello, Ill., USA, Sept. 2007.

[63] A. Khisti and G. W. Wornell, 'Secure transmission with multiple antennas I: The MISOME wiretap channel', *IEEE Trans. Inf. Theory,* vol. 56, No. 7, p3088-3104, July 2010.

[64] D. Klinc, J. Ha, S. McLaughlin, J. Barros and B.-J. Kwak, 'LDPC codes for the Gaussian wiretap channel', *IEEE Inf. Theory Wkshp 2009 (ITW 2009)*, Sicily, Oct. 2009.

[65] M. Kobayashi, M. Debbah, and S. Shamai, 'Secured communication over frequency-selective fading channels: A practical Vandermonde precoding', *EURASIP J. Wireless Commun. and Networking, special issue on wireless physical layer security,* June 2009.

[66] J. Körner and K. Marton, 'Comparison of two noisy channels', *Topics in Inf. Theory, Keszthely, Hungary, 1975, Colloquia Math. Soc. Janos Bolyai,* North-Holland, pp. 411–423, 1977.

[67] G. Kramer, 'Topics in multi-user information theory', *Foundations and Trends in Commun. and Inf. Theory,* Vol. 4, Issue 4-5, 2007.

[68] D. Krithivasan and S. Pradhan, 'A proof of the existence of good lattices', *Technical Report,* University of Michigan, July 2007. [online] available: http://www.eecs.umich.edu/techreports/systems/cspl/cspl-384.pdf

[69] L. Lai and H. El Gamal, 'The relay-eavesdropper channel: cooperation for secrecy', *IEEE Trans. Inf. Theory,* vol. 54, No. 9, p4005-4019, Sept. 2008.

[70] S.K. Leung-Yan-Cheong and M.E. Hellman, 'The Gaussian wire-tap channel', *IEEE Trans. Inf. Theory,* vol. 24, no. 4, p 451-456, July 1978.

[71] J. Li, A.P. Petropulu and S. Weber, 'On cooperative relaying schemes for wireless physical layer security,' *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4985–4997, Oct. 2011.

[72] Y. Liang, G. Kramer, H.V. Poor and S. Shamai, 'Compound wire-tap channels', *EURASIP J. Wireless Commun. and Networking, special issue on physical layer security,* 2009.

[73] Y. Liang, H.V. Poor and S. Shamai, 'Information theoretic security', *Foundations and Trends in Commun. and Inf. Theory,* Vol. 5, Issue 4-5, 2009.

[74] W.-C. Liao, T.-H. Chang, W.-K. Ma and C.-Y. Chi, 'QoS-based transmit beamforming in the presence of eavesdroppers: an optimized artificial-noise-aided approach', *IEEE Trans. Signal Process.,* vol. 59, no. 3, p 1202-1216, March 2011.

[75] S.-C. Liu, T.-H. Chang, Y.-L. Liang, Y.-W. Peter Hong and C.-Y. Chi, 'On the impact of quantized channel feedback in guaranteeing secrecy with artificial noise: the noise leakage problem', *IEEE Trans. Wireless Commun.,* vol. 10, no. 3, p 901-914, March 2011.

[76] R. Liu, Y. Liang, H.V. Poor and P. Spasojevic, 'Secure nested codes for type II wire-tap channels', *IEEE Inf. Theory Workshop 2007 (ITW 2007)*, Sept. 2007.

[77] R. Liu, T. Liu, H.V. Poor and S. Shamai, 'MIMO Gaussian broadcast channels with confidential messages', *IEEE Symp. Inf. Theory (ISIT 2009)* Seoul, Korea, June 28–July 3, 2009.

[78] R. Liu, I. Marić, P. Spasojević and R. Yates, 'Discrete memoryless interference and broadcast channels with confidential messages: Secrecy rate regions', *IEEE Trans. Inf. Theory,* vol. 54, no. 6, p 2493-2507, June 2008.

[79] T. Liu, P. Moulin and R. Koetter, 'On error exponents of modulo lattice additive noise channels', *IEEE Trans. Inf. Theory,* vol. 52, no. 2, pp. 454-471, Feb. 2006.

[80] R. Liu, H.V. Poor, P. Spasojevic and Y. Liang, "Nested codes for secure transmission", Proc. *PIMRC 2008*, Sept. 2008.

[81] T. Liu and S. Shamai, 'A note on secrecy capacity of the multi-antenna wiretap channel', *IEEE Trans. Inf. Theory,* vol. 55, no. 6, pp. 25472553, June 2009.

[82] J. Lotspiech, 'Broadcast encryption', *in Multimedia security technologies for digital rights management,* W. Zeng *et al* eds., Academic Press, 2006.

[83] H. Mahdavifar and A. Vardy, 'Achieving the secrecy capacity of wiretap channels using polar codes', *IEEE Trans. Inf. Theory,* vol. 57, no. 10, Oct. 2011.

[84] K. Marton, 'A coding theorem for the discrete memoryless broadcast channel', *IEEE Trans. Inf. Theory,* vol. 25, no. 3, p 306-311, 1979.

[85] U. Maurer, 'Secret-key agreement by public discussion based on common information', *IEEE Trans. Inf. Theory,* vol. 39, no. 7, p 733-742, July 1993.

[86] U. Maurer and S. Wolf, 'Information-theoretic key agreement: from weak to strong secrecy for free', *Proc. EUROCRYPT 2000 on Advances in Cryptology,* in Lecture Notes in Computer Science vol. 1807, pp. 352–358, 2000.

[87] A. Mazzeo, ed., *Proc. IEEE (Special issue on Cryptography and Security),* vol. 94, No. 2, Feb. 2006.

[88] A. Mukherjee and A.L. Swindelhurst, 'Robust beamforming for security in MIMO wiretap channels with imperfect CSI', *IEEE Trans. Signal Process.,* vol. 59, no. 1, p 351-360, Jan. 2011.

[89] A. Mukherjee and A.L. Swindelhurst, 'Securing Multi-Antenna Two-Way Relay Channels With Analog Network Coding Against Eavesdroppers', *Proc. IEEE Signal Process. Adv. Wireless Commun.,* Marrakech, Morocco, Jun. 2010.

[90] C. Nair and A. El Gamal, 'The capacity region of a class of 3-receiver broadcast channels with degraded message sets', *IEEE Int. Symp. Inf. Theory 2008 (ISIT 2008),* Toronto, July 6 –11, 2008.

[91] C. Nair and A. El Gamal, 'The capacity region of a class of three-receiver broadcast channels with degraded message sets', *IEEE Trans. Inf. Theory,* vol. 55, no. 10, pp. 4479–4493, Oct. 2009.

[92] C. Nair and Z.V. Wang, 'The capacity region of a the three-receiver less noisy broadcast channels', *IEEE Trans. Inf. Theory,* vol. 57, no. 7, pp. 4058–4062, July 2011.

[93] W. Nam, S. Y. Chung and Y. Lee, 'Capacity of the Gaussian Two-Way Relay Channel to Within 1/2 Bit', *IEEE Trans. Inf. Theory,* vol. 56, no. 11, pp. 5488-5494, Nov. 2010.

[94] B. Nazer and M. Gastpar, 'Compute-and-forward: harnessing interference through structured codes', *IEEE Trans. Inf. Theory,* vol. 57, no. 10, pp. 6463–6486, Oct. 2011.

[95] T. Oechtering, C. Schnurr, I. Bjelakovic and H. Boche, 'Broadcast capacity region of two-phase bidirectional relaying', *IEEE Trans. Inf. Theory,* vol. 54, No. 1, p454-458, Jan. 2008.

[96] F. Oggier and B. Hassibi, 'The secrecy capacity of the MIMO wiretap channel', *IEEE Trans. Inf. Theory,* vol. 57, no. 8, Aug. 2011.

[97] F. Oggier, P. Solé and J. C. Belfiore, 'Lattice codes for the wiretap Gaussian channel: construction and analysis', *submitted to IEEE Trans. Inf. Theory,* March 2011. [Online] http://arxiv.org/abs/1103.4086v1

[98] Y. Oohama, 'Relay channels with confidential messages', *submitted to IEEE Trans. Inf. Theory,* March 2007.

[99] L. Ozarow and A. Wyner, 'Wire-tap channel II', *Bell Syst. Tech. J.,* vol. 63, pp. 2135-2157, Dec. 1984.

[100] L. Pérez-Freire and F. Pérez-González, 'Security of Lattice-Based Data Hiding Against the Watermarked-Only Attack', *IEEE Trans. Inf. Forens. Security,* vol. 3, no. 4, pp. 593-610, Dec. 2008.

[101] G. Poltyrev, 'On coding without restrictions for the AWGN channel', *IEEE Trans. Inf. Theory,* vol. 40, no. 2, pp. 409-417, March 1994.

[102] H. Qin, X. Chen, Y. Sun, M. Zhao and J. Wang, 'Optimal power allocation for joint beamforming and artificial noise design in secure wireless communications', *IEEE Int. Conf. Commun. (ICC 2011),* Kyoto, Japan, 5–9 June, 2011.

[103] V. Rathi, M. Andersson, R. Thobaden, J. Kliwer, M. Skoglund, 'Performance analysis and design of two edge type LDPC codes for the BEC wiretap channel,' submitted to *IEEE Trans. Inf. Forens. Security*, Sept. 2010. [Online] Available: http://arxiv.org/abs/1009.4610

[104] Sadaf Salehkalaibar and Mohammad Reza Aref, 'The capacity region of a class of 3-Receiver broadcast channels with two eavesdroppers', *IEEE Symp. Inf. Theory 2011 (ISIT 2011),* St. Petersburg, Russia, 31 July–5 Aug. 2011.

[105] S. Shafiee, N. Liu and S. Ulukus, 'Towards the Secrecy Capacity of the Gaussian MIMO Wire-Tap Channel: The 2-2-1 Channel', *IEEE Trans. Inf. Theory,* vol. 55, no. 9, pp. 4033-4039, Sept. 2009.

[106] C. E. Shannon, 'Communication theory of secrecy systems', *Bell Syst. Tech. J.,* vol. 28, pp. 656715, 1949.

[107] A. Subramaniam, A. Thangaraj, M. Bloch and S. McLaughlin, 'Strong secrecy on the binary erasure wiretap channel using large-girth LDPC codes', *IEEE Trans. Inf. Forens. Security*, vol. 6, no. 3, p 585-594, Sept. 2011.

[108] A.T. Suresh, A. Subramaniam, A. Thangaraj, M. Bloch and S. McLaughlin, 'Strong secrecy for erasure wiretap channels', *IEEE Inf. Theory Wkshp 2010 (ITW 2010)*, Dublin, 30th Aug.-3 Sept. 2010.

[109] V. Tarokh, A. Vardy and K. Zeger, 'Universal bound on the performance of lattice codes', *IEEE Trans. Inf. Theory,* vol. 45, no. 2, pp. 670-681, March 1999.

[110] A. Thangaraj, S. Dihidar, A. Calderbank and S. McGaughlin, 'Application of LDPC codes to the wiretap channel', *IEEE Trans. Inf. Theory,* vol. 53, no. 8, pp. 2933-2945, Aug. 2007.

[111] J. Vilela, M. Bloch, J. Barros and S. McLaughlin, 'Wireless secrecy regions with friendly jamming', *IEEE Trans. Inf. Forens. Security,* vol. 6, no. 2, pp. 256-265, June 2011.

[112] J. Vilela, P. Pinto and J. Barros, 'Position-based jamming for enhanced wireless secrecy', *IEEE Trans. Inf. Forens. Security,* vol. 6, no. 3, pp. 616-627, Sept. 2011.

[113] A. Wiesel, Y. C. Eldar and A. Beck, 'Maximum likelihood estimation in linear models with a Gaussian model matrix,' *IEEE Signal Process. Lett.*, vol. 13, no. 5, pp. 292–295, May 2006.

[114] M.P. Wilson, K. Narayanan, H. Pfister and A. Sprintson, 'Joint physical layer coding and network coding for bidirectional relaying', *IEEE Trans. Inf. Theory,* vol. 56, no. 11, pp. 5641-5654, Nov. 2010.

[115] A.D. Wyner, 'The wire-tap channel', *Bell Syst. Tech. J.,* vol. 54, no. 8, p 1355-1387, 1975.

[116] J. Xu, Y. Cao, and B. Chen, 'Capacity bounds for broadcast channels with confidential messages', *IEEE Trans. Inf. Theory,* vol. 55, no. 10, pp. 4529–4542, Oct. 2009.

[117] R. Yeung, 'A first course in Information Theory', *Springer,* 2002.

[118] Abdellatif Zaidi and Luc Vandendorpe, 'Coding Schemes for Relay-Assisted Information Embedding', *IEEE Trans. Inf. Forens. Security,* vol. 4, no. 1, pp. 70–86, March 2009.

[119] R. Zamir and M. Feder, 'On universal quantization by randomized uniform/ lattice quantizer', *IEEE Trans. Inf. Theory,* vol. 38, pp. 428-436, March 1992.

[120] R. Zamir, S. Shamai and U. Erez, 'Nested linear/ lattice codes for structured multiterminal binning', *IEEE Trans. Inf. Theory,* vol. 48, no. 6, pp. 1250-1276, June 2002.

[121] R. Zamir, 'Lattices are everywhere', in Proc. *Inf. Theory and Applicat. 2009 (ITA 2009)*, Univ. Calif. San Diego, Jan. 2009.

[122] R. Zamir, 'How to generate a simple dither', *Proc. IEEE 26th Conv. of Elect. and Electron. Engineers in Israel*, 2010.

[123] J. Zhang and M. C. Gursoy, 'Collaborative relay beamforming for secrecy', submitted, Dec. 2009. [Online]. Available: http://arxiv.org/abs/0910.4132.

[124] J. Zhang and M. C. Gursoy, 'An Achievable Rate Region for Imperfectly-Known Two-Way Relay Fading Channels', *IEEE Int. Symp. Inf. Theory 2011 (ISIT 2011),* St. Petersburg, Russia, 28 July–5 Aug. 2011.

[125] R. Zhang, Y.C. Liang, C.C. Choy and S. Cui, 'Optimal beamforming for two-way multi-antenna relay channel with analogue network coding', *IEEE J. Sel. Areas Commun.,* vol. 27, no. 5, pp. 699-712, 2009.

[126] G. Zheng, L.C. Choo and K.K. Wong, 'Optimal cooperative jamming to enhance physical layer security using relays', *IEEE Trans. Signal Process.,* vol. 59, no. 3, pp. 1317-1322, March 2011.