Technische Universität München
Lehrstuhl für Theoretische Informationstechnik

Capacity Results for Classes of Wiretap Channels

Dipl.-Phys. Jochen Sommerfeld

Vollständiger Abdruck der von der Fakultät für Elektrotechnik und Informationstechnik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender:	Univ.-Prof. Dr. sc. techn. Gerhard Kramer
Prüfer der Dissertation:	
	1. Univ.-Prof. Dr.-Ing. Dr. rer. nat. Holger Boche
	2. Univ.-Prof. Dr. rer. nat. Michael M. Wolf

Die Dissertation wurde am 17.04.2012 bei der Technischen Universität München eingereicht und durch die Fakultät für Elektrotechnik und Informationstechnik am 22.02.2013 angenommen.

Zusammenfassung

Wir betrachten zwei Kommunikationssysteme unter dem Blickwinkel informationstheoretischer Sicherheit. Beide Systeme sind zeitdiskret, gedächtnislos und abhängig vom Kanalzustand. Der Compound Kanal besteht aus einer endlichen beziehungsweise unendlichen Menge von Kanälen, die sowohl dem Sender als auch dem Empfänger bekannt sind. Beiden ist aber nicht bekannt, welcher Kanal zur Übertragung der aktuellen Nachricht benutzt wird. Im Gegensatz zum Compound Kanal kann sich der Kanalzustand eines beliebig variierenden Kanals (Arbitrarily Varying Channel AVC) in einer beliebigen, aber den Teilnehmern unbekannten Weise, von Zeitschritt zu Zeitschritt innerhalb der Übertragung eines einzelnen Kodewortes ändern. In beiden Szenarien fordern wir Sicherheit vor dem Abhören (eavesdropping) durch eine möglicherweise existierende dritte Partei. Das entsprechende Modell für sichere Datenübertragung wird durch einen sogenannten Wiretap Kanal beschrieben. Für eine realistischere Untersuchung von praktischen drahtlosen Kommunikationssystemen betrachten wir den Wiretap Kanal bezüglich Kanalunkenntnis. Die resultierenden Modelle bezeichnen wir als Compound Wiretap Kanal und als beliebig variierenden Wiretap Kanal (Arbitrarily Varying Wiretap Channel AVWC). Für beide Systeme leiten wir Ergebnisse zur Sicherheitskapazität ab, die hinsichtlich der mittleren Fehlerwahrscheinlichkeit und des strengen Sicherheitskriteriums definiert ist.

Wir leiten eine untere Grenze an die Sicherheitskapazität des Compound Wiretap Kanals mit Kenntnis des Kanalzustandes beim Sender ab, welche mit der generellen oberen Schranke der Sicherheitskapazität eines beliebigen Compound Wiretap Kanals übereinstimmt. Somit können wir für diesen Fall einen vollständigen Kodierungssatz angeben und zwar für ein starkes Sicherheitskriterium und für einen Decoder, der unempfindlich ist gegenüber dem Effekt der Randomisierung in der Kodierung am Sender. Das enthebt uns der Notwendigkeit, den Randomisierungsparameter ebenfalls zu dekodieren, was innerhalb dieses Modells im allgemeinen nicht möglich wäre. Weiterhin leiten wir für den Fall, das der Kanalzustand nicht bekannt ist, eine untere Schranke für die Sichrheitskapazität und ebenso einen “multi-letter”-

Ausdruck für die Sicherheitskapazität ab.

Für den beliebig variierenden Wiretap Kanal AVWC leiten wir für den Fall eines "besten" Kanals zum Mithörer (Eavesdropper) eine untere Schranke an die Sicherheitskapazität für Randomcodes ab. Wir zeigen, dass die Sicherheitskapazität des AVWC für deterministische Codes identisch mit der Sicherheitskapazität für Randomcodes ist, einen nicht symmetrisierbaren Kanal zum legitimen Empfänger vorausgesetzt. Somit folgt, dass obige Schranke an die Sicherheitskapazität unter der entsprechenden Voraussetzung auch für deterministische Codes gültig ist. Die Beweismethode beruht auf der "elimination technique" von Ahlswede für "single-user" AVC's. Für den allgemeinen AVWC leiten wir weiterhin eine obere Schranke für die Sicherheitskapazität ab, woraus ein "multi-letter"-Ausdruck für die Sicherheitskapazität für den Fall eines besten Kanals zum Eavesdropper resultiert.

Abstract

We consider two communication systems which are time-discrete and memoryless, both depend on a state, in terms of information-theoretic secure data transmission. The compound channel consists of a finite or infinite set of channels which is known to both the sender and the receiver, but unfortunately it is not known which channel is in use for any codeword transmission. In contrast the state of an arbitrarily varying channel may change in an arbitrary but also unknown manner from letter to letter in the transmission of any single codeword. For both scenarios we require secrecy from eavesdropping by a possibly existing third party. The associated model of secure data transmission (or transmission of private messages) is described by a wiretap channel. For a more realistic investigation of practical wireless communication systems we consider the wiretap channel under channel uncertainty. We call the resulting models the *compound wiretap channel* and the *arbitrarily varying wiretap channel*. For both systems we derive results for the secrecy capacity, which is defined in terms of the average error probability and the strong secrecy criterion.

We derive a lower bound on the secrecy capacity of the compound wiretap channel with channel state information at the transmitter which matches the general upper bound on the secrecy capacity of general compound wiretap channels and thus establishing a full coding theorem in this case. We achieve this for the strong secrecy criterion and with a decoder that is robust against the effect of randomisation in the encoding. This relieves us from the need of decoding the randomisation parameter which is in general not possible within this model. Moreover we prove a lower bound on the secrecy capacity of the compound wiretap channel without channel state information and derive a multi-letter expression for the capacity in this communication scenario.

For the arbitrarily varying wiretap channels AVWC we derive a lower bound on the random code secrecy capacity in the case of a “best” channel to the eavesdropper. We show that, provided that the channel to the legitimate receiver is not symmetrisable, the deterministic code secrecy capacity of the AVWC equals the random code

secrecy capacity. With the same assumption we can derive, that the above lower bound is also valid for the deterministic code secrecy capacity. The proof of the identity is based on the "elimination technique" of Ahlwede for single-user AVC's. For the general AVWC we further give an upper bound on the secrecy capacity, which yields a multi-letter expression for the secrecy capacity in the case of a best channel to the eavesdropper.

Acknowledgements

First of all, I thank my advisor Professor Holger Boche for his support and motivation, and for giving me the opportunity to work on many interesting topics of information theory. I also would like to express my gratitude to Professor Michael Wolf for serving as the second referee of this dissertation.

I would like to thank Igor Bjelaković for teaching me information theory and a lot of mathematical techniques, for many suggestions and his support throughout the last years.

My thanks goes to all colleagues at the Technische Universität München and Technische Universität Berlin and especially to Peter Jung for sharing the room and for many interesting discussions.

My gratitude goes to my family and all my friends for their support and encouragement.

Contents

Chapter 1

Introduction

Apart from ensuring reliable transmission it has become necessary to guarantee secure transmission in wireless communication systems. In general the information transmission can be affected by passive and active attacks. While the intention of the former is to eavesdrop on the legitimate participants and to obtain information about the transmitted messages but without modifying it, the latter is interested in manipulating the messages or to disrupt the transmission to the legitimate recipients. So the aim of theoretical concepts in communication theory of protecting information transmission is to guarantee confidentiality or secrecy against possible unauthorised eavesdropping participants and integrity against attempts of tampering with the channels by a jammer. Conventional approaches to ensure secrecy, and therewith to protect against unwanted attempts of eavesdropping the data transmission, rely on cryptographic methods. Before the transmission of a message, it will be encrypted in a cyphertext based on a key and conversely the intended receivers have to decrypt the cyphertext after receiving the transmission. Basically it will be distinguished between symmetric key encryption, where the legitimate participants share a common secret key, and asymmetric key encryption, where the transmitter uses a public key for encryption whereas the legitimate receiver uses a private key, corresponding to the public key, for decryption of the cyphertext. The first method requires a high effort in the key management, the latter is based on a high computational complexity and are not provably perfectly secure. In wireless networks further problems arise due to the easy accessibility and its possibly decentralized realisation, which additionally complicates the handling of the key.

As a consequence of the growing presence of wireless and mobile networks in practically all areas of data transmission, in recent years more attention has been paid to information theoretical concepts. Information theory provides approaches achieving secrecy from eavesdropping and integrity from jamming without any assumptions about computational complexity and resources. Different to cryptographic approaches information theoretic models do not rely on the use of secret keys and instead make use of the noise and the fluctuations of the channel caused by the physical medium. Then the differences between the channels of legitimate and unauthorised participants can be used in the coding procedure to keep transmitted information hidden from the illegitimate participant. This basic approach has already been applied in the pioneering work of Wyner [Wyn75] and [CK78], where they have introduced the terminology of confidentiality and of the so-called *wiretap channel* in information theory. More precisely, the wiretap channel describes a communication system, which consists of a pair of channels with common input alphabet, where a confidential or private message is sent over the first channel to a legitimate receiver, and an eavesdropper, which observes the output of a second channel should be kept as ignorant as possible of the message sent.

In this thesis we extend the wiretap channel to a model where the legitimate users suffer from channel uncertainty, which gives a more realistic description of wireless systems. First we consider the *compound wiretap channels* where the channel realisation for the transmission of the whole actual codeword is not known but only that the channel realisation belongs to a given and known set of channels. If the channel realisation varies from symbol to symbol of the transmitted codeword unknown to the legitimate users, the resulting model is the *arbitrarily varying wiretap channel AVWC*. This second model has an additional informational theoretic security aspect, namely that it can be seen as a model, in which a jammer manipulates the transmission by changing the channel state in every time step in a way unknown to the legitimate parties. So the AVWC is a model which combines both passive attacks by an eavesdropper and active attacks by a jammer. In both models we use the *strong secrecy criterion* as the measure of the information theoretic secrecy to derive lower and upper bounds on the *secrecy capacity* as the supremum of achievable rates at which reliable transmission under perfect secrecy is possible. In special cases we can give explicit expressions for the secrecy capacity.

In the remaining part of the introduction we give an overview on the basic ideas of the information theoretic approach to secure data transmission in wireless communication systems and introduce the wiretap channel, the most basic information-theoretic model for achieving secrecy from eavesdropping. Further we briefly review two well accepted models for channel uncertainty, the compound channel and the arbitrarily varying channel, which will be combined in the later chapters with the wiretap channel to simulate realistic transmission of confidential messages in wireless communication systems. We will give the definitions of both of the models and recall the existing coding results.

1.1 Information Theoretic Security

In [Sha49] Shannon has introduced the notion of *perfect secrecy* in information theoretic analysis of a cryptosystem. In this system an eavesdropper had the possibility to intercept the cyphertext in which a source message is encrypted. In this work the notion of the *equivocation* was established as a measure of the eavesdropper's uncertainty about the message conditioned on its observation of the cyphertext. Different to cryptographic systems, information theoretical approaches for secure transmission make use of the different probabilistic description of the channels to the legitimate receiver and the eavesdropper caused by the transmission over a noisy medium. This together with an additional randomisation by a stochastic encoding procedure guarantees reliable transmission to the legitimate users while the eavesdropper is kept as ignorant as possible of the message sent. This basic approach was first used by Wyner in [Wyn75] where he introduced the wiretap channel. He adopted the notion of the equivocation as a measure of secrecy, which rely on the conditional entropy of a random variable, which is uniformly distributed on the message set, given the output at the eavesdropper. So he established a model in which it was possible to consider information transmission under a reliability and a *weak* secrecy criterion and to define the *secrecy capacity* as the largest achievable rate at which reliable transmission under the secrecy constraint is possible.

Finally, in contrast to cryptographic systems information theoretical approaches achieve guaranteed secure transmission without the high effort of key management and computational complexity.

1.1.1 The Wiretap Channel

The so-called wiretap channel consists of an input alphabet A and two discrete memoryless channels. $W : A \to \mathcal{P}(B)$ represents the communication link to the legitimate receiver and $V : A \to \mathcal{P}(C)$ is referred to as the channel to the eavesdropper, B and C are finite sets. To transmit a message $j \in \mathcal{J}_n$ over the channel W^n the sender make use of a stochastic encoder $E : \mathcal{J}_n \to \mathcal{P}(A^n)$ and the receiver decodes the received sequence by a collection of mutually disjoint decoding sets $\{D_j \subset B^n : j \in \mathcal{J}_n\}$. If we define the random variable Z^n as the outcome of the channel V^n and define a random variable J uniformly distributed on the message set the eavesdropper should obtain no significant information about J by observing its output. Then an (n, J_n) code $\mathcal{C}_n$ for the wiretap channel is a system $\{(E(\cdot|j), D_j) : \ j = 1, \ldots, J_n\}$, where the $E(\cdot|j)$ are probability distributions on A^n and the D_j are the mutually disjoint decoding sets. The average error probability is defined by

$$\bar{e}(\mathcal{C}_n) = \frac{1}{J_n} \sum_{j \in \mathcal{J}_n} \sum_{x^n \in A^n} E(x^n|j) W^n(D_j^c|x^n). \tag{1.1}$$

Definition 1.1. *We call a positive number R_S an achievable secrecy rate for the wiretap channel if there exists a sequence $\mathcal{C}_n$ of (n, J_n) codes such that*

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n \geq R_S,$$

and

$$\lim_{n\to\infty} \bar{e}(\mathcal{C}_n) = 0 \quad \textit{and} \quad \lim_{n\to\infty} I(J; Z^n) = 0. \tag{1.2}$$

Here J denotes the random variable uniformly distributed on the message set $\mathcal{J}_n$ and Z_n is the resulting random variable at the output of eavesdropper's channel V^n. Then the secrecy capacity C_S is defined as the largest achievable secrecy rate. For the definition of the achievable secrecy rates we have used the strong secrecy criterion given in the second term of (1.2). Wyner, who has introduced the wiretap channel in [Wyn75] as an information theoretic model to transmit confidential messages without the use of any key, and later Csiszár and Körner in [CK78] used the *equivocation* as a measure of secrecy. The equivocation rate as well as the weak secrecy criterion given by $\lim_{n\to\infty} \frac{1}{n} I(J; Z^n) = 0$ rely on the conditional entropy $H(J|Z^n)$ which measures the eavesdropper's uncertainty about the message J after observing its output Z^n. In Section 2.2 we show that the operational meaning of the strong secrecy criterion (1.2) is that the average error probability of every decoding strategy the eavesdropper might select tends to 1 as soon as $J_n \to \infty$.

For a discrete memoryless wiretap channel it was shown in [CK78], [AC93] that the following holds for the weaker notion of secrecy.

Theorem 1.2. *The secrecy capacity C_S of a general wiretap channel is given by*

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

where U is an auxiliary random variable and $U \to X \to (YZ)$ denotes a Markov chain.

In [Csi96] Csiszàr has proved the achievabilty part of the theorem under the strong secrecy criterion (1.2). Hence, with the proof of the converse in the earlier work of [AC93], the statement of the theorem is still valid for the strong secrecy criterion. In [Wyn75] Wyner has shown that under the assumption, that the channel to the eavesdropper V is a degraded version of the channel to the legitimate user W, the following statement is valid under the weak secrecy criterion.

Theorem 1.3. *Under the Markov chain condition $X \to Y \to Z$ the secrecy capacity C_S of a wiretap channel is given by*

$$C_S = \max_{p \in \mathcal{P}(A)} (I(X;Y) - I(X;Z)).$$

1.2 State-dependent Transmission under Channel Uncertainty

The main object of this thesis is to extend the model of the wiretap channel to models where, in addition, the legitimate users suffer from channel uncertainty, which gives a more realistic description of wireless communication scenarios. First we consider the *compound wiretap channels* where the channel realisation for the transmission of the whole actual codeword is not known except that the channel realisation belongs to a given and known set of channels. If the channel realisation varies from symbol to symbol of the transmitted codeword unknown to the legitimate users, the resulting model is the *arbitrarily varying wiretap channel AVWC*. This second model has an additional informational theoretic security aspect, namely that it can be seen as a model, in which a jammer manipulates the transmission by changing the channel state in every time step in a way unknown to the legitimate parties. So the AVWC is a model which combines both passive attacks by an eavesdropper and active attacks by a jammer. In both models we use the *strong secrecy criterion* as the measure for information theoretic secrecy to derive results for the *secrecy capacity* as the supremum of achievable rates at which reliable transmission under perfect secrecy is possible.

Here we briefly review two well accepted models for channel uncertainty, the compound channel and the arbitrarily varying channel, which will be combined then in later chapters with the wiretap channel to simulate realistic transmission of confidential messages in wireless communication systems. We will give the definitions of both of the models and recall the existing coding results.

1.2.1 The Compound Channel

If the channel realisation, unknown but as an element of a known set of channels, remains unchanged during the transmission of any codeword we call the resulting model of the information transmission a compound channel [BBT59], [Wol60], [Wol78]. The current channel realisation then will be represented as the channel state. Formally, let two finite sets A, B be the input respective output alphabet and $\{W_t : A \to \mathcal{P}(B) : \ t \in \theta\}$ the set of channels with an arbitrary index set θ. For $x^n \in A^n$, $y^n \in B^n$ and an index $t \in \theta$ the transmission of a single word is described by the n-extension of the channel $W_t : A \to \mathcal{P}(B)$

$$W_t^n(y^n|x^n) = \prod_{i=1}^{n} W_t(y_i|x_i).$$

Then the compound channel is described by the family of channels $\{(W_t^n : A^n \to B^n) : t \in \theta\}_{n\in\mathbb{N}}$. Now let the message set be $\mathcal{J}_n = \{1, \ldots, J_n\}$. Then a (n, J_n) code $\mathcal{C}_n$ consists of an encoder defining the codewords $\{x_j^n\}_{j\in\mathcal{J}_n}$ and mutually disjoint decoding sets $\{D_j \subset B^n : j \in \mathcal{J}_n\}$. The maximal error probability of the code $\mathcal{C}_n$ is defined by

$$e(\mathcal{C}_n) = \sup_{t\in\theta} \max_{j\in\mathcal{J}_n} W_t^n(D_j^c|x_j^n).$$

A positive number R is called an achievable rate if

$$\liminf_{n\to\infty} \frac{1}{n} J_n \geq R \quad \text{and} \quad \lim_{n\to\infty} e(\mathcal{C}_n) = 0,$$

and the capacity C is defined as the supremum of all achievable rates R. The code definition can be modified by replacing the maximal error probability criterion by the average error probability criterion

$$e(\mathcal{C}_n) = \sup_{t\in\theta} \frac{1}{J_n} \sum_{j=1}^{J_n} W_t^n(D_j^c|x_j^n).$$

Then the following coding theorem for the compound channel with respect to both the maximum and average error criterion and with the weak converse was proved in [BBT59], [Wol60].

Theorem 1.4. *The capacity of the compound channel is given by*

$$C(\theta) = \max_{p\in\mathcal{P}(A)} \inf_{t\in\theta} I(p, W_t).$$

Actually, Wolfowitz proved in [Wol60], [Wol78] the coding theorem with the strong converse and with respect to the maximum error probability.

Furthermore, in the special case, where the channel state is known to the transmitter before the transmission of a codeword, the capacity changes to

$$C(\theta) = \inf_{t\in\theta} \max_{p\in\mathcal{P}(A)} I(p, W_t) = \inf_{t\in\theta} C(t),$$

as the smallest capacity of all involved single channels $t \in \theta$.

1.2.2 Arbitrarily Varying Channels

Formally, let two finite sets A, B be the input respective output alphabet and let the elements s of a not necessary finite set S denote the state of the channel $W_s : A \to \mathcal{P}(B)$. For $x^n \in A^n$, $y^n \in B^n$ and a state sequence $s^n = (s_1, \dots, s_n) \in S^n$ the transmission of a single word is described by

$$W^n(y^n|x^n, s^n) := \prod_{i=1}^{n} W(y_i|x_i, s_i) := \prod_{i=1}^{n} W_{s_i}(y_i|x_i).$$

Then an arbitrarily varying channel AVC is defined as the sequence $\{\mathcal{W}^n\}_{n\in\mathbb{N}}$ of the family of channels $\mathcal{W}^n := \{W^n(\cdot|\cdot, s^n) : s^n \in S^n\}$. Now for the message set $\mathcal{J}_n = \{1, \dots, J_n\}$ a (n, J_n) code $\mathcal{C}_n$ consists of an encoder defining the codewords $\{x_j^n\}_{j\in\mathcal{J}_n}$ and mutually disjoint decoding sets $\{D_j \subset B^n : j \in \mathcal{J}_n\}$. The average error probability of the code $\mathcal{C}_n$ is defined by

$$e(\mathcal{C}_n) = \sup_{s^n\in S^n} \frac{1}{J_n} \sum_{j=1}^{J_n} W^n(D_j^c|x_j^n, s^n).$$

A positive number R is called an achievable rate if

$$\liminf_{n\to\infty} \frac{1}{n} J_n \geq R \quad \text{and} \quad \lim_{n\to\infty} e(\mathcal{C}_n) = 0.$$

For the arbitrarily varying channel AVC we further need the concept of random codes. A (n, J_n, μ, Γ) random code $\mathcal{C}_n^{\text{ran}}$ is a collection of $|\Gamma|$ deterministic (n, J_n) codes $\mathcal{C}_n^\gamma = \{\big((x_j^n)^\gamma, D_j^\gamma\big) : j \in \mathcal{J}_n\}$, where $\gamma \in \Gamma$ is chosen at random according to a distribution μ on Γ. Then the mean average error probability of the random code $\mathcal{C}_n^{\text{ran}}$ is described by

$$\bar{e}(\mathcal{C}_n^{\text{ran}}) = \sup_{s^n\in S^n} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{\gamma\in\Gamma} W^n((D_j^\gamma)^c|(x_j^n)^\gamma, s^n)\mu(\gamma),$$

and the definitions of the achievable rates and the random code capacity follow accordingly. In [BBT60] it was shown that for random codes the capacity for the arbitrarily varying channel $\{\mathcal{W}^n\}_{n\in\mathbb{N}}$ is given by

$$C_{\text{ran}} = \max_{p\in\mathcal{P}(A)} \min_{W\in\overline{\mathcal{W}}} I(p,W),$$

where $\overline{\mathcal{W}}$ denotes the convex hull of the set of channels $\{W_s : s \in S\}$. See also [AW69], [CK81] for a completely different derivation of the same result. Because $\mathcal{P}(A)$ and $\overline{\mathcal{W}}$ are convex compact sets and $I(p,W)$ is continuous and convex in W and concave in p we can apply the Minimax-Theorem (Sion's version) to obtain

$$C_{\text{ran}} = \min_{W\in\overline{\mathcal{W}}} C(W),$$

Unfortunately, because the code that is used for the transmission of a single codeword is chosen at random, reliable transmission can be guaranteed only if the outcome of the random experiment is available to both the transmitter and the receiver. With his so-called *elimination technique* Ahlswede showed in [Ahl78] that the deterministic code capacity C equals the random code capacity C_{ran} or is zero otherwise. Then a necessary and sufficient condition for the capacity C to be positive was given in [CN88] in terms of the definition of a symmetrisable AVC.

Definition 1.5 ([Eri85],[CN88])**.** *An AVC is symmetrisable if for some channel* $U : A \to S$

$$\sum_{s\in S} W(y|x,s)U(s|x') = \sum_{s\in S} W(y|x',s)U(s|x)$$

for all $x, x' \in A, y \in B$.

Ericson proved in [Eri85] nonsymmetrisability as a necessary condition for the capacity C to be positive but could not prove this as a sufficient condition. Instead he was referring to a result of Ahlswede, who showed in [Ahl78] the existence of a pair of distributions p_1, $p_2 \in \mathcal{P}(A)$, such that for any pair of distributions q_1, $q_2 \in \mathcal{P}(S)$

$$\sum_{x,s} p_1(x)q_1(s)W(y|x,s) \neq \sum_{x,s} p_2(x)q_2(s)W(y|x,s)$$

for at least one $y \in B$, as a sufficient condition for $C > 0$. Finally the authors of [CN88] found that the deterministic code capacity C of the AVC is strictly positive if and only if the AVC is not symmetrisable. In a previously published book [CK81] the authors used the elimination technique presented in a more descriptive way to derive the capacity result for $C > 0$. In a first step, the *random code reduction*, a new random code is constructed by selecting only a small number of deterministic codes of the original capacity achieving random code.

Lemma 1.6 ([CK81])**.** *Let $\mathcal{W}$ be a finite family of channels $W : A \to \mathcal{P}(B)$ and $\mathcal{C}^{\text{ran}}$ a random code which consists of a set of codes $\{\mathcal{C}(\gamma)\}_{\gamma\in\Gamma}$ and a probability distribution μ on Γ. Then for any ε and K satisfying*

$$\varepsilon > 2\log(1+\bar{e}(\mathcal{C}^{\text{ran}}), \quad K > \frac{2}{\varepsilon}(1+\log|\mathcal{W}|)$$

there exist K codes $\{\mathcal{C}(\gamma_i)\}_{i\in\{1,\dots,K\}} \in \{\mathcal{C}(\gamma)\}_{\gamma\in\Gamma}$, such that

$$\frac{1}{K}\sum_{i=1}^{K} e(\mathcal{C}(\gamma_i)) < \varepsilon$$

for all $W \in \mathcal{W}$.

In a second step called the *randomness elimination* they reduced this random code to a deterministic one by prefixing each member of the random code by a short sequence to inform the receiver which code $\mathcal{C}(\gamma_i)$, $i = 1,\dots,K$ is selected for transmission. Provided that the deterministic code capacity $C > 0$, and that the number K of codes $\mathcal{C}(\gamma_i)$ could be kept small enough that it causes no essential loss in rate, this approach results in the final theorem.

Theorem 1.7 ([Ahl78],[CN88])**.** *$C > 0$ if and only if the AVC is not symmetrisable. If $C > 0$, then*

$$C = C_{\text{ran}} = \max_{p\in\mathcal{P}(A)} \min_{W\in\overline{\mathcal{W}}} I(p,W) = \min_{W\in\overline{\mathcal{W}}} \max_{p\in\mathcal{P}(A)} I(p,W).$$

In [CN88] the same capacity result was derived directly for deterministic codes without extracting it from correlated random codes.

Until now there exist no general capacity results for the arbitrarily varying channel AVC with respect to the maximum error probability. The capacity was determined for AVC's with a binary output in [AW70]. Further partial results was given in [AW80]. In [Ahl70] Ahlswede has shown that the general solution under the maximum error criterion is connected to Shannon's zero error capacity problem [Sha56].

The arbitrarily varying channel AVC usually is used in information theory as a model of practical systems, which mirrors the impossibility to know the actual state of a channel, because it changes from one time step to the next in an unknown manner. Nevertheless it can be seen as well as a model, in which a possible third party, the jammer, manipulates the transmission by changing the channel state arbitrarily from time step to time step but unknown to the legitimate parties. The case that he randomises over all states is also included in this model [Ahl78]. Thus the AVC can be seen as a model of an active attack to disrupt the transmission between the legitimate users.

1.3 Contributions and Outline of the Thesis

In this thesis we consider a wiretap channel where the legitimate users suffer from channel uncertainty, which gives a more realistic description of practical wireless systems. We require reliable transmission to the legitimate receiver and at the same time secrecy against a potential eavesdropper. This will be realised by the validity of the strong secrecy criterion. The transmission to the legitimate receiver and the eavesdropper will be described by families of pairs of channels with common input alphabet and possibly different output alphabets.

First we consider the *compound wiretap channel* where the exact channel realisation (the pair of channels) for the transmission of the codeword is not known but only that the channel realisation belongs to a given and known set of channels. The channel realisation remains fixed for the whole transmission of a codeword. Additionally we assume that the eavesdropper always knows which channel is in use.

If the channel realisation (the pair of channels) varies arbitrarily from time step to time step during the transmission of a codeword in a way unknown to the legitimate users, the resulting model is the *arbitrarily varying wiretap channel AVWC*. This second model has an additional informational theoretic security aspect, namely that it can be seen as a model, in which a jammer manipulates the transmission by changing the channel state in every time step in a manner unknown to the legitimate parties. So the AVWC is a model which combines both passive attacks by an eavesdropper and active attacks by a jammer.

In Chapter 2 we define the compound wiretap channel under the strong secrecy criterion. We consider different communication scenarios. In the first the transmitter has perfect channel state information (CSI), in the second the legitimate users have no CSI at all and in a third the transmitter has only knowledge of the channel state to the legitimate receiver. In the last two scenarios we consider the special case where the channels to the eavesdropper are degraded versions of those to the legitimate receiver. According to that we derive lower and upper bounds on the secrecy capacity. In special cases we can provide the secrecy capacity as an explicit expression, but we show that in general it is possible to determine the capacity as a multi-letter expression. Parts of the results are published in [BBS11a] and [BBS11b].

In Chapter 3 we introduce the arbitrarily varying wiretap channel AVWC as a model, which combines passive and active attacks on the security of communication, so that the transmission of messages has to be protected from eavesdropping, modelled by a wiretapper, and against a possible jammer realised by the unknown variation of the channel state. We will give a condition under which the random code secrecy capacity equals the deterministic code capacity. Under the assumption of a ”best” channel to the eavesdropper we establish a lower bound on the secrecy capacity and give a multi-letter expression for the secrecy capacity, which holds for both random and deterministic codes. By adding an additional structure to the channels to the legitimate user we can determine the secrecy capacity under these special assumptions. Parts of the results will be published in [BBS12].

Chapter 2

Compound Wiretap Channels

Compound wiretap channels are among the simplest non-trivial models incorporating the requirement of security against a potential eavesdropper while at the same time the legitimate users suffer from channel uncertainty. They may be considered therefore as a starting point for theoretical investigation tending towards applications, for example, in wireless systems, a fact explaining an alive research activity in this area in recent years (cf. [LKPS08], [BL08] and references therein). In this chapter we give capacity results for different scenarios of channel state information under a strong secrecy criterion and the maximum error probability criterion. Parts of the results relying on [BBS11a], [BBS11b]. In Chapter 3 we make use of these results to derive capacity results for arbitrarily varying wiretap channels, a more realistic communication model, which, apart from eavesdropping, takes into account an active adversarial jamming situation.

A compound wiretap channel is described by finite families of pairs of channels $\mathfrak{W} = \{(W_t, V_t) : t = 1, \ldots, T\}$ with common input alphabet and possibly different output alphabets. The legitimate users control W_t and the eavesdropper observes the output of V_t and both the channels are coupled by the channel state t. The term compound here refers to the fact, that the legitimate users do not know which channel pair (W_t, V_t) is in use for the actual codeword transmission, or have imperfect knowledge of the actual channel state. Nevertheless they know the whole set of pairs of channels $\mathfrak{W}$. In particular, we will be dealing with two communication scenarios. In the first one the transmitter is informed about the index t (channel state information (CSI) at the transmitter) while in the second the transmitter has no information about that index at all (no CSI). In both scenarios the eavesdropper knows and the legitimate receiver does not know the channel state. Along the way we will comment what our results look like when applied to widely used class of models of the form $\mathfrak{W} = \{(W_t, V_s) : t = 1, \dots, T, s = 1, \dots, S\}$ with $T \neq S$ which are special cases of the model we are dealing with in this thesis.

Our contributions are summarised as follows: In [LKPS08] a general upper bound on the capacity of compound wiretap channel as the minimum secrecy capacity of the involved wiretap channels was given. We prove in Section 2.3.1 that the models whose secrecy capacity matches this upper bound contain all compound wiretap channels with CSI at the transmitter. At the same time we achieve this bound with a substantially stronger security criterion employed already in [Csi96], [MW00], [CWY04], and [Dev05]. Indeed, our security proof follows closely that developed in [Dev05] for single wiretap channel with classical input and quantum output. In order to achieve secrecy we follow the common approach according to which randomised encoding is a permissible operation. The impact of randomisation at the legitimate decoder's site is usually compensated by communicating to her/him the outcome of the random experiment performed. However, in the case of compound wiretap channel with CSI at the transmitter this strategy does not work as is illustrated by an example in Section 2.4.1. We resolve this difficulty by developing a decoding strategy which is independent of the particular channel realisation and is insensitive to randomisation while decoding just at the optimal secrecy rate for all channels $\{W_t : t = 1, \dots, T\}$ simultaneously.

Moreover, a slight modification of our proofs allows us to determine the capacity of the compound wiretap channel without CSI by a (non-computable) multi-letter expression. This is the content of Section 2.3.2. We should mention, however, that the traditional proof strategy of sending the pair consisting of message and randomisation parameter to the legitimate receiver works as well in the case where the transmitter has no CSI. The lower bound on the secrecy capacity, we will proof under the strong secrecy criterion, we will use for parts of the secrecy results for arbitrarily varying wiretap channels in Chapter 3. The lower bound on the secrecy capacity as well the as the multi-letter expression were given earlier in [LKPS08] respective in [BL08] for weaker secrecy criteria but without detailed proofs.

In Section 2.4.2 we give an example of compound wiretap channel such that both the set of channels to the legitimate receiver and to the eavesdropper are convex but whose secrecy capacities with CSI and without CSI at the transmitter are different. Indeed the former is positive while the latter is equal to 0.

Section 2.3.3 is devoted to the practically important model $\mathfrak{W} = \{(W_t, V_s) : t = 1, \dots, T, s = 1, \dots, S\}$ with the assumption that the transmitter has CSI for the T-part but has no CSI for the S-part of the channel. Here again we provide a multi-

letter expression for the capacity. Additionally, we give a computable description of the secrecy capacity in the case where the channels to the eavesdropper are degraded versions of those to the legitimate receiver.
Our results are easily extended to arbitrary sets (even uncountable) of wiretap channels via standard approximation techniques [BBT59].

2.1 Definitions

Let A, B, C be finite sets and $\theta = \{1, \dots, T\}$ an index set. We consider two families of channels $W_t : A \to \mathcal{P}(B)$[1], $V_t : A \to \mathcal{P}(C)$, $t \in \theta$, which we collectively abbreviate by $\mathfrak{W}$ and call the compound wiretap channel generated by the given families of channels. Here the first family represents the communication link to the legitimate receiver while the output of the latter is under control of the eavesdropper. In the rest of the chapter expressions like $W_t^{\otimes n}$ or $V_t^{\otimes n}$ stand for the n-th memoryless extension of the stochastic matrices W_t, V_t.
An (n, J_n) code for the compound wiretap channel $\mathfrak{W}$ consists of a stochastic encoder $E : \mathcal{J}_n \to \mathcal{P}(A^n)$ (a stochastic matrix) with a message set $\mathcal{J}_n := \{1, \dots, J_n\}$ and a collection of mutually disjoint decoding sets $\{D_j \subset B^n : j \in \mathcal{J}_n\}$. The maximum error probability of a (n, J_n) code $\mathcal{C}_n$ is given by

$$e(\mathcal{C}_n) := \max_{t \in \theta} \max_{j \in \mathcal{J}_n} \sum_{x^n \in A^n} E(x^n|j) W_t^{\otimes n}(D_j^c|x^n). \tag{2.1}$$

I.e. neither the sender nor the receiver have CSI.
If channel state information is available at the transmitter the notion of (n, J_n) code is modified in that the encoding may depend on the channel index while the decoding sets remain universal, i.e. independent of the channel index t. The probability of error in (2.1) changes to

$$e_{\mathrm{CSI}}(\mathcal{C}_n) := \max_{t \in \theta} \max_{j \in \mathcal{J}_n} \sum_{x^n \in A^n} E_t(x^n|j) W_t^{\otimes n}(D_j^c|x^n).$$

We assume throughout the chapter that the eavesdropper always knows which channel is in use.

Definition 2.1. *A non-negative number R is an achievable secrecy rate for the compound wiretap channel $\mathfrak{W}$ with or without CSI respectively if there is a sequence*

[1] $\mathcal{P}(B)$ denotes the set of probability distributions on B.

$(\mathcal{C}_n)_{n\in\mathbb{N}}$ of (n, J_n) codes such that

$$\lim_{n\to\infty} e(\mathcal{C}_n) = 0 \textit{ resp. } \lim_{n\to\infty} e_{\mathrm{CSI}}(\mathcal{C}_n) = 0,$$

$$\liminf_{n\to\infty} \frac{1}{n}\log J_n \geq R,$$

and

$$\lim_{n\to\infty} \max_{t\in\theta} I(J; Z_t^n) = 0, \tag{2.2}$$

where J is an uniformly distributed random variable taking values in $\mathcal{J}_n$ and Z_t^n are the resulting random variables at the output of eavesdropper's channel $V_t^{\otimes n}$. The secrecy capacity in either scenario is given by the largest achievable secrecy rate and is denoted by $C_S(\mathfrak{W})$ and $C_{S,CSI}(\mathfrak{W})$.

2.2 Hints on Operational Meaning of Strong Secrecy

A weaker and widely used security criterion is obtained if we replace (2.2) by $\lim_{n\to\infty} \max_{t\in\theta} \frac{1}{n} I(J; Z_t^n) = 0$. We prefer to follow [Csi96], [CWY04], and [Dev05] and require the validity of (2.2). A nice discussion on interrelation of several secrecy criteria is contained in [BL08]. We confine ourselves to giving some hints on the operational meaning of the requirement (2.2). To this end we restrict our attention to the case where the transmitter has no CSI in order to simplify our notation. The case of compound wiretap channel with CSI at the transmitter can be treated accordingly. Set

$$\varepsilon_n := \max_{t\in\theta} I(J; Z_t^n) \text{ with } \lim_{n\to\infty} \varepsilon_n = 0.$$

Then Pinsker's inequality implies that

$$||p_{JZ_t^n} - p_J \otimes p_{Z_t^n}|| \leq c\sqrt{\varepsilon_n} \quad \forall t \in \theta, \tag{2.3}$$

with a positive universal constant c, where $||\cdot||$ is the variational distance. Suppose that the eavesdropper chooses for each $t \in \theta$ decoding sets $\{K_{j,t} \subset C^n : j \in \mathcal{J}_n\}$ with $C^n = \bigcup_{j\in\mathcal{J}_n} K_{j,t}$. We will lower bound the average error probability (and consequently the maximum error probability) for every choice of the decoding rule the eavesdropper might make. Set

$$e_{\mathrm{av}}(t) := \frac{1}{J_n} \sum_{j\in\mathcal{J}_n} \sum_{x^n\in A^n} E(x^n|j) V_t^{\otimes n}(K_{j,t}^c|x^n).$$

Then

$$\begin{aligned} e_{\mathrm{av}}(t) &= \sum_{j\in\mathcal{J}_n} p_{JZ_t^n}(\{j\}\times K_{j,t}^c) = p_{JZ_t^n}\big(\bigcup_{j\in\mathcal{J}_n}\{j\}\times K_{j,t}^c\big) \\ &\geq p_J\otimes p_{Z_t^n}\Big(\bigcup_{j\in\mathcal{J}_n}\{j\}\times K_{j,t}^c\Big) - c\sqrt{\varepsilon_n} \\ &= \sum_{j\in\mathcal{J}_n} p_J\otimes p_{Z_t^n}\big(\{j\}\times K_{j,t}^c\big) - c\sqrt{\varepsilon_n} = \frac{1}{J_n}\sum_{j\in\mathcal{J}_n} p_{Z_t^n}(K_{j,t}^c) - c\sqrt{\varepsilon_n} \\ &= \frac{J_n-1}{J_n} - c\sqrt{\varepsilon_n} = 1-\frac{1}{J_n} - c\sqrt{\varepsilon_n}, \end{aligned} \tag{2.4}$$

where in the first and the third line we have used the fact that the sets $\{j\}\times K_{j,t}^c$, $j\in\mathcal{J}_n$, are mutually disjoint, the second line follows from (2.3), and in the fourth line we merely observed that for any non-negative numbers $a_1,\ldots,a_J$ with $\sum_{j=1}^J a_j = 1$ we have $\sum_{j=1}^J(1-a_j) = J-1$. Consequently, the average (and hence maximum) error probability of every decoding strategy the eavesdropper might select tends to 1 as soon as $J_n\to\infty$. It should be remarked, however, that although for the vast majority of messages the eavesdropper will be in error there is still a possibility left that she/he can decode a small fraction of them correctly. As will follow from the proofs below we will have $\varepsilon_n = 2^{-na}$, $a>0$, and $J_n = 2^{nR}$, $R>0$, if the secrecy capacity is positive so that the speed of convergence in (2.4) will be exponential.

Notice that (2.3) means that the random variables Z_t^n at the output of the channel to the eavesdropper are almost independent of the random variable J embodying the messages to be transmitted to the legitimate receiver. Therefore it is heuristically convincing that our criterion (2.2) offers secrecy to some extent for communication tasks going beyond the transmission of messages. To demonstrate this by an example we introduce, based on [AD89], the notion of identification attack as follows. Suppose that for each fixed $t\in\theta$ and any $j\in\mathcal{J}_n$ there is a subset $K_{j,t}\subset C^n$ on the eavesdropper's output alphabet where now the sets $K_{j,t}$ need not necessarily be mutually disjoint. With $E:\mathcal{J}_n\to\mathcal{P}(A^n)$ being the stochastic encoder used to transmit messages to the legitimate receiver we can write down the identification errors of first and second kind (cf. [AD89] for further explanation of this code concept) for the eavesdropper's channel as

$$\sum_{x^n\in A^n} E(x^n|j)V_t^{\otimes n}(K_{j,t}^c|x^n), \tag{2.5}$$

and

$$\sum_{x^n\in A^n} E(x^n|i)V_t^{\otimes n}(K_{j,t}|x^n) \tag{2.6}$$

for $j, i \in \mathcal{J}_n$, $i \neq j$.

One possible interpretation of this attack, again based on [AD89], is that on the eavesdropper's side of the channel there are persons $F_1, \ldots, F_{J_n}$ observing the output of the channel. The sole interest of F_j is whether or not the message j has been sent to the legitimate receiver. Thus F_j performs the hypothesis test represented by $K_{j,t}$ based on his/her knowledge of $t \in \theta$ and (2.5), (2.6) are just the errors of the first resp. second kind for that hypothesis test.

Let us define for $j \in \mathcal{J}_n$

$$g(j,t) := \sum_{x^n \in A^n} \left(E(x^n|j) V_t^{\otimes n}(K_{j,t}^c|x^n) + \frac{1}{J_n - 1} \sum_{\substack{i=1 \\ i \neq j}}^{J_n} E(x^n|i) V_t^{\otimes n}(K_{j,t}|x^n) \right)$$

which is a number in $[0, 2]$.

Notice that if

$$g(j,t) \geq 1 - \eta$$

for some $\eta \in (0,1)$ then either

$$\sum_{x^n \in A^n} E(x^n|j) V_t^{\otimes n}(K_{j,t}^c|x^n) \geq \frac{1-\eta}{2},$$

or there is at least one $i \neq j$ with

$$\sum_{x^n \in A^n} E(x^n|i) V_t^{\otimes n}(K_{j,t}|x^n) \geq \frac{1-\eta}{2},$$

or both, so that no reliable identification of message j can be guaranteed. We show now that under assumption of (2.2) we have

$$\frac{1}{J_n} \sum_{j=1}^{J_n} g(j,t) \geq 1 - \eta_n, \quad \eta_n = o(n^0) \tag{2.7}$$

so that at most a fraction $\frac{2}{3}(1 + \eta_n)$ of $j \in \mathcal{J}_n$ can satisfy the inequality

$$g(j,t) < \frac{1}{2}.$$

This last assertion is readily seen from (2.7) by applying Markov's inequality to the set

$$F := \{ j \in \mathcal{J}_n : 2 - g(j,t) > \frac{3}{2} \}.$$

In order to prove (2.7), note that for any $t \in \theta$

$$\begin{aligned}
\frac{1}{J_n}\sum_{j=1}^{J_n} g(j,t) &= \sum_{j=1}^{J_n}\Big(p_{JZ_t^n}(\{j\}\times K_{j,t}^c) + \frac{1}{J_n-1}p_{JZ_t^n}(\{j\}^c \times K_{j,t})\Big) \\
&= p_{JZ_t^n}(\bigcup_{j\in\mathcal{J}_n}\{j\}\times K_{j,t}^c) + \frac{1}{J_n-1}\sum_{j=1}^{J_n} p_{JZ_t^n}(\{j\}^c \times K_{j,t}) \\
&\geq p_J \otimes p_{Z_t^n}(\bigcup_{j\in\mathcal{J}_n}\{j\}\times K_{j,t}^c) + \frac{1}{J_n-1}\sum_{j=1}^{J_n} p_J \otimes p_{Z_t^n}(\{j\}^c \times K_{j,t}) \\
&\qquad -c\sqrt{\varepsilon_n} - c\frac{J_n}{J_n-1}\sqrt{\varepsilon_n}
\end{aligned}$$

where in the third line we have used (2.3). If we now insert $p_J(\{j\}^c) = \frac{J_n-1}{J_n}$, we obtain finally

$$\begin{aligned}
\frac{1}{J_n}\sum_{j=1}^{J_n} g(j,t) &\geq \frac{1}{J_n}\sum_{j=1}^{J_n}\big(p_{Z_t^n}(K_{j,t}^c) + p_{Z_t^n}(K_{j,t})\big) - c\sqrt{\varepsilon_n}\frac{2J_n-1}{J_n-1} \\
&= 1 - c\sqrt{\varepsilon_n}\frac{2J_n-1}{J_n-1}.
\end{aligned}$$

Besides the attempts of the eavesdropper to decode or identify messages we can introduce attacks corresponding to each communication task introduced in [Ahl08]. It would be interesting, not only from the mathematical point of view, to see against which of them and to what extent secrecy can be guaranteed by the condition (2.2).

2.3 Capacity Results

Now we will give the capacity results for three different scenarios. In the first the transmitter has perfect channel state information (CSI). In the second no channel state information are available. In the third case we allow that the state of the channel to the legitimate user and that of the channel to the eavesdropper can be chosen independently. In this scenario the transmitter should have knowledge of the channel state to the legitimate receiver but the channel state to the eavesdropper is unknown. In all cases the legitimate receiver has no CSI, whereas the eavesdropper always knows which channel is in use. For all proofs concerning the capacity results we will use some properties of *typical* and *conditionally typical* sequences as they were treated in [CK81] by Csiszár and Körner (cf. Appendix A).

2.3.1 CSI at the Transmitter

First we will give the capacity result in the case in which the transmitter has full knowledge of the channel state (CSI) while the legitimate receiver has no information about the channel state. So the transmitter can adapt the stochastic encoder to the specific channel realisations, whereas the decoding sets must be chosen independent of the possible channel realisations. The main result in this section is the following theorem.

Theorem 2.2. *The secrecy capacity of the compound wiretap channel* $\mathfrak{W}$ *with CSI at the transmitter is given by*

$$C_{S,CSI}(\mathfrak{W}) = \min_{t\in\theta} \max_{U\to X\to (YZ)_t} (I(U,Y_t) - I(U,Z_t)).$$

Here X is a random variable with probability distribution in $\mathcal{P}(A)$ and U is an auxiliary random variable with range equals A, such that $U, X, (YZ)_t$ form a Markov chain $U \to X \to (YZ)_t$ in this order. Then the maximum refers to all random variables satisfying the Markov chain condition such that X is connected with Y_t respective Z_t by the channels W_t respective V_t for every $t \in \theta$.

Notice first that the inequality

$$C_{S,CSI}(\mathfrak{W}) \leq \min_{t\in\theta} \max_{U\to X\to (YZ)_t} (I(U,Y_t) - I(U,Z_t))$$

is trivially true since we cannot exceed the secrecy capacity of the worst wiretap channel in the family $\mathfrak{W}$. This has been already pointed out in [LKPS08].
The rest of this section is devoted to the proof of the achievability. For this we need the following lemma which is a standard result from large deviation theory.

Lemma 2.3. *(Chernoff bounds) Let* $Z_1, \dots, Z_L$ *be i.i.d. random variables with values in* $[0,1]$ *and expectation* $\mathbb{E}Z_i = \mu$, *and* $0 < \epsilon < \frac{1}{2}$. *Then it follows that*

$$Pr\left\{\frac{1}{L}\sum_{i=1}^{L} Z_i \notin [(1\pm\epsilon)\mu]\right\} \leq 2\exp\left(-L\cdot\frac{\epsilon^2\mu}{3}\right),$$

where $[(1\pm\epsilon)\mu]$ *denotes the interval* $[(1-\epsilon)\mu, (1+\epsilon)\mu]$.

Proof. The proof is given in [DP09] (cf. Theorem 1.1) and in [AW02]. □

Proof. (of the Theorem) For $p \in \mathcal{P}(A)$, $V : A \to \mathcal{P}(C)$, $x^n \in A^n$, and $\delta > 0$ let $\mathcal{T}^n_{pt,\delta}$ the set of p_t-typical sequences on A^n, and $\mathcal{T}^n_{V_t,\delta}(x^n)$ the set of conditionally typical sequences given x^n on C^n. For the properties of typical and conditional typical sequences see A.2.

It suffices to prove that $\min_{t\in\Theta}(I(X_t,Y_t)-I(X_t,Z_t))$ for $(XYZ)_t$ as above is an achievable secrecy rate. Then we will have shown that $R=\min_{t\in\Theta}(I(U_t,Y_t)-I(U_t,Z_t))$, with $U_t\to X_t\to(YZ)_t$ form a Markov chain, is an achievable secrecy rate (cf. [CK81] page 409). We choose $p_1,\dots,p_T\in\mathcal{P}(A)$ and define new probability distributions on A^n by

$$p'_t(x^n):=\begin{cases}\frac{p_t^{\otimes n}(x^n)}{p_t^{\otimes n}(\mathcal{T}^n_{p_t,\delta})} & \text{if } x^n\in\mathcal{T}^n_{p_t,\delta},\\ 0 & \text{otherwise}\end{cases}. \tag{2.8}$$

Define then for $z^n\in C^n$, $x^n\in A^n$

$$\tilde{Q}_{t,x^n}(z^n)=V_t^n(z^n|x^n)\cdot\mathbf{1}_{\mathcal{T}^n_{V_t,\delta}(x^n)}(z^n)$$

on C^n. Additionally, we set for $z^n\in C^n$

$$\Theta'_t(z^n)=\sum_{x^n\in\mathcal{T}^n_{p_t,\delta}}p'_t(x^n)\tilde{Q}_{t,x^n}(z^n). \tag{2.9}$$

Now let $S:=\{z^n\in C^n:\Theta'_t(z^n)\geq\epsilon\alpha_t\}$ where $\epsilon:=2^{-nc'\delta^2}$ (cf. Lemma A.8) and $\alpha_t:=2^{-n(H(p_tV_t)+f_1(\delta))}$ (cf. (A.28) in Lemma A.9, computed with respect to p_t and V_t). By Lemma A.9 the support of Θ'_t has cardinality $\leq\alpha_t^{-1}$ since for each $x^n\in\mathcal{T}^n_{p_t,\delta}$ it holds that $\mathcal{T}^n_{V_t,\delta}(x^n)\subset\mathcal{T}^n_{p_tV_t,2|A|\delta}$, which implies that $\sum_{z^n\in S}\Theta_t(z^n)\geq 1-2\epsilon$, if

$$\Theta_t(z^n)=\Theta'_t(z^n)\cdot\mathbf{1}_S(z^n)$$

and

$$Q_{t,x^n}(z^n)=\tilde{Q}_{t,x^n}(z^n)\cdot\mathbf{1}_S(z^n). \tag{2.10}$$

Now for each $t\in\theta$ define $J_n\cdot L_{n,t}$ i.i.d. random variables $X^{(t)}_{jl}$ with $j\in[J_n]:=\{1,\dots,J_n\}$ and $l\in[L_{n,t}]:=\{1,\dots,L_{n,t}\}$ each of them distributed according to p'_t with

$$J_n = \left\lfloor 2^{n[\min_{t\in\theta}(I(p_t,W_t)-I(p_t,V_t))-\tau]}\right\rfloor \tag{2.11}$$

$$L_{n,t} = \left\lfloor 2^{n[I(p_t,V_t)+\frac{\tau}{4}]}\right\rfloor \tag{2.12}$$

for $\tau>0$. Moreover we suppose that the random matrices $\{X^{(t)}_{j,l}\}_{j\in[J_n],l\in[L_{n,l}]}$ and $\{X^{(t')}_{j,l}\}_{j\in[J_n],l\in[L_{n,l}]}$ are independent for $t\neq t'$. Now it is obvious from (2.9) and the definition of the set S that for any $z^n \in S$, $\Theta_t(z^n) = \mathbb{E}Q_{t,X_{jl}^{(t)}}(z^n) \geq \epsilon\alpha_t$ if $\mathbb{E}$ is the expectation value with respect to the distribution p_t'. Define further $\beta_t := 2^{-n(H(V_t|p_t)-f_2(\delta))}$ (cf. (A.29) in Lemma A.9). For the random variables $\beta_t^{-1}Q_{t,X_{jl}^{(t)}}(z^n)$ define the event

$$\bigcap_{z^n \in C^n} \left\{ \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} \beta_t^{-1} Q_{t,X_{jl}^{(t)}}(z^n) \in [(1 \pm \epsilon)\beta_t^{-1}\Theta_t(z^n)] \right\}, \tag{2.13}$$

which equals in probability the following event

$$\iota_j(t) := \bigcap_{z^n \in C^n} \left\{ \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} Q_{t,X_{jl}^{(t)}}(z^n) \in [(1 \pm \epsilon)\Theta_t(z^n)] \right\}. \tag{2.14}$$

Then keeping in mind that $\Theta_t(z^n) \geq \epsilon\alpha_t$ for all $z^n \in S$, it follows with (2.13) that for all $j \in [J_n]$ and for all $t \in \theta$

$$\Pr\{(\iota_j(t))^c\} \leq 2|C|^n \exp\Big(- L_{n,t} \frac{2^{-n[I(p_t,V_t)+g(\delta)]}}{3} \Big) \tag{2.15}$$

by Lemma 2.3, Lemma A.9, and our choice $\epsilon = 2^{-nc'\delta^2}$ with $g(\delta) := f_1(\delta) + f_2(\delta) + 3c'\delta^2$. Making $\delta > 0$ sufficiently small we have for all sufficiently large $n \in \mathbb{N}$

$$L_{n,t} 2^{-n[I(p_t,V_t)+g(\delta)]} \geq 2^{n\frac{\tau}{8}}.$$

Thus, for this choice of δ the RHS of (2.15) is double exponential in n uniformly in $t \in \theta$ and can be made smaller than ϵJ_n^{-1} for all $j \in [J_n]$ and all sufficiently large $n \in \mathbb{N}$. I.e.

$$\Pr\{(\iota_j(t))^c\} \leq \epsilon J_n^{-1} \quad \forall t \in \theta. \tag{2.16}$$

Let us turn now to the coding part of the problem. Let $p_t' \in \mathcal{P}(A^n)$ be given as in (2.8). We abbreviate $\mathcal{X} := \{X^{(t)}\}_{t\in\theta}$ for the family of random matrices $X^{(t)} = \{X_{jl}^{(t)}\}_{j\in[J_n],l\in[L_{n,t}]}$ whose components are i.i.d. according to p_t'. Further, as we have supposed, let $X^{(t)}$ and $X^{(t')}$ independent for $t \neq t'$.We will show now how the reliable transmission of the message $j \in [J_n]$ can be achieved when randomising over the index $l \in L_{n,t}$ without any attempt to decode the randomisation parameter at the legitimate receiver (see section 2.4.1). To this end let us define for each $j \in [J_n]$ a random set

$$D_j'(\mathcal{X}) := \bigcup_{s\in\theta} \bigcup_{k\in[L_{n,s}]} \mathcal{T}^n_{W_s,\delta}(X_{jk}^{(s)}),$$

and the subordinate random decoder $\{D_j(\mathcal{X})\}_{j\in[J_n]} \subseteq B^n$ is given by

$$D_j(\mathcal{X}) := D'_j(\mathcal{X}) \cap \Bigg(\bigcup_{\substack{j'\in[J_n]\\ j'\neq j}} D'_{j'}(\mathcal{X}) \Bigg)^c. \tag{2.17}$$

Thus the decoding set for a message $j \in \mathcal{J}_n$ is defined by the output sequences which are W_s-typical conditioned on the input $X_{jk}^{(s)}$ for all channels $s \in \theta$, and the union over all $k \in [L_{n,s}]$ makes the decoder robust to the effect of randomisation. The second part in (2.17) excludes the output sequences that are W_s-typical conditioned on inputs that are different in the message index, such that the decoding sets are mutually disjoint. Consequently we can define the random average probabilities of error for a specific channel $t \in \theta$ by

$$\lambda_n^{(t)}(\mathcal{X}) := \frac{1}{J_n} \sum_{j\in[J_n]} \frac{1}{L_{n,t}} \sum_{l\in[L_{n,t}]} W_t^{\otimes n}((D_j(\mathcal{X}))^c | X_{jl}^{(t)}). \tag{2.18}$$

Now (2.17) implies for each $t \in \theta$ and $l \in [L_{n,t}]$

$$\begin{aligned} &W_t^{\otimes n}((D_j(\mathcal{X}))^c|X_{jl}^{(t)}) \leq W_t^{\otimes n}((D'_j(\mathcal{X}))^c|X_{jl}^{(t)}) + W_t^{\otimes n}(\bigcup_{\substack{j'\in[J_n]\\ j'\neq j}} D'_{j'}(\mathcal{X})|X_{jl}^{(t)}) \\ &\leq W_t^{\otimes n}(\bigcap_{s\in\theta} \bigcap_{k\in[L_{n,s}]} (\mathcal{T}^n_{W_s,\delta}(X_{jk}^{(s)}))^c|X_{jl}^{(t)}) + \sum_{\substack{j'\in[J_n]\\ j'\neq j}} \sum_{s\in\theta} \sum_{k\in[L_{n,s}]} W_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k}^{(s)})|X_{jl}^{(t)}) \\ &\leq W_t^{\otimes n}((\mathcal{T}^{\otimes n}_{W_t,\delta}(X_{jl}^{(t)}))^c|X_{jl}^{(t)}) + \sum_{\substack{j'\in[J_n]\\ j'\neq j}} \sum_{s\in\theta} \sum_{k\in[L_{n,s}]} W_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k}^{(s)})|X_{jl}^{(t)}), \end{aligned} \tag{2.19}$$

where the first and second inequality follow by the union bound and the third one follows by the monotonicity of the probability. Next, if we average over all random codebooks, we obtain by Lemma A.8 and the independence of all involved random variables

$$\begin{aligned} &\mathbb{E}_{\mathcal{X}}(W_t^{\otimes n}((D_j(\mathcal{X}))^c|X_{jl}^{(t)})) \\ &\leq (n+1)^{|A||B|} \cdot 2^{-nc\delta^2} + \sum_{\substack{j'\in[J_n]\\ j'\neq j}} \sum_{s\in\theta} \sum_{k\in[L_{n,s}]} \mathbb{E}_{X_{j'k}^{(s)}} \mathbb{E}_{X_{jl}^{(t)}} W_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k}^{(s)})|X_{jl}^{(t)}). \end{aligned} \tag{2.20}$$

We shall find now for $j' \neq j$ an upper bound on the inner expectation of the second term

$$\begin{aligned}\mathbb{E}_{X_{jl}^{(t)}} W_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k}^{(s)})|X_{jl}^{(t)}) &= \sum_{x^n\in A^n} p'_t(x^n)W_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k}^{(s)})|x^n)\\ &\leq \sum_{x^n\in A^n} \frac{p_t^{\otimes n}(x^n)}{p_t^{\otimes n}(\mathcal{T}^n_{p_t,\delta})} W_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k}^{(s)})|x^n) \\ &= \frac{q_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k}^{(s)}))}{p_t^{\otimes n}(\mathcal{T}^n_{p_t,\delta})}.\end{aligned} \tag{2.21}$$

The first inequality follows by the definition of p'_t in (2.8), and in the third line $q_t \in \mathcal{P}(B)$ denotes the output distribution which is generated by p_t and W_t. By Lemma A.8 and by Lemma A.10 for any $t, s \in \theta$ we have

$$\begin{aligned} p_t^{\otimes n}(\mathcal{T}^n_{p_t,\delta}) &\geq 1-(n+1)^{|A|}\cdot 2^{-nc\delta^2} \\ q_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k}^{(s)})) &\leq (n+1)^{|A||B|}\cdot 2^{-n(I(p_s,W_s)-f(\delta))},\end{aligned} \tag{2.22}$$

with a universal $f(\delta) > 0$ satisfying $\lim_{\delta\to 0} f(\delta) = 0$, since $X_{j'k}^{(s)} \in \mathcal{T}^n_{p_s,\delta}$ with probability 1. Thus inserting this into (2.21) we obtain

$$\mathbb{E}_{X_{jl}^{(t)}} W_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k}^{(s)})|X_{jl}^{(t)}) \leq \frac{(n+1)^{|A||B|}}{1-(n+1)^{|A|}\cdot 2^{-nc\delta^2}}\cdot 2^{-n(I(p_s,W_s)-f(\delta))},$$

and consequently for the outer expectation in (2.20)

$$\mathbb{E}_{X_{j'k}^{(s)}}\mathbb{E}_{X_{jl}^{(t)}} W_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k}^{(s)})|X_{jl}^{(t)}) \leq \frac{(n+1)^{|A||B|}}{1-(n+1)^{|A|}\cdot 2^{-nc\delta^2}}\cdot 2^{-n(I(p_s,W_s)-f(\delta))} \tag{2.23}$$

for all $s, t \in \theta$, all $j' \neq j$, and all $l \in [L_{n,t}]$, $k \in [L_{n,s}]$. Now by defining $\nu_n(\delta) := (n+1)^{|A||B|}\cdot 2^{-nc\delta^2}$ and $\mu_n(\delta) := 1-(n+1)^{|A|}\cdot 2^{-nc\delta^2}$, thus for each $t \in \theta$, $l \in [L_{n,t}]$, and $j \in [J_n]$ (2.20) and (2.23) lead to

$$\begin{aligned}&\mathbb{E}_{\mathcal{X}}(W_t^{\otimes n}((D_j(\mathcal{X}))^c|X_{jl}^{(t)}))\\ &\leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} J_n \sum_{s\in\theta} L_{n,s} 2^{-n(I(p_s,W_s)-f(\delta))}\\ &\leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} J_n \sum_{s\in\theta} 2^{-n(I(p_s,W_s)-I(p_s,V_s)-f(\delta)-\frac{\tau}{4})}\\ &\leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} T\cdot J_n\cdot 2^{-n(\min_{s\in\theta}(I(p_s,W_s)-I(p_s,V_s))-f(\delta)-\frac{\tau}{4})}\\ &\leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} T\cdot 2^{-n(\tau-f(\delta)-\frac{\tau}{4})}\end{aligned}$$

and hence

$$\mathbb{E}_{\mathcal{X}}(W_t^{\otimes n}((D_j(\mathcal{X}))^c|X_{jl}^{(t)})) \leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} T \cdot 2^{-n\frac{\tau}{2}}, \tag{2.24}$$

where we have used the definition of J_n and $L_{n,s}$ in (2.11), (2.12), and we have chosen $\delta > 0$ small enough to ensure that $\tau - f(\delta) - \frac{\tau}{4} \geq \frac{\tau}{2}$. Defining $a = a(\delta, \tau) := \frac{\min\{c\delta^2, \frac{\tau}{4}\}}{2}$ we can find $n(\delta, \tau, |A|, |B|) \in \mathbb{N}$ such that for all $n \geq n(\delta, \tau, |A|, |B|)$

$$\mathbb{E}_{\mathcal{X}}(W_t^{\otimes n}((D_j(\mathcal{X}))^c|X_{jl}^{(t)})) \leq T \cdot 2^{-na}$$

holds for all $t \in \theta$, $l \in [L_{n,t}]$, and $j \in [J_n]$. Consequently, for any $t \in \theta$ we obtain by the definition of the error probability in (2.18) the mean average error probability, where the expectation is with respect to the distribution of the random codebook,

$$\mathbb{E}_{\mathcal{X}}(\lambda_n^{(t)}(\mathcal{X})) \leq T \cdot 2^{-na}.$$

Additionally we define for any $t \in \theta$ an event

$$\iota_0(t) = \{\lambda_n^{(t)}(\mathcal{X}) \leq \sqrt{T} 2^{-n\frac{a}{2}}\}. \tag{2.25}$$

Then using the Markov inequality applied to $\lambda_n^{(t)}(\mathcal{X})$ along with (2.25), we obtain that

$$\Pr\{(\iota_0(t))^c\} \leq \sqrt{T} 2^{-n\frac{a}{2}}. \tag{2.26}$$

Set

$$\iota := \bigcap_{t\in\theta} \bigcap_{k=0}^{J_n} \iota_k(t). \tag{2.27}$$

Then with (2.16), (2.26), and applying the union bound we obtain

$$\begin{aligned} \Pr\{\iota^c\} &\leq \sum_{t\in\theta} \sum_{k=0}^{J_n} \Pr\{(\iota_k(t))^c\} \leq T \cdot \epsilon + T^{\frac{3}{2}} \cdot 2^{-n\frac{a}{2}} \\ &\leq T^2 \cdot 2^{-nc''} \end{aligned}$$

for a suitable positive constant $c'' > 0$ and all sufficiently large $n \in \mathbb{N}$.
Hence, we have shown that for each $t \in \theta$ there exist realisations $\{(x_{jl}^{(t)})_{j\in[J_n], l\in[L_{n,t}]} : t \in \theta\} \in \iota$ of $\mathcal{X}$. Now, denoting by $\| \cdot \|$ the variational distance

$$||p - q|| := \sum_{x\in A} |p(x) - q(x)|$$

for $p, q \in A$, we show that the secrecy level is fulfilled uniformly in $t \in \theta$ for any particular $\{(x_{jl}^{(t)})_{j\in[J_n],l\in[L_{n,t}]} : t \in \theta\} \in \iota$.

$$\begin{aligned} &\left\| \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} V_t^n(\cdot|x_{jl}^{(t)}) - \Theta_t(\cdot) \right\| \\ &\leq \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} \left\| V_t^n(\cdot|x_{jl}^{(t)}) - \tilde{Q}_{t,x_{jl}^{(t)}}(\cdot) \right\| + \left\| \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} \left(\tilde{Q}_{t,x_{jl}^{(t)}}(\cdot) - Q_{t,x_{jl}^{(t)}}(\cdot) \right) \right\| + \qquad (2.28) \\ &\qquad + \left\| \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} Q_{t,x_{jl}^{(t)}}(\cdot) - \Theta_t(\cdot) \right\| \leq 5\epsilon. \end{aligned}$$

In the first term the functions $V_t^n(\cdot|x_{jl}^{(t)})$ and $\tilde{Q}_{t,x_{jl}^{(t)}}(\cdot)$ differ if $z^n \notin \mathcal{T}^n_{p_t V_t, 2|A|\delta}$, so it makes a contribution of ϵ to the bound. In the second term $\tilde{Q}_t$ and Q_t are different for $z^n \notin S$ and because $\iota_j(t)$ and $\sum_{z^n \in S} \Theta_t(z^n) \geq 1 - 2\epsilon$ imply that

$$\frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} \sum_{z^n \in S} Q_{t,x_{jl}^{(t)}}(z^n) \geq 1 - 3\epsilon,$$

the second term is bounded by 3ϵ. The third term is bounded by ϵ which follows directly from (2.14).

For any $\{(x_{jl}^{(t)})_{j\in[J_n],l\in[L_{n,t}]} : t \in \theta\} \in \iota$ with the corresponding decoding sets $\{D_j : j \in [J_n]\}$ it follows by construction that

$$\frac{1}{J_n} \sum_{j\in[J_n]} \frac{1}{L_{n,t}} \sum_{l\in[L_{n,t}]} W_t^{\otimes n}(D_j^c|x_{jl}^{(t)}) \leq \sqrt{T} \cdot 2^{-na'} \qquad (2.29)$$

is fulfilled for all $t \in \theta$ with $a' > 0$, which means that we have found a (n, J_n) code with average error probability tending to zero for $n \in \mathbb{N}$ sufficiently large for any channel realisation. Now by a standard expurgation scheme we show that this still holds for the maximum error probability. We define the set

$$G_t := \{j \in J_n : \frac{1}{L_{n,t}} \sum_{l\in[L_{n,t}]} W_t^{\otimes n}(D_j^c|x_{jl}^{(t)}) \leq \sqrt{\eta}\} \qquad (2.30)$$

with $\eta := \sqrt{T} \cdot 2^{-na'}$ and denote its complement as $B_t := G_t^c$ and the union of all complements as

$$B = \bigcup_{t\in\theta} B_t = \bigcup_{t\in\theta} G_t^c.$$

Then (2.29) and (2.30) imply that

$$\eta \geq \frac{1}{J_n} \sum_{j \in [J_n]} \frac{1}{L_{n,t}} \sum_{l \in [L_{n,t}]} W_t^{\otimes n}(D_j^c | x_{jl}^{(t)}) \geq \frac{|B_t|}{J_n} \sqrt{\eta}$$

for all $t \in \theta$ and by the union bound it follows that

$$|B| \leq \sum_{t \in \theta} |B_t| \leq T \cdot \sqrt{\eta} \cdot J_n.$$

After removing all $j \in B$ (which are at most a fraction of $T^{\frac{5}{4}} 2^{-n\frac{a'}{2}}$ of J_n) and relabeling we obtain a new $(n, \tilde{J}_n)$ code $(E_j, D_j)_{j \in [\tilde{J}_n]}$ without changing the rate. The maximum error probability of the new code fulfills for sufficiently large $n \in \mathbb{N}$

$$\max_{t \in \theta} \max_{j \in [\tilde{J}_n]} \frac{1}{L_{n,t}} \sum_{l \in [L_{n,t}]} W_t^{\otimes n}(D_j^c | x_{jl}^{(t)}) \leq T^{\frac{1}{4}} \cdot 2^{-n\frac{a'}{2}}.$$

On the other hand, if we set

$$\hat{V}_t^n(z^n | (j,l)) := V_t^n(z^n | x_{jl}^{(t)}) \tag{2.31}$$

and further define

$$\begin{aligned} \hat{V}_{t,j}^n(z^n) &= \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} \hat{V}_t^n(z^n | (j,l)) \,, && (2.32) \\ \bar{V}_t^n(z^n) &= \frac{1}{\tilde{J}_n} \sum_{j=1}^{\tilde{J}_n} \hat{V}_{t,j}^n(z^n), && (2.33) \end{aligned}$$

we obtain for all $j \in [\tilde{J}_n]$, $t \in \theta$ and with $\epsilon = 2^{-nc'\delta^2}$

$$\|\hat{V}_{t,j}^n - \bar{V}_t^n\| \leq \|\hat{V}_{t,j}^n - \Theta_t\| + \|\Theta_t - \bar{V}_t^n\| \leq 10\epsilon,$$

where we have used the convexity of the variational distance and (2.28) which still applies by our expurgation procedure. For a uniformly distributed random variable J taking values in the set $\{1, \dots, \tilde{J}_n\}$ we obtain with Lemma 2.7 of [CK81] (uniform continuity of the entropy function)

$$\begin{aligned} I(J; Z_t^n) &= \sum_{j=1}^{J_n} \frac{1}{\tilde{J}_n} (H(\bar{V}_t^n) - H(\hat{V}_{t,j}^n)) = H(Z_t^n) - H(Z_t^n | J) \\ &\leq -10\epsilon \log(10\epsilon) + 10n\epsilon \log |C| \end{aligned}$$

uniformly in $t \in \theta$ (for $10\epsilon \leq e^{-1}$). Hence the strong secrecy level of the definition 2.1 holds uniformly in $t \in \theta$. Using standard arguments (cf. [CK81] page 409) we then have shown the achievability of the secrecy rate

$$R_S = \min_{t\in\theta} \max_{U\to X\to(YZ)_t} (I(U, Y_t) - I(U, Z_t)). \tag{2.34}$$

□

Remark. Note that in the case that $\mathfrak{W} := \{W_t, V_s : t = 1, \dots T, s = 1, \dots S\}$ with $S \neq T$ and the pair (s, t) known to the transmitter prior to transmission nothing new happens. A slight modification of the arguments presented above shows that

$$C_{S,CSI}(\mathfrak{W}) = \min_{(t,s)} \max_{U\to X\to(Y_t Z_s)} (I(U, Y_t) - I(U, Z_s)).$$

2.3.2 No CSI

In the previous section we have assumed that the channel state is known to the transmitter. We now consider the case where neither the transmitter nor the receiver has knowledge of the channel state. Thus both the encoder and the decoding sets must be chosen independent of the channel realisation. We will prove that

Theorem 2.4. *For the secrecy capacity $C_S(\mathfrak{W})$ of the compound wiretap channel $\mathfrak{W}$ without CSI it holds that*

$$C_S(\mathfrak{W}) \geq \max_{p\in\mathcal{P}(A)} (\min_{t\in\theta} I(p, W_t) - \max_{t\in\theta} I(p, V_t)).$$

Proof. Caused by the lack of channel knowledge we use a stochastic encoder independent of the channel realisation. For any $p \in \mathcal{P}(A)$ let $p' \in \mathcal{P}(A^n)$ be the distribution given by

$$p'(x^n) := \begin{cases} \frac{p^{\otimes n}(x^n)}{p^{\otimes n}(\mathcal{T}^n_{p,\delta})} & \text{if } x^n \in \mathcal{T}^n_{p,\delta}, \\ 0 & \text{otherwise.} \end{cases}$$

Then analogously to the case with CSI we define $\tilde{Q}_{t,x^n}(z^n)$, $Q_{t,x^n}(z^n)$, and $\Theta'_t(z^n)$, $\Theta_t(z^n)$ for $z_n \in C^n$ but now with respect to the distribution p' (cf. (2.9) and (2.10)). Consequently, $\Theta'(\cdot)$ has support only on $\mathcal{T}^n_{pV_t,2|A|\delta}$, and $Q_{t,x^n}(\cdot)$ and $\Theta(\cdot)$ only on the set $S := \{z^n \in C^n : \Theta'_t(z^n) \geq \epsilon\alpha_t\}$, where now $\alpha_t := 2^{-n(H(pV_t)+f_1(\delta))}$. Furthermore $\Theta(z^n) \geq \epsilon\alpha_t$ for all $z^n \in S$. Now define $J_n \cdot L_n$ i.i.d random variables X_{jl} according to the distribution p' independent of $t \in \theta$ with $j \in [J_n]$ and $l \in [L_n]$ with

$$\begin{aligned} J_n &= \lfloor 2^{n[\min_t I(p,W_t) - \max_t I(p,V_t) - \tau]} \rfloor && (2.35) \\ L_n &= \lfloor 2^{n[\max_t I(p,V_t) + \frac{\tau}{4}]} \rfloor && (2.36) \end{aligned}$$

for $\tau > 0$. Now because $\Theta_t(z^n) = \mathbb{E}Q_{t,X_{jl}} \geq \epsilon\alpha_t$ for all $z^n \in S$ we define the event $\iota_j(t)$ as in (2.14) for the random variables $\beta_t^{-1}Q_{t,X_{jl}}$ with $\beta_t := 2^{-n(H(V_t|)-f_2(\delta))}$

$$\iota_j(t) = \bigcap_{z^n \in C^n} \left\{ \frac{1}{L_n} \sum_{l=1}^{L_n} Q_{t,X_{jl}}(z^n) \in [(1 \pm \epsilon)\Theta_t(z^n)] \right\},$$

but considering the difference that the random variables X_{jl} are independent of the channel state. Then analogously to (2.15) we obtain that

$$\Pr\{(\iota_j(t))^c\} \leq 2|C|^n \exp\Big(- L_n \frac{2^{-n(I(p,V_t)+g(\delta))}}{3} \Big)$$

by Lemma 2.3 and Lemma A.9. Notice that, because the sender does not know which channel is used, we need the maximum in the definition of L_n. Thus the right-hand side is a double exponential in n and can be made smaller than ϵJ_n^{-1} for all j and for all $t \in \theta$ and sufficiently large n.

Now let J_n and L_n be defined as stated above, and let $X^n = \{X_{jl}\}_{j\in[J_n], l\in[L_n]}$ be the set of i.i.d. random variables each of them distributed according to p' independent of $t \in \theta$. As in the case of CSI we can show that reliable transmission of the message $j \in [J_n]$ can be achieved. To this end define now the random decoder $\{D_j(X^n)\}_{j\in[J_n]} \subseteq B^n$ as in (2.17) but with

$$D'_j(X^n) := \bigcup_{s\in\theta} \bigcup_{k\in[L_n]} \mathcal{T}^n_{W_s,\delta}(X_{jk}),$$

and thus

$$D_j(X^n) := D'_j(X^n) \cap \Bigg(\bigcup_{\substack{j'\in[J_n] \\ j'\neq j}} D'_{j'}(X^n) \Bigg)^c. \tag{2.37}$$

changes only in the indepency of the random codewords from the channel realisation. Then we can define the random average probability of error for a specific channel $\lambda_n^{(t)}(X^n)$ analogously to (2.18) as

$$\lambda_n^{(t)}(X^n) := \frac{1}{J_n} \sum_{j\in[J_n]} \frac{1}{L_n} \sum_{l\in[L_n]} W_t^{\otimes n}((D_j(X^n))^c | X_{jl}). \tag{2.38}$$

Notice that now both X^n and L_n do not depend on $t \in \theta$ and this holds throughout the entire proof. Then we can give the bound in (2.19) now by

$$\begin{aligned} &W_t^{\otimes n}((D_j(X^n))^c|X_{jl}) \\ &\qquad \leq W_t^{\otimes n}((\mathcal{T}^{\otimes n}_{W_t,\delta}(X_{jl}))^c|X_{jl}) + \sum_{\substack{j' \in [J_n] \\ j' \neq j}} \sum_{s \in \theta} \sum_{k \in [L_n]} W_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k})|X_{jl}). \end{aligned}$$

We can bound the first term in the inequality by $\nu_n(\delta) := (n+1)^{|A||B|} \cdot 2^{-nc\delta^2}$ (see (2.20)). If we average over all codebooks we get

$$\begin{aligned} &\mathbb{E}_{X^n}(W_t^{\otimes n}((D_j(X^n))^c|X_{jl})) \\ &\qquad \leq \nu_n(\delta) + \sum_{\substack{j' \in [J_n] \\ j' \neq j}} \sum_{s \in \theta} \sum_{k \in [L_n]} \mathbb{E}_{X_{j'k}} \mathbb{E}_{X_{jl}} W_t^{\otimes n}(\theta^n_{W_s,\delta}(X_{j'k})|X_{jl}). \end{aligned} \tag{2.39}$$

Because $X_{j'k} \in \mathcal{T}^n_{p,\delta}$ with probability 1 by its definition for all $j' \in [J_n]$, $k \in [L_n]$, by the same reasoning as in (2.21) and (2.22) we can give an upper bound on

$$\begin{aligned} \mathbb{E}_{X_{jl}} W_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k})|X_{jl}) &\leq \frac{q_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k}))}{p^{\otimes n}(\mathcal{T}^n_{p,\delta})} \\ &\leq \frac{(n+1)^{|A||B|}}{1-(n+1)^{|A|} \cdot 2^{-nc\delta^2}} \cdot 2^{-n(I(p,W_s)-f(\delta))}, \end{aligned}$$

in which $q_t^{\otimes n}$ denotes the output distribution generated by the conditional distribution $W_t^{\otimes n}$ and the input distribution $p^{\otimes n}$. Thus we can upper bound the expectation in the second term of the RHS of (2.39)

$$\mathbb{E}_{X_{j'k}} \mathbb{E}_{X_{jl}} W_t^{\otimes n}(\mathcal{T}^n_{W_s,\delta}(X_{j'k})|X_{jl}) \leq \frac{(n+1)^{|A||B|}}{1-(n+1)^{|A|} \cdot 2^{-nc\delta^2}} \cdot 2^{-n(I(p,W_s)-f(\delta))}$$

for all $t \in \theta$, all $j' \neq j$ and all $k, l \in [L_n]$ with a universal $f(\delta) > 0$ satisfying $\lim_{\delta \to 0} f(\delta) = 0$. Additionally we define $\mu_n(\delta) := 1-(n+1)^{|A|} \cdot 2^{-nc\delta^2}$. Then (2.39) changes to

$$\begin{aligned} \mathbb{E}_{X^n}(W_t^{\otimes n}((D_j(X^n))^c|X_{jl})) &\leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} J_n \sum_{s \in \theta} L_n \cdot 2^{-n(I(p,W_s)-f(\delta))} \\ &\leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} T \cdot J_n L_n \cdot 2^{-n(\min_s I(p,W_s)-f(\delta))} \\ &\leq \nu_n(\delta) + T \cdot 2^{-n\frac{\tau}{2}} \end{aligned}$$

by the definition of J_n and L_n in (2.35), (2.36) and by choosing $\delta > 0$ small enough that $\tau - \frac{\tau}{4} - f(\delta) \geq \frac{\tau}{2}$. Now by defining $a = a(\delta,\tau) := \frac{\min\{c\delta^2, \frac{\tau}{2}\}}{2}$ and the definition of the error probability there exist a $n(\delta,\tau,|A|,|B|) \in \mathbb{N}$, such that the last inequality results in the upper bound

$$\mathbb{E}_{X^n}(\lambda_n^{(t)}(X^n)) \leq T \cdot 2^{-na}$$

for any $t \in \theta$ and $n > n(\delta,\tau,|A|,|B|) \in \mathbb{N}$.
Now we define the event $\iota_0(t)$ for any $t \in \theta$ and the event ι as in (2.25) and (2.27) but with the difference that the input is independent of the channel realisation. So by the same reasoning we end in

$$\Pr\{\iota^c\} \leq T^2 \cdot 2^{-nc''} \tag{2.40}$$

for a constant $c'' > 0$ and all sufficiently large $n \in \mathbb{N}$, which implies that there exist realisations $\{x_{jl}\}$ of $\{X_{jl}\}$ such that $x_{jl} \in \iota$ for all $j \in [J_n]$ and $l \in [L_n]$. Then analogously to (2.28) we get for any channel $t \in \theta$

$$\left\| \frac{1}{L_n} \sum_{l=1}^{L_n} V_t^n(\cdot|x_{jl}) - \Theta_t(\cdot) \right\| \leq 5\epsilon$$

differs from the former only by L_n in place of $L_{n,t}$. Hence, following the same arguments subsequent to (2.29), we have shown that there is a sequence of $(n, \tilde{J}_n)$ codes for which

$$\max_{t\in\theta} \max_{j\in[\tilde{J}_n]} \frac{1}{L_n} \sum_{l\in[L_n]} W_t^{\otimes n}(D_j^c|x_{jl}) \leq T^{\frac{1}{4}} \cdot 2^{-n\frac{a'}{2}}$$

holds for sufficiently large $n \in \mathbb{N}$, and the strong secrecy level is fulfilled for every channel $t \in \theta$ by

$$\|\hat{V}_{t,j}^n - \bar{V}_t^n\| \leq 10\epsilon$$

($\hat{V}_{t,j}^n, \bar{V}_t^n$ defined as in (2.32), (2.33)) and thus by

$$I(J; Z_t^n) \leq -10\epsilon \log(10\epsilon) + 10n\epsilon \log|C|$$

which tends to zero for $n \to \infty$ uniformly in $t \in \theta$. □

We turn now to the converse of Theorem 2.4. Actually, we give only a multiletter formula of the upper bound of the secrecy rates. First we need the following lemma.

Lemma 2.5. *Let $\mathfrak{W} = \{(W_t, V_t) : t \in \theta\}$ be an arbitrary compound wiretap channel* *without CSI. Then*

$$\lim_{n\to\infty}\frac{1}{n}\max_{U\to X^n\to Y_t^n Z_t^n}(\inf_{t\in\theta} I(U;Y_t^n)-\sup_{t\in\theta} I(U;Z_t^n))$$

exists and we have

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{n}\max_{U\to X^n\to Y_t^n Z_t^n}(\inf_{t\in\theta} I(U;Y_t^n)-\sup_{t\in\theta} I(U;Z_t^n))\\ &=\sup_{n\in\mathbb{N}}\frac{1}{n}\max_{U\to X^n\to Y_t^n Z_t^n}(\inf_{t\in\theta} I(U;Y_t^n)-\sup_{t\in\theta} I(U;Z_t^n)).\end{aligned}$$

Proof. The proof is based on Fekete's lemma [Fek23]. Consequently, if we apply the lemma to the sequence $(a_n)_{n\in\mathbb{N}}$ defined by

$$a_n := \max_{U\to X^n\to Y_t^n Z_t^n}(\inf_{t\in\theta} I(U;Y_t^n)-\sup_{t\in\theta} I(U;Z_t^n))$$

it suffices to show that the inequality

$$a_{n+m}\geq a_n+a_m$$

holds for all $n,m\in\mathbb{N}$. This will be done by considering two independent Markov chains $U_1\to X^n\to(Y_t^n,Z_t^n)$ and $U_2\to\hat{X}^m\to(\hat{Y}_t^m,\hat{Z}_t^m)$ and setting $U:=(U_1,U_2)$, $X^{n+m}:=(X^n,\hat{X}^m)$ and $(Y_t^{n+m},Z_t^{n+m}):=((Y_t^n,\hat{Y}_t^m),(Z_t^n,\hat{Z}_t^m))$. By definition

$$\begin{aligned}a_{n+m}&\geq\inf_{t\in\theta} I(U;Y_t^{n+m})-\sup_{t\in\theta} I(U;Z_t^{n+m})\\ &\geq\inf_{t\in\theta} I(U_1;Y_t^n)+\inf_{t\in\theta} I(U_2;\hat{Y}_t^m)-\sup_{t\in\theta} I(U_1;Z_t^n)-\sup_{t\in\theta} I(U_2;\hat{Z}_t^m).\end{aligned}$$

By the independence of the two Markov chains mentioned above and because apart from that these Markov chains were arbitrary we can conclude that

$$a_{n+m}\geq a_n+a_m$$

holds for all $n,m\in\mathbb{N}$. □

Then we can formulate the following

Proposition 2.6. *The secrecy capacity of the compound wiretap channel in the case of no CSI $C_S(\mathfrak{W})$ is upper bounded by*

$$C_S(\mathfrak{W})\leq\lim_{n\to\infty}\frac{1}{n}\max_{U\to X^n\to Y_t^n Z_t^n}(\inf_{t\in\theta} I(U;Y_t^n)-\sup_{t\in\theta} I(U;Z_t^n)).$$

Proof. Let $(\mathcal{C}_n)_{n\in\mathbb{N}}$ be any sequence of (n, J_n) codes such that with

$$\sup_{t\in\theta}\frac{1}{J_n}\sum_{j=1}^{J_n}\sum_{x^n\in A^n}E(x^n|j)W_t^{\otimes n}(D_j^c|x^n) =: \varepsilon_{1,n}, \tag{2.41}$$

and

$$\sup_{t\in\theta} I(J; Z_t^n) =: \varepsilon_{2,n}$$

it holds that $\lim_{n\to\infty}\varepsilon_{1,n} = 0$ and $\lim_{n\to\infty}\varepsilon_{2,n} = 0$, where J denotes the random variable which is uniformly distributed on the message set $\{1,\ldots,J_n\}$. Let us denote by $\hat{J}$ the random variable with values in $\{1,\ldots,J_n\}$ determined by the Markov chain $J\to X^n\to Y_t^n\to\hat{J}$ where the first transition is governed by E, the second by $W_t^{\otimes n}$, and the last by the decoding rule. Then we have for any $t\in\theta$

$$\begin{aligned}\log J_n &= H(J) = I(J;\hat{J}) + H(J|\hat{J})\\ &\leq I(J;Y_t^n) + H(J|\hat{J}),\end{aligned} \tag{2.42}$$

where the inequality follows from the data processing inequality. Then using Fano's inequality we find that

$$H(J|\hat{J}) \leq 1 + \varepsilon_{1,n}\log J_n$$

with (2.41). Thus we can rewrite inequality (2.42) as

$$(1-\varepsilon_{1,n})\log J_n \leq I(J;Y_t^n) + 1$$

for all $t\in\theta$. On the other hand we have for every $t\in\theta$

$$I(J;Y_t^n) = I(J;Y_t^n) - \sup_{t\in\theta} I(J;Z_t^n) + \varepsilon_{2,n}$$

where we have used the validity of the secrecy criterion stated above. Then the last two inequalities imply that for any $t\in\theta$

$$(1-\varepsilon_{1,n})\log J_n \leq I(J;Y_t^n) - \sup_{t\in\theta} I(J;Z_t^n) + \varepsilon_{2,n}. \tag{2.43}$$

Since the LHS of (2.43) does not depend on t we arrive at

$$(1-\varepsilon_{1,n})\log J_n \leq \max_{U\to X^n\to Y_t^n Z_t^n}(\inf_{t\in\theta} I(U;Y_t^n) - \sup_{t\in\theta} I(U;Z_t^n)) + \varepsilon_{2,n},$$

which concludes the proof after dividing by $n\in\mathbb{N}$, taking $\limsup$ and taking into account the assertion of Lemma 2.5. □

Remark. Following the same arguments subsequent to (2.34) concerning the use of the channels defined by $P_{Y_t|T} = W_t \cdot P_{X|T}$ and $P_{Z_t|T} = V_t \cdot P_{X|T}$ instead of W_t and V_t and applying the assertion of Theorem 2.4 to the n-fold product of channels W_t and V_t, we can give the coding theorem for the multiletter case. The capacity of the compound wiretap channel in the case of no CSI is

$$C_S(\mathfrak{W}) = \lim_{n\to\infty} \frac{1}{n} \max_{U\to X^n\to Y_t^n Z_t^n} (\inf_{t\in\theta} I(U;Y_t^n) - \sup_{t\in\theta} I(U;Z_t^n)).$$

Let us consider now the case $\mathfrak{W} := \{W_t, V_s : t = 1, \dots T, s = 1, \dots S\}$ with $S \neq T$ and the pair (s,t) unknown to both the transmitter and the legitimate receiver. Additionally we assume that each V_s is a degraded version of every W_t, which is characterised by

$$V_s(z|x) = \sum_{y\in B} W_t(y|x) D_{(t,s)}(z|y), \tag{2.44}$$

for all $x \in A$, $z \in C$, if D is defined as the stochastic matrix $D : B \to \mathcal{P}(C)$. Then we have the following

Lemma 2.7. *Let $p \in \mathcal{P}(A)$, $W : A \to \mathcal{P}(B)$, $V : A \to \mathcal{P}(C)$, and assume that V is a degraded version of W. Then $I(X;Y|Z)$ is a concave with respect to the input distribution $p_X = p$.*

Proof. Let X, Y, Z be random variables with values in A, B, C respectively distributed according to

$$\Pr(X = x, Y = y, Z = z) := p_{XYZ}(x,y,z) = p(x)W(y|x)D(z|y) \tag{2.45}$$

for all $x \in A, y \in B, z \in Z$. Because

$$I(X;Y|Z) = H(Y|Z) - H(Y|X,Z)$$

the proof is based on the two assertions

1. $H(Y|Z)$ depends concavely on p_X, and

2. $H(Y|X,Z)$ is an affine function of p_X.

First, $H(Y|Z)$ is a concave function with respect to p_{YZ} by the log-sum inequality (cf. [CK81] Lemma 3.1). Then because p_{XYZ} depends affinely on p_X by (2.45), so does p_{YZ}, and the first assertion follows. For the second consider that (2.44) and (2.45) imply that

$$p_{Y|X,Z}(y|x,z) = \frac{W(y|x)D(z|y)}{V(z|x)}$$

for every input distribution p_X, any $y \in B$ and all $x \in A, z \in C$ with $p_{XZ}(x,z) > 0$. Then we have

$$H(Y|X,Z) = \sum_{x \in A, z \in C} p_{XZ}(x,z) H\left(\frac{W(\cdot|x)D(z|\cdot)}{V(z|x)}\right)$$

showing that $H(Y|X,Z)$ is an affine function of p_{XZ} which in turn depends affinely on p_X. □

Now because the random variables X, Y_t, Z_s (Y_t, Z_s the channel outputs of W_t and V_s resp.) form a Markov chain for all $t \in \theta$ and $s \in \mathcal{S}$, we obtain that

$$I(X; Y_t|Z_s) = I(X, Y_t) - I(X, Z_s). \tag{2.46}$$

By virtue of Theorem 2 of [AC93] we can show that for the secrecy rate it holds that

$$R_S \leq \frac{1}{n} \sum_{i=1}^{n} I(X_i; Y_{i,t}|Z_{i,s}) + \epsilon'$$

for any channel $(t,s) \in \theta \times \mathcal{S}$ and $\epsilon' > 0$. The concavity of $I(X; Y_t|Z_s)$ with respect to the input distributions $p \in \mathcal{P}(A)$ together with (2.46) then imply the converse part of Theorem 2.4, that

$$R_S \leq \max_{p \in \mathcal{P}(A)} \min_{(t,s)} (I(p, W_t) - I(p, V_s)).$$

Now we can state the following

Proposition 2.8. *If V_s is a degraded version of W_t for all $s \in \mathcal{S}$ and $t \in \theta$ the capacity of the compound wiretap channel is given by*

$$\begin{aligned} C_S(\mathfrak{W}) &= \max_{p \in \mathcal{P}(A)} \min_{(t,s)} (I(p, W_t) - I(p, V_s)) \\ &= \max_{p \in \mathcal{P}(A)} (\min_t I(p, W_t) - \max_s I(p, V_s)). \end{aligned}$$

Remark. This result was obtained in [LKPS08] with a weaker notion of secrecy.

2.3.3 Channel State to the Legitimate Receiver Is Known at the Transmitter

We now consider the case, in which the transmitter has knowledge of the channel state to the legitimate receiver $t \in \theta$, but the channel state to the eavesdropper $s \in \mathcal{S}$ is unknown. We will denote this kind of channel state information by CSI$_t$. Consequently we get for each $t \in \theta$ possible channel realisations $\mathfrak{W}_t := \{(W_t, V_s) : s = 1, \dots S\}$. Then we can describe the compound channel as $\mathfrak{W} = \cup_{t\in\theta}\mathfrak{W}_t$.

Theorem 2.9. *For the secrecy capacity $C_{S,CSI_t}(\mathfrak{W})$ of the compound wiretap channel with CSI_t it holds that*

$$C_{S,CSI_t}(\mathfrak{W}) \geq \min_{t\in\theta} \max_{p\in\mathcal{P}(A)} (I(p, W_t) - \max_{s\in\mathcal{S}} I(p, V_s)).$$

Proof. Adapted to the channel realisation W_t define

$$p'_t(x^n) := \begin{cases} \frac{p_t^{\otimes n}(x^n)}{p_t^{\otimes n}(\mathcal{T}^n_{p_t,\delta})} & \text{if } x^n \in \mathcal{T}^n_{p_t,\delta}, \\ 0 & \text{otherwise.} \end{cases} \tag{2.47}$$

for arbitrary input distributions $p_1, \dots, p_T \in \mathcal{P}(A)$. Now define for $z^n \in C^n$ and $s \ \in \mathcal{S}$

$$\tilde{Q}_{s,x^n}(z^n) = V_s^n(z^n|x^n) \cdot \mathbf{1}_{\mathcal{T}^n_{V_s,\delta}(x^n)}(z^n)$$

on C^n. Additionally, we set for $z^n \in C^n$

$$\Theta'_s(z^n) = \sum_{x^n\in\mathcal{T}^n_{p_t,\delta}} p'_t(x^n)\tilde{Q}_{s,x^n}(z^n).$$

Now let $S := \{z^n \in C^n : \Theta'_s(z^n) \geq \epsilon\alpha_{t,s}\}$ where $\epsilon = 2^{-nc'\delta^2}$ and $\alpha_{t,s}$ is from (A.28) similar to the former cases but computed with respect to p_t and V_s. Then the support of Θ'_s has cardinality $\leq \alpha_{t,s}^{-1}$, which implies that $\sum_{z^n\in S}\Theta_s(z^n) \geq 1 - 2\epsilon$. Analogously to (2.10) define $\Theta_s(z^n)$ and $Q_{s,x^n}(z^n)$ with support on S and further

$$J_n = \lfloor 2^{n[\min_t(I(p_t,W_t) - \max_s I(p_t,V_s)) - \tau]} \rfloor \tag{2.48}$$

$$L_{n,t} = \lfloor 2^{n[\max_s I(p_t,V_s) + \frac{\tau}{4}]} \rfloor. \tag{2.49}$$

As in the case of CSI define random matrices $\{X^{(t)}_{jl}\}_{j\in[J_n],l\in[L_{n,t}]}$ such that the random variables $X^{(t)}_{jl}$ where i.i.d. according to p'_t. We suppose additionally that $\{X^{(t)}_{jl}\}_{j,l}$ and $\{X^{(t')}_{jl}\}_{j,l}$ are independent for $t \neq t'$. For any $z^n \in S$ it follows that

$\Theta_s(z^n) = \mathbb{E}Q_{s,X^{(t)}_{jl}}(z^n) \geq \epsilon\alpha_{t,s}$, if $\mathbb{E}$ is the expectation value with respect to the distribution p'_t. For the random variables $\beta^{-1}_{t,s}Q_{s,X^{(t)}_{jl}}(z^n)$ define the event

$$\iota_j(s,t) = \bigcap_{z^n \in C^n} \left\{ \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} Q_{s,X^{(t)}_{jl}}(z^n) \in [(1 \pm \epsilon)\Theta_s(z^n)] \right\}.$$

Then it follows that for all $j \in [J_n]$ and for all $s \in \mathcal{S}$ it holds for each $t \in \theta$

$$\Pr\{(\iota_j(s,t))^c\} \leq 2|C|^n \exp\Big(- L_{n,t} \frac{2^{-n[I(p_t,V_s)+g(\delta)]}}{3} \Big)$$

by Lemma 2.3, Lemma A.9, Thus the RHS is double exponential in n uniformly in $s \in \mathcal{S}, t \in \theta$ (guaranteed by the maximum in s in the definition of $L_{n,t}$) and can be made smaller than ϵJ_n^{-1} for all $j \in [J_n]$ and all sufficiently large n. Now the coding part of the problem is similar to the case with CSI. Let $p'_t \in \mathcal{P}(A^n)$ be given as in (2.47). We abbreviate $\mathcal{X} := \{X^{(t)}\}_{t\in\theta}$ for the family of random matrices $X^{(t)} = \{X^{(t)}_{jl}\}_{j\in[J_n],l\in[L_{n,t}]}$ whose components are i.i.d. according to p'_t. We will show how reliable transmission of the message $j \in [J_n]$ can be achieved. To this end define now the random decoder $\{D_j(\mathcal{X})\}_{j\in[J_n]} \subseteq B^n$ as in (2.17) and with

$$D'_j(\mathcal{X}) := \bigcup_{r\in\theta} \bigcup_{k\in[L_{n,r}]} \mathcal{T}^n_{W_r,\delta}(X^{(r)}_{jk}),$$

and the random average probabilities of error for a specific channel $\lambda^{(t)}_n(\mathcal{X})$ as in (2.18) by

$$\lambda^{(t)}_n(\mathcal{X}) := \frac{1}{J_n} \sum_{j\in[J_n]} \frac{1}{L_{n,t}} \sum_{l\in[L_{n,t}]} W^{\otimes n}_t((D_j(\mathcal{X}))^c|X^{(t)}_{jl}).$$

As in (2.19) we get for each $t \in \theta$ and $l \in [L_{n,t}]$

$$\begin{aligned} &W^{\otimes n}_t((D_j(\mathcal{X}))^c|X^{(t)}_{jl}) \\ &\quad \leq W^{\otimes n}_t((\mathcal{T}^{\otimes n}_{W_t,\delta}(X^{(t)}_{jl}))^c|X^{(t)}_{jl}) + \sum_{\substack{j'\in[J_n]\\ j'\neq j}} \sum_{r\in\theta} \sum_{k\in[L_{n,r}]} W^{\otimes n}_t(\mathcal{T}^n_{W_r,\delta}(X^{(r)}_{j'k})|X^{(t)}_{jl}), \end{aligned}$$

Then by Lemma A.8 we can bound the first term of the right hand side, such that together with the independence of all involved random variables we end up with

$$\begin{aligned} &\mathbb{E}_{\mathcal{X}}(W^{\otimes n}_t((D_j(\mathcal{X}))^c|X^{(t)}_{jl})) \\ &\quad \leq (n+1)^{|A||B|} \cdot 2^{-nc\delta^2} + \sum_{\substack{j'\in[J_n]\\ j'\neq j}} \sum_{r\in\theta} \sum_{k\in[L_{n,r}]} \mathbb{E}_{X^{(r)}_{j'k}} \mathbb{E}_{X^{(t)}_{jl}} W^{\otimes n}_t(\mathcal{T}^n_{W_r,\delta}(X^{(r)}_{j'k})|X^{(t)}_{jl}). \end{aligned}$$

We shall find now for $j' \neq j$ by the same reasoning as in (2.21) and (2.22) an upper bound on

$$\begin{aligned}\mathbb{E}_{X_{jl}^{(t)}} W_t^{\otimes n}(\mathcal{T}_{W_r,\delta}^n(X_{j'k}^{(r)})|X_{jl}^{(t)}) &\leq \frac{q_t^{\otimes n}(\mathcal{T}_{W_r,\delta}^n(X_{j'k}^{(r)}))}{p_t^{\otimes n}(\mathcal{T}_{p_t,\delta}^n)} \\ &\leq \frac{(n+1)^{|A||B|}}{1-(n+1)^{|A|}\cdot 2^{-nc\delta^2}} \cdot 2^{-n(I(p_r,W_r)-f(\delta))}\end{aligned}$$

for all $r, t \in \theta$, all $j' \neq j$, and all $l \in [L_{n,t}], k \in [L_{n,r}]$. Now by defining $\nu_n(\delta) := (n+1)^{|A||B|} \cdot 2^{-nc\delta^2}$ and $\mu_n(\delta) := 1-(n+1)^{|A|} \cdot 2^{-nc\delta^2}$ thus for each $t \in \theta$, $l \in [L_{n,t}]$, and $j \in [J_n]$ the last inequality leads to

$$\begin{aligned}&\mathbb{E}_{\mathcal{X}}(W_t^{\otimes n}((D_j(\mathcal{X}))^c|X_{jl}^{(t)})) \\ &\leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} J_n \sum_{r\in\theta} L_{n,r} 2^{-n(I(p_r,W_r)-f(\delta))} \\ &\leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} T J_n \cdot 2^{-n(\min_{r\in\theta}(I(p_r,W_r)-\max_s I(p_r,V_s))-f(\delta)-\frac{\tau}{4})} \\ &\leq \nu_n(\delta) + \frac{(n+1)^{|A||B|}}{\mu_n(\delta)} T \cdot 2^{-n\frac{\tau}{2}}\end{aligned}$$

where we have used the definitions of J_n and $L_{n,r}$ in (2.48), (2.49) and we have chosen $\delta > 0$ small enough to ensure that $\tau - f(\delta) - \frac{\tau}{4} \geq \frac{\tau}{2}$. Defining $a = a(\delta, \tau) := \frac{\min\{c\delta^2, \frac{\tau}{2}\}}{2}$ we can find $n(\delta, \tau, |A|, |B|) \in \mathbb{N}$ such that for all $n \geq n(\delta, \tau, |A|, |B|)$ we end in

$$\mathbb{E}_{\mathcal{X}}(\lambda_n^{(t)}(\mathcal{X})) \leq T \cdot 2^{-na}.$$

for any $t \in \theta$. To give a bound on the average probability of error we define the event $\iota_0(t)$ for any $t \in \theta$ as in (2.25) and the event

$$\iota := \bigcap_{t\in\theta} \bigcap_{s\in\mathcal{S}} \bigcap_{k=0}^{J_n} \iota_k(t,s)$$

differs from (2.27) only by the intersection of the unknown channel states $s \in \mathcal{S}$. Thus we can conclude that

$$\begin{aligned}\Pr\{\iota^c\} &\leq S \cdot T \cdot \epsilon + S \cdot T^{\frac{3}{2}} \cdot 2^{-n\frac{a}{2}} \\ &\leq S \cdot T^2 \cdot 2^{-nc''}\end{aligned}$$

holds for a suitable positive constant $c'' > 0$ and all sufficiently large $n \in \mathbb{N}$, and we have shown that for each $t \in \theta$ there exist realisations $\{(x_{jl}^{(t)})_{j\in[J_n],l\in[L_{n,t}]} : t \in \theta\} \in \iota$

of $\mathcal{X}$. By the same reasoning as in (2.28) we get for any channel realisation $t \in \theta$ to the legitimate receiver

$$\left\| \frac{1}{L_{n,t}} \sum_{l=1}^{L_{n,t}} V_s^n(\cdot|x_{jl}^{(t)}) - \Theta_s(\cdot) \right\| \leq 5\epsilon$$

for each of the unknown channels $s \in \mathcal{S}$ to the eavesdropper. Now, because for any $t \in \theta$ we have a different codeword set $\{x_{jl}^{(t)}\}$, we slightly change the definition in (2.31) to

$$\hat{V}_{(s,t)}^n(z^n|(j,l)) := V_s^n(z^n|x_{jl}^{(t)})$$

and accordingly to $\hat{V}_{(s,t),j}^n$ and $\bar{V}_{(s,t)}^n$ in (2.32), (2.33) in that way, that these distributions are defined separately for each codeword set $t \in \theta$. Thus we get, that

$$\|\hat{V}_{(s,t),j}^n - \bar{V}_{(s,t)}^n\| \leq 10\epsilon$$

is fulfilled for all $s \in \mathcal{S}$ for each individual channel $t \in \theta$ to the legitimate receiver. Hence, using the same expurgation scheme as in the previous sections we have shown that there is a sequence of $(n, \tilde{J}_n)$ codes for which

$$\max_{t \in \theta} \max_{j \in [\tilde{J}_n]} \frac{1}{L_{n,t}} \sum_{l \in [L_{n,t}]} W_t^{\otimes n}(D_j^c|x_{jl}^{(t)}) \leq T^{\frac{1}{4}} \cdot 2^{-n\frac{a'}{2}}$$

holds for sufficiently large $n \in \mathbb{N}$, and the strong secrecy level is fulfilled for every channel $t \in \theta$ by

$$I(J; Z_s^n) \leq -10\epsilon \log(10\epsilon) + 10n\epsilon \log |C|$$

which tends to zero for $n \to \infty$ for all channels $s \in \mathcal{S}$ to the eavesdropper. Thus

$$R_S = \min_{t \in \theta} \max_{p \in \mathcal{P}(A)} (I(p, W_t) - \max_{s:(s,t) \in \mathcal{S} \times \theta} I(p, V_s))$$

is an achievable secrecy rate for the compound wiretap channel $\cup_{t \in \theta} \mathfrak{W}_t$ in the case where the channel state to the legitimate receiver is known at the transmitter. □

Remark. By considering the converse of Theorem 2.9, we get for each $t \in \theta$ possible channel realisations $\mathfrak{W}_t := \{(W_t, V_s) : s = 1, \dots S\}$. Then we can describe the compound channel as $\mathfrak{W} = \cup_{t \in \theta} \mathfrak{W}_t$. In accordance to the case of no CSI for each $t \in \theta$ we obtain that

$$C_S(\mathfrak{W}_t) = \lim_{n \to \infty} \frac{1}{n} \max_{U \to X^n \to Y_t^n Z_s^n} (I(U; Y_t^n) - \sup_{s \in \mathcal{S}} I(U; Z_s^n)).$$

Proposition 2.10. *The secrecy capacity of the compound wiretap channel in the case where only the channel state to the legitimate receiver is known at the transmitter* $C_{S,CSI_t}(\mathfrak{W})$ *is given by*

$$C_{S,CSI_t}(\mathfrak{W}) = \inf_{t\in\theta} C_S(\mathfrak{W}_\mathrm{t}).$$

Now, additionally let us assume that each V_s is a degraded version of every W_t for $s \in \mathcal{S}$ and $t \in \theta$. Then as shown in Lemma 2.7 $I(X;Y_t|Z_s)$ is a concave function with respect to the input distribution $p_X = p$. In particular this still holds for $\min_{s\in\mathcal{S}} I(X;Y_t|Z_s)$. Now because the random variables X, Y_t, Z_s form a Markov chain for all $t \in \theta$ and $s \in \mathcal{S}$ and

$$\min_{s\in\mathcal{S}} I(X;Y_t|Z_s) = I(X,Y_t) - \max_{s\in\mathcal{S}} I(X,Z_s),$$

for any $t \in \theta$ we get the upper bound on the secrecy rate as the secrecy capacity of a single channel W_t with S channels to the eavesdropper. Then we can conclude

Proposition 2.11. *The secrecy capacity of the channel where only the channel states to the legitimate receiver are known and the channels to the eavesdropper are degraded versions of those to the legitimate receiver is given by*

$$C_{S,CSI_t}(\mathfrak{W}) = \min_{t\in\theta} \max_{p\in\mathcal{P}(A)} (I(p,W_t) - \max_{s\in\mathcal{S}} I(p,V_s)).$$

2.3.4 Compound Wiretap Channel with $C_S = C_{S,CSI}$

Let $\mathfrak{W} := \{W_t, V_s : t = 1, \dots T, s = 1, \dots S\}$ with $S \neq T$ and the pair (t,s) unknown to both the transmitter and the legitimate receiver. In addition let us assume that

$$\exists\, \hat{t} \in \theta\, \forall\, t \in \theta\, \exists\, U_t : \quad W_{\hat{t}} = U_t W_t, \tag{2.50}$$

which means that $W_{\hat{t}}$ is a degraded version of all channel W_t with $t \neq \hat{t}$. We further assume that

$$\exists\, \hat{s} \in \mathcal{S}\, \forall\, s \in \mathcal{S}\, \exists\, \hat{U}_s : \quad V_s = \hat{U}_s V_{\hat{s}}, \tag{2.51}$$

which means that all V_s with $s \neq \hat{s}$ are degraded versions of $V_{\hat{s}}$. Then we can show that the capacity of this channel equals the capacity of the same channel with CSI at the transmitter, e.g.

$$C_S(\mathfrak{W}) = C_{S,CSI}(\mathfrak{W}).$$

First, by Theorem 2.4 it holds that

$$C_S(\mathfrak{W}) \geq \max_{M \to X \to (Y_t Z_s)} \min_{(t,s)} (I(M, Y_t) - I(M, Z_s)), \tag{2.52}$$

where M is an auxiliary random variable, such that $M, X, (Y_t, Z_s)$ form a Markov chain $M \to X \to (Y_t Z_s)$ in this order. Now let

$$p^*_{MX} = \arg \max_{M \to X \to (Y_{\hat{t}} Z_{\hat{s}})} (I(M, Y_{\hat{t}}) - I(M, Z_{\hat{s}}))$$

the joint distribution of M and X that achieves capacity for the single wiretap channel $(W_{\hat{t}}, V_{\hat{s}})$. Because the capacity of the compound wiretap channel $\mathfrak{W}$ is less than or equal the capacity of each single channel we obtain

$$\begin{aligned} C_{S,CSI}(\mathfrak{W}) &\leq I(p^*_M, W_{\hat{t}} \cdot P^*_{X|M}) - I(p^*_M, V_{\hat{s}} \cdot P^*_{X|M}) = C_S(W_{\hat{t}}, V_{\hat{s}}) \\ &\leq I(p^*_M, U_t(W_t \cdot P^*_{X|M})) - I(p^*_M, \hat{U}_s(V_{\hat{s}} \cdot P^*_{X|M})) \\ &\leq I(p^*_M, W_t \cdot P^*_{X|M}) - I(p^*_M, V_s \cdot P^*_{X|M}) \end{aligned} \tag{2.53}$$

for all $(s,t) \in \mathcal{S} \times \theta$ because of (2.50), (2.51). Then by the last inequality it follows that

$$\begin{aligned} I(p^*_M, W_{\hat{t}} \cdot P^*_{X|M}) - I(p^*_M, V_{\hat{s}} \cdot P^*_{X|M}) &= \min_{(s,t)} (I(p^*_M, W_t \cdot P^*_{X|M}) - I(p^*_M, V_s \cdot P^*_{X|M})) \\ &\leq \max_{M \to X \to (Y_t Z_s)} \min_{(t,s)} (I(M, Y_t) - I(M, Z_s)) \end{aligned}$$

Now taking into account (2.52) and (2.53) we end in

$$C_{S,CSI}(\mathfrak{W}) \leq C_S(\mathfrak{W})$$

and therewith for this channel the lower bound of the capacity without CSI matches the capacity of the compound wiretap channel with CSI.

2.4 Examples

In this section we provide some examples which display some striking features of compound wiretap channels as opposed to the usual compound channels. Our first example shows clearly that for compound wiretap channels with CSI at the transmitter the strategy of sending both the message and the randomisation parameter does not work. The second one demonstrates that even in the case where the sets of channels to the legitimate receiver and the eavesdropper both are convex, we can have

$$C_{S,CSI}(\mathfrak{W}) > 0 \text{ and } C_S(\mathfrak{W}) = 0,$$

as opposed to the case of the usual compound channel where the Minimax-Theorem applies.

In the following we use some simple facts which we state here without proof.

Fact 1. The binary entropy function

$$h(x) := -x \log x - (1-x)\log(1-x), \quad x \in [0,1],$$

is strictly increasing on $[0, \frac{1}{2}]$.

Fact 2. Let $\eta \in [0,1]$ and set

$$D_\eta := \begin{pmatrix} 1-\eta & \eta \\ \eta & 1-\eta \end{pmatrix}.$$

Then for every $\tau, \tau' \in [0,1]$ it follows that

$$D_\tau D_{\tau'} = D_{\tau+\tau'-2\tau\tau'}.$$

Moreover, if $\tau, \tau' \in (0, \frac{1}{2})$ then

$$\tau + \tau' - 2\tau\tau' \in (0, \frac{1}{2})$$

and

$$\tau + \tau' - 2\tau\tau' > \tau, \tau'.$$

Fact 3. For $\tau, t \in [0,1]$

$$(1-t)D_0 + tD_\tau = D_{t\tau}.$$

2.4.1 Example 1

Consider a compound wiretap channel $\mathfrak{W} = \{(W_t, V_t) : t = 0, 1\}$ in the case of CSI at the transmitter. First we define the channels to the legitimate receiver and to the eavesdropper for $t = 0$ by

$$W_0 = D_\eta, \ \eta \in [0, \frac{1}{2}), \ \eta \approx 0, \qquad V_0 := D_\tau W_0, \ \tau \in [0, \frac{1}{2}), \ \tau \approx 0,$$

and for $t = 1$, $\hat{\tau} \in (0, 1/2]$

$$W_1 := D_{\hat{\tau}} V_0 = D_{\hat{\tau}} D_{\tau} W_0, \qquad V_1 := \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}.$$

Hence V_0 and W_1 are degraded versions of W_0 and

$$I(p, V_1) = 0, \quad \forall p \in \mathcal{P}(A)$$

by definition of V_1. Now for every $p \in \mathcal{P}(A)$ we can choose τ small enough, that

$$I(p, W_0) - I(p, V_0) < I(p, W_1).$$

Now with $p_0 = (\frac{1}{2}, \frac{1}{2})$, $\nu > 0$ and because we have CSI at the transmitter we have by the defining equations (2.11) and (2.12)

$$\begin{aligned} J_n &= 2^{n[I(p_0,W_0) - I(p_0,V_0)) - \nu]} \\ L_{n,0} &= 2^{n[I(p_0,V_0) + \frac{\nu}{4}]} \end{aligned}$$

such that we obtain

$$J_n L_{n,0} = 2^{n[I(p_0,W_0) - \frac{3\nu}{4}]}.$$

But for $\hat{\tau}$ close to $1/2$ it holds then that

$$I(p_0, W_0) - \frac{3\nu}{4} > I(p_0, W_1) = \max_{p \in \mathcal{P}(A)} I(p, W_1) = C_{CSI}\{W_0, W_1\},$$

where $C_{CSI}\{W_0, W_1\}$ is the capacity of a compound channel with CSI at the transmitter. Hence we have shown, that we can achieve reliable transmission of the message $j \in [J_n]$, but identifying both the message and the randomizing indices is not possible for all pairs $j \in [J_n]$ and $l \in [L_{n,t}]$. This is in contrast to the case where we have only one channel to both the legitimate receiver and the eavesdropper (cf. [Dev05], [Csi96]).

2.4.2 Example 2

Now, for $\eta, \tau \in (0, \frac{1}{2})$ we set

$$\begin{aligned} W_0 &= D_{\eta}, & V_0 &:= D_{\tau} W_0 = D_{\eta + \tau - 2\eta\tau}, \\ W_1 &:= D_{\tau} V_0 = D_{2\tau - 2\tau^2} W_0, & V_1 &:= D_{\tau} W_1. \end{aligned} \tag{2.54}$$

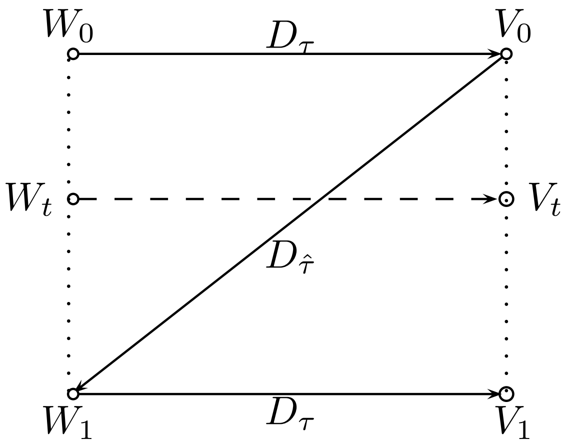

Figure 2.1: Compound wiretap channel $\mathfrak{W} := \{(W_t, V_t) : t \in [0,1]\}$ of Ex. 2

Notice that V_0 is a degraded version of W_0, W_1 of V_0, and V_1 of W_1. Next we define for $t \in [0,1]$

$$\begin{aligned} W_t &:= (1-t)W_0 + tW_1 \\ &= \big[(1-t)D_0 + tD_{2\tau-\tau^2}\big]W_0, \end{aligned} \tag{2.55}$$

and

$$\begin{aligned} V_t &:= (1-t)V_0 + tV_1 \\ &= D_\tau\big[(1-t)D_0 + tD_{2\tau-2\tau^2}\big]W_0 \\ &= D_\tau W_t \end{aligned} \tag{2.56}$$

By the definition, the set of channels to the legitimate receiver $\{W_t\}$ and the set of channels to the eavesdropper $\{V_t\}$ both are convex. Nevertheless we will show now, that for the compound wiretap channel $\mathfrak{W} := \{(W_t, V_t) : t \in [0,1]\}$ we have

$$C_{S,CSI}(\mathfrak{W}) > 0, \quad C_S(\mathfrak{W}) = 0.$$

To this end, note that by (2.55), fact 3, and fact 2 we have

$$W_t = D_{t(2\tau-\tau^2)}D_\eta = D_{f(t,\eta,\tau)}$$

with

$$f(t,\eta,\tau) := \eta + t(2\tau - 2\tau^2) - 2\eta t(2\tau - 2\tau^2) \in (0, \frac{1}{2}). \tag{2.57}$$

Similarly from (2.56) and fact 2 we obtain

$$V_t = D_\tau D_{f(t,\eta,\tau)} = D_{\tau+f(t,\eta,\tau)-2\tau f(t,\eta,\tau)}$$

Additionally from (2.57) and fact 2 we get

$$\begin{aligned} \tau + f(t,\eta,\tau) - 2\tau f(t,\eta,\tau) &\in (0,\frac{1}{2}) \quad \text{and} \\ \tau + f(t,\eta,\tau) - 2\tau f(t,\eta,\tau) &> f(t,\eta,\tau). \end{aligned} \tag{2.58}$$

Taking $p = (1/2, 1/2)$ we obtain for every $t \in [0,1]$

$$I(p, W_t) - I(p, V_t) = h(\tau + f(t,\eta,\tau) - 2\tau f(t,\eta,\tau)) - h(f(t,\eta,\tau)) > 0$$

where the last inequality follows from fact 1 and (2.58). Thus we have shown that

$$C_{S,CSI}(\mathfrak{W}) > 0$$

holds by Theorem 2.2. In order to show that $C_S(\mathfrak{W}) = 0$, we have to employ our multiletter converse in the case of no CSI, Proposition 2.6. First, a simple algebra shows that for any $t \in [0,1]$

$$V_t = ((1-t)D_0 + tD_{2\tau-2\tau^2})V_0$$

by (2.56) and thus each V_t is a degraded version of V_0. Let us now consider the Markov chain $U \to X^n \to (Y_t^n, Z_t^n)$ where the transition from the random variable U to Y_t^n is governed by $P_{Y_t^n|U} = V_t^{\otimes n} \cdot P_{X^n|U}$ for all $t \in [0,1]$. Then we obtain that each $P_{Y_t^n|U}$ is a degraded version of $P_{Y_0^n|U} = V_0^{\otimes n} \cdot P_{X^n|U}$, and the data processing inequality implies that for each $n \in \mathbb{N}$

$$\max_{t\in[0,1]} I(U, Y_t^n) = I(U, Y_0^n) \tag{2.59}$$

for all distributions P_{UX^n} that satisfy the Markov chain condition $U \to X^n \to (Y_t^n, Z_t^n)$.

On the other hand, since $W_1 = D_\tau V_0$ we obtain for the matrix $P_{Z_1^n|U} = W_1^{\otimes n} \cdot P_{X^n|U}$ by the data processing inequality and (2.59) for all $n \in \mathbb{N}$

$$I(U, Z_1^n) - \max_{t\in[0,1]} I(U, Y_t^n) = I(U, Z_1^n) - I(U, Y_0^n) \le 0,$$

for all P_{UX^n}. Then Proposition 2.6 implies that

$$C_S(\mathfrak{W}) = 0$$

as desired.

Chapter 3

Arbitrarily Varying Wiretap Channels

Models of communication systems connecting the requirement of security against a potential eavesdropper and reliable information transmission to legitimate receivers which suffer from channel uncertainty, have been received much interest in current research. One of the simplest models of channel uncertainty are compound channels, where the channel realisations remains fixed during the whole transmission of a codeword, were subject of the previous chapter. In contrast, in the model of an arbitrarily varying wiretap channel AVWC the channel state to both the legitimate receiver and the eavesdropper varies from symbol to symbol in an unknown and arbitrary manner. Thus apart from eavesdropping the model can simulate an active adversarial jamming situation in which the jammer chooses the states. Then, in addition, reliable transmission to the legitimate receiver must be guaranteed in the presence of the jammer.

The arbitrarily varying wiretap channels AVWC will be described by families of pairs of channels $\mathfrak{W} = \{(W_{s^n}, V_{s^n}) : s^n \in S^n\}$ with common input alphabets and possibly different output alphabets, where $s^n \in S^n$ denotes the state sequence during the transmission of a codeword. The legitimate users are connected via W_{s^n} and the transmitter observes the output of V_{s^n}. In our communication scenario the legitimate users have no channel state information at all. We derive capacity results for the AVWC $\mathfrak{W}$ under the average error probability criterion and a strong secrecy criterion. As it was emphasized in Section 1.2.2 there is no full capacity result for an ordinary AVC for the maximum error criterion apart from partial results for specific alphabets given in [AW70], [AW80]. Because our proof techniques are partially based on methods that are used for the ordinary AVC, we must leave the investigation of the corresponding problem for the AVWC for future work.

Two fundamental techniques, discovered by Ahlswede, will play a crucial role in this chapter. In [Ahl78] he developed the *elimination technique* to derive the deterministic code capacity for ordinary AVCs under the average error probability criterion. He showed that it is either zero or equals its random code capacity, which is called Ahlswede's dichotomy for single user AVCs. Actually, it seems to be the more practicable way to derive first explicit capacity results for random codes. With the so-called *robustification technique* [Ahl86] in turn he could link random codes for the AVC to deterministic codes for compound channels. Because the AVWC combines both the wiretap channel and the AVC it is not surprising that we can use the aforementioned techniques to derive capacity results capacity for the AVWC. The challenge then is to integrate the strong secrecy criterion in both the elimination and the robustification technique, approaches both were developed to guarantee a reliability criterion. As it was shown in Section 2.2 the secrecy criterion ensures that the average error probability of every decoding strategy of the eavesdropper in the limit tends to one.

In Section 3.2.2 we show that provided that the channel to the legitimate receiver is non-symmetrisable, the deterministic code secrecy capacity equals the random code secrecy capacity, even if the random code secrecy capacity is unknown, and to give a condition when it is greater than zero. Thus we establish a result for the AVWC that is similar to that of the dichotomy result for ordinary AVCs. The proof uses the *elimination technique* of [Ahl78] for single user AVC's, which is composed of the *random code reduction* and the *elimination of randomness* [CK81], which we will adapt both to the scenario of an arbitrarily varying wiretap channel with the strong secrecy criterion.

In Section 3.2.3 we give a lower bound on the random code secrecy capacity in the special case of a "best" channel to the eavesdropper.The proof is based on the *robustification technique* by Ahlswede [Ahl86] and the approach we used in the proof of Theorem 2.4 for the compound wiretap channel. As a consequence of the result of the previous section the lower bound on the random code secrecy capacity can be achieved by a deterministic code, if we assume that the channel to the legitimate receiver is not symmetrisable.

In section 3.2.4 we give a single-letter upper bound on the deterministic code secrecy capacity, which corresponds to the upper bound of the secrecy capacity of a compound wiretap channel. Therewith it is possible to give a capacity result for the arbitrarily varying wiretap channel in the case of a "worst" channel to the legitimate receiver and a best channel to the eavesdropper. Moreover, by establishing a multiletter upper bound on the secrecy capacity we can conclude to a multiletter expression of the secrecy capacity of the AVWC in the special case of a best channel to the eavesdropper.

The lower bound on the secrecy capacity as well as other results were given earlier in [Mol09] for a weaker secrecy criterion, but the proof techniques for the stronger secrecy criterion differ significantly, especially in the achievability part for random codes.

3.1 Definitions

Let A, B, C be finite sets and consider a not necessarily finite family of channels $W_s : A \to \mathcal{P}(B)$[1], where $s \in S$ denotes the state of the channel. Now, given $s^n = (s_1, s_2, \dots, s_n) \in S^n$ we define the stochastic matrix

$$W^n(y^n|x^n, s^n) = \prod_{i=1}^{n} W(y_i|x_i, s_i) := \prod_{i=1}^{n} W_{s_i}(y_i|x_i) \tag{3.1}$$

for all $y^n = (y_1, \dots, y_n) \in B^n$ and $x^n = (x_1, \dots, x_n) \in A^n$. An arbitrarily varying channel is then defined as the sequence $\{\mathcal{W}^n\}_{n=1}^{\infty}$ of the family of channels $\mathcal{W}^n = \{W^n(\cdot|\cdot, s^n) : s^n \in S^n\}$. Now let $\mathcal{W}^n$ represent the communication link to a legitimate receiver to which the transmitter wants to send a private message, such that a possible second receiver should be kept as ignorant as possible of the message sent. We call this receiver the eavesdropper, which observes the output of a second family of channels $\mathcal{V}^n = \{V^n(\cdot|\cdot, s^n) : s^n \in S^n\}$ with an analogue definition of $V^n(\cdot|\cdot, s^n)$ as in (3.1) for $V_s : A \to \mathcal{P}(C)$, $s \in S$. Then we denote the families of pairs of channels with common input by $\mathfrak{W} = \{(W_{s^n}, V_{s^n}) : s^n \in S^n\}$ and call it the arbitrarily varying wiretap channel. In addition, we assume that the state sequence s^n is unknown to the legitimate participants, whereas the eavesdropper always knows which channel is in use.

A (n, J_n) code $\mathcal{C}_n$ for the arbitrarily varying wiretap channel $\mathfrak{W}$ consists of a stochastic encoder $E : \mathcal{J}_n \to \mathcal{P}(A^n)$ (a stochastic matrix) with a message set $\mathcal{J}_n := \{1, \dots, J_n\}$ and a collection of mutually disjoint decoding sets $\{D_j \subset B^n : j \in \mathcal{J}_n\}$. Then the average error probability of a code $\mathcal{C}_n$ is defined by

$$e(\mathcal{C}_n) := \max_{s^n \in S^n} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{x^n \in A^n} E(x^n|j) W_{s^n}^{\otimes n}(D_j^c|x^n) \ . \tag{3.2}$$

[1] $\mathcal{P}(B)$ denotes the set of probability distributions on B.

A *correlated* (n, J_n, Γ, μ) *random code* $\mathcal{C}_n^{\text{ran}}$ for the arbitrarily varying wiretap channel is given by a family of wiretap codes $\{\mathcal{C}_n(\gamma)\}_{\gamma\in\Gamma}$ together with a random experiment choosing γ according to a distribution μ on Γ. The mean average error probability of the random code $\mathcal{C}_n^{\text{ran}}$ is defined analogously to the ordinary one but with respect to the random experiment choosing γ by

$$\bar{e}(\mathcal{C}_n^{\text{ran}}) := \max_{s^n\in S^n} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{\gamma\in\Gamma} \sum_{x^n\in A^n} E^{\gamma}(x^n|j) W_{s^n}^{\otimes n}((D_j^{\gamma})^c|x^n)\mu(\gamma) \ .$$

Definition 3.1. *A non-negative number R is an achievable secrecy rate for the AVWC $\mathfrak{W}$, if there is a sequence $(\mathcal{C}_n)_{n\in\mathbb{N}}$ of (n, J_n) codes such that*

$$\lim_{n\to\infty} e(\mathcal{C}_n) = 0 \ ,$$

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n \geq R \ ,$$

and

$$\lim_{n\to\infty} \max_{s^n\in S^n} I(p_J; V_{s^n}^n) = 0 \ , \tag{3.3}$$

where J is a uniformly distributed random variable taking values in $\mathcal{J}_n$ and $I(p_J; V_{s^n}^n)$ is the mutual information of J and the output variable Z^n of the eavesdropper's channel $V_{s^n}^n$. The secrecy capacity then is given as the supremum of all achievable secrecy rates R_S and is denoted by $C_S(\mathfrak{W})$.

Analogously we define the secrecy rates and the secrecy capacity $C_{S,\text{ran}}(\mathfrak{W})$ for random codes, if we replace $\mathcal{C}_n$ by $\mathcal{C}_n^{\text{ran}}$ in the above definition.

Definition 3.2. *A non-negative number R_S is an achievable secrecy rate for correlated random codes for the AVWC $\mathfrak{W}$, if there is a sequence $(\mathcal{C}_n^{\text{ran}})_{n\in\mathbb{N}}$ of (n, J_n, Γ, μ) codes such that*

$$\lim_{n\to\infty} \bar{e}(\mathcal{C}_n^{\text{ran}}) = 0 \ ,$$

$$\liminf_{n\to\infty} \frac{1}{n} \log J_n \geq R_S \ ,$$

and

$$\lim_{n\to\infty} \max_{s^n\in S^n} \sum_{\gamma\in\Gamma} I(p_J, V_{s^n}^n; \mathcal{C}(\gamma))\mu(\gamma) = 0 \ ,$$

where $I(p_J, V_{s^n}^n; \mathcal{C}(\gamma))$ is the mutual information according to the code $\mathcal{C}(\gamma)$, $\gamma \in \Gamma$ chosen according to the distribution μ. The secrecy capacity then is given as the supremum of all achievable secrecy rates R_S and is denoted by $C_{S,\text{ran}}(\mathfrak{W})$.

3.2 Capacity Results

3.2.1 Preliminaries

In what follows we use the notation as well as some properties of types of sequences, and of *typical* and *conditionally typical* sequences, which are summarised in the appendix sections A.1 and A.2. The set of sequences $s^n \in \mathcal{S}^n$ of type $q \in \mathcal{P}(S)$ is denoted by $\mathcal{T}_q^n$, the set of all types by $\mathcal{P}_0(n, S)$. For $p \in \mathcal{P}(A)$, $W : A \to \mathcal{P}(B)$, $x^n \in A^n$, and $\delta > 0$ we denote by $\mathcal{T}_{p,\delta}^n$ the set of p-typical sequences and by $\mathcal{T}_{W,\delta}^n(x^n)$ the set of W- (conditionally) typical sequences given x^n in the sense of [CK81].

For the optimal random coding strategy of the AVWC with the strong secrecy criterion we need the *robustification technique* by Ahlswede [Ahl86] which is formulated as a further lemma. Therefor let Σ_n be the group of permutations acting on $(1, 2, \ldots, n)$. Then every permutation $\sigma \in \Sigma_n$ induces a bijection $\pi \in \Pi_n$ defined by $\pi : \mathcal{S}^n \to \mathcal{S}^n$ with $\pi(s^n) = (s_{\sigma(1)}, \ldots, s_{\sigma(n)})$ for all $s^n = (s_1, \ldots, s_n) \in \mathcal{S}^n$ and Π_n denotes the group of these bijections.

Lemma 3.3. *(Robustification technique) If a function $f : \mathcal{S}^n \to [0, 1]$ satisfies*

$$\sum_{s^n \in \mathcal{S}^n} f(s^n) q(s_1) \cdot \ldots \cdot q(s_n) \geq 1 - \gamma \tag{3.4}$$

for all $q \in \mathcal{P}_0(n, \mathcal{S})$ and some $\gamma \in [0, 1]$, then

$$\frac{1}{n!} \sum_{\pi \in \Pi_n} f(\pi(s^n)) \geq 1 - 3 \cdot (n+1)^{|\mathcal{S}|} \cdot \gamma \quad \forall s^n \in \mathcal{S}^n \ . \tag{3.5}$$

Proof. The proof is given in [Ahl86]. □

To reduce the random code for the AVWC $\mathfrak{W}$ to a deterministic code we need the concept of symmetrisability, which was established for ordinary AVC's in the following representation by [Eri85], [CN88].

Definition 3.4. *[CN88] An AVC is symmetrisable if for some channel $U : A \to S$*

$$\sum_{s \in S} W(y|x, s) U(s|x') = \sum_{s \in S} W(y|x', s) U(s|x) \tag{3.6}$$

for all $x, x' \in A$, $y \in B$.

A new channel defined by (3.6) then would be symmetric with respect to all $x, x' \in A$. The authors of [CN88] proved the following statement which is an concretion of

Ahlswede's dichotomy result for single-user AVC, which states that the deterministic code capacity C is either $C = 0$ or equals the random code capacity.

Theorem 3.5. *[CN88] $C > 0$ if and only if the AVC is not symmetrisable. If $C > 0$, then*

$$C = \max_{p \in \mathcal{P}(A)} \min_{W \in \bar{\mathcal{W}}} I(p, W) \tag{3.7}$$

Here the RHS gives the random code capacity and $\bar{\mathcal{W}}$ denotes the convex closure of all channels W_s with $s \in S$, S finite or countable. A more detailed treatment about AVC's is given in Section 1.2.2. Anyway, we will show that there exist an analogous result for the secrecy capacity of the arbitrarily varying wiretap channels AVWC. But here only the symmetrisability of the channel to the eavesdropper will decide weather the deterministic code secrecy equals the random code capacity. In addition, only in some special cases the random code capacity can determined explicitly.

3.2.2 Deterministic Code Secrecy Capacity

To derive the deterministic code secrecy capacity from the random code secrecy capacity, in this subsection we give a counterpart of Theorem 3.5 for arbitrarily varying wiretap channels AVWC . Because in a random coding strategy the code $\mathcal{C}(\gamma)$ that is used for the transmission of a single message is subjected to a random selection from a family $\{\mathcal{C}(\gamma)\}_{\gamma \in \Gamma}$ of codes, reliable transmission can only be guaranteed if the outcome of the random experiment can be shared by both the transmitter and the receiver. In the absence of *common randomness* between the legitimate participants one way to inform the receiver about the code that is chosen is to add a short prefix to the actual codeword. Provided that the number of codes is small enough, the transmission of these additional prefixes causes no essential loss in rate. In the following we use the *elimination technique* by Ahlswede [Ahl78] which has introduced the above approach to derive deterministic codes from random codes for determining capacity of arbitrarily varying channels.

Theorem 3.6. 1. *Assume that for the random code secrecy capacity of the AVWC $\mathfrak{W}$ it holds that $C_{S,\text{ran}}(\mathfrak{W}) > 0$. Then the deterministic code secrecy capacity $C_S(\mathfrak{W})$ of the AVWC $\mathfrak{W}$ equals its random code capacity $C_{S,\text{ran}}(\mathfrak{W})$*

$$C_S(\mathfrak{W}) = C_{S,\text{ran}}(\mathfrak{W}), \tag{3.8}$$

if and only if the channel to the legitimate receiver is non-symmetrisable.

2. *If $C_{S,\text{ran}}(\mathfrak{W}) = 0$ it always holds that $C_S(\mathfrak{W}) = 0$.*

First, if the channel to the legitimate receiver is symmetrisable then the deterministic code capacity of the channel to the legitimate receiver equals zero by Theorem 3.5 and no reliable transmission of messages is possible. Hence the deterministic code secrecy capacity of the arbitrarily varying wiretap channel also equals zero although the random code secrecy capacity could be greater than zero. So we can restrict to the case in which the channel to the legitimate receiver is not symmetrisable. If $C_S(\mathfrak{W}) = C_{S,\text{ran}}(\mathfrak{W}) > 0$, then the channel to the legitimate receiver must be nonsymmetrisable. For the other direction, because the secrecy capacity of the AVWC $\mathfrak{W}$ cannot be greater than the random code secrecy capacity it suffices to show that $C(\{W_{s^n}\}) > 0$ implies that $C_S(\mathfrak{W}) \geq C_{S,\text{ran}}(\mathfrak{W})$. Here $C(\{W_{s^n}\})$ denotes the capacity of the arbitrarily varying channels to the legitimate receiver without secrecy. The proof is given in the following two paragraphs *Random code reduction* and *Elimination of randomness*, where we have used two lemmas and their proofs of [CK81] for single user AVC's and adapted it to arbitrarily varying channels under the strong secrecy criterion. Note that in contrast to Theorem 3.5 no actual value of the random code secrecy capacity is given, and for the proof we need only the assumption of a random code with $C_{S,\text{ran}}(\mathfrak{W}) > 0$. Actually, as we will see later, we can not give a computational description of the random code secrecy capacity except for special cases.

3.2.2.1 Random Code Reduction

We first reduce the random code $\mathcal{C}^{\text{ran}}$ to a new random code selecting only a small number of deterministic codes from the former by random selection according to a distribution μ from the family of codes $\{\mathcal{C}(\gamma)\}_{\gamma\in\Gamma}$, and averaging over this codes gives a new random code with a constant small mean average error probability, which additionally fulfills the secrecy criterion.

Lemma 3.7. *(Random Code Reduction) Let $\mathcal{C}(\mathcal{Z})$ be a random code for the AVWC $\overline{\mathfrak{W}}$ consisting of a family $\{\mathcal{C}(\gamma)\}_{\gamma\in\Gamma}$ of wiretap codes where γ is chosen according to the distribution μ of $\mathcal{Z}$. Then let*

$$\bar{e}(\mathcal{C}_n^{\text{ran}}) = \max_{s^n} \mathbb{E}_\mu e(\mathcal{C}(\mathcal{Z})) \leq \lambda_n \quad \textit{and,} \quad \max_{s^n} \mathbb{E}_\mu I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z})) \leq \epsilon_n' \ . \tag{3.9}$$

Then for any ϵ and K satisfying

$$\epsilon > 4 \max\{\lambda_n, \epsilon_n'\} \quad \textit{and} \quad K > \frac{2n \log |A|}{\epsilon}(1 + n \log |S|) \tag{3.10}$$

there exist K deterministic codes $\mathcal{C}_i$, $i = 1, \ldots, K$, chosen from the random code by *random selection, such that*

$$\frac{1}{K}\sum_{i=1}^{K} e(s^n|\mathcal{C}_i) \leq \epsilon \quad \textit{and} \quad \frac{1}{K}\sum_{i=1}^{K} I(p_J, V_{s^n}; \mathcal{C}_i) \leq \epsilon \tag{3.11}$$

for all $s^n \in S^n$.

Proof. The proof is analogue to that of Lemma 6.8 [CK81], where a similar assertion in terms of the maximal probability of error for the single user AVC without secrecy criterion is established. Cf. also [Ahl78]. Let $\mathcal{Z}$ be the random variable distributed according to μ on Γ for the (n, J_n, Γ, μ) random code. Now consider K independent repetitions of the random experiment of code selections according to μ and call the according random variables $\mathcal{Z}_i$, $i \in \{1, \ldots, K\}$. Then for any $s^n \in S^n$ it holds that

$$\begin{aligned}
&\Pr\Big\{\frac{1}{K}\sum_{i=1}^{K} e(s^n|\mathcal{C}(\mathcal{Z}_i)) \geq \epsilon \quad \text{or} \quad \frac{1}{K}\sum_{i=1}^{K} I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}_i)) \geq \epsilon\Big\} \\
&\leq \Pr\Big\{\exp\sum_{i=1}^{K}\frac{e(s^n|\mathcal{C}(\mathcal{Z}_i))}{n\log|A|} \geq \exp\frac{K\epsilon}{n\log|A|}\Big\} \\
&\qquad\qquad + \Pr\Big\{\exp\sum_{i=1}^{K}\frac{I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}_i))}{n\log|A|} \geq \exp\frac{K\epsilon}{n\log|A|}\Big\},
\end{aligned}$$

and further by Markov's inequality it holds that

$$\begin{aligned}
&\Pr\Big\{\frac{1}{K}\sum_{i=1}^{K} e(s^n|\mathcal{C}(\mathcal{Z}_i)) \geq \epsilon \quad \text{or} \quad \frac{1}{K}\sum_{i=1}^{K} I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}_i)) \geq \epsilon\Big\} \\
&\leq \exp\Big(-\frac{K\epsilon}{n\log|A|}\Big)\mathbb{E}\exp\sum_{i=1}^{K}\frac{e(s^n|\mathcal{C}(\mathcal{Z}_i))}{n\log|A|} \\
&\qquad\qquad + \exp\Big(-\frac{K\epsilon}{n\log|A|}\Big)\mathbb{E}\exp\sum_{i=1}^{K}\frac{I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}_i))}{n\log|A|} \quad .
\end{aligned}$$

Now because of the independency of the random variables $\mathcal{Z}_i$ and because all $\mathcal{Z}_i$ are distributed as $\mathcal{Z}$ and we have $\exp t \leq 1 + t$, for $0 \leq t \leq 1$ (exp to the base 2), we can give the following upper bounds

$$\Big(\mathbb{E}\exp\frac{e(s^n|\mathcal{C}(\mathcal{Z}))}{n\log|A|}\Big)^K \leq \Big(1 + \mathbb{E}\frac{e(s^n|\mathcal{C}(\mathcal{Z}))}{n\log|A|}\Big)^K \leq \Big(1 + \frac{\lambda_n}{n\log|A|}\Big)^K$$

and

$$\Big(\mathbb{E}\exp\frac{I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}))}{n\log|A|}\Big)^K \leq \Big(1 + \mathbb{E}\frac{I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}))}{n\log|A|}\Big)^K \leq \Big(1 + \frac{\epsilon'_n}{n\log|A|}\Big)^K \quad .$$

Hence we obtain for any $s^n \in S^n$

$$\begin{aligned}
&\Pr\Big\{\frac{1}{K}\sum_{i=1}^{K} e(s^n|\mathcal{C}(\mathcal{Z}_i)) \geq \epsilon \quad \text{or} \quad \frac{1}{K}\sum_{i=1}^{K} I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}_i)) \geq \epsilon\} \\
&\quad \leq \exp\Big[-K\Big(\frac{\epsilon}{n\log|A|} - \log(1+\frac{\lambda_n}{n\log|A|})\Big)\Big] \\
&\qquad\qquad\qquad + \exp\Big[\Big(-K(\frac{\epsilon}{n\log|A|} - \log(1+\frac{\epsilon_n'}{n\log|A|})\Big)\Big] \\
&\quad \leq 2\exp\Big[-K\Big(\frac{\epsilon}{n\log|A|} - \log(1+\max\{\frac{\lambda_n}{n\log|A|}, \frac{\epsilon_n'}{n\log|A|}\})\Big)\Big] \ .
\end{aligned}$$

Then

$$\begin{aligned}
&\Pr\Big\{\frac{1}{K}\sum_{i=1}^{K} e(s^n|\mathcal{C}(\mathcal{Z}_i)) \leq \epsilon \text{ and } \frac{1}{K}\sum_{i=1}^{K} I(p_J, V_{s^n}; \mathcal{C}(\mathcal{Z}_i)) \leq \epsilon, \forall s^n \in S^n\Big\} \\
&\quad \geq 1 - 2|S|^n \exp\Big[-K\Big(\frac{\epsilon}{n\log|A|} - \log(1+\max\{\frac{\lambda_n}{n\log|A|}, \frac{\epsilon_n'}{n\log|A|}\})\Big)\Big],
\end{aligned} \tag{3.12}$$

which is strictly positive, if we choose

$$\begin{aligned}
&\epsilon \geq 2n\log|A|\log(1+\max\{\frac{\lambda_n}{n\log|A|}, \frac{\epsilon_n'}{n\log|A|}\}) \\
&\text{and} \qquad\qquad K \geq \frac{2\log|A|}{\epsilon}(n + n^2\log|S|) \ .
\end{aligned} \tag{3.13}$$

Now because for $0 \leq t \leq 1$ and log to the base 2 it holds that $t \leq \log(1+t) \leq 2t$, we increase the lower bound for choosing ϵ if

$$\epsilon \geq 4\max\{\lambda_n, \epsilon_n'\} \ .$$

and with (3.13) the assertion of (3.12) still holds. Thus we have shown that there exist K realisations $\mathcal{C}_i := \mathcal{C}(\mathcal{Z}_i = \gamma_i)$, $\gamma_i \in \Gamma$, $i \in \{1,\dots,K\}$ of the random code, which build a new random code with uniform distribution on these codes $\mathcal{C}_i$ with mean average error probability and mean secrecy criterion fulfilled by (3.11). □

Now, if we assume that the channel to the legitimate receiver is non-symmetrisable, which means that $C(\{W_{s^n}\}) > 0$, and that there exist a random code $\mathcal{C}_n^{\text{ran}}$ that achieves the random code capacity $C_{S,\text{ran}}(\mathfrak{W}) > 0$, then there exist a sequence of random (n, J_n, Γ, μ) codes with average error probability

$$\lim_{n\to\infty}\max_{s^n\in S^n} \frac{1}{J_n}\sum_{j=1}^{J_n}\sum_{\gamma\in\Gamma}\sum_{x^n\in A^n} E^{\gamma}(x^n|j)\cdot W_{s_n}^{\otimes n}((D_j^{\gamma})^c|x^n)\mu(\gamma) = 0 \ ,$$

$$\liminf_{n\to\infty}\frac{1}{n}\log J_n \to C_{S,\mathrm{ran}}(\mathfrak{W}) > 0,$$

and

$$\lim_{n\to\infty}\max_{s^n\in S^n}\sum_{\gamma\in\Gamma} I(p_J; V^n_{s^n}; \mathcal{C}(\gamma))\mu(\gamma) = 0. \tag{3.14}$$

Then on account of the random code reduction lemma there exist a sequence of random (n, J_n) codes consisting of n^3 deterministic codes (cf. (3.10)) chosen from the former random code, and it holds for any $\epsilon > 0$ and sufficiently large n that

$$\max_{s^n\in S^n}\frac{1}{J_n}\sum_{j=1}^{J_n}\frac{1}{n^3}\sum_{i=1}^{n^3}\sum_{x^n\in A^n} E^i(x^n|j)W^{\otimes n}_{s_n}((D^i_j)^c|x^n) \le \epsilon$$

and

$$\max_{s^n\in S^n}\frac{1}{n^3}\sum_{i=1}^{n^3} I(p_J; V^n_{s^n}; \mathcal{C}_i) \le \epsilon, \tag{3.15}$$

where $\mathcal{C}_i = \{(E^i_j, D^i_j), j \in \mathcal{J}_n\}$, $i = 1, \dots, n^3$, and E^i is the stochastic encoder of the deterministic wiretap code. Then the reduced random code consists of the family of codes $\{\mathcal{C}_i\}_{i\in\{1,\dots,n^3\}}$ together with the uniform distribution $\mu'(i) = \frac{1}{n^3}$ for all $i \in \{1, \dots, n^3\}$.

3.2.2.2 Elimination of randomness

Now, to construct a deterministic code we make use of Theorem 6.11 in [CK81]). If there exist a deterministic code and $C(\{W_{s^n}\}) > 0$ then there exist a code

$$\{x^{k_n}_i, F_i \subset B^{k_n} : i = 1, \dots n^3\} \tag{3.16}$$

where $x^{k_n}_i$ is chosen according to an encoding function $f_i : \{1, \dots, n^3\} \to A^{k_n}$ with $\frac{k_n}{n} \to 0$ as $n \to \infty$ with error probability

$$\frac{1}{n^3}\sum_{i=1}^{n^3} W^{\otimes k_n}(F^c_i|x^{k_n}_i, s^{k_n}) \le \epsilon \qquad \text{for all } s^{k_n} \in S^{k_n}\ . \tag{3.17}$$

If we now compose a new deterministic code for the AVWC $\mathfrak{W}$ by prefixing the codewords of each $\mathcal{C}_i$

$$\{f_iE^i_j, F_i \times D^i_j : i = 1, \dots, n^3, j \in [J_n]\} =: \mathcal{C}\ , \tag{3.18}$$

the decoder is informed of which encoder E^i is in use for the actual message j if he identifies the prefix correctly. Note that for the transmission of the prefix only the reliability is of interest, because it contains no information about the message $j \in \mathcal{J}_n$ to be sent. Now the new codewords has a length of $k_n + n$, transmit a message from $\{1, \dots, n^3\} \times \mathcal{J}_n$, whereat the channel which is determined by the state sequence $s^{k_n+n} \in S^{k_n+n}$ yields an average error probability of

$$\begin{aligned} \bar{\lambda}_n(\mathcal{C}, W^{\otimes(k_n+n)}_{s^{k_n+n}}) &\leq \frac{1}{n^3 J_n} \sum_{i=1}^{n^3} \sum_{j \in [J_n]} (\lambda_i + \lambda_j(i)) \\ &\leq \frac{1}{n^3} \sum_{i=1}^{n^3} \lambda_i + \frac{1}{n^3} \sum_{i=1}^{n^3} e_n(s^n, \mathcal{C}_i) \leq 2\epsilon. \end{aligned} \tag{3.19}$$

Here, for each $s^{k_n} \in S^{k_n}$ λ_i means the error probability for transmitting i from $\{1, \dots, n^3\}$ encoded in $x_i^{k_n}$ by $W^{k_n}_{s^{k_n}}$ followed by the transmission of j, where the codeword is chosen according to the stochastic encoder E^i_j, over the last n channel realisations determined by s^n with error probability $\lambda_j(i)$. This construction is possible due to the memorylessness of the channel.

Now if we turn to the security part of the transmission problem it is easily seen that for the code $\mathcal{C}$ defined in (3.18)

$$\begin{aligned} p^{\mathcal{C}}_{JZ^{k_n+n}_{s^{k_n+n}}}(j, z^{k_n+n}) &= \frac{1}{J_n} \frac{1}{n^3} \sum_{i=1}^{n^3} V^{\otimes k_n}_{s^{k_n}}(\hat{z}^{k_n} | x_i^{k_n}) \sum_{x^n} E^i(x^n | j) V^n_{s^n}(z^n | x^n) \\ &= \frac{1}{n^3} \sum_{i=1}^{n^3} V^{\otimes k_n}_{s^{k_n}}(\hat{z}^{k_n} | x_i^{k_n}) \cdot p^{\mathcal{C}_i}_{JZ^n_{s^n}} \quad , \end{aligned} \tag{3.20}$$

where $\hat{z}^{k_n}$ are the first k_n components of z^{k_n+n}. With (3.20) and the representation of the mutual information by the information divergence we obtain that

$$\begin{aligned} &D(p^{\mathcal{C}}_{JZ^{k_n+n}_{s^{k_n+n}}} || p_J \otimes p^{\mathcal{C}}_{Z^{k_n+n}_{s^{k_n+n}}}) \\ &= D\Big(\frac{1}{n^3} \sum_{i=1}^{n^3} V^{\otimes k_n}_{s^{k_n}}(\hat{z}^{k_n} | x_i^{k_n}) p^{\mathcal{C}_i}_{JZ^n_{s^n}} \Big\| \frac{1}{n^3} \sum_{i=1}^{n^3} V^{\otimes k_n}_{s^{k_n}}(\hat{z}^{k_n} | x_i^{k_n}) p_J \otimes p^{\mathcal{C}_i}_{Z^n_{s^n}} \Big) \\ &\leq \frac{1}{n^3} \sum_{i=1}^{n^3} D\big(V^{\otimes k_n}_{s^{k_n}}(\hat{z}^{k_n} | x_i^{k_n}) p^{\mathcal{C}_i}_{JZ^n_{s^n}} \big\| V^{\otimes k_n}_{s^{k_n}}(\hat{z}^{k_n} | x_i^{k_n}) p_J \otimes p^{\mathcal{C}_i}_{Z^n_{s^n}} \big) \\ &= \frac{1}{n^3} \sum_{i=1}^{n^3} D\big(p^{\mathcal{C}_i}_{JZ^n_{s^n}} \| p_J \otimes p^{\mathcal{C}_i}_{Z^n_{s^n}} \big) = \frac{1}{n^3} \sum_{i=1}^{n^3} I(p_J, V^n_{s^n}; \mathcal{C}_i) \leq \epsilon \end{aligned} \tag{3.21}$$

for all $s^n \in S^n$ and $n \in \mathbb{N}$ sufficiently large, where the first inequality follows because for two probability distributions p, q the relative entropy $D(p\|q)$ is a convex function in the pair (p, q) and the last inequality follows by the random code reduction lemma.

Because $\frac{k_n}{n} \to 0$ as $n \to \infty$

$$\lim_{n\to\infty} \frac{1}{k_n + n} \log(n^3 J_n) = \lim_{n\to\infty} (\frac{1}{n} \log J_n + \frac{1}{n} \log(n^3)) = \lim_{n\to\infty} \frac{1}{n} \log J_n \ , \tag{3.22}$$

$\mathcal{C}_n$ is a deterministic (n, J_n) code which achieves the same rates as the random code $\mathcal{C}_n^{\text{ran}}$ and so the random code capacity $C_{S,\text{ran}}$ as given in (3.14), provided that the channel to the legitimate receiver is not symmetrisable.

Thus, with $\{1, \ldots, J_n\}$ as the message set, $\mathcal{C}_n$ is a deterministic $(n + o(n), n^3 \cdot J_n)$ code with average error probability bounded for all $s^{k_n+n} \in S^{k_n+n}$ as in (3.19) and which fulfills the strong secrecy criterion as in (3.21), and which achieves the random code secrecy capacity $C_{S,\text{ran}}$ of the arbitrarily varying wiretap channels AVWC $\mathcal{W}$ which implies that $C_S = C_{S,\text{ran}}$. This concludes the proof.

Note that in the case in which the channel to the legitimate receiver is not symmetrisable and we know that the deterministic code secrecy capacity $C_S(\mathfrak{W})$ equals zero we can conclude that the random code secrecy capacity $C_{S,\text{ran}}(\mathfrak{W})$ equals zero.

3.2.3 Random Code Construction

In this section we will derive a lower bound on the random code secrecy capacity in the case of a best channel to the eavesdropper. We use the *robustification technique* by Ahlswede [Ahl86] in combination with the the approach we used in the proof of Theorem 2.4 for the compound wiretap channel without any channel state information at the legitimate participants. With the result of the previous subsection we can show that this lower bound can also be applied to the deterministic code secrecy capacity.

First let us define the convex hull of the set of channels $\{W_s : s \in S\}$ by the set of channels $\{W_q : q \in \mathcal{P}(S)\}$, where W_q is defined by

$$W_q(y|x) = \sum_{s\in S} W(y|x,s)q(s), \tag{3.23}$$

for all possible distributions $q \in \mathcal{P}(S)$. Accordingly we define V_q and its convex hull $\{V_q : q \in \mathcal{P}(S)\}$. Then denote the convex closure of the set of channels $\{(W_s, V_s) : s \in S\}$ by $\overline{\overline{\mathfrak{W}}} := \{(W_q, V_q) : q \in \mathcal{P}(\tilde{S}), \tilde{S} \subseteq S, \tilde{S} \text{ is finite}\}$. Occasionally we restrict q to be from the set of all types $\mathcal{P}_0(n, S)$ of state sequences $s^n \in S^n$.

Lemma 3.8. *The secrecy capacity $C_S(\mathfrak{W})$ of the arbitrarily varying wiretap channel AVWC $\mathfrak{W}$ equals the secrecy capacity of the arbitrarily varying wiretap channel $\overline{\overline{\mathfrak{W}}}$.*

Proof. The proof was given for an ordinary arbitrarily varying channel AVC without secrecy criterion in [CK81], and for an AVWC with the weak secrecy criterion in [Mol09]. Let $\tilde{W}_1, \dots, \tilde{W}_n$ be averaged channels as defined in (3.23) and a channel $W^n_{\tilde{q}} : A^n \to \mathcal{P}(B^n)$ with $\tilde{q} = \prod_{i=1}^n q_i$, $\tilde{q} \in \mathcal{P}(S^n)$, $q_i \in \mathcal{P}(S)$ defined by

$$W^n_{\tilde{q}}(y^n|x^n) = \prod_{i=1}^n \tilde{W}_i(y_i|x_i) = \prod_{i=1}^n W_{q_i}(y_i|x_i) = \sum_{s^n \in S^n} W^{\otimes n}(y^n|x^n, s^n)\tilde{q}(s^n)$$

If we now use the same (n, J_n) code $\mathcal{C}_n$ defined by the same pair of encoder and decoding sets as for the AVWC $\mathfrak{W}$ the error probability for transmission of a single codeword by the channel $W^n_{\tilde{q}}$ is given by

$$\sum_{x^n \in A^n} E(x^n|j) W^n_{\tilde{q}}(D^c_j|x^n) = \sum_{s^n \in S^n} \tilde{q}(s^n) \sum_{x^n \in A^n} E(x^n|j) W^{\otimes n}_{s^n}(D^c_j|x^n),$$

and we can bound the average error probability by

$$\begin{aligned} &\frac{1}{J_n}\sum_{j=1}^{J_n}\sum_{x^n \in A^n} E(x^n|j) W^n_{\tilde{q}}(D^c_j|x^n) \\ &\leq \max_{s^n \in S^n} \frac{1}{J_n}\sum_{j=1}^{J_n}\sum_{x^n \in A^n} E(x^n|j) W^{\otimes n}_{s^n}(D^c_j|x^n) = e(\mathcal{C}_n) \ . \end{aligned}$$

Otherwise, because $\mathfrak{W}$ is a subset of $\overline{\mathfrak{W}}^n$, which is the closure of the set of channels $(W^n_{\tilde{q}}, V^n_{\tilde{q}})$, the opposite inequality holds for the channel $W^n_{\tilde{q}}$ that maximizes the error probability. Because $V^n_{\tilde{q}}$ is defined analogously to $W^n_{\tilde{q}}$, we can define for the (n, J_n) code

$$\hat{V}(z^n|j) := \sum_{x^n \in A^n} E(x^n|j) V^n_{\tilde{q}}(z^n|x^n) \tag{3.24}$$

for all $z^n \in C^n$, $j \in \mathcal{J}_n$. Then

$$\hat{V}(z^n|j) = \sum_{s^n \in S^n} \tilde{q}(s^n) \sum_{x^n \in A^n} E(x^n|j) V^n_{s^n}(z^n|x^n) = \sum_{s^n \in S^n} \tilde{q}(s^n) \hat{V}^n_{s^n}(z^n|j) \tag{3.25}$$

and because of the convexity of the mutual information in the channel $\hat{V}$ and (3.25) it holds that

$$I(J, Z^n_{\tilde{q}}) \leq \sum_{s^n \in S^n} \tilde{q}(s^n) I(J; Z^n_{s^n}) \leq \sup_{s^n} I(J, Z^n_{s^n}). \tag{3.26}$$

Now because $\{\hat{V}^n_{s^n}(z^n|j) : s^n \in S^n\}$ is a subset of $\{\hat{V}(z^n|j) : \tilde{q} \in \mathcal{P}(S^n)\}$ we end in

$$\sup_{\tilde{q} \in \mathcal{P}(S^n)} I(J, Z^n_{\tilde{q}}) = \sup_{s^n} I(J, Z^n_{s^n}) \ ,$$

which concludes the proof. □

Now we can proceed in the construction of the random code of the AVWC $\mathfrak{W}$.

Definition 3.9. *We call a channel to the eavesdropper a best channel if there exist a channel $V_{q^*} \in \{V_q : q \in \mathcal{P}(S)\}$ such that all other channels from $\{V_q : q \in \mathcal{P}(S)\}$ are degraded versions of V_{q^*}. If we denote the output of any channel V_q, $q \in \mathcal{P}(S)$ by Z_q it holds that*

$$X \to Z_{q^*} \to Z_q, \quad \forall q \in \mathcal{P}(S). \tag{3.27}$$

Proposition 3.10. *Provided that there exist a best channel to the eavesdropper, for the random code secrecy capacity $C_{S,ran}(\mathfrak{W})$ of the AVWC $\mathfrak{W}$ it holds that*

$$C_{S,ran}(\mathfrak{W}) \geq \max_{p\in\mathcal{P}(A)} (\min_{q\in\mathcal{P}(S)} I(p, W_q) - \max_{q\in\mathcal{P}(S)} I(p, V_q)). \tag{3.28}$$

Proof. The proof is based on Ahlswedes *robustification technique* [Ahl86] and is divided in two parts:
step 1): The set

$$\overline{\mathcal{W}} := \{(W_q^{\otimes n}, V_q^{\otimes n}) : q \in \mathcal{P}(S)\}$$

corresponds to a compound wiretap channel indexed by the set of all possible distributions $q \in \mathcal{P}(S)$ on the set of states S. First we show, that there exist a deterministic code for the compound wiretap channel $\overline{\mathcal{W}}$ that achieves the lower bound on the random code secrecy capacity of the AVWC $\mathfrak{W}$ given in (3.28).

In Section 2.3.2 Theorem 2.4 and in [BBS11b] it was shown that for a compound wiretap channel $\{(W_t, V_t) : t \in \theta\}$ without channel state information at the legitimate receivers the secrecy capacity is bounded from below by

$$C_{S,\mathrm{comp}} \geq \max_{p\in\mathcal{P}(A)} (\min_{t\in\theta} I(p, W_t) - \max_{t\in\theta} I(p, V_t)). \tag{3.29}$$

In accordance with the proof of (3.29) in Section 2.3.2 we define a set of i.i.d. random variables $\{X_{jl}\}_{j\in[J_n],l\in[L_n]}$ each according to the distribution $p' \in \mathcal{P}(A^n)$ with

$$p'(x^n) := \begin{cases} \frac{p^{\otimes n}(x^n)}{p^{\otimes n}(\mathcal{T}^n_{p,\delta})} & \text{if } x^n \in \mathcal{T}^n_{p,\delta}, \\ 0 & \text{otherwise,} \end{cases} \tag{3.30}$$

for any $p \in \mathcal{P}(A)$, and where J_n and L_n are defined as

$$J_n := \lfloor 2^{n[\inf_{q\in\mathcal{P}(S)} I(p,W_q) - \sup_{q\in\mathcal{P}(S)} I(p,V_q) - \tau]} \rfloor \tag{3.31}$$

$$L_n := \lfloor 2^{n[\sup_{q\in\mathcal{P}(S)} I(p,V_q) + \frac{\tau}{4}]} \rfloor \tag{3.32}$$

with $\tau > 0$. Now we assume that there exist a best channel to the eavesdropper V_{q^*} in contrast to the proof of Theorem 2.4. Hence by the definition of V_{q^*} in (3.27) and because the mutual Information $I(p, V)$ is convex in V and every member of $\{V_q\}_{q\in\mathcal{P}(S)}$ is a convex combination of the set $\{V_s\}_{s\in S}$, it holds that

$$I(p, V_{q^*}) = \sup_{q\in\mathcal{P}(S)} I(p, V_q) = \sup_s I(p, V_s) \tag{3.33}$$

for all $p \in \mathcal{P}(A)$. Note that because of (3.33) for $|S| \leq \infty$ $V_{q^*} \in \{V_s : s \in S\}$, which means that q^* is a one-point distribution.

By the definition of the compound channel $\overline{\mathcal{W}}$ the channels to the eavesdropper are of the form

$$V_q^{\otimes n}(z^n|x^n) := \prod_{i=1}^{n} V_q(z_i|x_i) \tag{3.34}$$

for all $q \in \mathcal{P}(S)$. Then following the same approach as in the proof of Theorem 2.4 we define

$$\tilde{Q}_{q,x^n}(z^n) = V_q^{\otimes n}(z^n|x^n) \cdot \mathbf{1}_{\mathcal{T}^n_{V_q,\delta}(x^n)}(z^n),$$

and

$$\Theta'_q(z^n) = \sum_{x^n\in\mathcal{T}^n_{p,\delta}} p'(x^n)\tilde{Q}_{q,x^n}(z^n). \tag{3.35}$$

for all $z^n \in C^n$. Now let $\mathcal{B} := \{z^n \in C^n : \Theta'_q(z^n) \geq \epsilon\alpha_q\}$ where $\epsilon = 2^{-nc'\delta^2}$ (cf. Lemma A.8) and α_q is defined according to the channel V_q by $\alpha_q := 2^{-n(H(pV_q)+f_1(\delta))}$. By Lemma A.9 the support of Θ'_q has cardinality $\leq \alpha_q^{-1}$ since for each $x^n \in \mathcal{T}^n_{p,\delta}$ it holds that $\mathcal{T}^n_{V_q,\delta}(x^n) \subset \mathcal{T}^n_{pV_q,2|A|\delta}$, which implies that $\sum_{z^n\in\mathcal{B}} \Theta_q(z^n) \geq 1 - 2\epsilon$, if

$$\begin{aligned} \Theta_q(z^n) &= \Theta'_q(z^n) \cdot \mathbf{1}_{\mathcal{B}}(z^n) \quad \text{and} \\ Q_{q,x^n}(z^n) &= \tilde{Q}_{q,x^n}(z^n) \cdot \mathbf{1}_{\mathcal{B}}(z^n). \end{aligned} \tag{3.36}$$

Now it is obvious from (3.35) and the definition of the set $\mathcal{B}$ that for any $z^n \in \mathcal{B}$ $\Theta_q(z^n) = \mathbb{E}Q_{q,X_{jl}}(z^n) \geq \epsilon\alpha_q$ if $\mathbb{E}$ is the expectation value with respect to the distribution p'. For the random variables $\beta_q^{-1}Q_{q,X_{jl}}(z^n)$ define the event

$$\iota_j(q) = \bigcap_{z^n\in C^n} \left\{ \frac{1}{L_n} \sum_{l=1}^{L_n} Q_{q,X_{jl}}(z^n) \in [(1 \pm \epsilon)\Theta_q(z^n)] \right\}, \tag{3.37}$$

and keeping in mind that $\Theta_q(z^n) \geq \epsilon\alpha_q$ for all $z^n \in \mathcal{B}$ it follows that for all $j \in [J_n]$

and for all $s \in S$

$$\Pr\{(\iota_j(q))^c\} \leq 2|C|^n \exp\Big(- L_n \frac{2^{-n[I(p,V_q)+g(\delta)]}}{3}\Big) \tag{3.38}$$

by Lemma 2.3, Lemma A.9, and our choice $\epsilon = 2^{-nc'\delta^2}$ with $g(\delta) := f_1(\delta) + f_2(\delta) + 3c'\delta^2$. Making $\delta > 0$ sufficiently small we have for all sufficiently large $n \in \mathbb{N}$

$$L_n 2^{-n[I(p,V_q)+g(\delta)]} \geq 2^{n\frac{\tau}{8}}.$$

Thus, for this choice of δ the RHS of (3.38) is double exponential in n uniformly in $q \in \mathcal{P}(S)$ and can be made smaller than ϵJ_n^{-1} for all $j \in [J_n]$ and all sufficiently large $n \in \mathbb{N}$. I.e.

$$\Pr\{(\iota_j(q))^c\} \leq \epsilon J_n^{-1}, \quad \forall q \in \mathcal{P}(S). \tag{3.39}$$

Now we will show that we can achieve reliable transmission to the legitimate receiver governed by $\{(W_q^{\otimes n} : q \in \mathcal{P}(S)\}$ for all messages $j \in [J_n]$ when randomising over the index $l \in L_n$ but without the need of decoding $l \in [L_n]$. To this end define $\mathcal{X} = \{X_{jl}\}_{j\in[J_n],l\in[L_n]}$ to be the set of random variables with X_{jl} are i.i.d. according to p' defined in (3.30). Define now the random decoder $\{D_j(\mathcal{X})\}_{j\in[J_n]} \subseteq B^n$ analogously as in (2.37). Then it was shown in the proof of Theorem 2.4 that there exist a sequence of (n, J_n) codes for the compound wiretap channel in the particular case without CSI with arbitrarily small mean average error

$$\mathbb{E}_{\mathcal{X}}(\lambda_n^{(q)}(\mathcal{X})) \leq 2^{-na}$$

for all $q \in \mathcal{P}(S)$ and sufficiently large $n \in \mathbb{N}$. Additionally we define for each $q \in \mathcal{P}(S)$

$$\iota_0(q) = \{\lambda_n^{(q)}(\mathcal{X})) \leq 2^{-n\frac{a}{2}}\} \tag{3.40}$$

and set

$$\iota := \bigcap_{q\in\mathcal{P}_0(n,S)} \bigcap_{j=0}^{J_n} \iota_j(q). \tag{3.41}$$

Then with (3.39), (3.40) and applying the union bound we obtain

$$\Pr\{\iota^c\} \leq 2^{-nc}$$

for a suitable positive constant $c > 0$ and all sufficiently large $n \in \mathbb{N}$ (Cf. (2.40)). Hence, we have shown that there exist realisations $\{x_{jl}\}$ of $\{X_{jl}^n\}_{j\in[J_n],l\in[L_n]}$ such that $x_{jl} \in \iota$ for all $j \in [J_n]$ and $l \in [L_n]$. Now following the same argumentation as in Section 2.3.2 we obtain that there is a sequence of (n, J_n) codes that for all codewords $\{x_{jl}\}$ it follows by construction that

$$\frac{1}{J_n} \sum_{j \in [J_n]} \frac{1}{L_n} \sum_{l \in [L_n]} W_q^{\otimes n}(D_j^c | x_{jl}) \leq 2^{-na'} \tag{3.42}$$

is fulfilled for $n \in \mathbb{N}$ sufficiently large and for all $q \in \mathcal{P}(S)$ with $a' > 0$. So we have found a (n, J_n) code with average error probability upper bounded by (3.42). Further, for the given code and a random variable J uniformly distributed on the message set $\{1, \ldots, J_n\}$ it holds that

$$I(p_J; V_q^{\otimes n}) \leq \epsilon' \tag{3.43}$$

uniformly in $q \in \mathcal{P}(S)$. Both (3.42) and (3.43) ensure that in the scenario of the compound wiretap channel $\overline{\mathcal{W}}$ the legitimate receiver can identify each message j from the message set $\{1, \ldots, J_n\}$ with high probability, while at the same time the eavesdropper receives almost no information about it. That is, that all numbers R_S with

$$R_S \leq \inf_{q \in \mathcal{P}(S)} I(p, W_q) - \sup_{q \in \mathcal{P}(S)} I(p, V_q) \tag{3.44}$$

are achievable secrecy rates of the compound wiretap channel $\overline{\mathcal{W}}$.
step 2): *Robustification* : In the second step we derive from the deterministic (n, J_n) code for the above mentioned compound wiretap channel $\overline{\mathcal{W}}$ a (n, J_n) ranodm code $\mathcal{C}_n^{\text{ran}}$ for the AVWC $\mathfrak{W}$, which achieves the same secrecy rates. We note first that by (3.33) and (3.43)

$$\max_{s^n \in S^n} I(p_J, V_{s^n}) = I(p_J, V_{q^*}^{\otimes n}) \leq \epsilon', \tag{3.45}$$

which means, that, due to the assumption of a best channel to the eavesdropper, the code achieving the secrecy rate for the best channel to the eavesdropper fulfills the secrecy criterion for a channel with any state sequence $s^n \in S^n$. Now, as already mentioned we use the robustification technique (cf. Lemma 3.3) to derive from the deterministic code $\mathcal{C}_{\overline{\mathcal{W}}} = \{x_{jl}, D_j : j \in [J_n], l \in [L_n]\}$ of the compound wiretap channel $\overline{\mathcal{W}}$ the random code for the AVWC $\mathfrak{W}$. With (3.42) it holds that

$$\frac{1}{J_n} \sum_{j \in [J_n]} \frac{1}{L_n} \sum_{l \in [L_n]} \sum_{s^n \in S^n} W^{\otimes n}(D_j | x_{jl}, s^n) q^{\otimes n}(s^n) \geq 1 - 2^{-na'} \tag{3.46}$$

for all $q^{\otimes n} = \prod_{i=1}^n q$ and in particular for all $q \in \mathcal{P}_0(n, S)$. Now let $\pi \in \Pi_n$ be the bijection on S^n induced by the permutation $\sigma \in \Sigma_n$. Since (3.4) is fulfilled with

$$f(s^n) = \frac{1}{J_n} \sum_{j\in[J_n]} \frac{1}{L_n} \sum_{l\in[L_n]} W^{\otimes n}(D_j|x_{jl}, s^n) \tag{3.47}$$

it follows from (3.5) that

$$\frac{1}{n!} \sum_{\pi\in\Pi_n} \frac{1}{J_n} \sum_{j\in[J_n]} \frac{1}{L_n} \sum_{l\in[L_n]} W^{\otimes n}(D_j|x_{jl}, \pi(s^n)) \geq 1 - (n+1)^{|S|} 2^{-na'} \tag{3.48}$$

for all $s^n \in S^n$. Hence by defining $\mathcal{C}^\pi := \{\pi^{-1}(x^n_{jl}), \pi^{-1}(D_j)\}$ as a member of a family of codes $\{\mathcal{C}^\pi\}_{\pi\in\Pi_n}$ together with a random variable K distributed according to μ as the uniform distribution on Π_n, (3.48) is equivalent to

$$\mathbb{E}_\mu(\bar{\lambda}_n(\mathcal{C}^K, W^n_{s^n})) \leq (n+1)^{|S|} 2^{-na'} =: \lambda_n \tag{3.49}$$

with $\bar{\lambda}_n(\mathcal{C}^\pi, W^n_{s^n})$ as the respective average error probability for $K = \pi$ and it holds for all $s^n \in S^n$. Thus we have shown that

$$\mathcal{C}^{\text{ran}}_n := \{(\pi^{-1}(x_{jl}), \pi^{-1}(D_j)) : j \in [J_n], l \in [L_n], \pi \in \Pi_n, \mu\} \tag{3.50}$$

is a (n, J_n, Π_n, μ) random code for the AVC channel $\mathcal{W}^n = \{W_{s^n} : s^n \in S^n\}$ with the mean average error probability $\mathbb{E}_\mu(\bar{\lambda}_n(\mathcal{C}^K, W^n_{s^n}))$ upper bounded by λ_n as in (3.49).

Now it is easily seen that

$$p^{\mathcal{C}^\pi}_{JZ^n_{q^*}}(j, z^n) = \frac{1}{J_n} \frac{1}{L_n} \sum_{l=1}^{L_n} V^{\otimes n}_{q^*}(\pi^{-1}(z^n)|\pi^{-1}(x_{jl})) = p_{JZ^n_{q^*}}. \tag{3.51}$$

Actually, it still holds that

$$p^{\mathcal{C}^r}_{JZ^n_{q^*}}(j, z^n) = \frac{1}{n!} \sum_{\pi\in\Pi_n} p^{\mathcal{C}^\pi}_{JZ^n_{q^*}}(j, z^n) = p_{JZ^n_{q^*}} \quad . \tag{3.52}$$

With (3.51) and the representation of the mutual information by the information divergence we obtain from (3.45)

$$\begin{aligned} \mathbb{E}_\mu(D(p^{\mathcal{C}^K}_{JZ^n_{q^*}} || p_J \otimes p^{\mathcal{C}^K}_{Z^n_{q^*}})) &= \frac{1}{n!} \sum_{\pi\in\Pi_n} D(p^{\mathcal{C}^\pi}_{JZ^n_{q^*}} || p_J \otimes p^{\mathcal{C}^\pi}_{Z^n_{q^*}}) \\ &= \frac{1}{n!} \sum_{\pi\in\Pi_n} D(p_{JZ^n_{q^*}} || p_J \otimes p_{Z^n_{q^*}}) = I(p_J, V^{\otimes n}_{q^*}) \leq \epsilon' \ . \end{aligned}$$

Thus we have constructed a random (n, J_n, Π_n, μ) code $\mathcal{C}^{\text{ran}}_n$ with mean average error probability bounded for all $s^n \in S^n$ as in (3.49) and which fulfills the strong secrecy criterion almost surely, provided that there exist a best channel to the eavesdropper. By the construction of the random code it follows that the secrecy rates given by (3.44) for the compound wiretap channel $\overline{\mathcal{W}}$ achieved by the deterministic code $\mathcal{C}_{\overline{\mathcal{W}}}$ are achievable secrecy rates for the AVWC $\mathfrak{W}$ with random code $\mathcal{C}_n^{\text{ran}}$. That is, we have shown that all rates R_S with

$$R_S \leq \max_{p\in\mathcal{P}(A)} \big(\min_{q\in\mathcal{P}(S)} I(p, W_q) - \max_{q\in\mathcal{P}(S)} I(p, V_q)\big) \ . \tag{3.53}$$

are achievable secrecy rates of the arbitrarily varying wiretap channel AVWC with a best channel to the eavesdropper with random code $\mathcal{C}_n^{\text{ran}}$. □

As a consequence of the proposition we can give the following result for the deterministic code secrecy capacity.

Corollary 3.11. *The deterministic code secrecy capacity of the arbitrarily varying wiretap channel $\mathfrak{W}$, provided that there exists a best channel to the eavesdropper and under the assumption that the channel to the legitimate receiver is non-symmetrisable, is lower bounded by*

$$C_S(\mathfrak{W}) \geq \max_{p\in\mathcal{P}(A)} \big(\min_{q\in\mathcal{P}(S)} I(p, W_q) - \max_{q\in\mathcal{P}(S)} I(p, V_q)\big) \ .$$

Proof. Combine the assertions of Proposition 3.10 and Theorem 3.6. □

3.2.4 Upper Bounds on the Capacity of the AVWC $\mathfrak{W}$ and a Multiletter Coding Theorem

In this section we give an upper bound on the secrecy capacity of the AVWC $\mathfrak{W}$ which corresponds to the bound for the compound wiretap channel built by the same family of channels. In addition we give the proof of the multiletter converse of the AVWC $\mathfrak{W}$.

Theorem 3.12. *The secrecy capacity of the arbitrarily varying wiretap channel AVWC $\mathfrak{W}$ is upper bounded,*

$$C_S(\mathfrak{W}) \leq \min_{q\in\mathcal{P}(S)} \max_{U\to X\to(YZ)_q} (I(U, Y_q) - I(U, Z_q)) \ . \tag{3.54}$$

Proof. By Lemma 3.8 the capacity of the AVWC $\mathfrak{W}$ equals the capacity of the AVWC $\overline{\mathfrak{W}}$. Obviously, the set $\overline{\mathcal{W}} = \{(W_q^{\otimes n}, V_q^{\otimes n}) : q \in \mathcal{P}(S)\}$ which describes a compound wiretap channel is a subset of $\overline{\mathfrak{W}}^n = \{(W_{\tilde{q}}^n, V_{\tilde{q}}^n) : \tilde{q} \in \mathcal{P}(S^n), \tilde{q} = \prod_{i=1}^n q_i\}$. Now, because we can upper bound the secrecy capacity of the AVWC $\overline{\mathfrak{W}}$ by the secrecy capacity of the worst wiretap channel in the family $\overline{\mathfrak{W}}^n$, together with the foregoing we can upper bound it by the capacity of the worst channel of the compound channel $\overline{\mathcal{W}}$. Hence,

$$\begin{aligned} C_S(\mathfrak{W}) = C_S(\overline{\mathfrak{W}}) &\leq \inf_{\tilde{q}} C_S((W^n_{\tilde{q}}, V^n_{\tilde{q}})) \\ &\leq \inf_{q} C_S((W^n_{q}, V^n_{q})) = \inf_{q} C_S(W_q, V_q) \ , \end{aligned}$$

The minimum is attained because of the continuity of $C_S(W_q, V_q)$ on the compact set $\overline{\mathfrak{W}}$. □

Remark 3.13. *Consider the special case of an AVWC $\mathfrak{W} = \{(W_{s^n}, V_{r^n}) : s^n \in S_1^n,\ r^n \in S_2^n\}$, where both the state of the main channel $s \in S_1$ and the state of the eavesdropper's channel $r \in S_2$ in every time step can be chosen independently. In addition let us assume that there exist a channel $W_{q_1^*} \in \{W_{q_1} : q_1 \in \mathcal{P}(S_1)\}$, which is a degraded version of all other channels from $\{W_{q_1} : q_1 \in \mathcal{P}(S_1)\}$, and a best channel to the eavesdropper $V_{q_2^*}$ from the set $\{V_{q_2} : q_2 \in \mathcal{P}(S_2)\}$ (cf. Definition 3.9). Then in accordance with Section 2.3.4 the lower bound on the secrecy capacity given in Corollary 3.11 matches the upper bound from Theorem 3.12. Thus we can conclude that under the assumption, that the channel to legitimate receiver is not symmetrisable, the capacity of the AVWC $\mathfrak{W}$ is given by*

$$C_S(\mathfrak{W}) = \max_{U \to X \to (Y_{q_1^*} Z_{q_2^*})} (I(U, Y_{q_1^*}) - I(U, Z_{q_2^*})) \ .$$

Now in addition to Theorem 3.12 we give a multiletter formula of the upper bound of the secrecy rates. Therefore we need the following lemma used in analogy to Lemma 2.5.

Lemma 3.14. *For the arbitrarily varying wiretap channel AVWC $\mathfrak{W}^n$ the limit*

$$\lim_{n\to\infty} \frac{1}{n} \max_{U \to X^n \to (Y^n Z^n)_{\tilde{q}}} (\inf_{\tilde{q} \in \mathcal{P}(S^n)} I(U, Y^n_{\tilde{q}}) - \sup_{\tilde{q} \in \mathcal{P}(S^n)} I(U, Z^n_{\tilde{q}}))$$

exists.

The proof is carried out in analogy to Lemma 2.5 and therefore omitted.

Theorem 3.15. *The secrecy capacity of the arbitrarily varying wiretap channel*

AVWC $\mathfrak{W}$ is upper bounded by

$$C_S(\mathfrak{W}) \leq \lim_{n\to\infty} \frac{1}{n} \max_{U\to X^n\to(Y^nZ^n)_{\tilde{q}}} \big(\inf_{\tilde{q}\in\mathcal{P}(S^n)} I(U, Y^n_{\tilde{q}}) - \sup_{\tilde{q}\in\mathcal{P}(S^n)} I(U, Z^n_{\tilde{q}})\big) \ , \tag{3.55}$$

where $\tilde{q} = \prod_{i=1}^n q_i$, $q_i \in \mathcal{P}(S)$ and $Y^n_{\tilde{q}}, Z^n_{\tilde{q}}$ are the outputs of the channels $W^n_{\tilde{q}}$ and $V^n_{\tilde{q}}$ respective.

Proof. Let $(\mathcal{C}_n)_{n\in\mathbb{N}}$ be any sequence of (n, J_n) codes such that with

$$\sup_{s^n\in S^n} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{x^n\in A^n} E(x^n|j) W^{\otimes n}_{s_n}(D^c_j|x^n) =: \varepsilon_{1,n} \text{ and, } \sup_{s^n\in S^n} I(J, Z^n_{s^n}) =: \varepsilon_{2,n}$$

it holds that $\lim_{n\to\infty} \varepsilon_{1,n} 0 =$ and $\lim_{n\to\infty} \varepsilon_{2,n}$, where J denotes the random variable which is uniformly distributed on the message set $\mathcal{J}_n$. Because of Lemma 3.8 we obtain that for the same sequences of (n, J_n) codes

$$\lim_{n\to\infty} \sup_{\tilde{q}\in\mathcal{P}(S^n)} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{x^n\in A^n} E(x^n|j) W^n_{\tilde{q}}(D^c_j|x^n) = \lim_{n\to\infty} \varepsilon_{1,n} = 0 \tag{3.56}$$

and

$$\lim_{n\to\infty} \sup_{\tilde{q}\in\mathcal{P}(S^n)} I(J, Z^n_{\tilde{q}}) = \lim_{n\to\infty} \varepsilon_{2,n} = 0 \ . \tag{3.57}$$

Now let us denote another random variable by $\hat{J}$ with values in $\mathcal{J}_n$ determined by the Markov chain $J \to X^n \to Y^n_{\tilde{q}} \to \hat{J}$, where the first transition is governed by E, the second by $W^n_{\tilde{q}}$, and the last by the decoding rule. Now the proof is analogue to the proof of Proposition 2.6 in Section 2.3.2. For any $\tilde{q} \in \mathcal{P}(S^n)$ we have from data processing and Fano's inequality

$$(1 - \varepsilon_{1,n}) \log J_n \leq I(J, Y^n_{\tilde{q}}) + 1.$$

We then use the validity of the secrecy criterion (3.57) to derive

$$(1 - \varepsilon_{1,n}) \log J_n \leq I(J, Y^n_{\tilde{q}}) - \sup_{\tilde{q}} I(J, Z^n_{\tilde{q}}) + \varepsilon_{2,n} + 1$$

for any $\tilde{q} \in \mathcal{P}(S^n)$. Since the LHS does not depend on $\tilde{q}$ we end in

$$(1 - \varepsilon_{1,n}) \log J_n \leq \max_{U\to X^n\to Y^n_{\tilde{q}} Z^n_{\tilde{q}}} (\inf_{\tilde{q}} I(U, Y^n_{\tilde{q}}) - \sup_{\tilde{q}} I(U, Z^n_{\tilde{q}})) + \varepsilon_{2,n} + 1 \ .$$

Dividing by $n \in \mathbb{N}$ and taking $\limsup$ concludes the proof. □

Now if we consider the set $\overline{\mathcal{W}} = \{(W_q^{\otimes n}, V_q^{\otimes n}) : q \in \mathcal{P}(S)\}$ as a subset of $\overline{\mathfrak{W}}^n = \{(W_{\tilde{q}}^n, V_{\tilde{q}}^n) : \tilde{q} \in \mathcal{P}(S^n), \tilde{q} = \prod_{i=1}^n q_i\}$ and the same sequence $(\mathcal{C}_n)_{n\in\mathbb{N}}$ of (n, J_n) codes for the AVWC $\mathfrak{W}$ for which (3.56) and (3.57) holds, we can conclude that

$$\lim_{n\to\infty} \sup_{q\in\mathcal{P}(S)} \frac{1}{J_n} \sum_{j=1}^{J_n} \sum_{x^n\in A^n} E(x^n|j) W_q^{\otimes n}(D_j^c|x^n) \leq \lim_{n\to\infty} \varepsilon_{1,n} \tag{3.58}$$

and

$$\lim_{n\to\infty} \sup_{q\in\mathcal{P}(S)} I(J, Z_q^n) \leq \lim_{n\to\infty} \varepsilon_{2,n} \ , \tag{3.59}$$

with $\varepsilon_{1,n}$ and $\varepsilon_{2,n}$ as above. Then we can conclude with the same argumentation as in the previous proof,

Corollary 3.16. *The secrecy capacity of the arbitrarily varying wiretap channel AVWC $\mathfrak{W}$ is upper bounded by*

$$C_S(\mathfrak{W}) \leq \lim_{n\to\infty} \frac{1}{n} \max_{U\to X^n\to(Y^nZ^n)_q} \big(\inf_{q\in\mathcal{P}(S)} I(U, Y_q^n) - \sup_{q\in\mathcal{P}(S)} I(U, Z_q^n)\big) \ ,$$

where $q \in \mathcal{P}(S)$ and Y_q^n, Z_q^n are the outputs of the channels $W_q^{\otimes n}$ and $V_q^{\otimes n}$ respective.

Now, using standard arguments concerning the use of the channels defined by $P_{Y_q|U} = W_q \cdot P_{X|U}$ and $P_{Z_q|U} = V_q \cdot P_{X|U}$ instead of W_q and V_q and applying the assertion of Corollary 3.11 to the n-fold product of channels W_q and V_q, we are able to give the coding theorem for the multiletter case of the AVWC with a best channel to the eavesdropper.

Theorem 3.17. *Provided that there exist a best channel to the eavesdropper, the multiletter secrecy capacity $C_S(\mathfrak{W})$ of the AVWC $\mathfrak{W}$ is given by*

$$C_S(\mathfrak{W}) = \lim_{n\to\infty} \frac{1}{n} \max_{U\to X^n\to(Y^nZ^n)_q} \big(\inf_{q\in\mathcal{P}(S)} I(U, Y_q^n) - \sup_{q\in\mathcal{P}(S)} I(U, Z_q^n)\big) \ ,$$

if the channel to the legitimate receiver is not symmetrisable, and is zero otherwise.

Chapter 4

Conclusion

In this thesis we have considered two models of a discrete memoryless wiretap channel under channel uncertainty. If the exact channel realisation is unknown to the legitimate participants, but fixed during the transmission of a codeword, we call the resulting model the compound wiretap channel. In this model the users only know that the channel realisation belongs to a set of channels, which is known to them. If the channel state varies from time step to time step in the transmission of a codeword in an arbitrary and unknown manner the resulting model is called the arbitrarily varying wiretap channel. For both models we could derive secrecy capacity results under a strong secrecy criterion.

For the compound wiretap channel we could show that the secrecy capacity of compound wiretap channels with CSI at the transmitter matches the general upper bound on the capacity of compound wiretap channels given in an obvious way as the minimum secrecy capacity of all involved wiretap channels. We achieved this by using a coding technique developed by [Dev05], modified to the compound channel, and by developing a decoding strategy, which relieved us from the need of decoding the randomisation parameter used in the encoding procedure. With a slightly modified proof we could derive a lower bound for the secrecy capacity and determine a multi-letter expression for the secrecy capacity in the case of no CSI. If there exist a channel to the legitimate receiver such that all other channels to the legitimate receiver are more capable than that, together with a channel to the eavesdropper that is more capable than all other channels to the eavesdropper we could show that the lower bound on the secrecy capacity equals the general upper bound, actually in the case of no CSI. In the case of no CSI as well as in the case where the transmitter only knows the channel state to the legitimate receiver we gave a computable description of the secrecy capacity, if the channels to the eavesdropper are degraded versions of those to the legitimate receiver.

With the model of the arbitrarily varying wiretap channel AVWC it was possible to connect two problems of secure communication. In addition to the passive attack of eavesdropping the transmission, modelled by a wiretapper, the AVWC can be seen as a model, in which a jammer manipulates the transmission by an active attack, which is realised by the arbitrary change of the channel state in every time step. We could show that, under the assumption of a not symmetrisable channel to the legitimate user, the deterministic code secrecy capacity equals the random code secrecy capacity, if the channel is symmetrisable, then the deterministic code secrecy capacity is definitely zero. This result can be seen as a counterpart of the dichotomy result for single user AVC's, which was established by Ahlswede in [Ahl78]. Nevertheless, due to the large complexity of the model of the AVWC it was not possible to derive a computable expression for the secrecy capacity in the general case. In the proof we adapted the elimination technique for AVC's from [Ahl78] to random codes for the arbitrarily varying wiretap channels AVWC, which fulfill both the average error criterion and the strong secrecy criterion, which was carried out earlier by the authors of [Mol09] for a weaker secrecy criterion. Further, in the case of a "best" channel to the eavesdropper we could derive a lower bound on the deterministic code secrecy capacity, where in the proof we have used the robustification technique by Ahlswede [Ahl86] combined with our secrecy results for a compound wiretap channel. By establishing upper bounds on the secrecy capacity, we could give a multi-letter letter coding theorem in the case of a best channel to the eavesdropper. Finally, in the special case of the AVWC, where both the state of the main channel and the state of the eavesdropper's channel can be chosen independently combined with the assumption of the existence of a worst channel to the legitimate receiver and a best channel to the eavesdropper, we could give a full coding theorem.

Appendix A

Types and Typical Sequences

In the majority of the proofs of our coding results we make use of methods based on the concept of types and typical sequences. In this chapter we summarise some simple combinatorial lemmas and their proofs to give a short overview of the properties of types and typical sequences.

A.1 Types

Definition A.1. *The type of a sequence $x^n = (x_1, \dots, x_n) \in A^n$ (or its empirical distribution) is the probability distribution $p_{x^n} \in \mathcal{P}(A)$ defined by*

$$p_{x^n}(a) := \frac{N(a|x^n)}{n} \tag{A.1}$$

for all $a \in A$, where

$$N(a|x^n) := |\{i \in \{1, \dots, n\} : \; x_i = a\}| \tag{A.2}$$

denotes the number of the occurrences of a in x^n.
The set of all sequence $x^n \in A^n$ of type p is denoted by $\mathcal{T}_p^n$. A distribution $p \in \mathcal{P}(A)$ is called a type of sequences in A^n if $\mathcal{T}_p^n \neq \emptyset$. Further we define

$$\mathcal{P}_0(n, A) = \{p \in \mathcal{P} : \; p \text{ is type of sequences in } A^n\}. \tag{A.3}$$

Lemma A.2. *The number of different types of sequences in A^n is bounded by*

$$|\mathcal{P}_0(n, A)| < (n+1)^{|A|}. \tag{A.4}$$

Joint types of pairs of sequences are defined analogously to types.

Definition A.3. *The joint type of two sequences $x^n \in A^n$ and $y^n \in B^n$ is the distribution $p_{x^n,y^n} \in \mathcal{P}(A \times B)$ defined by*

$$p_{x^n,y^n} := \frac{N(a,b|x^n,y^n)}{n} \tag{A.5}$$

for all $a \in A$, $b \in B$, where

$$N(a,b|x^n,y^n) := |\{i \in \{1,\dots,n\} : \ (x_i,y_i) = (a,b)\}| \tag{A.6}$$

Joint types of sequences x^n, y^n can be described by types of $x^n \in A^n$ and a stochastic matrix $V : A \to \mathcal{P}(B)$ in the following way. Therefor let $V : A \to \mathcal{P}(B)$ some stochastic matrix with

$$p_{x^n,y^n} = p_{x^n}(a)V(b|a), \qquad (a \in A, b \in B). \tag{A.7}$$

It can be shown that $V(\cdot|a)$ exist for all a with $p_{x^n}(a) > 0$ and that it is uniquely determined by p_{x^n,y^n} and $P_{x^n}(a)$. Then with the stochastic matrices we can define the conditional types.

Definition A.4. *A sequence $y^n \in B^n$ has conditional type $V : A \to \mathcal{P}(B)$ given $x^n \in A^n$ if*

$$N(a,b|x^n,y^n) = N(a|x^n)V(b|a) \qquad (a \in A, b \in B). \tag{A.8}$$

For a given $x^n \in A^n$ and $V : A \to \mathcal{P}(B)$ the set

$$\mathcal{T}_V(x^n) := \{y^n \in B^n : \ y^n \textit{ has conditional type } V \textit{ given } x^n\} \tag{A.9}$$

is called the V-shell of x^n.

In the following we need the quantity $D(V\|W|p)$ which is called the conditional informational divergence. More precisely, let $V, W : A \to \mathcal{P}(B)$ stochastic matrices and $p \in \mathcal{P}(A)$ a probability distribution. Then we define

$$D(V\|W|p) := \sum_{a \in A} p(a)D(V(\cdot|a)\|W(\cdot|a)), \tag{A.10}$$

and for every $a \in A$, $D(V(\cdot|a)\|W(\cdot|a))$ is the informational divergence of the two probability distributions $V(\cdot|a)$, $W(\cdot|a) \in \mathcal{P}(B)$. The next lemma, that is mentioned here without proof, is part of Lemma 2.6 in [CK81].

Lemma A.5. *For a given x^n and stochastic matrices $V, W : A \to \mathcal{P}(B)$, such that*

$\mathcal{T}_V(x^n) \neq \emptyset$, it holds that

$$W^n(\mathcal{T}_V(x^n)|x^n) \leq 2^{-nD(V\|W|p_{x^n})}. \tag{A.11}$$

A.2 Typical Sequences

In what follows we use only strong typical sequences, whose definition relies on the empirical distribution of the sequences (cf. [CK81]) in contrast to the entropy-typical sequences.

Definition A.6. *Let $p \in \mathcal{P}(A)$ and $\delta > 0$. A sequence $x^n \in A^n$ is called typical (p-typical) with constant δ if*

$$|p_{x^n}(a) - p(a)| \leq \delta \qquad (a \in A), \tag{A.12}$$

and if $p(a) = 0$ then $p_{x^n}(a) = 0$ ($p_{x^n} \ll p$). Then $\mathcal{T}^n_{p,\delta}$ denotes the set of p-typical sequences $x^n \in A^n$.

From the definition it is obviously that

$$\mathcal{T}^n_{p,\delta} = \bigcup_{\hat{p} \in \mathcal{P}_0(n,A)} \{\mathcal{T}^n_{\hat{p}} : |\hat{p}(a) - p(a)| \leq \delta, \forall a \in A, \hat{p} \ll p\} \tag{A.13}$$

Definition A.7. *Let $W : A \to \mathcal{P}(B)$ a stochastic matrix and $\delta > 0$. A sequence $y^n \in B^n$ is called W-typical under the condition $x^n \in A^n$ with constant δ, if*

$$\left| \frac{1}{n} N(a, b|x^n, y^n) - \frac{1}{n} N(a, x^n) W(b|a) \right| \leq \delta \qquad (a \in A, b \in B) \tag{A.14}$$

and in addition, $W(b|a) = 0$ implies $N(a, b|x^n, y^n) = 0$ for all $a \in A$, $b \in B$. Then $\mathcal{T}_{W,\delta}(x^n)$ denotes the set of all W-typical sequences $y^n \in B^n$ under the condition $x^n \in A^n$ with constant δ.

The basic properties of these sets, that are needed in the most coding theorems, are summarised in the following lemmas.

Lemma A.8. *Fixing $\delta > 0$, for every $p \in \mathcal{P}(A)$ and $W : A \to \mathcal{P}(B)$ we have*

$$\begin{aligned} p^{\otimes n}(\mathcal{T}^n_{p,\delta}) &\geq 1 - (n+1)^{|A|} 2^{-nc\delta^2} \\ W^{\otimes n}(\mathcal{T}^n_{W,\delta}(x^n)|x^n) &\geq 1 - (n+1)^{|A||B|} 2^{-nc\delta^2} \end{aligned}$$

for all $x^n \in A^n$ *with* $c = 1/(2\ln 2)$*. In particular, there is* $n_0 \in \mathbb{N}$ *such that for each* $\delta > 0$ *and* $p \in \mathcal{P}(A)$*,* $W : A \to \mathcal{P}(B)$

$$\begin{aligned} p^{\otimes n}(\mathcal{T}^n_{p,\delta}) &\geq 1 - 2^{-nc'\delta^2} && \text{(A.15)} \\ W^{\otimes n}(\mathcal{T}^n_{W,\delta}(x^n)|x^n) &\geq 1 - 2^{-nc'\delta^2} && \text{(A.16)} \end{aligned}$$

holds with $c' = \frac{c}{2}$*.*

Proof. (A.15) and (A.16) follow directly from the two foregoing inequalities in the lemma. We prove (A.16), then the proof of (A.15) follows, if we choose A as a set consisting only of one element. Now let $x^n \in A^n$, $W : A \to \mathcal{P}(B)$ and $\delta > 0$ be given. Let $\mathcal{T}_{W,\delta}(x^n)$ be the set of all W-typical sequences $y^n \in B^n$ under condition x^n defined by (A.14). By analogy with (A.13) it can be shown that there exist V-shells $\mathcal{T}_{V_1}(x^n), \ldots, \mathcal{T}_{V_M} \neq \emptyset$ with

$$\begin{aligned} &\mathcal{T}_{W,\delta}(x^n) = \bigcup_{i=1}^{M} \mathcal{T}_{V_i}(x^n) \text{ and} \\ &M \leq (n+1)^{|A||B|}, \end{aligned} \tag{A.17}$$

where the distributions $V_i(b|a)$ are determined by

$$N(a, b|x^n, y^n) = N(a, x^n)V_i(b|a) \tag{A.18}$$

for all specific $y^n \in \mathcal{T}_{W,\delta}(x^n)$.

Now we consider the set $(\mathcal{T}_{W,\delta}(x^n))^c$, which is given by

$$\begin{aligned} (\mathcal{T}_{W,\delta}(x^n))^c = \{y^n \in B^n :& (\exists a \in A, b \in B : |\frac{N(a, b|x^n, y^n)}{n} - \frac{N(a|x^n)}{n} W(b|a)| > \delta) \\ &\text{or } (\exists a \in A, b \in B : N(a, b|x^n, y^n) > 0 \text{ and } W(b|a) = 0)\} \end{aligned} \tag{A.19}$$

First assume that for an $y^n \in B^n$ there exist $\bar{a} \in A$, $\bar{b} \in B$

$$N(\bar{a}, \bar{b}|x^n, y^n) > 0 \text{ and } W(\bar{b}|\bar{a}) = 0. \tag{A.20}$$

Then we obtain that

$$W^n(y^n|x^n) = \prod_{i=1}^{n} W(y_i|x_i) = \prod_{\substack{a \in A \\ b \in B}} W(b|a)^{N(a,b|x^n,y^n)},$$

and hence

$$W^n(y^n|x^n) = W(\bar{b}|\bar{a})^{N(\bar{a},\bar{b}|x^n,y^n)} \cdot \prod_{\substack{a\in A\setminus\{\bar{a}\}\\ b\in B\setminus\{\bar{b}\}}} W(b|a)^{N(a,b|x^n,y^n)} = 0, \tag{A.21}$$

which means that $y^n \in B^n$ under the assumptions that stated above does not exist. Thus the complement of $\mathcal{T}_{W,\delta}(x^n)$ is determined by the set

$$\{y^n \in B^n : (\exists a \in A, b \in B : |\frac{N(a,b|x^n,y^n)}{n} - \frac{N(a|x^n)}{n}W(b|a)| > \delta)\},$$

which is a subset of

$$B_{n,\delta}(x^n) := \{y^n \in B^n : \sum_{\substack{a\in A\\ b\in B}} |\frac{N(a,b|x^n,y^n)}{n} - \frac{N(a|x^n)}{n}W(b|a)| > \delta)\}.$$

As in (A.17) we can find V-shells $\mathcal{T}_{V_i} \neq \emptyset$, $i = 1,\dots,M$ with $M \leq (n+1)^{|A||B|}$, and

$$B_{n,\delta}(x^n) = \bigcup_{i=1}^{M} \mathcal{T}_{V_i}(x^n), \tag{A.22}$$

and thus to every $y^n \in B_{n,\delta}(x^n)$ there exist a V_i, $i = 1,\dots,M$ with

$$\sum_{\substack{a\in A\\ b\in B}} |\frac{N(a|x^n)}{n}V_i(b|a) - \frac{N(a|x^n)}{n}W(b|a)| > \delta. \tag{A.23}$$

Now with $p_{x^n} := \frac{N(a|x^n)}{n}$ we define the following probability distributions on $A \times B$ by

$$\begin{aligned} (p_{x^n} \circ V_i)(a,b) &:= p_{x^n}(a)V_i(b|a) \qquad (i \in \{1,\dots,M\}),\\ (p_{x^n} \circ W)(a,b) &:= p_{x^n}(a)W(b|a) \end{aligned} \tag{A.24}$$

for all $a \in A, b \in B$, and make use of the following equality, that is

$$\begin{aligned} D(V_i||W|p_{x^n}) &= \sum_{(a,b)\in A\times B} p_{x^n}(a)V_i(b|a)\log\frac{V_i(b|a)}{W(b|a)}\\ &= \sum_{(a,b)\in A\times B} p_{x^n}(a)V_i(b|a)\log\frac{p_{x^n}(a)V_i(b|a)}{p_{x^n}(a)W(b|a)}\\ &= D(p_{x^n}\circ V_i||p_{x^n}\circ W), \end{aligned} \tag{A.25}$$

for all $i \in \{1, \ldots, M\}$. Then we can summarise

$$\begin{aligned} W^n((\mathcal{T}_{W,\delta}(x^n))^c|x^n) \leq \sum_{i=1}^{M} W^n(B_{n,\delta}|x^n) &\leq \sum_{i=1}^{M} W^n(\mathcal{T}_{V_i}(x^n)|x^n) \\ &\leq \sum_{i=1}^{M} 2^{-nD(V_i||W|p_{x^n})}, \end{aligned} \tag{A.26}$$

where the second inequality follows by (A.22) and the last inequality from Lemma A.5. Now

$$\begin{aligned} W^n((\mathcal{T}_{W,\delta}(x^n))^c|x^n) &\leq \sum_{i=1}^{M} 2^{-nD(p_{x^n} \circ V_i||p_{x^n} \circ W)} \\ &\leq \sum_{i=1}^{M} 2^{-n\frac{1}{2\ln 2}\delta^2} \\ &\leq (n+1)^{|A||B|} 2^{-nc\delta^2}, \end{aligned} \tag{A.27}$$

wherein the second inequality follows from (A.23) and Pinsker's inequality and the last by the upper bound of M and with $c := \frac{1}{2\ln 2}$, which concludes the proof. □

Recall that for $p \in \mathcal{P}(A)$ and $W : A \to \mathcal{P}(B)$, $pW \in \mathcal{P}(B)$ denotes the output distribution generated by p and W and that $x^n \in \mathcal{T}^n_{p,\delta}$ and $y^n \in \mathcal{T}^n_{W,\delta}(x^n)$ imply that $y^n \in \mathcal{T}^n_{pW,2|A|\delta}$.

Lemma A.9. *Let $x^n \in \mathcal{T}^n_{p,\delta}$, then for $V : A \to \mathcal{P}(C)$*

$$\begin{aligned} |\mathcal{T}^n_{pV,2|A|\delta}| &\leq \alpha^{-1} \\ V^n(z^n|x^n) &\leq \beta \quad \textit{for all} \quad z^n \in \mathcal{T}^n_{V,\delta}(x^n) \end{aligned}$$

hold where

$$\alpha = 2^{-n(H(pV)+f_1(\delta))} \tag{A.28}$$

$$\beta = 2^{-n(H(V|p)-f_2(\delta))} \tag{A.29}$$

with universal $f_1(\delta), f_2(\delta) > 0$ satisfying $\lim_{\delta\to\infty} f_1(\delta) = 0 = \lim_{\delta\to\infty} f_2(\delta)$.

Proof. Cf. Lemma 2.13 and Lemma 2.6 from [CK81]. □

In addition we need a further lemma which will be used to determine the rates at which reliable transmission to the legitimate receiver is possible.

Lemma A.10. *Let $p, \tilde{p} \in \mathcal{P}(A)$ and two stochastic matrices $W, \widetilde{W} : A \to \mathcal{P}(B)$ be given. Further let $q, \tilde{q} \in \mathcal{P}(B)$ be the output distributions, the former generated by p and W and the latter by $\tilde{p}$ and $\widetilde{W}$. Fix $\delta \in (0, \frac{1}{4|A||B|})$. Then for every $n \in \mathbb{N}$*

$$q^{\otimes n}(\mathcal{T}^n_{\widetilde{W},\delta}(\tilde{x}^n)) \leq (n+1)^{|A||B|} 2^{-n(I(\tilde{p},\widetilde{W})-f(\delta))}$$

for all $\tilde{x}^n \in \mathcal{T}^n_{\tilde{p},\delta}$ and

$$q^{\otimes n}(\mathcal{T}^n_{W,\delta}(x^n)) \leq (n+1)^{|A||B|} 2^{-n(I(p,W)-f(\delta))}$$

for all $x^n \in \mathcal{T}^n_{p,\delta}$ holds for a universal $f(\delta) > 0$ and $\lim_{\delta \to 0} f(\delta) = 0$.

Proof. The proof can be found in [WBOB10] but is given here for the sake of completeness. Let $\tilde{x}^n \in \mathcal{T}^n_{\tilde{p},\delta}$ and $y^n \in \mathcal{T}^n_{\widetilde{W},\delta}(\tilde{x}^n)$. Then with the empirical distribution $p_{y^n}(b) = \frac{N(b|y^n)}{n}$, $b \in B$ it follows by Lemma 2.6 in [CK81] that

$$q^n(y^n) = 2^{-n(D(p_{y^n}||q)+H(p_{y^n}))} \leq 2^{-nH(p_{y^n})}, \tag{A.30}$$

where the inequality holds, since $D(p_{y^n}||q) \geq 0$. By Lemma 2.10 in [CK81], because $\tilde{x}^n \in \mathcal{T}^n_{\tilde{p},\delta}$ and $y^n \in \mathcal{T}^n_{\widetilde{W},\delta}(\tilde{x}^n)$, it follows that $y^n \in \mathcal{T}^n_{\tilde{q},2|X|\delta}$ and thus

$$\sum_{b \in B} |p_{y^n}(b) - \tilde{q}(b)| \leq 2|A||B|\delta \tag{A.31}$$

By the continuity of the entropy function it follows by 2.7 in [CK81] that

$$|H(p_{y^n}) - H(\tilde{q})| \leq -2|A||B|\delta \log \frac{2|A||B|\delta}{|B|} =: \varphi(\delta) \tag{A.32}$$

with $\lim_{\delta \to 0} \varphi(\delta) = 0$. By the last two inequalities we obtain that

$$q^n(\mathcal{T}^n_{\widetilde{W},\delta}(\tilde{x}^n)) \leq |\mathcal{T}^n_{\widetilde{W},\delta}(\tilde{x}^n)| 2^{-n(H(\tilde{q})-\varphi(\delta)}. \tag{A.33}$$

By the proof of Lemma 2.13 it follows that

$$|\mathcal{T}^n_{\widetilde{W},\delta}(\tilde{x}^n)| \leq (n+1)^{|A||B|} 2^{n(H(\widetilde{W}|\tilde{p})+\psi(\delta))} \tag{A.34}$$

with $\psi(\delta) > 0$ and $\lim_{\delta \to 0} \psi(\delta) = 0$. Then from (A.33) by defining $f(\delta) := \varphi(\delta) + \psi(\delta)$ we end up with

$$q^n(\mathcal{T}^n_{\widetilde{W},\delta}(\tilde{x}^n)) \leq (n+1)^{|A||B|} 2^{-n(I(\tilde{p},\widetilde{W})-f(\delta))} \tag{A.35}$$

The second assertion follows if we replace $\widetilde{W}$ by W and $\tilde{p}$ by p throughout the proof. □

Bibliography

[AC93] R. Ahlswede and I. Csiszar, *Common randomness in information theory and cryptography-part I: Secret sharing*, IEEE Transactions on Information Theory **39** (1993), no. 4, 1121–11132.

[AD89] R. Ahlswede and G. Dueck, *Identification via channels*, IEEE Transactions on Information Theory **35** (1989), no. 1, 15–29.

[Ahl70] R. Ahlswede, *A note on the existence of the weak capacity for channels with arbitrarily varying channel probability functions and its relation to shannon's zero error capacity*, The Annals of Mathematical Statistics **41** (1970), no. 3, 1027–1033.

[Ahl78] ______, *Elimination of correlation in random codes for arbitrarily varying channels*, Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete **44** (1978), 159–175.

[Ahl86] Rudolf Ahlswede, *Arbitrarily varying channels with states sequence known to the sender*, IEEE Transactions on Information Theory **32** (1986), no. 5, 621–629.

[Ahl08] R. Ahlswede, *General theory of information transfer: updated*, General Theory of Information Transfer and Combinatorics, Special Issue of Discrete Applied Mathematics **156** (2008), no. 9, 1348–1388.

[AW69] R. Ahlswede and J. Wolfowitz, *Correlated decoding for channels with arbitrarily varying channel probability functions*, Information and Control **14** (1969), no. 5, 457–473.

[AW70] ______, *The capacity of a channel with arbitrarily varying channel probability functions and binary output alphabet*, Z. Wahrscheinlichkeitstheorie verw. Gebiete **15** (1970), 186–194.

[AW80] ———, *A method of coding and its application to arbitrarily varying channels*, Journal of Combinatorics, Information and System Sciences **5** (1980), no. 1, 10–35.

[AW02] R. Ahlswede and A: Winter, *Strong converse for identification via quantum channels*, IEEE Transactions on Information Theory **48** (2002), no. 3, 569–579.

[BBS11a] I. Bjelaković, H. Boche, and J. Sommerfeld, *Capacity results for compound wiretap channels*, Proc. IEEE Information Theory Workshop (2011), 60–64.

[BBS11b] ———, *Secrecy results for compound wiretap channels*, accepted for publication in Problems of Information Transmission, 2011, Available from: `http://arxiv.org/abs/1106.2013v1`.

[BBS12] ———, *Capacity results for arbitrarily varying wiretap channels*, accepted for publication in LNCS Volume in Memory of Rudolf Ahlswede, 2012.

[BBT59] D. Blackwell, L. Breiman, and A. J. Thomasian, *The capacity of a class of channels*, Ann. Math. Stat. **30** (1959), 1229–1241.

[BBT60] D. Blackwell, L. Breiman, and A.J. Thomasian, *The capacity of certain channel classes under random coding*, The Annals of Mathematical Statistics **31** (1960), 558–567.

[BL08] M. Bloch and J.N. Laneman, *On the secrecy capacity of arbitrary wiretap channel*, Forty-Sixth Annual Allerton Conference, Allerton House, Illinois, USA (2008).

[CK78] I. Csiszar and J. Körner, *Broadcast channels with confidential messages*, IEEE Transactions on Information Theory **24** (1978), no. 3, 339–348.

[CK81] I. Csiszar and J. Körner, *Information theory: Coding theorems for discrete memoryless systems*, Akademiai Kiado, 1981.

[CN88] I. Csiszar and P. Narayan, *The capacity of the arbitrarily varying channel revisited: Positivity, constraints*, IEEE Transactions on Information Theory **34** (1988), no. 2, 181–193.

[Csi96] I. Csiszar, *Almost independence and secrecy capacity*, Problems of Information Transmission **32** (1996), no. 1, 40–47.

[CWY04] N Cai, A. Winter, and R.W. Yeung, *Quantum privacy and quantum wiretap channel*, Problems of Information Transmission **40** (2004), no. 4, 318–336.

[Dev05] Igor Devetak, *The Private Classical Capacity and Quantum Capacity of a Quantum Channel*, IEEE Transactions on Information Theory **51** (2005), no. 51, 44–55.

[DP09] Devdatt P. Dubhashi and Alessandro Panconesi, *Concentration of measure for the analysis of randomized algorithms*, Cambridge University Press, 2009.

[Eri85] T. Ericson, *Exponential error bounds for random codes in the arbitrarily varying channel*, IEEE Transactions on Information Theory **31** (1985), no. 1, 42–48.

[Fek23] M. Fekete, *Über die Verteilung der Wurzeln bei gewissen algebraischen Gleichungen mit ganzzahligen Koeffizienten*, Mathematische Zeitschrift **17** (1923), 228–249.

[LKPS08] Y. Liang, G. Kramer, H.V. Poor, and S. Shamai, *Compound Wiretap Channels*, EURASIP Journal on Wireless Communications and Networking (2008).

[Mol09] Ebrahim MolavianJazi, *Secure communications over arbitrarily varying wiretap channels*, Master's thesis, Graduate School of the University of Notre Dame, 2009.

[MW00] U.M. Maurer and S. Wolf, *Information-theoretic key agreement: From weak to strong secrecy for free*, in Advances in Cryptology-Eurocrypt 2000, Lecture Notes in Computer Science **1807** (2000), 351–368.

[Sha49] C.E. Shannon, *Communication theory of secrecy systems*, The Bell Systems Technical Journal **28** (1949), 656–715.

[Sha56] ———, *The zero error capacity of a noisy channel*, IRE Trans. Information Theory IT-2 (1956), 8–19.

[WBOB10] Rafael F. Wyrembelski, I. Bjelaković, T. J. Oechtering, and H. Boche, *Optimal coding strategies for bidirectional broadcast channels under channel uncertainty*, IEEE Transactions on Communications **58** (2010), no. 10, 2984–2994.

[Wol60] J. Wolfowitz, *Simultaneous channels*, Arch. Rational Mech. Anal. **4** (1960), no. 4, 371–386.

[Wol78] ———, *Coding theorems of information theory*, 3rd ed., Springer-Verlag, 1978.

[Wyn75] A.D. Wyner, *The wire-tap channel*, The Bell System Tech. J. **54** (1975), no. 8, 1355–1387.