Watermarking Identification Codes with Related Topics on Common Randomness

Rudolf Ahlswede and Ning Cai *

Abstract

Watermarking identification codes were introduced by Y. Steinberg and N. Merhav. In their model they assumed that

(1) the attacker uses a single channel to attack the watermark and both, the iformation hider and the decoder, know the attack channel;

(2) the decoder either completely he knows the covertext or knows nothing about it.

Then instead of the first assumption they suggested to study more robust models and instead of the second assumption they suggested to consider the case where the information hider is allowed to send a secret key to the decoder according to the covertext.

In response to the first suggestion in this paper we assume that the attacker chooses an unknown (for both information hider and decoder) channel from a set of channels or a compound channel, to attack the watermark. In response to the second suggestion we present two models. In the first model according to the output sequence of covertext the information hider generates side information componentwise as the secret key. In the second model the only constraint to the key space is an upper bound for its rate.

We present lower bounds for the identification capacities in the above models, which include the Steinberg and Merhav results on lower bounds. To obtain our lower bounds we introduce the corresponding models of common randomness. For the models with a single channel, we obtain the capacities of common randomness. For the models with a compound channel, we have lower and upper bounds and the differences of lower and upper bounds are due to the exchange and different orders of the max–min operations.

Index Terms: Watermarking, identification, compound channel, common randomness.

*This paper is supported in part by INTAS 00–738

1 Introduction

Watermarking technique is a way to embed secret information into a given message, say image, that cannot be removed nor deciphered without access to a secret key. It can be used to protect copy right. Watermarking is now a major activity in audio, image, and video processing and standardization efforts for JPEG–2000, MPEG–4 and Digital Video Disks are underway.

One way to analyse watermarking problems is to regard them as communication systems e.g., [16], [20], [27], [28], [29], [30], and [32]. In these systems the messages, which are called covertext, are generated by an information source. An information hider, whom we often call encoder because of his role in the system, has full access to the information source of covertexts and the set of secret messages. These secret messages are independent of the covertext, they are uniformly generated from the set, and will be called watermark. The role of the information hider, or encoder, is to embedd the watermark in the covertext. When the embedding changes the covertext, it disturbs the message. To guarantee the quality of the watermarked message, we certainly would like not too much distortion. That is, for a given distortion measure, the distortion between the original covertext and the watermarked message in average may not exceed a given constant. An attacker wants to remove the watermark from the watermarked message without distorting the message too much i.e., the distortion between the covertext and the message corrupted by the attacker is not too large with respect to a certain distortion measure. Finally a decoder tries to recover the watermark from the corrupted message correctly with high probability. As the attacker is allowed to use a random strategy, we assume that the attacker uses a noisy chanel to attack the watermark. Depending on the models the attacker may choose various channels and the encoder and decoder share different resources (e.g., secret key, side information, etc.).

Among huge contributions on watermarking we here briefly review two of them. In [28] P. Moulin and J.A. O'sullivan obtained the capacity for the watermarking codes under the assumptions that the covertexts are generated from a memoryless source, the distortions are sum–type and the attack channels are compound channels whose states are known to the decoder but unknown to the encoder. The strategies of encoder–decoder and attacker are discussed as well.

Identification codes for noisy channels were introduced by R. Ahlswede and G. Dueck for the situation in which the receiver needs to identify whether the coming message equals a specified one. If not, then they don't care what it is [11]. It turned out that this weaker requirement dramatically increased the sizes of messages sets which could be handled: double exponential grown in the block lengths of codes. Identification is much faster than transmission!

Y. Steinberg and N. Merhav notice that in most cases people check watermarks in order to identify them (e.g. copyright) rather than recognize them and so they introduced identification codes to watermarking models [32]. In their models the attack channels are single memoryless channels. That means the attacker's random strategy is known by information hider (encoder) and decoder. They notice that the assumption is not robust and so suggested to study more robust models. As to the resources shared by encoders and decoders they consider two cases, the decoder either completely know the covertext or he knows nothing about it. (In all cases the attacker must not know the covertext because otherwise there would be no safe watermarking.)

By considering common randomness between encoder and decoder, they obtained lower bounds to the capacities of watermarking identification in both cases and the upper bounds easily followed from a theorem in [31]. The lower and upper bounds are tight in the former case but not in the latter case. As Y. Steinberg and N. Merhav only studied two extremal cases, they suggested to consider the more general case, that the decoder may obtain partial information, about the covertext, say key, from the encoder via a secure noiseless channel. The exponent of error probability was discussed as well.

In the present paper we deal with these two problems. But before turning to our result, we draw readers' attention to common randomness, which – as noticed in [12] – plays a central role in identification problems. It does so also in [32] and here R. Ahlswede and G. Dueck discovered in [12] that common randomness shared by encoder and decoder can be used to construct identification codes and therefore the rate of common randomness (in the sense of first order of logarithm) is not larger than the rate of identification codes (in the sense of the second order of logarithm). In general the capacities of common randomness shared by the encoder and the decoder may be smaller than the capacities of identification. Examples for discrete channels and Gaussian channels were presented in [14] and [17] respectively. Notice that the sizes of the input alphabets of the former channel is growing super exponentially as the length of codes and the sizes of the input alphabets of the latter is infinity. In fact it is seen from [31] that for any channel, whose input alphabet is exponentially increasing in the case that strong converse holds, the rates of common randomness and identification codes are the same.

The topic of common randomness has been become more and more popular e.g., [6], [9], [10], [23], [26], [33], [34], etc. Common randomness may be applied to cryptography, (e.g., [9], [18], [23], [26]), identification (e.g., [5], [11], [12], [10], [15], [18]), and arbitrarily varying channels (e.g., [1], [2], [8], [10]). For the first two applications the rates are important and the distributions of common randomnesses are required nearly uniformly. For cryptography certain secure conditions additionally needed. For the last application one has to face in the difficulty made by the jammer and find a smart way to generate the common randomness.

Now let us return to the two suggestions by Steinberg and Merhav. For the first suggestion we assume in our models, attackers are allowed to choose a channel arbitrarily from a set of memoryless channels to attack watermarks and neither encoders nor decoders know the attack channels. This is known as compound channel in Information Theory.

The assumption makes our models slightly more robust than that in [28] since in [28] the decoders are supposed to know the attack channels.

For the second suggestion we set up two models. In out first model we assume the encoder generates a random variable at time t according to component at time t of the output sequence of covertext source and certain probability and sends it to decoder via a securet channel. In this case the "key" actually is a side information of covertext shared by encoder and decoder. We obtain the first and the second models in [32] if we choose the side information equal to covertext almost surely and independent of covertext respectively. So our first model contain both models in [32]. In our second model the encoder is allowed to generate a key according to the covertext (but independently on watermark) in arbitrary way and sends the key to decoder through a securet chanel with rate R_K. Obviously in our second model the key can be generated in a more general way than in our first model. For all combinations of above assumptions, we obtain lower bounds to the identification capacitite, which contains both lower bounds in [32] as special cases.

To obtain our lower bounds to identification capacities, for each combination, we introduce a corresponding model of common randomness and obtain lower and upper bound to its capacity. For the single channel the two bound is closed for compound channel the gap between two bounds is up to the order of max–min. In addition, we show a lower bound to common randomness in [32] in fact is tight, which supports a conjecture in [32].

We must point out that our assumption of compound attack channels is still far from the most robust and practical assumption although according to our knowledge, it is most robust and practical existing assumption in this area. Actually the attacker has much more choices.

- He does not necessarily use a memoryless channel and stead he can chooses a channel with finite memory.
- The attacker may change the states time by time i.e., he may use an arbitrarily varying channel.
- The attacker knows output of the channel; even at time t, he know the output at time $t' > t$, since all outputs in fact are shosen by hemself/herself. So the attacker may use this information to choose attack channel. This clearly makes the attack much more efficient.

So there is still a long way for us to achieve the most practical results and it previde a wide space for future research.

The rest part of the paper is organized as follows. In the next section we present the notation used in the paper. Our models and results are stated in Section 3 and Section 4 respectively. The direct parts of coding theorems of common randomness are proven in Section 5 and their converse parts are proven in Section 6. In Section 7 we briefly review the observation in [12] on the relation of identification and common randomness and therefore the lower bounds to the identification capacities from capacities of common randomness. Finally the converse theorem for a model in [32] is proven in Section 8.

2 The Notation

Our notation in this paper is fairly standard. log and ln stand for the logarithms with bases 2 and e respectively and a^z is often written as $exp_a[z]$. The random variables will be denoted by capital letters L, U, V, X, Y, Z etc. and their domains are often denoted by the correspondent script letters $\mathcal{L}, \mathcal{U}, \mathcal{V}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}$ etc. But in some special cases it may be exceptional. When we denote a set by a script letter (for example, $\mathcal{X}$), its element is often denoted by the corresponding lower letter (for example x). $\mathcal{X}^n$ is the nth cartesian power of the set $\mathcal{X}$ and $x^n = (x_1, x_2, \ldots x_n)$ is the sequence of length n. $Pr\{\mathcal{E}\}$ is the probability of that the event $\mathcal{E}$ occurs and $\mathbf{E}[\cdot]$ is the operator of expectation. P_X, P_{XY}, $P_{Z|X}$ etc. will stand for the distribution of random variable X , the joint distribution of the random variables (X, Y), the conditional distribution of random variable Z under the condition that X is given respectively. When we write a probability distribution as P^n, we mean that it is a product distribution of P and similarly a discrete memoryless channel of length n with stochastic matrix W is written as W^n.

Throughout this paper $\mathcal{T}_U^n$, $\mathcal{T}_{UV}^n, \mathcal{T}_{U|VL}^n(v^n l^n)$ etc. will denote the sets of typical, joint typical, and conditional typical sequences and the corresponding sets of δ- typical, joint typical, and conditional typical sequences are written as $\mathcal{T}_U^n(\delta)$, $\mathcal{T}_{UV}^n(\delta), \mathcal{T}_{U|VL}^n(v^n l^n, \delta)$ etc.. We always understand these sets are not empty when we use the notation. When we introduce a set of typical sequences (for example, say $\mathcal{T}_Z^n$), it is understood that the correspondent random variable(s) (i.e., Z in the example) with the (joint) type as distribution (P_Z) is introduced at the same time. For a subset $\mathcal{A}$ of sequences of length n we write $\mathcal{A}_U = \mathcal{A} \cap \mathcal{T}_U^n$ and analogously $\mathcal{A}_{UV}, \mathcal{A}_{U|VL}(v^n l^n), \mathcal{A}_U(\delta), \mathcal{A}_{UV}(\delta), \mathcal{A}_{U|VL}(v^n l^n, \delta)$ etc. $|\mathcal{T}_U^n|$ and the common values of $|\mathcal{T}_{U|L}^n(l^n)|$, $l^n \in \mathcal{T}_L^n$ some times are written as t_U, $t_{U|L}$ etc. respectively (the length n of the sequences are understood by the context). Analogously $t_U(\delta)$, $t_{Y|X}(\delta)$ etc, also are used.

3 The Models

3.1 Watermarking Identification Codes

In this subsection, we state our models for the simpler case that the attacker choose a single channel to attack the watermark and both the encoder (information hider) and the decoder know the attack channel. In the next subsection, we introduce the corresponding models of common randomness. In the last subsection of the section, we assume the attack chooses a channel unknown by both encoder and decoder from a set of channels and replace the single channel by a compound channel.

Let $\mathcal{V}$ be a finite set, and V be a random variable taking values in $\mathcal{V}$. Then the covertext is assumed to be generated by an memoryless information source $\{V^n\}_{n=1}^{\infty}$ with generic V. The watermark is uniformly chosen from a finite set $\{1, 2, \ldots, M\}$ independently on the convext. The encoder is fully accessed the covertext and source of watermark and encodes the outputs of covertext v^n and of watermark m jointly to a sequence $x^n\big(= x^n(v^n, m)\big)$

with the same length of sequence of covertext. The attack use a single discrete memoryless channel W to attack the watermarked sequence x^n i.e., to change x^n to y^n with probability $W^n(y^n|x^n) = \prod_{t=1}^{n} W(y_t|x_t)$. Usually for practical reason people assume that v^n, x^n, and y^n are chosen from the same finite alphabet, but for convenience of notation we assume they are from finite alphabets $\mathcal{V}$, $\mathcal{X}$, and $\mathcal{Y}$ respectively. The encoding mapping in general disturbs the covertext. To measure the distortion, we introduce a sum type distortion measure, watermarking distortion measure (WD–measure) ρ, such that for all $v^n = (v_1, \ldots, v_n) \in \mathcal{V}^n$, $x^n = (x_1, \ldots, x_n) \in \mathcal{X}^n$,

$$\rho(v^n, x^n) = \sum_{t=1}^{n} \rho(v_t, x_t), \tag{1}$$

where for all $v \in \mathcal{V}$, $x \in \mathcal{X}$ $0 \le \rho(v, x) \le \Delta$, for a positive constant Δ.

By definition, there should be certain distortion constraint to the output of attack channel. But nowwe are given a memoryless attack channel and we may omit the constraint simply by assum that the attack channel satisfies the constraint automatically. This clearly does not loss generality. Next we have to set up the key–resources shared by encoder and decoder, according to which we distinguish our watermarking identification codes into watermarking identification codes with side information (WIDCSI) and watermarking identification codes with secure key (WIDCK) as follows.

Watermarking identification codes with side information (WIDCSI): In the first case, we assume that the encoder can generate "a component of a key", $L_t = l_t$ at the time t according to the current output of covertext $V_t = v_t$ and a given conditional distribution $P_{L|V}(\cdot|v)$. That is, the sender generates a sequence $L^n = (L_1, L_2, \ldots, L_n) = l^n = (l_1, l_2, \ldots, l_n)$ with probability $P^n_{L|V}(l^n|v^n)$ if the source outputs a sequence v^n of covertext and then sends it to the decoder. The latter try to recover the watermark from the invalidated message by the attacker with the help of the side information $L^n = l^n$. In this case the key-resource is actually governed by the conditional distribution $P_{L|V}$ or equivalently the joint probability distribution P_{VL}. So it can be understood as a pure side information at both sides of encoder and decoder instead of a " secure key". That is, if $\{V^n\}_{n=1}^{\infty}$ is a memoryless covertext with generic V, and $\{L^n\}_{n=1}^{\infty}$ is a side information observed by both encoder and decoder, then $\{(V^n, L^n)\}$ is a correlated memoryless source with generic (V, L). Thus the decoder can learn some thing about the covertext from the side information whereas the attacker knows nothing about it. A WIDCSI code becomes a "watermarking identification code with side information at transmitter and receiver" in [32] when V and L have the same alphabet and equal to each other almost surely and it becomes a "watermarking identification code with side information at the transmitter only" in [32] if V and L are independent.So the two codes defined in [32] is really the extreme cases of WIDCI codes.

Watermarking identification codes with secure key (WIDCK): In this case we assume the encoder may generate a key $K_n = K_n(v^n)$ according to the whole output sequence $V^n = v^n$ of the random covertext V^n in an arbitrary way and send it to the decoder through a secure (noiseless) channel so that the attacker has absolutely has no knowledge about the covertext (except its distribution) nor the key. Since for given output v^n of the covertext the encoder may generate the K_n randomly, a WIDCSI code is a special WIDCK code. We shall see that in general the latter is more powerful. Notice that a deterministic key function of output of covertext is a special random key. Finally of course the size of the key must be constraint. We require it exponentially increasing with the length of the code and its rate upper bounded by the key rate R_K. When the key rate is larger than the covertext entropy $H(V)$ the encoder certainly may inform the receiver the output of covertext. However "the rest part" of the key may serve as a common randomness between the communicators which increases the identification capacity (see [12], [10], and [32]).

Thus an $(n, R, \lambda_1, \lambda_2, D_1)$ WIDCSI code is a system $\{Q_m, \mathcal{D}_m(l^n) : \ l^n \in \mathcal{L}^n, \ m \in \mathcal{M}\}$ for $\mathcal{M} = \{1, 2, \ldots, M\}$ satisfying the following conditions.

- $Q_m, m = 1, 2, \ldots, M$ are stochastic matrices $Q_m : \ \mathcal{V}^n \times \mathcal{L}^n \longrightarrow \mathcal{X}^n$ such that for $m = 1, 2, \ldots, M$,

$$\sum_{v^n \in \mathcal{V}^n, l^n \in \mathcal{L}^n} P^n_{VL}(v^n, l^n) \sum_{x^n \in \mathcal{X}^n} Q_m(x^n | v^n, l^n) \rho(v^n, x^n) \leq D_1, \tag{2}$$

where P_{VL} is the joint distribution of the generic (V, L).

- For all $l^n \in \mathcal{L}^n, m \in \mathcal{M}$, $\mathcal{D}_m(l^n) \subset \mathcal{Y}^n$ and for all $m \in \mathcal{M}$,

$$\sum_{v^n \in \mathcal{V}^n, l^n \in \mathcal{L}^n} P^n_{VL}(v^n, l^n) \sum_{x^n \in \mathcal{X}^n} Q_m(x^n | v^n, l^n) W^n(\mathcal{D}_m(l^n) | x^n) > 1 - \lambda_1, \tag{3}$$

and for all $m, m' \in \mathcal{M}, m \neq m'$,

$$\sum_{v^n \in \mathcal{V}^n, l^n \in \mathcal{L}^n} P^n_{VL}(v^n, l^n) \sum_{x^n \in \mathcal{X}^n} Q_m(x^n | v^n, l^n) W^n(\mathcal{D}_{m'}(l^n) | x^n) < \lambda_2. \tag{4}$$

λ_1 and λ_2 is called the errors of the first and the second kinds of the code

- The rate of the code is

$$R = \log \log M. \tag{5}$$

Watermarking identification codes with secure key (WIDCK):

Next we define WIDCK code. Let $\{V^n\}_{n=1}^{\infty}$ be a memoryless covertext with generic V and alphabet $\mathcal{V}$, the attack channel W be memoryless, and WD-measure ρ be as (1). Then an $(n, R, R_K, \lambda_1, \lambda_2, D_1)$ WIDCK code is a system $\{Q^*_m, \mathcal{D}^*_m(k_n), W_{K_n} : \ m \in \mathcal{M}, k_n \in \mathcal{K}_n\}$ for $\mathcal{M} = \{1, 2, \ldots, M\}$ satisfying the following conditions.

- $\mathcal{K}_n$ is a finite set, which will be called the key book, with

$$\frac{1}{n} \log |\mathcal{K}_n| \leq R_K. \tag{6}$$

R_K will be called key rate.

– W_{K_n} is a stochastic matrix, $W_{K_n} : \mathcal{V}^n \longrightarrow \mathcal{K}_n$. The output random variable will be denoted by K_n when the random covertext V^n is input to the channel W_{K_n} i.e., the pair of random variables (V^n, K_n) have joint distribution $P_{V^n K^n}(v^n, k^n) = P_V^n(v^n) W_{K_n}(k_n|v^n)$, $v^n \in \mathcal{V}$ $k_n \in \mathcal{K}_n$. In particular K_n may be a deterministic function of output of covertext and in this case we write $K(\cdot)$ as a function defined on $\mathcal{V}^n$. Note that the choice of K_n does NOT depend on the message $m \in \mathcal{M}$ since the key should independent of the protected message.

– $Q_m^*, m = 1, 2, \ldots, M$ are stochastic matrices from $\mathcal{V}^n \times \mathcal{K}_n$ to $\mathcal{X}^n$, (the alphabet of the input of the attack channel), such that

$$\sum_{v^n \in \mathcal{V}^n} P_V^n(v^n) \sum_{k_n \in \mathcal{K}_n} W_{K_n}(k_n|v^n) \sum_{x^n \in \mathcal{X}^n} Q_m^*(x^n|v^n, k_n)\rho(v^n, x^n) \leq D_1. \tag{7}$$

– For all $k_n \in \mathcal{K}_n, m \in \mathcal{M}$, $\mathcal{D}_m(k_n) \subset \mathcal{Y}^n$ and for all $m \in \mathcal{M}$, the error of first kind

$$\sum_{v^n \in \mathcal{V}^n} P_V^n(v^n) \sum_{k_n \in \mathcal{K}_n} W_{K_n}(k_n|v^n) \sum_{x^n \in \mathcal{X}^n} Q_m^*(x^n|v^n, k_n) W^n(\mathcal{D}_m(k_n)|x^n) > 1 - \lambda_1, \tag{8}$$

and for all $m, m' \in \mathcal{M}$ $m \neq m'$,

$$\sum_{v^n \in \mathcal{V}^n} P_V^n(v^n) \sum_{k_n \in \mathcal{K}_n} W_{K_n}(k_n|v^n) \sum_{x^n \in \mathcal{X}^n} Q_m^*(x^n|v^n, k_n) W^n(\mathcal{D}_{m'}(k_n)|x^n) < \lambda_2. \tag{9}$$

– Finally the rate of the code is defined in (5).

The capacities of the codes of the two types are defined in the standard way and denoted by $C_{WIDSI}((V, L), W, D_1)$ and $C_{WIDK}(V, W, R_K, D_1)$ respectively, where (V, L) and V are the generic of memoryless correlated source and source respectively, W is an attack memoryless channel, R_K is the key rate, and D_1 is the distortion criterion.

3.2 The Common Randomness

We speak of the common randomness between two (or among more than two) persons who share certain common resources, which may be correlated sources and/or (noisy or noiseless) channels. The common randomness between these two persons is just two random variables with common domain, which converges each other respect to probability. According to the resources different models are established.

For the purpose to build watermarking identification codes we need the following two kinds of common randomness. In the following two models of common randomness, the correlated source $\{(V^n, L^n)\}_{n=1}^{\infty}$ corresponds to the source of covertext and side information and the memnoreless channel W corresponds the attack channel in the models of watermarking idemtification. The K_n in the Model II corresponds the key in the model of WIDCK.

Model I: Two-source with a constraint noisy channel

Let $\{(V^n, L^n)\}_{n=1}^{\infty}$ be a correlated memoryless source with two components, alphabets $\mathcal{V}$ and $\mathcal{L}$, and generic (V, L). Assume that there are two persons, say sender (or encoder) and receiver (or decoder). The sender may observe the whole output of the source (V^n, L^n) whereas only the output of the component L^n is observable for the receiver. To establish common randomness the sender may send message through memoryless channels W with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ under certain constraint condition (specified below). The receiver is not allowed to send any message to the sender. The sender first chooses a channel code with set of codewords $\mathcal{U} \subset \mathcal{X}^n$ with the same length n as output sequence of the source and generates a random variable M, his/her "private randomness" taking values uniformly in a finite set $\mathcal{M}$ (, which is exponentially increasing as the length n of the source sequences increases) and independent of (V^n, L^n) of the output of the source. Assume a (sum type) distortion measure ρ in (1) and a criterion of distortion D_1 are given. According to the output $(V^n, L^n) = (v^n, l^n)$ of the source and the output of his/her private randomness $M = m$ the sender chooses a codeword $x_m(v^n, l^n) \in \mathcal{U}(\subset \mathcal{X}^n)$ such that the average of the distortion between the codeword and the component $V^n = v^n$ of the correlated source may not exceed D_1. Namely,

$$\frac{1}{n} \sum_{m \in \mathcal{M}} P_M(m) \sum_{v^n \in \mathcal{V}^n} \sum_{l^n \in \mathcal{L}^n} P_{VL}(v^n, l^n) \rho(x_m(v^n, l^n), v^n) \leq D_1. \tag{10}$$

The receiver receives an output sequence $y^n \in \mathcal{Y}^n$ with the probability $W^n(y^n | x_m(v^n, l^n))$ if the sender input the codeword $x_m(v^n, l^n)$ to the channel. We also allow to choose $x_m(v, l^n)$ as a random input sequence instead of deterministic one (it is more convenient in the proof). Finally for a finite set $\mathcal{A}$ which typically increases exponentially when the length n of the source increases, i. e., for a constant κ

$$\frac{1}{n} \log |\mathcal{A}| \leq \kappa, \tag{11}$$

the sender creates a random variable F with range $\mathcal{A}$, according to the outputs of (V^n, L^n) and M, through a function

$$F : \mathcal{V}^n \times \mathcal{L}^n \times \mathcal{M} \longrightarrow \mathcal{A} \tag{12}$$

and the receiver creates a random variable G according to the output of the channel W^n and the output of the component L^n of the source, through a function

$$G : \mathcal{L}^n \times \mathcal{Y}^n \longrightarrow \mathcal{A}. \tag{13}$$

After the terminology in [10] we called the pair of random variables (F, G) generated in the above way permissible and say that a permissible pair (F, G) represents λ-common randomness if

$$Pr\{F \neq G\} < \lambda. \tag{14}$$

Typically λ should be an arbitrarily small but positive real number when length n of source sequences is arbitrarily large. It is not hard to see that under the conditions (11) and (14) by Fano inequality, the entropy rates $\frac{1}{n}H(F)$ and $\frac{1}{n}H(G)$ are arbitrarily close if λ in (14) is arbitrarily small. This was observed in [10]. Thus we can choose any one from the pair of entropy rates, say $\frac{1}{n}H(F)$ as the rate of common randomness.

A pair of real numbers (r, D_1) is called achievable for common randomness if for arbitrary positive real numbers ϵ, λ, μ and sufficiently large n (depending on ϵ, λ and μ) there exists a λ-common randomness satisfying (10) – (14), such that

$$\frac{1}{n}H(F) > r - \epsilon \tag{15}$$

and

$$\sum_{a\in\mathcal{A}} \mid Pr\{F = a\} - \frac{1}{|\mathcal{A}|} \mid < \mu. \tag{16}$$

The last condition says that the common randomness is required to be nearly uniform and we call it nearly uniform condition. We set it for reducing the errors of second kind of identification codes. The set of achievable pairs is called common randomness capacity region. For fixed D_1 the common randomness capacity (CR-capacity) is $C_{CRI}((V, L), W, D_1) = \max\{r : (r, D_1) \text{ is achievable}\}$.

Notice that there is no limit to the amount of sender's private randomness in the present model and the next model, Model II. However because of the limit of the capacity of the channel the " extra" private randomness is useless.

We remark here that this model is different from the model (i) in [10] in three points. First, the channel connect the sender and receiver is noiseless with constraint that rate $\leq R$ in the model (i) of [10] whereas in general it is noisy in current model. More importantly, because of the requirement of distortion the source not only plays a role of " side information" but also a role of "constrainer". That is, to fight for reducing the distortion the sender has to choose codewords properly. This makes the transformation more difficult. To see that let us consider an extremal case that the component L^n of the source is a constant. In this case the source makes no difference at all in the model (i) of [10] and therefore the common randomness capacity is trivially equal to capacity of the channel. But in this case for the present model the source makes difference i.e., because of it the sender may not choose the codewords freely and therefore the common randomness is reduced. To obtain the CR-capacity region for this model is also absolutely non-trivial. Finally in this model the sender and receiver observe the output $(V^n, L^n) = (v^n, l^n)$ and $L^n = l^n$ respectively. The common randomness before the transmission, is equal to $H(L^n) = I(V^n, L^n; L^n)$ the mutual information between the two observations. So it seems to be not surprising our characterization in Theorem 4.1 is quite different from that in Theorem 4.1 of [10] and it cannot obtain simply by substituting rate of noiseless channel by the capacity of the noisy channel.

Model II: Two-source with a constraint noisy channel and a noiseless channel

It is clear that our goal to study the common randomness of the model I is for the construction of WIDCSI-codes. Next to study WIDCK codes we introduced the Model II of common randomness. Actually our model is a little more general than that we really need. That is, we add "the side information". But for this we need to do almost no more work. Thus to define the Model II we only add a noiseless channel between the sender and receiver based on th Model I.

Namely we assume that the correlated source $\{(V^n, L^n)\}_{n=1}^{\infty}$, the noisy channel W, the distortion constraint (10), and the sender's private randomness M are still available. Additionally the sender may send a message k_n from a set of message $\mathcal{K}_n$ with rate $\frac{1}{n}\log|\mathcal{K}| \leq R_K$ to the receiver via noiseless channel. Again R_K is called key rate. Of course k_n is necessarily to be a function of the outputs of the source and sender's private randomness i.e., $k_n = k_n(v^n, m)$ for $v^n \in \mathcal{V}^n,\ m \in \mathcal{M}$. More generally the sender may use randam strategies i.e., treats k_n is output of a channel W_K with input (v^n, m). To define the common randomness for this model we change (13) to

$$G : \mathcal{K}_n \times \mathcal{L}^n \times \mathcal{Y}^n \longrightarrow \mathcal{A}. \tag{17}$$

and keep the conditions (10), (11), (12), (14), (15), and (16) unchanged (but now the definition of function G has been changed due to the changing).

Analogously, one can define CR-capacity $C_{CRII}((V, L), W, R_K, D_1)$ for memoryless correlated source with generic (V, L), memoryless channel W, key rate R_K and the distortion criterion D_1 of this model.

3.3 The Models for Compound Channels

In this subsection we assume that the attacker employ a (stationary) memoryless channel from a family of channels satisfying attack distortion criterion to attack the watermak. Neither the sender nor receiver knows the which channel the attacker uses. These channels are known as compound channels in Information Theory. This assumption is slightly more robust and practical than that in [28] where the decoder has to know the attack channel in order to decode. In fact, according to our knowledege it is most robust assumption in this direstion.

A compound channel is just a family of memoryless channels $\mathcal{W} = \{W(\cdot|\cdot, s) : s \in \mathcal{S}\}$ with common input and output alphabet $\mathcal{X}$ and $\mathcal{Y}$ respectively. $\mathcal{S}$ is a index set which is called state set and its members are called states. An output sequence $y^n \in \mathcal{Y}^n$ is output with the probability

$$W^n(y^n|x^n, s) = \prod_{t=1}^{n} W(y_t|x_t, s)$$

when the channel is governed by the state s and $x^n \in \mathcal{X}^n$ is input.

Underlie assumption for the attacker to use a compound channel to attack a watermarking transmission or identification code is that the attacker knows the input distribution P_n generated by the code. He then may employ such a compound channel that for all $s \in \mathcal{S}$

$$\frac{1}{n}\sum_{x^n \in \mathcal{X}^n} P_n(x^n) \sum_{y^n \in \mathcal{Y}^n} W^n(y^n|x^n, s)\rho'(x^n, y^n) \leq D_2,$$

where ρ' is a sum type distortion measure, attack distortion measure (AD-measure), may or may not be identify to WD-measure ρ and D_2 is the attack distortion criterion. In particular when the codewords are generated by an i. i. d. input distributions so that the input distribution generated by the code is an i. i. d. distribution

$$P^n(x^n) = \prod_{i=1}^{n} P(x_t)$$

a compound channel such that for all $s \in \mathcal{S}$

$$\sum_{x \in \mathcal{X}} P(x) \sum_{y \in \mathcal{Y}^n} W(y|x,s)\rho'(x,y) \leq D_2$$

may be used. We always assume that all compound channels under the consideration satisfy the condition of distortion and do not worry it at all.

To adjust the models in the last two subsections to the compound channels the following modifications are necessary.

For *WIDCSI code* for compound channels: replace (3) and (4) by for all $l^n \in \mathcal{L}^n, m \in \mathcal{M}$, $\mathcal{D}_m(l^n) \subset \mathcal{Y}^n$ such that for all $m \in \mathcal{M}$, and $s \in \mathcal{S}$,

$$\sum_{v^n \in \mathcal{V}^n, l^n \in \mathcal{L}^n} P^n_{VL}(v^n, l^n) \sum_{x^n \in \mathcal{X}^n} Q_m(x^n|v^n, l^n) W^n(\mathcal{D}_m(l^n)|x^n, s) > 1 - \lambda_1, \tag{18}$$

and for all $m, m' \in \mathcal{M}$ $m \neq m'$, and $s \in \mathcal{S}$

$$\sum_{v^n \in \mathcal{V}^n, \ell^n \in \mathcal{L}^n} P^n_{VL}(v^n, \ell^n) \sum_{x^n \in \mathcal{X}^n} Q_m(x^n|v^n, \ell^n) W^n(\mathcal{D}_{m'}(l^n)|x^n, s) < \lambda_2 \tag{19}$$

respectively.

For *WIDCK* for compound channels: replace (8) and (9) by for all $k_n \in \mathcal{L}_n, m \in \mathcal{M}$, $\mathcal{D}_m(k_n) \subset \mathcal{Y}^n$ such that for all $m \in \mathcal{M}$, and $s \in \mathcal{S}$,

$$\sum_{v^n \in \mathcal{V}^n} P^n_V(v^n) \sum_{k_n \in \mathcal{K}_n} W_{K_n}(k_n|v^n) \sum_{x^n \in \mathcal{X}^n} Q^*_m(x^n|v^n, k_n) W^n(\mathcal{D}_m(k_n)|x^n, s) > 1 - \lambda_1, \tag{20}$$

and for all $m, m' \in \mathcal{M}$ $m \neq m'$, and $s \in \mathcal{S}$,

$$\sum_{v^n \in \mathcal{V}^n} P^n_V(v^n) \sum_{k_n \in \mathcal{K}_n} W_{K_n}(k_n|v^n) \sum_{x^n \in \mathcal{X}^n} Q^*_m(x^n|v^n, k^n) W^n(\mathcal{D}_{m'}(k_n)|x^n, s) < \lambda_2. \tag{21}$$

Here the fact that Q_m, Q^*_m, $\mathcal{D}_m(l^n)$ and $\mathcal{D}_m(k_n)$ are independent of the states governing the channels reflects the requirement that neither encoder nor decoder knows the states and that (18) – (21) hold for all $s \in \mathcal{S}$ is because the worst case to the encoder and decoder is considered.

For the *Common randomness in the models I and II:* for compound channels, replace (14) by
whenever any state s governs the channel,

$$Pr\{F \neq G|s\} < \lambda. \tag{22}$$

Again the functions F, G, codewords are independent of the states because the states are unknown for both encoder and the decoder.

Analogously, for compound channel $\mathcal{W}$ the corresponding capacities of watermarking identification codes and common randomness are denoted by $C_{WIDSI}((V,L),\mathcal{W},D_1)$, $C_{WIDK}(V,\mathcal{W},R_K,D_1)$, $C_{CRI}((V,L),\mathcal{W},D_1)$ and $C_{CRII}((V,L),\ \mathcal{W},R_K,D_1)$.

4 The Results

The Results on Common randomness

For given a correlated memoryless source $\{(V^n,L^n)\}_{n=1}^{\infty}$ whose generic has joint distribution P_{VL}, a memoryless channel W and distortion criterion D_1, let $\mathcal{Q}((V,L),W,D_1)$ be the set of random variable (V,L,U,X,Y) with domain $\mathcal{V}\times\mathcal{L}\times\mathcal{U}\times\mathcal{X}\times\mathcal{Y}$ and the following properties, where $\mathcal{U}$ is a finite set with cardinality $|\mathcal{U}|\leq|\mathcal{V}||\mathcal{L}||\mathcal{X}|$ and $\mathcal{X}$ and $\mathcal{Y}$ are input and output alphabets of the channel W respectively.

For all $v\in\mathcal{V},\ l\in\mathcal{L},\ u\in\mathcal{U},\ x\in\mathcal{X}$, and $y\in\mathcal{Y}$

$$\begin{aligned}Pr\{(V,L,U,X,Y)&=(v,l,u,x,y)\}\\&=P_{VLUXY}(v,l,u,x,y)\\&=P_{VL}(v,l)P_{UX|VL}(u,x|v,l)W(y|x).\end{aligned}\tag{23}$$

For the given distortion measure ρ

$$\mathbf{E}\rho(V,X)\leq D_1.\tag{24}$$

$$I(U;V,L)\leq I(U;L,Y).\tag{25}$$

Then we have the coding theorem of common randomness in the model I for single channel W.

Theorem 4.1

$$C_{CRI}((V,L),W,D_1)=\max_{(V,L,U,X,Y)\in\mathcal{Q}((V,L),W,D_1)}[I(U;L,Y)+H(L|U)].\tag{26}$$

For a given correlated source with generic (V,L) a channel W and positive real numbers R_K and D_1, we denote by $\mathcal{Q}^*((V,L),W,R_K,D_1)$ the set of random variables (V,L,U,X,Y) with domain as above and such that (23), (24) and

$$I(U;V,L)\leq I(U;L,Y)+R_K\tag{27}$$

hold. Then

Theorem 4.2

$$C_{CRII}((V,L),W,R_K,D_1)=\max_{(V,L,U,X,Y)\in\mathcal{Q}^*((V,L),W,R_K,D_1)}[I(U;L,Y)+H(L|U)]+R_K.\tag{28}$$

To state the coding theorem for compound channels we need new notation. For random variables (V, L, U, X) with alphabet $\mathcal{V}\times\mathcal{L}\times\mathcal{U}\times\mathcal{X}$ as above and the channel with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ respectively, denote by $Y(W)$ the random variable such that the joint distribution $P_{LVUXY(W)} = P_{LVUX}W$ (consequently, $LVU \leftrightarrow X \leftrightarrow Y$ form a Markov chain). For a compound channel $\mathcal{W}$ with set of stares $\mathcal{S}$ and a state $s \in \mathcal{S}$ we also write $Y(W(\cdot|\cdot, s)) = Y(s)$. With the notation we write

$$I(U; L, Y(\mathcal{W})) = \inf_{s\in\mathcal{S}} I(U; L, Y(s))$$

and

$$I(U; Y(\mathcal{W})|L) = \inf_{s\in\mathcal{S}} I(U; Y(s)|L).$$

Sometimes just for the convenience, we also write $Y(s)$ as $\tilde{Y}(s)$ when we substitute P_{LVUX} by $P_{\tilde{L}\tilde{V}\tilde{U}\tilde{X}}$ and similarly $\tilde{Y}(\mathcal{W})$. Then

$$I(U; L, Y(\mathcal{W})) = I(U; L) + I(U; Y(\mathcal{W})|L). \tag{29}$$

Now for a compound channel we define $\mathcal{Q}_1((V, L), \mathcal{W}, D_1)$ as the set of random variables (V, L, U, X) such that its marginal distribution for the first two components is equal to the distribution P_{VL} and (24) and

$$I(U; V, L) \leq I(U; L, Y(\mathcal{W})) \tag{30}$$

hold. Analogously to set $\mathcal{Q}^*((V, L), W, R_K, D_1)$ we define $\mathcal{Q}_1^*((V, L), \mathcal{W}, R_K, D_1)$ the set of random variables (V, L, U, X) such that its marginal distribution for the first two components is equal to the distribution P_{VL} and (24) and

$$I(U; V, L) \leq I(U; L, Y(\mathcal{W})) + R_K. \tag{31}$$

hold. Then

Theorem 4.3

$$\begin{aligned}\sup_{(V,L,U,X)\in\mathcal{Q}_1((V,L),\mathcal{W},D_1)} &[I(U; L, Y(\mathcal{W})) + H(L|U)] \leq C_{CRI}((V, L), \mathcal{W}, D_1)\\ \leq \inf_{W\in\mathcal{W}} \max_{(V,L,U,X,Y)\in\mathcal{Q}((V,L),W,D_1)} &[I(U; L, Y) + H(L|U)].\end{aligned} \tag{32}$$

Theorem 4.4

$$\begin{aligned}\sup_{(V,L,U,X)\in\mathcal{Q}_1^*((V,L),\mathcal{W},R_K,D_1)} &[I(U; L, Y(\mathcal{W})) + H(L|U)] + R_K \leq C_{CRII}((V, L), \mathcal{W}, R_K, D_1)\\ \leq \inf_{W\in\mathcal{W}} \max_{(V,L,U,X,Y)\in\mathcal{Q}^*((V,L),W,R_K,D_1)} &[I(U; L, Y) + H(L|U)] + R_K.\end{aligned} \tag{33}$$

Notice the gaps of lower and upper bounds in both Theorems 4.3 and 4.4 are due to the orders of inf–sup.

The Results on Watermarking Identification Codes

We shall use the same notation as in the above part. Moreover for above sets $\mathcal{V}, \mathcal{X}$ and $\mathcal{Y}$ and a finite set $\mathcal{U}$ with cardinality bounded by $|\mathcal{V}||\mathcal{X}|$, a memoryless source with generic V, a memoryless cannel W, and compound channel $\mathcal{W}$, we define the following sets. Let $\mathcal{Q}^{**}(V, W, R_K, D_1)$ be the set of random variables (V, U, X, Y) with domain $\mathcal{V} \times \mathcal{U} \times \mathcal{X} \times \mathcal{Y}$ such that for all $v \in \mathcal{V},\ u \in \mathcal{U},\ x \in \mathcal{X}$, and $y \in \mathcal{Y}$

$$P_{VUXY}(v, u, x, y) = P_V(v)P_{UX|V}(u, x|v)W(y|x), \tag{34}$$

$$I(U; V) \leq I(U; Y) + R_K, \tag{35}$$

and (24) hold. Let $\mathcal{Q}_1^{**}(V, \mathcal{W}, R_K, D_1)$ be set of random variables (V, U, X) with domain $\mathcal{V} \times \mathcal{U} \times \mathcal{X}$ such that for all $v \in \mathcal{V}, u \in \mathcal{U}$ and $x \in \mathcal{X}$,

$$P_{VUX}(v, u, x) = P_V(v)P_{UX|V}(u.x|v), \tag{36}$$

$$I(U; V) \leq I(U; Y(\mathcal{W})) + R_K, \tag{37}$$

and (24) hold, where $I(U; Y(\mathcal{W})) = \inf_{W \in \mathcal{W}} I(U; Y(W))$. In particular, when the second component L^n of the correlated source $\{(V^n, L^n)\}_{n=1}^{\infty}$ is a constant, $\mathcal{Q}^*((V, L), W, R_K, D_1)$ and $\mathcal{Q}_1^*((V, L), \mathcal{W}, R_K, D_1)$ become $\mathcal{Q}^{**}(V, W, R_K, D_1)$ and $\mathcal{Q}_1^{**}(V, \mathcal{W}, R_K, D_1)$ respectively.

Theorem 4.5

$$C_{WIDSI}((V, L), W, D_1) \geq \max_{(V,L,U,X,Y) \in \mathcal{Q}((V,L),W,D_1)} [I(U; L, Y) + H(L|U)]. \tag{38}$$

Theorem 4.6

$$C_{WIDK}(V, W, R_K, D_1) \geq \max_{(V,U,X,Y) \in \mathcal{Q}^{**}(V,W,R_K,D_1)} I(U; Y) + R_K. \tag{39}$$

Theorem 4.7

$$C_{WIDSI}((V, L), \mathcal{W}, D_1) \geq \sup_{(V,L,U,X) \in \mathcal{Q}_1((V,L),\mathcal{W},D_1)} [I(U; L, Y(\mathcal{W})) + H(L|U)]. \tag{40}$$

Theorem 4.8

$$C_{WIDK}(V, W, R_K) \geq \sup_{(V,U,X) \in \mathcal{Q}_1^{**}(V,\mathcal{W},R_K,D_1)} I(U; Y(\mathcal{W})) + R_K. \tag{41}$$

Note that in Theorems 4.6 and 4.8 one may add side information L^n, the second component of the correlated source and then one can obtain the corresponding lower bound almost does not change the proofs.

A result on Watermarking Transmission Code with a common Experiment Introduced by Steinberg-Merhav:

To construct watermarking identification code Y. Steinberg and N. Merhav in [32] introduced a code, which they call watermarking transmission code with common experiment, distortion measure ρ, and covertext P_V. They obtained there an inner bound to the its capacity region, which is sufficient for acheiving their goal. We shall show their bound is tight and therefore actually the capacity region. Their definition and result on it and our proof will be presented it the last section.

5 The Direct Theorems for Common Randomness

In this section we prove the direct parts of Theorems 4.1 – 4.4. Since a DMC can be regarded as a special compound channel with a single member (i.e., $|\mathcal{S}| = 1$), we only have to show the direct parts of Theorems 4.3 and 4.4. To this end we need the following three lemmas for n–type $P_{\tilde{V}\tilde{L}\tilde{U}}$ over the product set $\mathcal{V} \times \mathcal{L} \times \mathcal{U}$ of finite sets $\mathcal{V}, \mathcal{L}$ and $\mathcal{U}$.

Lemma 5.1 (Uniformly covering). *For $\ell^n \in \mathcal{T}^n_{\tilde{L}}$, let $U_i(\ell^n)$ $i = 1, 2, \ldots, \lfloor 2^{n\alpha} \rfloor$ be a sequence of independent random variables with uniform distribution over $\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)$ and for any $v^n \in \mathcal{T}^n_{\tilde{V}|\tilde{L}}(\ell^n)$ let $\hat{\mathcal{U}}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)$ be the random set $\{U_i(\ell^n) : i = 1, 2, \ldots, \lfloor 2^{n\alpha} \rfloor\} \cap \mathcal{T}^n_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)$. Then for all $\varepsilon \in (0, 1]$*

$$Pr\left\{\left| |\hat{\mathcal{U}}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)| - \lfloor 2^{n\alpha} \rfloor \frac{|\mathcal{T}^n_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)|}{|\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)|} \right| \geq \lfloor 2^{n\alpha} \rfloor \frac{|\mathcal{T}^n_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)|}{|\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)|}\varepsilon \right\} < 4 \cdot 2^{-\frac{\varepsilon^2}{4}2^{n\eta}} \tag{42}$$

for sufficiently large n if

$$\lfloor 2^{n\alpha} \rfloor > 2^{n\eta} \frac{|\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)|}{|\mathcal{T}^n_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)|}$$

Proof: Let

$$Z_i(v^n, \ell^n) = \begin{cases} 1 & \text{if } U_i(\ell^n) \in \mathcal{T}^n_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n), \\ 0 & \text{else,} \end{cases} \tag{43}$$

and $q = \frac{|\mathcal{T}^n_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)|}{|\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)|}$. Then $|\hat{\mathcal{U}}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)| = \sum_{i=1}^{\lfloor 2^{n\alpha} \rfloor} Z_i(v^n\ell^n)$ and for $i = 1, 2, \ldots, \lfloor 2^{n\alpha} \rfloor$

$$Pr\{Z_i(v^n\ell^n) = z\} = \begin{cases} q & \text{if } z = 1 \\ 1 - q & \text{if } z = 0 \end{cases} \tag{44}$$

by the definitions of $U_i(\ell^n)$ and $Z_i(v^n, \ell^n)$.

Then by Chernov's bound, we have that

$$
\begin{aligned}
&Pr\left\{\sum_{i=1}^{\lfloor 2^{n\alpha}\rfloor} Z_i(v^n\ell^n) \geq \lfloor 2^{n\alpha}\rfloor q(1+\varepsilon)\right\} \\
&\leq e^{-\frac{\varepsilon}{2}\lfloor 2^{n\alpha}\rfloor q(1+\varepsilon)}\ E\ e^{\frac{\varepsilon}{2}\sum\limits_{i=1}^{\lfloor 2^{n\alpha}\rfloor} Z_i(v^n,\ell^n)} \\
&= e^{-\frac{\varepsilon}{2}\lfloor 2^{n\alpha}\rfloor q(1+\varepsilon)} \prod_{i=1}^{\lfloor 2^{n\alpha}\rfloor} E\ e^{\frac{\varepsilon}{2}Z_i(v^n,\ell^n)} \\
&= e^{-\frac{\varepsilon}{2}\lfloor 2^{n\alpha}\rfloor q(1+\varepsilon)}[1+(e^{\frac{\varepsilon}{2}}-1)q]^{\lfloor 2^{n\alpha}\rfloor} \\
&\leq e^{-\frac{\varepsilon}{2}\lfloor 2^{n\alpha}\rfloor q(1+\varepsilon)}\left[1+\left(\frac{\varepsilon}{2}+\left(\frac{\varepsilon}{2}\right)^2\right)q\right]^{\lfloor 2^{n\alpha}\rfloor} \\
&\leq \exp_e\left\{-\frac{\varepsilon}{2}\lfloor 2^{n\alpha}\rfloor q(1+\varepsilon)+\frac{\varepsilon}{2}\lfloor 2^{n\alpha}\rfloor q\left(1+\frac{\varepsilon}{2}\right)\right\} \\
&= e^{-\frac{\varepsilon^2}{4}\lfloor 2^{n\alpha}\rfloor q} < 2e^{-\frac{\varepsilon^2}{4}2^{n\eta}}
\end{aligned} \tag{45}
$$

if $\lfloor 2^{n\alpha}\rfloor > 2^{n\eta}q^{-1}$.

Here the first inequality follows from Chernov's bound; the second equality holds by (44); the second inequality holds because $e^{\frac{\varepsilon}{2}} < 1+\frac{\varepsilon}{2}+\left(\frac{\varepsilon}{2}\right)^2$ a by the assymption that $\varepsilon < 1$, $e^{\frac{\varepsilon}{2}} < e^{\frac{1}{2}} < 2$; and the third inequality follows from the wellknown inequality $1+x < e^x$. Similarly one can obtain

$$
Pr\left\{\sum_{i=1}^{\lfloor 2^{n\alpha}\rfloor} Z_i(v^n\ell^n) \leq \lfloor 2^{n\alpha}\rfloor q(1-\varepsilon)\right\} < 2e^{-\frac{\varepsilon^2}{4}2^{n\eta}} \tag{46}
$$

if $\lfloor 2^{n\alpha}\rfloor > 2^{n\eta}q^{-1}$.

Finally we obtain the lemma by combining (45) and (46).

Lemma 5.2 (Packing). *Let $P_{\tilde{L}\tilde{U}}$ be an n–type, let $U_i(\ell^n)$, $i=1,2,\dots,\lfloor 2^{n\alpha}\rfloor$ be a sequence of independent random variables uniformly distributed on $\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)$ for an $\ell^n\in\mathcal{T}^n_{\tilde{L}}$, and let $\mathcal{Y}$ be a finite set. Then for all n–types $P_{\tilde{L}\tilde{U}\tilde{Y}}$ and $P_{\tilde{L}\tilde{U}\overline{Y}}$ with common marginal distributions $P_{\tilde{L}\tilde{U}}$ and $P_{\overline{Y}}=P_{\tilde{Y}}$, all i, $\gamma>0$ and sufficiently large n,*

$$
Pr\left\{\frac{1}{\lfloor 2^{n\alpha}\rfloor}\sum_{i=1}^{\lfloor 2^{n\alpha}\rfloor}\left|\mathcal{T}^n_{\tilde{Y}|\tilde{L}\tilde{U}}\left(\ell^n U_i(\ell^n)\right)\cap\left[\bigcup_{j\neq i}\mathcal{T}^n_{\overline{Y}|\tilde{L}\tilde{U}}\left(\ell^n U_j(\ell^n)\right)\right]\right| \geq t_{\tilde{Y}|\tilde{L}\tilde{U}}2^{-\frac{n}{2}\gamma}\right\} < 2^{-\frac{n}{2}\gamma} \tag{47}
$$

if $\lfloor 2^{n\alpha}\rfloor \leq \frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{U}|\tilde{L}\overline{Y}}}2^{-n\gamma}$.

Here $t_{\tilde{Y}|\tilde{L}\tilde{U}}$, $t_{\tilde{U}|\tilde{L}}$, and $t_{\tilde{U}|\tilde{L}\overline{Y}}$ are the common values of $|\mathcal{T}^n_{\tilde{Y}|\tilde{L}\tilde{U}}(\ell^n u^n)|$ for $(\ell^n,u^n)\in\mathcal{T}^n_{\tilde{L}\tilde{U}}$, $|\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)|$ for $\ell^n\in\mathcal{T}^n_{\tilde{L}}$, and $|\mathcal{T}^n_{\tilde{U}|\tilde{L}\overline{Y}}(\ell^n y^n)|$ for $(\ell^n,y^n)\in\mathcal{T}^n_{\tilde{L}\overline{Y}}$, respectively.

Proof: For $i = 1, 2, \dots, \lfloor 2^{n\alpha} \rfloor$, $y^n \in \mathcal{T}^n_{\overline{Y}} = \mathcal{T}^n_{\tilde{Y}}$, let

$$\hat{Z}_i(y^n) = \begin{cases} 1 & \text{if } y^n \in \bigcup_{j \neq i} \mathcal{T}_{\overline{Y}|\tilde{L}\tilde{U}}\big(\ell^n U_j(\ell^n)\big) \\ 0 & \text{else} \end{cases} \tag{48}$$

and for all $u^n \in \mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)$

$$S_i(u^n) = \left| \mathcal{T}^n_{\tilde{Y}|\tilde{L}\tilde{U}}(\ell^n u^n) \cap \left[\bigcup_{j \neq i} \mathcal{T}^n_{\overline{Y}|\tilde{L}\tilde{U}}\big(\ell^n U_j(\ell^n)\big) \right] \right|. \tag{49}$$

Then

$$S_i(u^n) = \sum_{y^n \in \mathcal{T}_{\tilde{Y}|\tilde{L}\tilde{U}(\ell^n u^n)}} \hat{Z}_j(y^n) \tag{50}$$

and

$$\begin{aligned} E\ \hat{Z}_i(y^n) &= Pr\left\{ y^n \in \bigcup_{j \neq i} \mathcal{T}^n_{\overline{Y}|\tilde{L}\tilde{U}}\big(\ell^n U_j(\ell^n)\big) \right\} \leq \sum_{j \neq i} Pr\big\{ y^n \in \mathcal{T}^n_{\overline{Y}|\tilde{L}\tilde{U}}\big(\ell^n U_j(\ell^n)\big) \big\} \\ &= \sum_{j \neq i} Pr\big\{ U_j(\ell^n) \in \mathcal{T}^n_{\tilde{U}|\tilde{L}\overline{Y}}(\ell^n y^n) \big\} = (2^{\lfloor n\alpha \rfloor} - 1) \frac{t_{\tilde{U}|\tilde{L}\overline{Y}}}{t_{\tilde{U}|\tilde{L}}} < 2^{-n\gamma} \end{aligned} \tag{51}$$

if $\lfloor 2^{n\alpha} \rfloor \leq \frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{U}|\tilde{L}\overline{Y}}} 2^{-n\gamma}$.

Hence by (50) and (51) we have that $E\ S_i(u^n) \leq t_{\tilde{Y}|\tilde{L}\tilde{U}} 2^{-n\gamma}$ and i.e., $E\big[S_i(U_i(\ell^n)) | U_i(\ell^n)\big] < t_{\tilde{Y}|\tilde{L}\tilde{U}} 2^{-n\gamma}$ (a.s.), so

$$E\ S_i\big(U_i(\ell^n)\big) = E\big\{ E\big[S_i(U_i(\ell^n)) | U_i(\ell^n) \big] \big\} < t_{\tilde{Y}|\tilde{L}\tilde{U}} 2^{-n\gamma}. \tag{52}$$

Thus by Markov's inequality we have that

$$Pr\left\{ \frac{1}{\lfloor 2^{n\alpha} \rfloor} \sum_{i=1}^{\lfloor 2^{n\alpha} \rfloor} S_i\big(U_i(\ell^n)\big) \geq t_{\tilde{Y}|\tilde{L}\tilde{U}} 2^{-\frac{n}{2}\gamma} \right\} < 2^{-\frac{n}{2}\gamma},$$

i.e., (47).

Lemma 5.3 (Multi–Packing). *Under the conditions of the previous lemma, let $U_{i,k}(\ell^n)$, $i = 1, 2, \dots, \lfloor 2^{n\beta_1} \rfloor$, $k = 1, 2, \dots, \lfloor 2^{n\beta_2} \rfloor$, be a sequence of independent random variables uniformly distributed on $\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)$ for a given $\ell^n \in \mathcal{T}^n_{\tilde{L}}$. Then for all n–types $P_{\tilde{L}\tilde{U}\tilde{Y}}$ and $P_{\tilde{L}\tilde{U}\overline{Y}}$ in the previous lemma*

$$Pr\left\{ \frac{1}{\lfloor 2^{n\beta_2} \rfloor} \sum_{k=1}^{\lfloor 2^{n\beta_2} \rfloor} \frac{1}{\lfloor 2^{n\beta_1} \rfloor} \sum_{i=1}^{\lfloor 2^{n\beta_1} \rfloor} \left| \mathcal{T}^n_{\tilde{Y}|\tilde{L}\tilde{U}}\big(\ell^n U_{i,k}(\ell^n)\big) \cap \left[\bigcup_{j \neq i} \mathcal{T}^n_{\overline{Y}|\tilde{L}\tilde{U}}\big(\ell^n U_{j,k}(\ell^n)\big) \right] \right| \geq t_{\tilde{Y}|\tilde{L}\tilde{U}} 2^{-n\eta} \right\} < 2^{-\frac{n}{2}\gamma} \tag{53}$$

if $\lfloor 2^{n\alpha} \rfloor \leq \frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{U}|\tilde{L}\overline{Y}}} 2^{-n\gamma}$.

Proof: For $u^n \in \mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)$, let

$$S_{i,k}(u^n) = \left| \mathcal{T}^n_{\tilde{Y}|\tilde{L}\tilde{U}}(\ell^n u^n) \cap \left[\bigcup_{j \neq i} \mathcal{T}^n_{\overline{Y}|\tilde{L}\tilde{U}}\big(\ell^n U_{j,k}(\ell^n)\big) \right] \right|.$$

Then we have shown in the proof to the previous lemma (c.f. (52))

$$E\ S_{i,k}\big(U_i(\ell^n)\big) < t_{\tilde{Y}|\tilde{L}\tilde{U}} 2^{-n\gamma}.$$

Thus (53) follows from Markov's inequality.

Now let us turn to the direct part of Theorem 4.3.

Lemma 5.4 (The Direct Part of Theorem 4.3). *For a compound channel* $\mathcal{W}$,

$$C_{CRI}\big((V,L),\mathcal{W},D_1\big) \geq \sup_{(V,L,U,X)\in Q_1\big((V,L),\mathcal{W},D_1\big)} \big[I\big(U;L,Y(\mathcal{W})\big) + H(L|U)\big]. \tag{54}$$

Proof: We have to show for a given correlated memoryless source with generic (V,L), a compound channel $\mathcal{W}$, $(V,L,U,X) \in Q_1\big((V,L),\mathcal{W},D_1\big)$ and sufficiently large n, the existence of the functions, F, G and $x_m(v^n,\ell^n)$ satisfying (10) – (13), (22), (15) and (16) with the rate arbitrarily close to $I\big(U;L,Y(\mathcal{W})\big) + H(L|U)$. Obviously the set of achievable rates of the common randomness is bounded and closed (i.e., compact). So without loss of generality, by uniform continuity of information quantities, we can assume that $E\rho(V,X) < D_1$, and $I(U;V,L) < I\big(U;L,Y(\mathcal{W})\big)$. Because $I(U;V,L) = I(U;L) + I(U;V|L)$ and $I\big(U;L,Y(\mathcal{W})\big) = I(U;L) + I\big(U;Y(\mathcal{W})|L\big)$, there exists a sufficiently small but positive constant ξ, such that

$$I\big(U;Y(\mathcal{W})|L\big) - I(U;V|L) > \xi. \tag{55}$$

Without loss of generality, we also assume P_U is an n–type to simplify the notation. Then for arbitrary $\varepsilon_1 > 0$, by uniform continuity of information quantities, we can find $\delta_1, \delta_2 > 0$ with the following properties.

(a) For all $\ell^n \in \mathcal{T}^n_L(\delta_1)$ with type $P_{\ell^n} = P_{\tilde{L}}$, there exists a $\delta' > 0$, such that $(v^n,\ell^n) \in \mathcal{T}^n_{VL}(\delta'_2)$ yields that $\mathcal{T}^n_{\hat{V}|\tilde{L}}(\ell^n) \subset \mathcal{T}^n_{\hat{V}|\tilde{L}}(\ell^n,\delta_2)$, where $P_{\hat{V}\tilde{L}}$ is the joint type of (v^n,ℓ^n) and $P_{\hat{V}\tilde{L}} = P_{\tilde{L}} P_{V|L}$.

We call a pair (v^n,ℓ^n) of sequences with $\ell^n \in \mathcal{T}^n_L(\delta_1)$, $(v^n,\ell^n) \in \mathcal{T}^n_{VL}(\delta_2)$, (δ_1,δ_2)–typical and denote the set of (δ_1,δ_2)–typical sequences by $\mathcal{T}^n(\delta_1,\delta_2)$.

Then we may require $\delta_2 \to 0$ as $\delta_1 \to 0$. Moreover (e.g., see [35]), there exist positive $\zeta_1 = \zeta_1(\delta_1)$, $\zeta_2 = \zeta_2(\delta_1,\delta_2)$, and $\zeta = \zeta(\delta_1,\delta_2)$ such that

$$P^n_L\big(\mathcal{T}^n_L(\delta_1)\big) > 1 - 2^{-n\zeta_1} \tag{56}$$

$$P^n_{V|L}\{v^n : (v^n, \ell^n) \in \mathcal{T}^n(\delta_1, \delta_2)|\ell^n\big) > 1 - 2^{-n\zeta_2} \tag{57}$$

for all $\ell^n \in \mathcal{T}^n_L(\delta_1)$ and

$$P^n_{VL}\big(\mathcal{T}^n(\delta_1, \delta_2)\big) > 1 - 2^{n\zeta}. \tag{58}$$

(b) For all $\ell^n \in \mathcal{T}^n_L(\delta_1)$ with type $P_{\ell^n} = P_{\tilde{L}}$ (say), one can find a joint type of sequences in $L^n \times \mathcal{U}^n$, say $P_{\tilde{L}\tilde{U}}$, with marginal distributions $P_{\tilde{L}}$ and P_U, sufficiently close to P_{LU}, (which will be specified below). We say that $P_{\tilde{L}\tilde{U}}$ is generated by the type $P_{\tilde{L}}$ of ℓ^n.

(c) For all $(v^n, \ell^n) \in \mathcal{T}^n(\delta_1, \delta_2)$ with joint type $P_{v^n\ell^n} = P_{\tilde{V}\tilde{L}}$ (say), one can find a joint type $P_{\tilde{V}\tilde{L}\tilde{U}}$ of sequences in $\mathcal{V}^n \times \mathcal{L}^n \times \mathcal{U}^n$ with marginal distributions $P_{\tilde{V}\tilde{L}}$ and $P_{\tilde{L}\tilde{U}}$ and sufficiently close to P_{VLU} (which will be specified below), where $P_{\tilde{L}\tilde{U}}$ is the type generated by $P_{\tilde{L}}$. We say $P_{\tilde{V}\tilde{L}\tilde{U}}$ is generated by the joint type $P_{\tilde{V}\tilde{L}}$ of (v^n, ℓ^n).

(d) For all (δ_1, δ_2)–typical sequences (v^n, ℓ^n) with joint type $P_{\tilde{V}\tilde{L}}$ (say) and the joint type $P_{\tilde{V}\tilde{L}\tilde{U}}$ generated by $P_{\tilde{V}\tilde{L}}$, we let $(\tilde{V}, \tilde{L}, \tilde{U}, \tilde{X})$ be random variables with joint distribution $P_{\tilde{V}\tilde{L}\tilde{U}\tilde{X}}$ such that for all $v \in \mathcal{V}$, $\ell \in \mathcal{L}$, $u \in \mathcal{U}$ and $x \in \mathcal{X}$

$$P_{\tilde{V}\tilde{L}\tilde{U}\tilde{X}}(v, \ell, u, x) = P_{\tilde{V}\tilde{L}\tilde{U}}(v, \ell, u)P_{X|VLU}(x|v, \ell, u), \tag{59}$$

and let $\big(\tilde{V}, \tilde{L}, \tilde{U}, \tilde{X}, \tilde{Y}(W)\big)$ be random variables with joint distribution $P_{\tilde{V}\tilde{L}\tilde{U}\tilde{X}\tilde{Y}(\mathcal{W})}$ such that for all $v \in \mathcal{V}$, $\ell \in \mathcal{L}$, $u \in \mathcal{U}$, $x \in \mathcal{X}$, and $y \in \mathcal{Y}$

$$P_{\tilde{V}\tilde{L}\tilde{U}\tilde{X}\tilde{Y}(W)}(v, \ell, u, x, y) = P_{\tilde{V}\tilde{L}\tilde{U}\tilde{X}}(v, \ell, u, x)W(x|y), \tag{60}$$

for any $W \in \mathcal{W}$ and $P_{\tilde{V}\tilde{L}\tilde{U}\tilde{X}}$ in (59). Then the following inequalities hold

$$E\rho(\tilde{V}, \tilde{X}) < D_1, \tag{61}$$

$$|H(\tilde{L}) - H(L)| < \varepsilon_1, \tag{62}$$

$$|I(\tilde{U}; \tilde{V}|\tilde{L}) - I(U; V|L)| < \varepsilon_1, \tag{63}$$

and

$$|I(\tilde{U}; \tilde{Y}(\mathcal{W})|\tilde{L}) - I(U; Y(\mathcal{W})|L)| < \varepsilon_1, \tag{64}$$

where $I(\tilde{U}; \tilde{Y}(\mathcal{W})|\tilde{L}) = \inf_{W\in\mathcal{W}} I(\tilde{U}; \tilde{Y}(W)|\tilde{L})$.

For arbitrarily small fixed ε_2 with $0 < \varepsilon_2 < \frac{1}{2}\xi$, for ξ in (55), we choose ε_1 (and consequently, δ_1, δ_2) so small that $\varepsilon_1 < \frac{1}{2}\varepsilon_2$ and an α such that

$$I(U; Y(\mathcal{W})|L) - \frac{\xi}{2} < \alpha < I(U; Y(\mathcal{W})|L) - \varepsilon_2 \tag{65}$$

and $M = 2^{n\alpha}$ (say) is an integer. Notice that by (65) we may choose α arbitrarily close to $I(U;Y(\mathcal{W})|L)-\varepsilon_2$ and therefore arbitrarily close to $I(U;Y(\mathcal{W})|L)$ by choosing ε_2 arbitrarily small. Then by (55), (63) and (65) we have that

$$\alpha > I(U;V|L) + \frac{\xi}{2} > I(\tilde{U};\tilde{V}|\tilde{L}) + \frac{\xi}{2} - \varepsilon_1 > I(\tilde{U};\tilde{V}|\tilde{L}) + \frac{\xi}{4}, \tag{66}$$

where the last inequality holds by our choice $\varepsilon_1 < \frac{1}{2}\varepsilon_2 < \frac{1}{4}\xi$, and by (64) and (65) we have

$$\alpha < I(\tilde{U};\tilde{Y}(\mathcal{W})|\tilde{L}) + \varepsilon_1 - \varepsilon_2 < I(\tilde{U};\tilde{Y}(\mathcal{W})|\tilde{L}) - \frac{\varepsilon_2}{2}. \tag{67}$$

Denote by $t_{\tilde{U}|\tilde{L}}$ and $t_{\tilde{U}|\tilde{V}\tilde{L}}$ the common values of $|\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)|$, $\ell^n \in \mathcal{T}^n_{\tilde{L}}$ and $|\mathcal{T}^n_{\tilde{U}|\tilde{V}\tilde{L}}(v^n,\ell^n)|$, $(v^n,\ell^n) \in \mathcal{T}^n_{\tilde{V}\tilde{L}}$, respectively. Then it is wellknown that $\frac{1}{n}\log\frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{U}|\tilde{V}\tilde{L}}}$ arbitrarily close to $I(\tilde{U};\tilde{V}|\tilde{L})$.

This means under our assumption that $\frac{1}{2}\varepsilon_2 < \frac{1}{4}\xi$, (66) implies that for all types $P_{\tilde{V}\tilde{L}\tilde{U}}$ generated by the joint types $P_{\tilde{V}\tilde{L}}$ of (δ_1,δ_2)–typical sequences

$$2^{\frac{n}{3}\varepsilon_2}\frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{U}|\tilde{V}\tilde{L}}} < 2^{n\alpha} = M. \tag{68}$$

Next we let $\mathcal{Q}_{\mathcal{W}}(\ell^n u^n,\tau)$ be the set of conditional type $P_{\overline{Y}|\tilde{L}\tilde{U}}$, for a pair (ℓ^n,u^n) of sequences such that there exists a $W \in \mathcal{W}$ with $\mathcal{T}^n_{\overline{Y}|\tilde{L}\tilde{U}}(\ell^n u^n) \subset \mathcal{T}^n_{\tilde{Y}(W)|\tilde{L}\tilde{U}}(\ell^n u^n,\tau)$, where $P_{\tilde{L}\tilde{U}}$ is the type of (ℓ^n,u^n) and $P_{\tilde{L}\tilde{U}\tilde{Y}(\mathcal{W})}$ is the marginal distribution of the distribution in (60). Then

$$\bigcup_{P_{\overline{Y}|\tilde{L}\tilde{U}} \in \mathcal{Q}_{\mathcal{W}}(\ell^n u^n,\tau)} \mathcal{T}^n_{\overline{Y}|\tilde{L}\tilde{U}}(\ell^n u^n,\tau) = \bigcup_{W\in\mathcal{W}} \mathcal{T}^n_{\tilde{Y}(\mathcal{W})|\tilde{L}\tilde{U}}(\ell^n u^n,\tau), \tag{69}$$

and

$$|\mathcal{Q}_{\mathcal{W}}(\ell^n u^n,\tau)| < (n+1)^{|\mathcal{L}||\mathcal{U}||\mathcal{Y}|}. \tag{70}$$

Again for the common values $t_{\tilde{U}|\tilde{L}}$ of $|\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)|$, $\ell^n \in \mathcal{T}^n_{\tilde{L}}$, $t_{\tilde{U}|\tilde{L}\overline{Y}}$ of $|\mathcal{T}^n_{\tilde{U}|\tilde{L}\overline{Y}}(\ell^n y^n)|$, $(\ell^n,y^n) \in \mathcal{T}^n_{\tilde{L}\overline{Y}}$, $\lim_{n\to\infty}\frac{1}{n}\log\frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{U}|\tilde{L}\overline{Y}}} = I(\tilde{U};\overline{Y}|\tilde{L})$.

Thus, (67) yields that for all $P_{\tilde{V}\tilde{L}\tilde{U}}$ generated by the joint type of (δ_1,δ_2)–typical sequences, $(\ell^n,u^n) \in \mathcal{T}^n_{\tilde{L}\tilde{U}}$, and $P_{\overline{Y}|\tilde{L}\tilde{U}} \in \mathcal{Q}_{\mathcal{W}}(\ell^n u^n,\tau)$,

$$M = 2^{n\alpha} < 2^{-\frac{n}{4}\varepsilon_2}\frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{U}|\tilde{L}\overline{Y}}}, \tag{71}$$

if we choose τ so small (depending on ε_2) that for all $P_{\overline{Y}|\tilde{L}\tilde{U}} \in \mathcal{Q}_{\mathcal{W}}(\ell^n u^n,\tau)$

$$I(\tilde{U};\overline{Y}|\tilde{L}) > I(\tilde{U};Y(\mathcal{W})|\tilde{L}) - \frac{1}{8}\varepsilon_2$$

(recalling that by its definition $I(\tilde{U};\tilde{Y}(\mathcal{W})|\tilde{L}) = \inf_{W\in\mathcal{W}} I(\tilde{U};\tilde{Y}(W)|\tilde{L})$).

Now we are ready to present our coding scheme at rate α, which may arbitrarily close to $I(U;Y(\mathcal{W})|L)$.

Coding Scheme:

1) Choosing Codebooks:

For all $\ell^n \in \mathcal{T}_L^n(\delta_1)$ with type $P_{\tilde{L}}$, $P_{\tilde{L}\tilde{U}}$ generated by $P_{\tilde{L}}$ (cf. condition (b) above), we apply Lemma 5.1 with $\eta = \frac{\varepsilon_2}{3}$ and Lemma 5.2 with $\gamma = \frac{\varepsilon_2}{4}$ to random choice. Then since the numbers of sequences v^n, ℓ^n and the number of n–joint types are increasing exponentially and polynomially respectively, for all $\ell^n \in \mathcal{T}_L^n(\delta_1)$ with type $P_{\tilde{L}\tilde{U}}$ generated by $P_{\tilde{L}}$, by (68), (71) we can find a subset $\mathcal{U}(\ell^n) \subset \mathcal{T}_{\tilde{U}|\tilde{L}}^n(\ell^n)$ with the following property if n is sufficiently large.

If $(v^n, \ell^n) \in \mathcal{T}^n(\delta_1, \delta_2)$ and has joint type $P_{\tilde{V}\tilde{L}}$ and $P_{\tilde{V}\tilde{L}\tilde{U}}$ is generated by $P_{\tilde{V}\tilde{L}}$ (cf. condition (c) above), then

$$\left| |\mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)| - M\frac{t_{\tilde{U}|\tilde{V}\tilde{L}}}{t_{\tilde{U}|\tilde{L}}} \right| < M\frac{t_{\tilde{U}|\tilde{V}\tilde{L}}}{t_{\tilde{U}|\tilde{L}}}\varepsilon \tag{72}$$

for any $\varepsilon > 0$ (with $\varepsilon \to 0$ as $n \to \infty$), where

$$\mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n) \triangleq \mathcal{U}(\ell^n) \cap \mathcal{T}_{\tilde{U}|\tilde{V}\tilde{L}}^n(v^n\ell^n). \tag{73}$$

For any $P_{\tilde{V}\tilde{L}\tilde{U}}$ generated by a joint type of (δ_1, δ_2)–typical sequence, (v^n, ℓ^n), and joint type $P_{\tilde{L}\tilde{U}\tilde{Y}}$ with marginal distribution $P_{\tilde{L}\tilde{U}}$ and any $P_{\overline{Y}|\tilde{L}\tilde{U}} \in \mathcal{Q}_{\mathcal{W}}\big(\ell^n u_{m'}^n(\ell^n), \tau\big)$ (notice that $\mathcal{Q}_{\mathcal{W}}(\ell^n u^n, \tau)$ depends on $(\ell^n u^n)$ only through their joint type $P_{\ell^n u^n}$!)

$$M^{-1}\sum_{m=1}^{M}\left| \mathcal{T}_{\tilde{Y}|\tilde{L}\tilde{U}}^n\big(\ell^n\tilde{u}_m^n(\ell^n)\big) \cap \left[\bigcup_{m'\neq m} \mathcal{T}_{\overline{Y}|\tilde{L}\tilde{U}}^n\big(\ell^n\tilde{u}_{m'}^n(\ell^n)\big)\right]\right| < 2^{-\frac{n}{8}\varepsilon_2}t_{\tilde{Y}|\tilde{L}\tilde{U}} \tag{74}$$

if we label the members of $\mathcal{U}(\ell^n)$ as $\tilde{u}_1^n(\ell^n), \tilde{u}_2^n(\ell^n), \ldots, \tilde{u}_M^n(\ell^n)$. Consequently by (70) and the fact that (ℓ^n, u^n), (ℓ'^n, u'^n) have the same type $\mathcal{Q}_{\mathcal{W}}(\ell^n u^n) = \mathcal{Q}_{\mathcal{W}}(\ell'^n u'^n)$,

$$M^{-1}\sum_{m=1}^{M}\left| \mathcal{T}_{\tilde{Y}|\tilde{L}\tilde{U}}^n\big(\ell^n, \tilde{u}_m^n(\ell^n)\big) \cap \left[\bigcup_{m'\neq m}\ \bigcup_{P_{\overline{Y}|\tilde{L}\tilde{U}} \in \mathcal{Q}_{\mathcal{W}}(\ell^n u_{m'}^n(\ell^n))} \mathcal{T}_{\overline{Y}|\tilde{L}\tilde{U}}^n\big(\ell^n u_{m'}^n(\ell^n)\big)\right]\right| < 2^{-\frac{n}{9}\varepsilon_2}t_{\tilde{Y}|\tilde{L}\tilde{U}}. \tag{75}$$

We call the subset $\mathcal{U}(\ell^n)$ the codebook for ℓ^n and its members $\tilde{u}_m^n(\ell^n)$ $m = 1, 2, \ldots, M$ codewords.

2) Choosing Input Sequence to Send through the Channel:

The sender chooses an input sequence $x^n \in \mathcal{X}^n$ according to the output (v^n, ℓ^n) of the correlated source observed by him and his private randomness as follows.

— In the case that outcome of the source is a (δ_1, δ_2)–typical sequence (v^n, ℓ^n) with joint type $P_{\tilde{V}\tilde{L}}$, the sender chooses a codeword in $\mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n, \ell^n)$ in (73) randomly uniformly (by using his private randomness), say

$$\tilde{u}_m(\ell^n) \in \mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n, \ell^n) \subset \mathcal{U}(\ell^n). \tag{76}$$

Then the sender chooses an input sequence $x^n \in \mathcal{X}^n$ with probability

$$P_{X|VLU}\big(x^n|v^n, \ell^n, \tilde{u}^n_m(\ell^n)\big) \tag{77}$$

by using the chosen $\tilde{u}^n_m(\ell^n)$ and his private randomness and sends it through the channel.

— In the other case i.e., a non–(δ_1, δ_2)–typical sequence is output, the sender chooses an arbitrarily fixed sequence, say x^n_e, and sends it through the channel.

— The codewords randomly chosen here and the random input of the channel generated here will be denoted by U'^n and X'^n in the part of analysis below.

3) Choosing the Common domain $\mathcal{A}$ of Functions F and G:

Let

$$J = \lfloor 2^{n(H(L)-2\varepsilon_1)} \rfloor \tag{78}$$

and let e be an abstract symbol (which stands for that "an error occurs"). Then we define

$$\mathcal{A} = \big\{\{1, 2, \ldots, M\} \times \{1, 2, \ldots, J\}\big\} \cup \{e\}. \tag{79}$$

4) Defining the Functions F and G:

To define functions F and G we first partition each $\mathcal{T}^n_{\tilde{L}} \subset \mathcal{T}^n_L(\delta_1)$ into J subsets with nearly equal size i.e., each subset has cardinality $\Big\lfloor \frac{|\mathcal{T}^n_{\tilde{L}}|}{J} \Big\rfloor$ or $\Big\lceil \frac{|\mathcal{T}^n_{\tilde{L}}|}{J} \Big\rceil$. Then we take the union of the jth subsets in the partitions over all $\mathcal{T}^n_{\tilde{L}} \subset \mathcal{T}^n_L(\delta_1)$ and obtain a subset $\mathcal{L}_j$ of $\mathcal{T}^n_L(\delta_1)$. That is for $j = 1, 2, \ldots, J$

$$|\mathcal{L}_j \cap \mathcal{T}^n_{\tilde{L}}| = \left\lfloor \frac{|\mathcal{T}^n_{\tilde{L}}|}{J} \right\rfloor \text{ or } \left\lceil \frac{|\mathcal{T}^n_{\tilde{L}}|}{J} \right\rceil. \tag{80}$$

4.1) Defining Function F:

The sender observes the output of the source and decides on the value of function F.

— In the case that the source outputs a (δ_1, δ_2)–typical sequence (v^n, ℓ^n), F takes value (m, j) if $\ell^n \in \mathcal{L}_j$, according to sender's private randomness $\tilde{u}_m(\ell^n)$ in (76) is choosen in the step 2) of the coding scheme.

— In the other case $F = e$.

4.2) Defining Function G:

The receiver observes the output ℓ^n of the component L^n (side information) of the correlated source and output of the channel y^n to decide on the value of function G. We use the abbreviation $\mathcal{Y}_m(\ell^n) = \bigcup_{P_{\overline{Y}|\tilde{L}\tilde{U}} \in \mathcal{Q}_{\mathcal{W}}(\ell^n \tilde{u}^n_m(\ell^n), \tau)} \mathcal{T}^n_{\overline{Y}|\tilde{L}\tilde{U}}\big(\ell^n \tilde{u}^n_m(\ell_n), \tau\big)$.

— In the case that $\ell^n \in \mathcal{T}_L^n(\delta_1)$ and that there exists an $m \in \{1, 2, \ldots, M\}$ such that $y^n \in \mathcal{Y}_m(\ell^n) \setminus \left\{ \bigcup_{m' \neq m} \mathcal{Y}_m(\ell^n) \right\}$ G takes value (m, j) if $\ell^n \in \mathcal{L}_j$. Notice that this m must be unique if it exists.

— In the other case $G = e$.

Analysis:

1) Distortion Criterion:

First we recall our assumption that the watermarking distortion measure ρ is bounded i.e.

$$0 \leq \rho \leq \Delta. \tag{81}$$

Then by (58)

$$\frac{1}{n} Pr\big((V^n, L^n) \notin \mathcal{T}_{VL}^n(\delta_2)\big) E\big[\rho(V'^n, X'^n) | (V^n, L^n \notin \mathcal{T}_{VL}^n(\delta_2)\big] < 2^{-n\xi}\Delta. \tag{82}$$

On the other hand, under the condition that $(V^n, L^n) \in \mathcal{T}_{\tilde{V}\tilde{L}}^n \subset \mathcal{T}_{VL}^n(\delta_2)$, by definition $(V^n, L^n, U'^n) \in \mathcal{T}_{\tilde{V}\tilde{L}\tilde{U}}^n$ with probability one for the joint type $P_{\tilde{V}\tilde{L}\tilde{U}}$ generated by $P_{\tilde{V}\tilde{L}}$. So, by (60), (61) and the definition of (U'^n, X'^n) we have that

$$\begin{aligned}
&\frac{1}{n} E\big[\rho(V'^n, X'^n) | (V^n, L^n) \in \mathcal{T}_{\tilde{L}\tilde{V}}^n\big] \\
&\quad = \sum_{(v,\ell,u) \in \mathcal{V} \times \mathcal{L} \times \mathcal{U}} P_{\tilde{V}\tilde{L}\tilde{U}}(v, \ell, u) \sum_x P_{X|VLU}(x|v, \ell, u) \rho(v, x) \\
&\quad = E\rho(\tilde{V}, \tilde{X}) < D_1.
\end{aligned} \tag{83}$$

Thus it follows from (82) and (83) that

$$\begin{aligned}
\frac{1}{n} E\rho(V^n, X'^n) &= Pr\big((V^n, L^n) \notin \mathcal{T}_{VL}^n(\delta_2)\big) E\big[\rho(V^n, X'^n) | (V^n, L^n) \notin \mathcal{T}_{VL}^n(\delta_2)\big] \\
&\quad + \sum_{\mathcal{T}_{\tilde{V}\tilde{L}}^n \subset \mathcal{T}_{VL}^n(\delta_2)} Pr\big((V^n, L^n) \in \mathcal{T}_{\tilde{V}\tilde{L}}^n\big) E\big[\rho(V^n, X'^n) | (V^n, L^n) \in \mathcal{T}_{\tilde{V}\tilde{L}}^n\big] \\
&\quad < D_1,
\end{aligned} \tag{84}$$

for sufficiently large n.

2) The Condition of Nearly Uniformity

By the definition of function F in the step 4.1) of the coding scheme, $Pr\{F = e\} \leq Pr\big\{(V^n, L^n) \notin \mathcal{T}_{VL}^n(\delta_2)\big\} = 1 - P_{VL}^n(\mathcal{T}_{VL}^n(\delta_2)\big\}$, and hence by (58),

$$|Pr\{F = e\} - |\mathcal{A}|^{-1}| \leq \max\{2^{-n\zeta}, |\mathcal{A}|^{-1}\} \longrightarrow 0 \ (n \to \infty). \tag{85}$$

Next fix an $\ell^n \in \mathcal{T}^n_L(\delta_1)$ with type $P_{\tilde{L}}$ (say), let $P_{\tilde{L}\tilde{U}}$ be the joint type generated by $P_{\tilde{L}}$, and let $\mathcal{Q}(\tilde{L}\tilde{U})$ be the set of joint types $P_{\tilde{V}\tilde{L}\tilde{U}}$ with marginal distribution $P_{\tilde{L}\tilde{U}}$ and generated by the joint type of some (δ_1,δ_2)–typical sequence. Then $Pr\{U'^n = u^n | L^n = \ell^n\} > 0$, only if $u^n \in \mathcal{U}(\ell^n) = \{\tilde{u}^n_m(\ell^n) : m = 1,2,\ldots,M\}$.

Moreover, for a (δ_1,δ_2)–typical sequence (v^n,ℓ^n) with joint type $P_{\tilde{V}\tilde{L}}$, $\tilde{u}^n_m(\ell^n) \in \mathcal{U}(\ell^n)$, by the coding scheme

$$
\begin{aligned}
&Pr\{V^n = v^n, U'^n = u^n_m(\ell^n) | L = \ell^n\} \\
&= \begin{cases} P^n_{V|L}(V^n = v^n|\ell^n)|\mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)|^{-1} & \text{if } u^n_m(\ell^n) \in \mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n) \\ 0 & \text{else.} \end{cases}
\end{aligned} \tag{86}
$$

Recalling (73), then we have that for all $\ell^n \in \mathcal{T}^n_{\tilde{L}} \subset \mathcal{T}^n_L(\delta_1)$, $\tilde{u}^n_m(\ell^n) \in \mathcal{U}(\ell^n)$

$$
Pr\{U'^n = \tilde{u}^n_m(\ell^n)|L = \ell^n\} = \sum_{P_{\tilde{V}\tilde{L}\tilde{U}} \in \mathcal{Q}(\tilde{L}\tilde{U})} \sum_{v^n \in \mathcal{T}^n_{\tilde{V}|\tilde{L}\tilde{U}}(\ell^n \tilde{u}^n_m(\ell^n))} P^n_{V|L}(v^n|\ell^n)|\mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)|^{-1}. \tag{87}
$$

By (72) we have that

$$
[M(1+\varepsilon)]^{-1} \frac{|\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)|}{|\mathcal{T}^n_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)|} < |\mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)|^{-1} < [M(1-\varepsilon)]^{-1} \frac{|\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)|}{|\mathcal{T}^n_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)|}. \tag{88}
$$

On the other hand,

$$
\begin{aligned}
&\sum_{P_{\tilde{V}\tilde{L}\tilde{U}} \in \mathcal{Q}(\tilde{L}\tilde{U})} \sum_{v^n \in \mathcal{T}^n_{\tilde{V}|\tilde{L}\tilde{U}}(\ell^n \tilde{u}^n_m(\ell^n))} P^n_{V|L}(v^n|\ell^n) \frac{|\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)|}{|\mathcal{T}^n_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)|} \\
&= \sum_{P_{\tilde{V}\tilde{L}\tilde{U}} \in \mathcal{Q}(\tilde{L}\tilde{U})} \sum_{v^n \in \mathcal{T}^n_{\tilde{V}|\tilde{L}\tilde{U}}(\ell^n \tilde{u}^n_m(\ell^n))} P^n_{V^n|L}(\mathcal{T}^n_{\tilde{V}|\tilde{L}}(\ell^n)|\ell^n) \frac{|\mathcal{T}^n_{\tilde{U}|\tilde{L}}(\ell^n)|}{|\mathcal{T}^n_{\tilde{V}|\tilde{L}}(\ell^n)||\mathcal{T}^n_{\tilde{U}|\tilde{V}\tilde{L}}(v^n,\ell^n)|} \\
&= \sum_{P_{\tilde{V}\tilde{L}\tilde{U}} \in \mathcal{Q}(\tilde{L}\tilde{U})} P^n_{V|L}\big(\mathcal{T}^n_{\tilde{V}|\tilde{L}}(\ell^n)|\ell^n\big) \\
&= Pr\{(V^n,\ell^n) \in \mathcal{T}^n(\delta_1,\delta_2)|\ell^n\},
\end{aligned} \tag{89}
$$

where the first equality holds because the value of $P^n_{V|L}(v^n|\ell^n)$ for given ℓ^n depends on v^n through the conditional type; the second equality hold by the fact that $\frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{V}|\tilde{L}} t_{\tilde{U}|\tilde{V}\tilde{L}}} = \frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{V}\tilde{U}|\tilde{L}}} = \frac{1}{t_{\tilde{V}|\tilde{L}\tilde{U}}}$; and the last equality holds because $P_{\tilde{V}\tilde{L}\tilde{U}}$ is generated by $P_{\tilde{V}\tilde{L}}$ uniquely (see its definition in condition (c)).

Thus by combining (57), (87) – (89), we obtain for an $\eta > 0$ with $\eta \to 0$ as $n \to \infty$, $\varepsilon \to 0$

$$
(1-\eta)M^{-1} < Pr\{U'^n = \tilde{u}^n_m(\ell^n)|L = \ell^n\} < (1+\eta)M^{-1}, \tag{90}
$$

for $\ell^n \in \mathcal{T}^n_{L(\delta_1)}$, $\tilde{u}^n_m(\ell^n) \in \mathcal{U}(\ell^n)$.

So for $m \in \{1, 2, \ldots, M\}$, $j \in \{1, 2, \ldots, J\}$,

$$\begin{aligned} Pr\big\{F = (m, j)\big\} &= Pr\big\{U'^n = \tilde{u}^n_m(L^n), L^n \in \mathcal{L}_j\big\} \\ &= \sum_{\ell^n \in \mathcal{L}_j} P^n_L(\ell^n) Pr\big\{U'^n = \tilde{u}^n_m(\ell^n) | L = \ell^n\big\} < (1+\eta) M^{-1} P^n_L(\mathcal{L}_j). \end{aligned} \tag{91}$$

Since $|\mathcal{T}^n_{\tilde{L}}| > 2^{n\left(H(\tilde{L}) + \frac{\varepsilon_1}{2}\right)}$ for sufficiently large n, by (62) and (78), we have that $\frac{|\mathcal{T}^n_{\tilde{L}}|}{J} > 2^{\frac{n}{2}\varepsilon_1}$ and hence by (80)

$$|\mathcal{L}_j \cap \mathcal{T}^n_{\tilde{L}}| \leq \left\lceil \frac{|\mathcal{T}^n_{\tilde{L}}|}{J} \right\rceil < \frac{|\mathcal{T}^n_{\tilde{L}}|}{J} + 1 < \frac{|\mathcal{T}^n_{\tilde{L}}|}{J} \left(1 + 2^{-\frac{n}{2}\varepsilon_1}\right).$$

Because the value of $P^n_L(\ell^n)$ depends on ℓ^n through its type, this means that

$$P^n_L(\mathcal{L}_j \cap \mathcal{T}^n_{\tilde{L}}) < J^{-1} P^n_L(\mathcal{T}^n_{\tilde{L}}) \left(1 + 2^{-\frac{n}{2}\varepsilon}\right)$$

and consequently

$$P^n_L(\mathcal{L}_k) < P^n_L\big(\mathcal{T}^n_L(\delta_1)\big) J^{-1} \left(1 + 2^{-\frac{n}{2}\varepsilon_1}\right) \tag{92}$$

which with (91) is followed by

$$Pr\big\{F = (m, j)\big\} < M^{-1} J^{-1} (1+\eta) \left(1 + 2^{-\frac{n}{2}\varepsilon_1}\right) P^n_L\big(\mathcal{T}^n_L(\delta_1)\big). \tag{93}$$

Similarly we have that

$$Pr\big\{F = (m, j)\big\} > M^{-1} J^{-1} (1-\eta) \left(1 - 2^{-\frac{n}{2}\varepsilon_1}\right) P^n_L\big(\mathcal{T}^n_L(\delta_1)\big). \tag{94}$$

Now (56), (93) and (94) together imply that for an $\eta' > 0$ with $\eta' \to 0$ as $n \to \infty$, $\eta \to 0$,

$$\sum_{(m,j)} |Pr\big\{F = (m, j)\big\} - |\mathcal{A}|^{-1}| < \eta', \tag{95}$$

which with (85) completes the proof of condition of nearly uniformity.

3) The Rate:

In (65) one can choose

$$\alpha > I\big(U; Y(\mathcal{W}) | L\big) - \varepsilon' \text{ for any } \varepsilon' \text{ with } \varepsilon_2 < \varepsilon' < \frac{1}{2}\xi.$$

Then by (58), (78), (79), (95), we know that for an $\eta'' > 0$ with $\eta'' \to 0$ as $n \to \infty$, $\eta' \to 0$

$$\begin{aligned} \frac{1}{n} H(F) > \frac{1}{n} \log |\mathcal{A}| - \eta'' &> I\big(U; Y(\mathcal{W}) | L\big) - \varepsilon' + H(L) - 2\varepsilon_1 - \eta' \\ &= I\big(U; Y(\mathcal{W}) | L\big) + I(U; L) + H(L|U) - \varepsilon' - 2\varepsilon_1 - \eta' \\ &= I\big(U; L, Y(\mathcal{W})\big) + H(L|U) - \varepsilon' - 2\varepsilon_1 - \eta', \end{aligned}$$

for sufficiently large n.

4) Estimation of Probability of Error:

In and only in the following three cases an error occurs.

Case 1:

The source outputs a non–(δ_1, δ_2)–typical sequence whose probability is less than $2^{-n\zeta}$ by (58).

Now we assume that a (δ_1, δ_2)–typical sequence (v^n, ℓ^n) with joint type $P_{\tilde{V}\tilde{L}}$ is output. So the sender first chooses a $\tilde{u}^n_m(\ell^n) \in \mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n, \ell^n)$, then an $x^n \in \mathcal{X}^n$ according to his private randomness and sends x^n through the channel. Consequently a $y^n \in \mathcal{Y}^n$ is output by the channel. Then in the following two cases an error occurs.

Case 2:

A codeword $\tilde{u}_m(\ell^n) \in \mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n) \subset \mathcal{U}^n(\ell^n)$ is chosen and an output sequence

$$y^n \notin \mathcal{Y}_m(\ell^n) = \bigcup_{P_{\overline{Y}|\tilde{L}\tilde{U}} \in \mathcal{Q}_{\mathcal{W}}(\ell^n, \tilde{u}_m(\ell^n))} \mathcal{T}^n_{\overline{Y}|\tilde{L}\tilde{U}}\big(\ell^n \tilde{u}_m(\ell^n), \tau\big)$$

is output of the channel. Suppose now $W \in \mathcal{W}$ governs the channel. Then by (59), and (60) the probability that $y^n \in \mathcal{Y}^n$ is output of the channel under the condition that $(V^n, L^n) = (v^n, \ell^n) \in \mathcal{T}^n(\delta_1, \delta_2)$ is output of the correlated source and $U'^n = \tilde{u}^n_m(\ell^n) \in \mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n, \ell^n)$ is chosen is

$$Pr\big\{Y'^n = y^n|(V^n, L^n) = (v^n, \ell^n), U'^n = \tilde{u}^n_m(\ell^n)\big\} = \sum_{x^n \in \mathcal{X}^n} P^n_{X|VLU}(x^n|v^n, \ell^n, \tilde{u}^n_m(\ell^n))W^n(y^n|x^n)$$

$$= P^n_{\tilde{Y}(\mathcal{W})|\tilde{V}\tilde{L}\tilde{U}}\big(y^n|v^n, \ell^n, \tilde{u}_m(\ell^n)\big). \tag{96}$$

On the other hand

$$\mathcal{T}^n_{\tilde{Y}(W)|\tilde{V}\tilde{L}\tilde{U}}\big(v^n\ell^n u^n_m(\ell^n), \tau\big) \subset \mathcal{T}^n_{\tilde{Y}(W)|\tilde{L}\tilde{U}}\big(\ell^n u^n_m(\ell^n), \tau\big) \subset \mathcal{Y}_m.$$

So the probability that such an error occurs vanishes exponentially as n grows.

Case 3:

A codeword $\tilde{u}^n_m(\ell^n)$ is chosen and a $y^n \in \mathcal{Y}_m \cap \Big[\bigcup_{m' \neq m} \mathcal{Y}_{m'}\Big]$ is output of the channel.

Now by (86), (88), (90), and simple calculation, we obtain that

$$\big[(1-\eta)(1-\varepsilon)\big]^{-1} P^n_{V|L}(v^n|\ell^n)\frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{U}|\tilde{V}\tilde{L}}} < Pr\big\{V^n = v^n|L^n = \ell^n, U'^n = \tilde{u}^n_m(\ell^n)\big\}$$
$$< \big[(1+\eta)(1+\varepsilon)\big]^{-1} P^n_{V|L}(v^n|\ell^n)\frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{U}|\tilde{V}\tilde{L}}} \tag{97}$$

for (δ_1, δ_2)–typical sequences (v^n, ℓ^n) with joint type $P_{\tilde{V}\tilde{L}}$ and $\tilde{u}_m(\ell^n) \in \mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n, \ell^n)$, where $P_{\tilde{V}\tilde{L}\tilde{U}}$ is the type generated by $P_{\tilde{V}\tilde{L}}$.

Moreover, since $t_{\tilde{U}|\tilde{V}\tilde{L}} = \frac{t_{\tilde{V}\tilde{U}|\tilde{L}}}{t_{\tilde{V}|\tilde{L}}}$, $t_{\tilde{U}\tilde{V}|\tilde{L}} = t_{\tilde{U}|\tilde{L}} t_{\tilde{V}|\tilde{L}\tilde{U}}$, and since for given ℓ^n, the value of $P^n_{V|L}(v^n|\ell^n)$ depends on v^n through the conditional type,

$$P^n_{V|L}(v^n|\ell^n)\frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{U}|\tilde{V}\tilde{L}}} = P^n_{V|L}(v^n|\ell^n)\frac{t_{\tilde{V}|\tilde{L}}}{t_{\tilde{V}|\tilde{L}\tilde{U}}} = P^n_{V|L}\big(\mathcal{T}^n_{\tilde{V}|\tilde{L}}(\ell^n)|\ell^n\big)\frac{1}{t_{\tilde{V}|\tilde{L}\tilde{U}}}. \tag{98}$$

Further it is well known that for all
$(v^n, \ell^n, u^n) \in \mathcal{T}^n_{\tilde{V}\tilde{L}\tilde{U}}$, $\lim_{n\to\infty} \frac{1}{n}\left(\log P^n_{\tilde{V}|\tilde{L}\tilde{U}}(v^n|\ell^n, u^n) - \log \frac{1}{t_{\tilde{V}|\tilde{L}\tilde{U}}}\right) = 0$.
So by (97) and (98), we have that

$$\begin{aligned} Pr\big\{V = v^n|L^n = \ell^n, U'^n = \tilde{u}^n_m(\ell^n)\big\} &< 2^{n\theta} P^n_{V|L}\big(\mathcal{T}^n_{\tilde{V}|\tilde{L}}(\ell^n)|\ell^n\big) P^n_{\tilde{V}|\tilde{L}\tilde{U}}\big(v^n|\ell^n, \tilde{u}^n_m(\ell^n)\big) \\ &\le 2^{n\theta} P^n_{\tilde{V}|\tilde{L}\tilde{U}}\big(v^n|\ell^n, \tilde{u}^n_m(\ell^n)\big) \end{aligned} \tag{99}$$

for (δ_1, δ_2)–typical sequences (v^n, ℓ^n) with type $P_{\tilde{V}\tilde{L}}$, $\tilde{u}^n_m \in \mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(\ell^n) \subset \mathcal{U}(\ell^n)$ and sufficiently large n, and a $\theta \to 0$ as $n \to \infty$.
We choose $\theta < \frac{1}{20}\varepsilon_2$.
Since $Pr\big\{(V^n, L^n) = (v^n, \ell^n), U'^n = u^n\big\} > 0$ only if (v^n, ℓ^n) is (δ_1, δ_2) typical and $u^n \in \mathcal{U}_{\tilde{U}|\tilde{V}\tilde{L}}(v^n, \ell^n)$, by (96) and (99) we have that

$$\begin{aligned} &Pr\big\{Y'^n = y^n|L^n = \ell^n, U'^n = \tilde{u}^n_m(\ell^n)\big\} \\ &= \sum_{v^n \in \mathcal{V}^n} Pr\big\{V^n = v^n|L^n = \ell^n, U'^n = \tilde{u}_m(\ell^n)\big\} Pr\big\{Y'^n = y^n|(V^n, L^n) = (v^n, \ell^n), U'^n = \tilde{u}^n_m(\ell^n)\big\} \\ &\le \sum_{v^n \in \mathcal{V}^n} 2^{n\theta} P^n_{\tilde{V}|\tilde{L}\tilde{U}}(v^n|\ell^n, u^n_m(\ell^n)\big\} P^n_{\tilde{Y}(W)|\tilde{V}\tilde{L}\tilde{U}}(y^n|v^n, \ell^n, u^n) \\ &\le 2^{n\theta} P^n_{\tilde{Y}(W)|\tilde{L}\tilde{U}}\big(y^n|\ell^n, u^n_m(\ell^n)\big) \end{aligned} \tag{100}$$

for $\ell^n \in \mathcal{T}^n_L(\delta_1)$, $\tilde{u}_m(\ell^n) \in \mathcal{U}(\ell^n)$ and $y^n \in \mathcal{Y}^n$ if $W \in \mathcal{W}$ governs the channel. Now we obtain an upper bound in terms of a product probability distribution $P^n_{\tilde{Y}(\mathcal{W})|\tilde{L}\tilde{U}}\big(y^n|\ell^n, u^n_m(\ell^n)\big)$ whose value depends on y^n through the conditional type. Consequently by (75) and (100) we have that for all $\ell^n \in \mathcal{T}^n_L(\delta_1)$, $\tilde{u}_m(\ell^n) \in \mathcal{U}(\ell^n)$ with joint type $P_{\tilde{L}\tilde{U}}$, $P_{\tilde{Y}|\tilde{L}\tilde{U}} \in \mathcal{Q}_{\mathcal{W}}\big(\ell^n \tilde{u}_m(\ell^n), \tau\big)$

$$\begin{aligned} &M^{-1}\sum_{m=1}^{M} Pr\left\{Y'^n \in \mathcal{T}^n_{\tilde{Y}|\tilde{L}\tilde{U}}\big(\ell^n, u^n_m(\ell^n)\big) \cap \left[\bigcup_{m'\neq m} \mathcal{Y}_{m'}(\ell^n)\right] |L^n = \ell^n, U'^n = \tilde{u}^n_m(\ell^n)\right\} \\ &\le 2^{n\theta} M^{-1} \sum_{m=1}^{M} P^n_{\tilde{Y}(W)|\tilde{L}\tilde{U}}\left\{\mathcal{T}^n_{\tilde{Y}|\tilde{L}\tilde{U}}\big(\ell^n u^n_m(\ell^n)\big) \cap \left[\bigcup_{m'\neq m} \mathcal{Y}_{m'}(\ell^n)\right] |\ell^n, u^n_m(\ell^n)\right\} \\ &\le 2^{n\theta} \cdot 2^{-\frac{n}{9}\varepsilon_2} P^n_{\tilde{Y}(W)|\tilde{L}\tilde{U}}\big\{\mathcal{T}^n_{\tilde{Y}|\tilde{L}\tilde{U}}\big(\ell^n, u^n_m(\ell^n)\big)|\ell^n, \tilde{u}^n_m(\ell^n)\big\} \le 2^{-n\left(\frac{1}{9}\varepsilon_2 - \theta\right)} < 2^{-\frac{n}{20}\varepsilon_2}, \end{aligned} \tag{101}$$

where the last inequality holds by our choice $\theta < \frac{\varepsilon_2}{20}$. Recalling

$$\mathcal{Y}_m(\ell^n) = \bigcup_{P_{\tilde{Y}|\tilde{L}\tilde{U}} \in \mathcal{Q}_{\mathcal{W}}(\ell^n \tilde{u}_m(\ell^n), \tau)} \mathcal{T}^n_{\tilde{Y}|\tilde{L}\tilde{U}}\big(\ell^n u^n_m(\ell^n)\big),$$

by the union bound and (101) we obtain that

$$M^{-1}\sum_{m=1}^{M} Pr\left\{Y'^n \in \mathcal{Y}_m(\ell^n) \cap \left[\bigcup_{m'\neq m} \mathcal{Y}_m(\ell^n)\right] | L^n = \ell^n, U'^n = \tilde{u}_m(\ell^n)\right\}$$
$$< (n+1)^{|\mathcal{L}||\tilde{U}||\mathcal{Y}|} 2^{-\frac{n}{20}\varepsilon_2}$$
$$< 2^{-\frac{n}{21}\varepsilon_2} \tag{102}$$

for $\ell^n \in \mathcal{T}_L^n(\delta_1)$, $\tilde{u}_m^n(\ell^n) \in \mathcal{U}(\ell^n)$ and sufficiently large n. Finally by (90) and (102) we obtain an upper bound to the probability that on error of this type occurs, under the condition $L^n = \ell^n \in \mathcal{T}_L^n(\delta_1)$.

$$\sum_{m=1}^{M} Pr\{U'^n = \tilde{u}_m(\ell^n)|L^n = \ell^n\} Pr\left\{Y'^n \in \mathcal{Y}_m(\ell^n) \cap \left[\bigcup_{m'\neq m} \mathcal{Y}_{m'}(\ell^n)\right] | L^n = \ell^n, U'^n = \tilde{u}_m(\ell^n)\right\}$$
$$< (1+\eta)\sum_{m=1}^{M} M^{-1} Pr\left\{Y'^n \in \mathcal{Y}_m(\ell^n) \cap \left[\bigcup_{m'\neq m} \mathcal{Y}_{m'}(\ell^n)\right] | L^n = \ell^n, U'^n = \tilde{u}_m(\ell^n)\right\}$$
$$< (1+\eta) 2^{-\frac{n}{21}\varepsilon_2}, \tag{103}$$

which completes the proof because by definition $\sum_{m=1}^{M} Pr\{U'^n = \tilde{u}_m(\ell^n)|L^n = \ell^n\} = 1$ for all $\ell^n \in \mathcal{T}_L^n(\delta_1)$.

Remark: Our model of identification becomes that in [32] if L takes a constant value with probability one. So our proof of the lemma above provides a new proof of Theorem 4 in [32] (as special case) without using the Gelfand–Pinsker Theorem in [24].

Corollary 5.1 (Direct Part Theorem 4.1): *For all single channels W*

$$C_{CRI}\big((V,L),W,D_1\big) \geq \max_{(V,L,U,X,Y)\in\mathcal{Q}((V,L),W,D_1)} \big[I(U;L,Y) + H(L|U)\big].$$

Lemma 5.5 (Direct Part of Theorem 4.4): *For all compound channels $\mathcal{W}$*

$$C_{CRII}\big((V,L),W,R_K,D_1\big) \geq \sup_{(V,L,U,X)\in\mathcal{Q}_1^*((V,L),\mathcal{W},R_K,D_1)} \big[I\big(U;L,Y(\mathcal{W})\big) + H(L|U)\big] + R_K. \tag{104}$$

Proof: By the same reason as in the proof of the previous lemma, it is sufficient for us to show the achievability of $I\big(U;L,Y(\mathcal{W})\big)+H(L|U)+R_K$ for (V,L,U,X) with $E\rho(V,X) < D_1$ and for some $\xi > 0$

$$I\big(U;Y(\mathcal{W})|L\big) + R_K - I(U;V|L) > \xi. \tag{105}$$

In the case $I\big(U;Y(\mathcal{W})|L\big) > I(U;V|L)$, by the previous lemma $I\big(U;LY(\mathcal{W})\big) + H(L|U)$ is achievable even if the noiseless channel is absent. So sender and receiver may generate $n\big(I(U;LY(\mathcal{W})) + H(L|U)\big)$ bits of common randomness and at the same time the sender sends R_K bits of his private randomness via the noiseless channel to the receiver to make additionally nR_K bits of common randomness. That is, the rate $I\big(U;L,Y(\mathcal{W})\big)+H(L|U)+R_K$ is achievable.

So, next we may assume that $I\big(U;Y(\mathcal{W})|L\big) \leq I(U;V;|L)$. Moreover we can assume $I\big(U;Y(\mathcal{W})|L\big) > 0$, because otherwise $I\big(U;L,Y(\mathcal{W})\big)+H(L|U)+R_K = I(U;L)+H(L|U)+R_K = H(L)+R_K$ is achievable as follows. We partition $\mathcal{T}_L^n(\delta_1)$ into $\mathcal{L}_j$, $j=1,2,\ldots,J$ as in the step 4) of the coding scheme in the proof of the previous lemma to get $n\big(H(L)-2\varepsilon_1\big)$ bits of common randomness and get other nR_K bits of common randomness by using the noiseless channel. Thus it is sufficient for us to assume that

$$0 < I\big(U;Y(\mathcal{W})|L\big) \leq I(U;V|L) < I\big(U;Y(\mathcal{W})|L\big) + R_K - \xi, \tag{106}$$

for a ξ with $0<\xi<R_K$.

We shall use (δ_1,δ_2)–typical sequences, the joint types $P_{\tilde{L}\tilde{U}}$ and $P_{\tilde{V}\tilde{L}\tilde{U}}$ generated by the types $P_{\tilde{L}}$ and $P_{\tilde{U}\tilde{L}}$ respectively, and the random variables $(\tilde{V},\tilde{L},\tilde{U},\tilde{X})$ and $\big(\tilde{V},\tilde{L},\tilde{U},\tilde{X},\tilde{Y}(\mathcal{W})\big)$ in (59) and (60) satisfying (61) – (64), which are defined in the conditions (a) – (d) in the proof of the previous lemma.

Instead of the choice α in (65) we now choose $\beta_1,\beta_2>0$ and $\beta_3\geq 0$ for arbitrarily small but fixed ε_2 with $0<\varepsilon_2<\frac{1}{2}\xi$ such that

$$I\big(U;Y(\mathcal{W})|L\big) - \frac{3}{2}\varepsilon_2 < \beta_1 < I\big(U;Y(\mathcal{W})|L\big) - \varepsilon_2, \tag{107}$$

$$I(U;V|L) - I\big(U;Y(\mathcal{W})|L\big) + \xi \leq \beta_2 \leq R_K \tag{108}$$

and

$$0 \leq \beta_3 = R_K - \beta_2. \tag{109}$$

Notice that the existence and positivity of β_2 are guaranteed by (106).

By adding both sides of the first inequalities in (107) and (108), we obtain that

$$\beta_1+\beta_2 > I(U;V|L) + \left(\xi - \frac{3}{2}\varepsilon_2\right), \tag{110}$$

and by the first inequality in (107) and the equality in (109) we have that

$$\beta_1+\beta_2+\beta_3 > I\big(U;Y(\mathcal{W})|L\big) + R_K - \frac{3}{2}\varepsilon_2. \tag{111}$$

Let $\xi - \frac{3}{2}\varepsilon_2 = 2\eta$ and rewrite (110) as

$$\beta_1+\beta_2 > I(U;V|L) + 2\eta. \tag{112}$$

Then $\eta > \frac{\xi}{8} > 0$ by our choice $\varepsilon_2 < \frac{1}{2}\xi$.

Next as in the proof to the previous lemma we fix an (arbitrary small) positive ε_2, η, choose ε_1 (and consequently δ_1, δ_2) sufficiently small so that $\varepsilon_1 < \min\left(\frac{1}{2}\varepsilon_2, \frac{1}{2}\eta\right)$. Then by (64) and the second inequality in (109) we have that

$$\beta_1 < I\big(\tilde{U}; \tilde{Y}(\mathcal{W})|\tilde{L}\big) - \frac{\varepsilon_2}{2}, \tag{113}$$

and by (65) and (112) we have that

$$\beta_1 + \beta_2 > I(\tilde{U}; \tilde{V}|\tilde{L}) + \frac{3}{2}\eta. \tag{114}$$

Without loss of generality we assume that $2^{n\beta_1}$, $2^{n\beta_2}$ and $2^{n\beta_3}$ are integers and denote by $M_1 = 2^{n\beta_1}$, $I = 2^{n\beta_2}$ and $K' = 2^{n\beta_3}$.

Then similarly as in the proof of the previous lemma, we have that for sufficiently large n, sufficiently small τ, all joint types $P_{\tilde{V}\tilde{L}\tilde{U}}$ generated by types of (δ_1, δ_2)–typical sequences and $\mathcal{Q}_{\mathcal{W}}(\ell^n u^n, \tau)$ in the proof of the previous lemma,

$$2^{n\eta}\frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{U}|\tilde{V}\tilde{L}}} < M_1 I \tag{115}$$

and

$$M_1 < 2^{-\frac{n}{3}\varepsilon_2}\frac{t_{\tilde{U}|\tilde{L}}}{t_{\tilde{U}|\tilde{L}\overline{Y}}}, \tag{116}$$

for all $P_{\overline{Y}|\tilde{L}\tilde{U}} \in \mathcal{Q}_{\mathcal{W}}(\ell^n u^n, \tau)$.

Coding Scheme:

1) Choosing the Codebook:

   We choose a codebook for all $\ell^n \in \mathcal{T}_L^n(\delta_1)$ in a similar way as in the step 1) of the coding scheme in the proof of the previous lemma. But we now use Lemm 5.1 for $\alpha = \beta_1 + \beta_2$ and Lemma 5.3 for $\gamma = \frac{\varepsilon_2}{3}$ instead of Lemmas 5.1 and 5.2. Thus by random choice we obtain subsets of $\mathcal{T}_U^n$ $\mathcal{U}^i(\ell^n) = \left\{\tilde{u}^n_{m,i}(\ell^n) : m = 1, 2, \ldots, M_1\right\}$ for $i = 1, 2, \ldots, I$ for all $\ell^n \in \mathcal{T}_L^n(\delta_1)$ such that for

$$\mathcal{U}^*(\ell^n) = \bigcup_{i=1}^{I} \mathcal{U}^i(\ell^n), \tag{117}$$

   and $\mathcal{U}^*_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n) = \mathcal{U}^*(\ell^n) \cap \mathcal{T}^n_{\tilde{U}|\tilde{V}\tilde{L}}(v^n, \ell^n)$, where $P_{\tilde{V}\tilde{L}\tilde{U}}$ is the type generated by the joint type $P_{\tilde{V}\tilde{L}}$ of (δ_1, δ_2)–sequences (v^n, ℓ^n) as before, and with an abuse of notation in the union in (117): counting it twice and labelling it as different elements $\tilde{u}^n_{m,i}(\ell^n)$ and $\tilde{u}^n_{m',i'}(\ell^n)$ if a codeword appears twice in it, the following holds.

$$\left|\mathcal{U}^*_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n) - M_1 I\frac{t_{\tilde{U}|\tilde{V}\tilde{L}}}{t_{\tilde{U}|\tilde{V}\tilde{L}}}\right| < M_1 I\frac{t_{\tilde{U}|\tilde{V}\tilde{L}}}{t_{\tilde{U}|\tilde{V}\tilde{L}}}\varepsilon, \tag{118}$$

and for $\mathcal{Q}_{\mathcal{W}}(\ell^n v^n, \tau)$ in the proof of the previous lemmas and any conditional type $P_{\tilde{Y}|\tilde{L}\tilde{U}}$,

$$I^{-1}\sum_{i=1}^{I} M_1^{-1}\sum_{m=1}^{M_1}\left|\mathcal{T}^n_{\tilde{Y}|\tilde{L}\tilde{U}}\big(\ell^n u^n_{m,i}(\ell^n)\big)\cap\left[\bigcup_{m'\neq m}\ \bigcup_{P_{\overline{Y}|\tilde{L}\tilde{U}}\in\mathcal{Q}_{\mathcal{W}}(\ell^n v^n)} \mathcal{T}^n_{\overline{Y}|\tilde{L}\tilde{U}}\big(\ell^n u^n_{m',i}(\ell^n)\big)\right]\right|$$
$$< 2^{-\frac{n}{7}\varepsilon_2} t_{\tilde{Y}|\tilde{L}\tilde{V}} \tag{119}$$

here (118) and (119) are analogous to (72) and (75) respectively, and are shown in an analogous way.

2) Choosing Inputs of the Channels:

In the current model, we have an additional noiseless channel with rate R_K except for the noisy channel which exists in the Model I. The sender chooses the inputs of the two channels as follows.

2.1) Choosing the Input Sequence of the Noisy Channel:

— In the case that the source outputs a (δ_1, δ_2)–typical sequence (v^n, ℓ^n) with joint type $P_{\tilde{V}\tilde{L}}$, by (118) for the type $P_{\tilde{V}\tilde{L}\tilde{U}}$ generated by $P_{\tilde{V}\tilde{L}}$, $\mathcal{U}^*_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n) \neq \varnothing$. Then similarly to the Step 2) of the coding scheme in the proof of the previous lemma, the sender randomly and uniformly chooses a member of $\mathcal{U}^*_{\tilde{U}|\tilde{V}\tilde{L}}(v^n\ell^n)$, say $\tilde{u}^n_{m,i}(\ell^n)$, and according to the probability $P_{X|VLU}\big(x^n|v^n, \ell^n, \tilde{u}^n_{m,i}(\ell^n)\big)$ chooses an input sequence x^n of the channel $\mathcal{W}$ and sends x^n through the channel.

— In the case that the output of the source is non–(δ_1, δ_2)–typical, the sender sends an arbitrary fixed sequence x^n_e through the channel.

2.2) Choosing the Input of the Noiseless Channel:

— In the case that a (δ_1, δ_1)–typical sequence (v^n, ℓ^n) with joint type $P_{\tilde{V}\tilde{L}}$ is output of the correlated channel, the sender first spends $\log I = n\beta_2$ bits to send the index $i \in \{1, 2, \ldots, I\}$ to the receiver via the noiseless channel if a codeword $\tilde{u}^n_{m,i}(\ell^n) \in \mathcal{U}^i(\ell^n) \subset \mathcal{U}^*(\ell^n)$ is chosen in the substep 2.1) in the current coding scheme, then he randomly and uniformly chooses a $k' \in \{1, 2, \ldots, K'\}$ independent of the output of the source and sends it through the noiseless channel by using the rest of $nR_K - n\beta_2 = n\beta_3 = \log K'$ bits.

— In the case that a non–(δ_1, δ_2)–typical sequence is output, the sender sends a constant message through the noiseless channel.

3) Choosing the Common Range $\mathcal{A}$ of Functions F and G:

Let J be as in (78) and

$$\mathcal{A} = \big[\{1, 2, \ldots, M_1\} \times \{1, 2, \ldots, I\} \times \{1, 2, \ldots, K'\} \times \{1, 2, \ldots, J\}\big] \cup \{e\}. \tag{120}$$

4) Defining the Functions G and G:

Partition $\mathcal{T}^n_L(\delta_1)$ into $\mathcal{L}_j$, $j = 1, 2, \ldots, J$ as in the step 4) of the coding scheme in the proof of the previous lemma and let $\mathcal{K}_n = \{1, 2, \ldots, I\} \times \{1, 2, \ldots, K'\}$.

4.1) Defining Function F:

The sender decides on the value of function F according to the output of the correlated source and his private randomness as follows.

— In the case that a (δ_1, δ_2)–typical sequence (v^n, ℓ^n) is output, F takes value (m, i, k', j) if $\ell^n \in \mathcal{L}_j$, $\tilde{u}^n_{m,i}(\ell^n) \in \mathcal{U}j(\ell^n) \cap \mathcal{U}^*_{\tilde{U}|\tilde{V}\tilde{L}}(v^n \ell^n)$ is chosen in step 2) of the current coding scheme, and k' is chosen for sending it via the noiseless channel in the last $n\beta_3$ bits (that means (i, k') is sent through the noiseless channel).

— In the other case $F = e$.

4.2) Defining Function G:

The receiver decides on the value of the function G according to the output $(i, k') \in \mathcal{K}_n$ of the noiseless channel, the output ℓ^n of the component L^n of the correlated source, and the output $y^n \in \mathcal{Y}^n$ of the noisy compound channel $\mathcal{W}$ as follows.

Let $\mathcal{Y}_{m.i}(\ell^n) = \bigcup_{P_{\overline{Y}|\tilde{L}\tilde{U}} \in \mathcal{Q}_{\mathcal{W}}(\ell^n \tilde{u}^n_{m,i}(\ell^n), \tau)} \mathcal{T}^n_{\overline{Y}|\tilde{L}\tilde{U}}\big(\ell^n u^n_m, i(\ell^n)\big)$ for $m = 1, 2, \ldots, M_1$, $i = 1, 2, \ldots, I$, and the type $P_{\tilde{L}\tilde{U}}$ generated by the type $P_{\tilde{L}}$ of $\ell^n \in \mathcal{L}_j \subset \mathcal{T}^n_L(\delta_1)$.

— In the case that (i, k') is output of the noiseless channel, $\ell^n \in \mathcal{T}^n_L(\delta_1)$ is output of the source, and there exists an $m \in \{1, 2, \ldots, M_1\}$ such that the output of the noisy compound channel $\mathcal{W}$, $y^n \in \mathcal{Y}_{m,i}(\ell^n) \smallsetminus \left\{ \bigcup_{m' \neq m} \mathcal{Y}_{m',i}(\ell^n) \right\}$, G takes value (m, i, k', j) if $\ell^n \in \mathcal{L}_j$.

— In the other case $G = e$.

Analysis:

1) – 3) Distortion Criterion, The Nearly Uniformity Condition, and the Rate.

One can verify the distortion criterion, the nearly uniformity condition and the rate

$$\frac{1}{n} \log H(F) > \beta_1 + \beta_2 + \beta_3 + o(1) = I\big(U; Y(\mathcal{W})|L\big) + R_K + o(1)$$

(c.f. (111)), and obtain analogous inequalities

$$(1-\eta)(M_1 I)^{-1} < Pr\big\{U'^n = u^n_{m,i}(\ell^n)|L = \ell^n\big\} < (1+\eta)(M_1 I)^{-1} \tag{121}$$

to the inequalities in (90) for $\ell^n \in \mathcal{T}^n_L(\delta_1)$, $u^n_{m,i}(\ell^n) \in \mathcal{U}^*(\ell^n)$ and random variable U'^n chosen by the sender in step 2) of the coding scheme in the same way as in parts 1) – 3) of the Analysis in the proof of the previous lemma except that the roles of $\mathcal{U}(\ell^n)$ and (72) there are played by $\mathcal{U}^*(\ell^n) = \bigcup_{i=1}^{I} \mathcal{U}^i(\ell^n)$ and (118). Notice that in those parts of the proof of the previous lemma (75) is not used, neither is (119) here correspondingly.

4) Estimation of Probability of Error:

   By the same reason as in the proof of the previous lemma, the probabilities of errors of the first two types, the error caused by that a non–(δ_1, δ_2)–typical sequence is output and the error caused by that $\tilde{u}_{m,i}(\ell^n)$ is chosen and $y^n \notin \mathcal{Y}_{m,i}(\ell^n)$ is output of the noisy compound channel exponentially vanish as n grows.

   Next by replacing $\mathcal{U}(\ell^n)$ and (75) by $\mathcal{U}^i(\ell^n)$ and (119), in the same way as in the proof of the previous lemma we now obtain

$$(M_1 I)^{-1} \sum_{i=1}^{I} \sum_{m=1}^{M_1} Pr \left\{ Y'^n \in \mathcal{Y}_{m,i}(\ell^n) \cap \left[\bigcup_{m' \neq m} \mathcal{Y}_{m',i}(\ell^n) \right] | L^n = \ell^n, U'^n = u^n_{m,i}(\ell^n) \right\}$$
$$< 2^{-\frac{n}{21}\varepsilon_2} \tag{122}$$

   instead of (102).

   Finally analogously to in the way to obtain (103) in the proof of the previous lemma from (90) and (102), we finish the proof by combining (121) and (122).

Corollary 5.2 (Direct Part of Theorem 4.2): *For all single channels W*

$$C_{CRII}\big((V,L), W, R_K, D_1\big) \geq \max_{(V,L,U,X,Y) \in \mathcal{Q}^*((V,L),W,R_K,D_1)} \big(I(U;L,Y) + H(L|U)\big) + R_K.$$

6 The Converse Theorems for Common Randomness

To obtain single letter characterizations for the converse parts of coding theorems for common randomness, we need a useful identity which appears in [22] (on page 314).

Lemma 6.1 *(Csizár-Körner) Let (A^n, B^n) be an arbitrary pair of random sequences and let C be an arbitrary random variable. Then*

$$H(A^n|C) - H(B^n|C) = \sum_{t=1}^{n} [H(A_t|A_{t+1}, A_{t+2}, \dots, A_n, B^{t-1}, C) - H(B_t|A_{t+1}, A_{t+2}, \dots, A_n, B^{t-1}, C)]. \tag{123}$$

Proof:

$$H(A^n|C)-(B^n|C)=\sum_{t=0}^{n-1}H(A_{t+1},A_{t+2},\ldots,A_n,B^t|C)-\sum_{t=1}^{n}H(A_{t+1},A_{t+2},\ldots,A_n,B^t|C)$$

$$=\sum_{t=1}^{n}H(A_t,A_{t+1},\ldots,A_n,B^{t-1}|C)-\sum_{t=1}^{n}H(A_{t+1},A_{t+2},\ldots,A_n,B^t|C)$$

$$=\sum_{t=1}^{n}[H(A_t,A_{t+1},\ldots,A_n,B^{t-1}|C)-H(A_{t+1},\ldots,A_n,B^{t-1}|C)]$$

$$-\sum_{t=1}^{n}[H(A_{t+1},A_{t+2},\ldots,A_n,B^t|C)-H(A_{t+1},\ldots,A_n,B^{t-1}|C)]$$

$$=\sum_{t=1}^{n}[H(A_t|A_{t+1},A_{t+2},\ldots,A_n,B^{t-1},C)-H(B_t|A_{t+1},A_{t+2},\ldots,A_n,B^{t-1},C)], \tag{124}$$

where $(A_{t+1},A_{t+2},\ldots,A_n,B^t)$ to be understood as A^n and B^n when $t=0$ and $t=n$ respectively.

Lemma 6.2 *(The converse part of Theorem 4.1)*
For single channel W,

$$C_{CRI}((V,L),W,D_1)\le max_{(V,L,U,X,Y)\in\mathcal{Q}((V,L),W,D_1)}[I(U;LY)+H(L|U)]. \tag{125}$$

Proof: Assume that for a source output of length n there are functions F and K such that for the channel W^n and the distortion measure (10) - (16) hold. Denote by X^n and Y^n the random input and output of the channel generated by the correlated source (V^n,L^n), sender's private randomness M, and the channel.

Then (10) be rewritten in terms of (V^n,X^n) as

$$\frac{1}{n}\mathbf{E}\rho(V^n,X^n)\le D_1 \tag{126}$$

Further by Fano inequality, (11) - (14), we have that

$$\begin{aligned}
&H(F)\\
\le&H(F)-H(F|G)+n\lambda\log\kappa+h(\lambda)\\
=&I(F;G)+n\lambda\log\kappa+h(\lambda)\\
\le&I(F;L^n,Y^n)+n\lambda\log\kappa+h(\lambda)\\
=&I(F;Y^n|L^n)+I(F;L^n)+n\lambda\log\kappa+h(\lambda)\\
\le&I(F;Y^n|L^n)+H(L^n)+n\lambda\log\kappa+h(\lambda)\\
=&I(F;Y^n|L^n)+\sum_{t=1}^{n}H(L_t)+n\lambda\log\kappa+h(\lambda)\\
=&\sum_{t=1}^{n}I(F;Y_t|L^n,Y^{t-1})+\sum_{t=1}^{n}H(L_t)+n\lambda\log\kappa+h(\lambda),
\end{aligned} \tag{127}$$

where $h(z) = -z\log z - (1-z)\log(1-z)$ for $z \in [0,1]$ is the binary entropy. Here the first inequality follows from the Fano inequality, (11), (12) and (14); the second inequality holds by (13); and the third equality holds because the source is memoryless. Since $I(F; V^n, L^n) \leq H(F)$, the first four lines in (127) is followed by

$$\begin{aligned}
0 &\leq I(F; L^n, Y^n) - I(F; V^n, L^n) + n\lambda\log\kappa + h(\lambda) \\
&\leq [I(F; Y^n|L^n) + I(F; L^n)] - [I(F; V^n|L^n) + I(F; L^n)] + n\lambda\log\kappa + h(\lambda) \\
&= I(F; Y^n|L^n) - I(F; V^n|L^n) + n\lambda\log\kappa + h(\lambda) \\
&= [H(Y^n|L^n) - H(Y^n|L^n, F)] - [H(V^n|L^n) - H(V^n|L^n, F)] + n\lambda\log\kappa + h(\lambda) \\
&= [H(Y^n|L^n) - H(V^n|L^n)] + [H(V^n|L^n, F) - H(Y^n|L^n, F)] + n\lambda\log\kappa + h(\lambda). \qquad (128)
\end{aligned}$$

To obtain a single letter characterization we substitute A^n, B^n and C in (123) by V^n, Y^n and (L^n, F) respectively and so

$$\begin{aligned}
&H(V^n|L^nF) - H(Y^n|L^nF) \\
&= \sum_{t=1}^{n} [H(V_t|V_{t+1}, V_{t+2}, \ldots, V_n, L^n, Y^{t-1}, F) - H(Y_t|V_{t+1}, V_{t+2}, \ldots, V_n, L^n, Y^{t-1}, F)]. \qquad (129)
\end{aligned}$$

Moreover because the source is memoryless, we have

$$H(V^n|L^n) = \sum_{t=1}^{n} H(V_t|L_t). \qquad (130)$$

We now substitute (128), (129); (130) and $H(Y^n|L^n) = \sum_{t=1}^{n} H(Y_t|L^n, Y^{t-1})$ into (127) and continue it,

$$
\begin{aligned}
0 &\leq \sum_{t=1}^{n}[H(Y_t|L^n, Y^{t-1}) - H(V_t|L_t)] + \sum_{t=1}^{n}[H(V_t|V_{t+1}, V_{t+2}, \dots, V_n, L^n, Y^{t-1}, F) \\
&\quad - H(Y_t|V_{t+1}, V_{t+2}, \dots, V_n, L^n, Y^{t-1}, F)] + n\lambda \log\kappa + h(\lambda) \\
&= \sum_{t=1}^{n}[H(Y_t|L^n, Y^{t-1}) - H(Y_t|V_{t+1}, V_{t+2}, \dots, V_n, L^n, Y^{t-1}, F)] \\
&\quad - \sum_{t=1}^{n}[H(V_t|L_t) - H(V_t|V_{t+1}, V_{t+2}, \dots, V_n, L^n, Y^{t-1}, F)] + n\lambda \log\kappa + h(\lambda) \\
&= \sum_{t=1}^{n} I(Y_t; V_{t+1}, V_{t+2}, \dots, V_n, F|L^n, Y^{t-1}) \\
&\quad - \sum_{t=1}^{n} I(V_t; V_{t+1}, V_{t+2}, \dots, V_n, L_1, L_2 \dots, L_{t-1}, L_{t+1}, \dots, L_n, Y^{t-1}, F|L_t) + n\lambda \log\kappa + h(\lambda) \\
&\leq \sum_{t=1}^{n}[I(Y_t; V_{t+1}, V_{t+2}, \dots, V_n L_1, L_2 \dots, L_{t-1}, L_{t+1}, \dots, L_n, Y^{t-1}, F|L_t)] \\
&\quad - \sum_{t=1}^{n} I(V_t; V_{t+1}, V_{t+2}, \dots, V_n, L_1, L_2 \dots, L_{t-1}, L_{t+1}, \dots, L_n, Y^{t-1}, F|L_t)] \\
&\quad + n\lambda \log\kappa + h(\lambda).
\end{aligned}
\tag{131}
$$

Let J be the random variable taking values in $\{1, 2, \dots, n\}$ uniformly, and

$$
U_J = (V_{J+1}, V_{J+2}, \dots, V_n, L_1, L_2 \dots, L_{J-1}, L_{J+1}, \dots, L_n, Y^{J-1}, F). \tag{132}
$$

Then J and (V_J, L_J) are independent i. e., $I(J; V_J, L_J) = 0$. Thus (131) is rewritten and continued in the following a few lines.

$$
\begin{aligned}
0 &\leq nI(U_J; Y_J|L_J, J) - nI(U_J; V_J|L_J, J) + n\lambda \log\kappa + h(\lambda) \\
&= n[I(U_J; L_J, Y_J|J) - I(U_J; L_J|J)] - [I(U_J; V_J, L_J|J) - I(U_J; L_J|J) + n\lambda \log\kappa + h(\lambda) \\
&= nI(U_J; L_J, Y_J|J) - nI(U_J; V_J, L_J|J) + n\lambda \log\kappa + h(\lambda) \\
&\leq nI(U_J, J; L_J, Y_J) - n[I(U_J, J; V_J, L_J) - I(J; V_J, L_J)] + n\lambda \log\kappa + h(\lambda) \\
&= nI(U_J, J; L_J, Y_J) - nI(U_J, J; V_J, L_J) + n\lambda \log\kappa + h(\lambda).
\end{aligned}
\tag{133}
$$

Next we denote by

$$
(V'', L'', U'', X'', Y'') = (V_J, L_J, U_J J, X_J, Y_J) \tag{$*$}
$$

for the uniformly distributed J and U_J in (132). Then, obviously (V'', L'') has the same probability distribution with the generic (V, L) of the correlated source, the conditional probability distribution $P_{Y''|X''} = W$, and $(V''L''U'', X'', Y'')$ forms a Markov Chain. Namely, the joint distribution of $(V'', L'', U'', X'', Y'')$ is $P_{V''L''U''X''Y''} = P_{VL}P_{U''X''|V''L''}W$. With the defined random variables, (126) is rewritten as

$$\mathbf{E}\rho(V'', X'') = \mathbf{E}[\mathbf{E}\rho(V'', X'')|J] = \mathbf{E}[\mathbf{E}\rho(V_J, X_J)|J] = \frac{1}{n}\mathbf{E}\rho(V^n, X^n) \leq D_1. \tag{134}$$

Moreover, by substituting $(*)$ in (133) and then dividing both sides of resulting inequality by n, we obtain that

$$0 \leq I(U''; L'', Y'') - I(U''; V'', L'') + o(1), \tag{135}$$

(as $\lambda \to 0$).

Because the set $\{P_{V,L,U,X,Y} : (V, L, U, X, Y) \in \mathcal{Q}((V, L), W, D_1)\}$ is a closed set, by (134) and (135) is is sufficient for us to complete the proof to show that

$$\frac{1}{n}H(F) \leq I(U''; L'', Y'') + H(L''|U'') + o(1)$$

for $\lambda \to 0$. This is done by dividing both sides of (127) by n and continuing it by the following few lines.

$$\begin{aligned}
&\frac{1}{n}H(F)\\
\leq&\frac{1}{n}\sum_{t=1}^{n} I(F; Y_t|L^n, Y^{t-1}) + \frac{1}{n}\sum_{t=1}^{n} H(L_t) + \lambda\log\kappa + \frac{1}{n}h(\lambda),\\
\leq&\frac{1}{n}\sum_{t=1}^{n} I(V_{t+1}, V_{t+2}, \dots, V_n, F; Y_t|L^n, Y^{t-1}) + \frac{1}{n}\sum_{t=1}^{n} H(L_t) + \lambda\log\kappa + \frac{1}{n}h(\lambda),\\
\leq&\frac{1}{n}\sum_{t=1}^{n} I(V_{t+1}, V_{t+2}, \dots, V_n, L_1, L_2 \dots, L_{t-1}, L_{t+1}, \dots, L_n, Y^{t-1}, F; Y_t|L_t)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{n}\sum_{t=1}^{n} H(L_t) + \lambda\log\kappa + \frac{1}{n}h(\lambda)\\
=&I(U_J;Y_J|L_J,J) + H(L_J|J) + \lambda\log\kappa + \frac{1}{n}h(\lambda)\\
\leq&I(U_J,J;Y_J|L_J) + H(L_J|J) + \lambda\log\kappa + \frac{1}{n}h(\lambda)\\
=&I(U_J,J;Y_J|L_J) + H(L_J) + \lambda\log\kappa + \frac{1}{n}h(\lambda)\\
=&I(U_J,J;Y_J|L_J) + I(U_J;L_J) + H(L_J|U_J) + \lambda\log\kappa + \frac{1}{n}h(\lambda)\\
\leq&I(U_J,J;Y_J|L_J) + I(U_J,J;L_J) + H(L_J|U_J) + \lambda\log\kappa + \frac{1}{n}h(\lambda)\\
=&I(U_J,J;L_J,Y_J) + H(L_J|U_J) + \lambda\log\kappa + \frac{1}{n}h(\lambda)\\
=&I(U'';L'',Y'') + H(L''|U'') + \lambda\log\kappa + \frac{1}{n}h(\lambda),
\end{aligned}
\tag{136}
$$

where the second equality holds because U_J is independent of J. Finally the upper bound to the size of $\mathcal{U}$ follows from the Support Lemma in [13] (as well on page 310 in the book [22]).

Lemma 6.3 *(The converse part of Theorem 4.2)*
For a single channel W,

$$
C_{CRI}((V,L),W,R_K,D_1) \leq \max_{(V,L,U,X,Y)\in\mathcal{Q}^*((V,L),W,R_K,D_1)} [I(U;L,Y) + H(L|U)] + R_K. \tag{137}
$$

Proof: Let $\{(V^n,L^n)\}_{n=1}^{\infty}$ be a correlated source with generic (V,L), W be a noisy channel, and R_K and D_1 be the key rate and the distortion criterion in the Model II of common randomness respectively. Let F and G be functions satisfying (10) - (12), (17), and (14) - (16) in the Model II of common randomness (for output sequence of source of length n). Denote by X^n and K_n inputs of noisy channel W^n and the noiseless channel chosen by the sender according to the output of the correlated source and his/her private randomness. Then (126) holds and similarly to (127) by Fano inequality, we have that

$$
\begin{aligned}
&H(F)\\
\leq&I(F;G) + n\lambda\log\kappa + h(\lambda)\\
\leq&I(F;Y^n,L^n,K_n) + n\lambda\log\kappa + h(\lambda)\\
=&I(F;Y^n,L^n) + I(F;K_n|Y^n,L^n) + n\lambda\log\kappa + h(\lambda)\\
=&I(F;Y^n|L^n) + I(F;L^n) + I(F;K_n|Y^n,L^n) + n\lambda\log\kappa + h(\lambda)\\
\leq&I(F;Y^n|L^n) + H(L^n) + H(K_n|Y^n,L^n) + n\lambda\log\kappa + h(\lambda)\\
\leq&I(F;Y^n|L^n) + H(L^n) + H(K_n) + n\lambda\log\kappa + h(\lambda)\\
\leq&I(F;Y^n|L^n) + H(L^n) + nR_K + n\lambda\log\kappa + h(\lambda)\\
=&\sum_{t=1}^{n} I(F;Y_t|L^n,Y^{t-1}) + \sum_{t=1}^{n} H(L_t) + nR_K + n\lambda\log\kappa + h(\lambda),
\end{aligned}
\tag{138}
$$

where the second inequality holds by (17). Analogously to (128) we have

$$
\begin{aligned}
0 &\leq I(F;Y^n,L^n,K_n) - I(F;V^n,L^n) + n\lambda\log\kappa + h(\lambda) \\
&= I(F;Y^n,L^n) - I(F;V^n,L^n) + I(F;K_n|Y^n,L^n) + n\lambda\log\kappa + h(\lambda) \\
&\leq I(F;Y^n,L^n) - I(F;V^n,L^n) + H(K_n|Y^n,L^n) + n\lambda\log\kappa + h(\lambda) \\
&\leq I(F;Y^n,L^n) - I(F;V^n,L^n) + nR_K + n\lambda\log\kappa + h(\lambda).
\end{aligned}
\tag{139}
$$

Note that we only used the basic properties of Shannon information measures, Lemma 6.1, and the assumption that the correlated source is memoryless in the estimation of $I(F;Y^n,L^n) - I(F;V^n,L^n)$ in the part of (128) - (131) and all these are available here. So we have the same estimation here i. e.,

$$
\begin{aligned}
&I(F;Y^n,L^n) - I(F;V^nL^n) \\
\leq &\sum_{t=1}^{n} I(Y_t;V_{t+1},V_{t+2},\ldots,V_n,L_1,L_2\ldots,L_{t-1},L_{t+1},\ldots,L_n,Y^{t-1},F|L_t) \\
&-\sum_{t=1}^{n} I(V_t;V_{t+1},V_{t+2},\ldots,V_n,L_1,L_2\ldots,L_{t-1},L_{t+1},\ldots,L_n,Y^{t-1},F|L_t) \\
&+ n\lambda\log\kappa + h(\lambda).
\end{aligned}
\tag{140}
$$

Let U_J and J be defined as in (132). Then (140) is rewritten as

$$
I(F;Y^n,L^n) - I(F;V^nL^n) \leq nI(U_J,J;L_J,Y_J) - nI(U_J,J;V_J,L_J) + n\lambda\log\kappa + h(\lambda).
\tag{141}
$$

Let (V'',L'',U'',X'',Y'') is defined as in the previous lemma. Then (134) and $P_{V''L''U''X''Y''} = P_{VL}P_{U''X''|V''L''}W$ are certainly fulfilled. But now (139) - (141) lead us to

$$
0 \leq I(U'';L'',Y'') - I(U'';V'',L'') + R_K + o(1).
\tag{142}
$$

In the same way as (136) we can show

$$
\begin{aligned}
&\sum_{t=1}^{n} I(F;Y_t|L^n,Y^{t-1}) + \sum_{t=1}^{n} H(L_t) + nR_K + n\lambda\log\kappa + h(\lambda) \\
\leq &nI(U'';L'',Y'') + nH(U''|L'') + n\lambda\log\kappa + h(\lambda)
\end{aligned}
\tag{143}
$$

which with (138) yields

$$
\frac{1}{n}H(F) \leq I(U'';L''Y'') + H(U''|L'') + R_K + \lambda\log\kappa + \frac{1}{n}h(\lambda).
$$

Again $|\mathcal{U}|$ is bounded by the Support Lemma. Thus our proof is finished.

Finally it immediately follows from Lemmas 6.2 and 6.3 that

Corollary 6.4 *For compound channel $\mathcal{W}$,* *1) (The converse part of Theorem 4.3:)*

$$C_{CRI}((V,L),\mathcal{W},D_1) \leq \inf_{W\in\mathcal{W}} \max_{(V,L,U,X,Y)\in\mathcal{Q}((V,L),W,D_1)} [I(U;L,Y)+H(L|U)] \tag{144}$$

and

2) (The converse part of Theorem 4.4:)

$$C_{CRII}((V,L),\mathcal{W},R_K,D_1) \leq \inf_{W\in\mathcal{W}} \max_{(V,L,U,X,Y)\in\mathcal{Q}^*((V,L),W,R_K,D_1)} [I(U;L,Y)+H(L|U)]+R_K. \tag{145}$$

7 Constructing Watermarking Identification Codes from Common Randomness

R. Ahlswede and G. Dueck found in [12] that a identification code with the same rate can be always obtained from the common randomness between a sender and and receiver under the condition
(*) *The sender can send a massage with arbitrarily small but positive rate (in the exponential sense).*
Thus under the condition (*) the capacity of identification is not smaller than that of common randomness. Note that the sets $\mathcal{Q}((V,L),W,D_1)$, $\mathcal{Q}^{**}(V,W,R_k,D_1)$, $\mathcal{Q}_1((V,L),\mathcal{W},D_1)$, and $\mathcal{Q}_1^{**}(V,\mathcal{W},R_k,D_1)$ are not empty implies the condition (*) in the Theorems 4.5, 4.6, 4.7, and 4.8 respectively. Consequently Theorems 4.5, 4.6, 4.7, and 4.8 follows from Theorems 4.1, 4.2, 4.3, and 4.4 respectively.

8 A Converse Theorem of a Watermaking Coding Theorem Due to Steinberg-Merhav

In order to construct identification codes in [32], Y. Steinberg and N. Merhav introduced the following code to build common randomness between sender and receiver and obtained an inner bound of the capacity region. This inner bound is sufficient for their goal. We shall show that it is as well tight. This would support their conjecture that the lower bound in their Theorem 4 ([32]) is tight although it does not imply it.

Let $\{V^n\}_{n=1}^{\infty}$ be a memoryless source with alphabet $\mathcal{V}$ and generic V and W be a noisy channel with input and output alphabets $\mathcal{X}$ and $\mathcal{Y}$ respectively. A pair of functions (f,g) is called an (n,M,J,δ,λ,D) watermarking transmission code with a common experiment, distortion measure ρ, distortion level D and covertext P_V if the followings are true.

— f is a function from $\mathcal{V}^n\times\{1,2,\dots,M\}$ to $\{1,2,\dots,J\}\times\mathcal{X}^n$.

— g is a function from $\mathcal{Y}^n$ to $\{1,2,\dots,J\}\times\{1,2,\dots,M\}$.

$$\frac{1}{M}\sum_{m=1}^{M}\sum_{v^n\in\mathcal{V}^n} P_V^n(v^n)W^n(\{y : g(y^n) = (f_J(v^n,m),m)\}|f_X(v^n,m)) \geq 1-\lambda, \tag{146}$$

where f_X and f_J are projections of f to $\mathcal{X}^n$ and $\{1,2,\ldots,J\}$ respectively.

$$\frac{1}{M}\sum_{m=1}^{M}\sum_{v^n\in\mathcal{V}^n} P_V^n(v^n)\rho(v^n, f_X(v^n,m)) \leq D. \tag{147}$$

For $m = 1,2,\ldots,M$, there exists a subset $\mathcal{B}^{(m)} \subset \{1,2,\ldots,J\}$ of cardinality $|\mathcal{B}^{(m)}| \geq J2^{-n\delta}$ such that

$$J^{-1}2^{-n\delta} \leq P_V^n\{f_J(V^n,m) = j\} \leq J^{-1}2^{n\delta} \tag{148}$$

for all j and

$$\sum_{j\in\mathcal{B}^{(m)}} P_V^n\{f_J(V^n,m) = j\} \geq 1-\lambda. \tag{149}$$

g serves as a decoding function here. (148) and (149) play the same role as nearly uniform condition in construction of identification codes from common randomness. In fact one can find the nearly uniform condition (16) is stronger but for the purpose to construct identification codes the conditions (148) and (149) are strong enough.

A pair (R_1, R_2) is called achievable with distortion D if for all positive reals δ, λ, and ϵ there is an $(n, M, J, \delta, \lambda, D)$ code defined as above such that

$$\frac{1}{n}\log M > R_1 - \epsilon \tag{150}$$

and

$$\frac{1}{n}\log J > R_2 - \epsilon. \tag{151}$$

The set of achievable pair of rates is called capacity region and denoted by $\mathcal{R}$. Denote by $\mathcal{R}^{(*)}$ the subset of pair of real numbers such that there exist random variables (V,U,X,Y) taking values in $\mathcal{V}\times\mathcal{U}\times\mathcal{X}\times\mathcal{Y}$ such that $|\mathcal{U}| \leq |\mathcal{Y}| + |\mathcal{X}|$, for all $v\in\mathcal{V}, u\in\mathcal{U}, x\in\mathcal{X}$ and $y\in\mathcal{Y}$,

$$P_{VUXY}(v,u,x,y) = P_V(v)P_{UX|V}(u,x|v)W(y|x),$$
$$\mathbf{E}\rho(V,X) \leq D,$$

$$0 \leq R_1 \leq I(U;Y) - I(U;V), \tag{152}$$

and

$$0 \leq R_2 \leq I(U;V). \tag{153}$$

It was shown in [32]

Theorem 8.1 *(Steinberg-Merhav)*

$$\mathcal{R}^* \subset \mathcal{R}. \tag{154}$$

We now show the opposite contained relation holds i. e.,

Theorem 8.2

$$\mathcal{R} \subset \mathcal{R}^*. \tag{155}$$

Proof: Let (f, g) be a pair of functions satisfying (146) - (151) for sufficiently large n (, which is specified later,) and Z_n be a random variable with uniform distribution over $\{1, 2, \dots, M\}$. Further let $f(V^n, Z_n) = (B_n, X^n)$, where B_n and X^n have ranges $\{1, 2, \dots, J\}$ and $\mathcal{X}^n$ respectively and Y^n be the random output of the channel W^n when X^n is input.

Then (148) and (149) are rewritten as

$$J^{-1}2^{-n\delta} \leq P_{B_n|Z_n}(j|m) \leq J^{-1}2^{n\delta} \tag{156}$$

for all $j \in \mathcal{B}^{(m)}$ and

$$P_{B_n|Z_n}(B_n \in \mathcal{B}^{(m)}|m) \geq 1 - \lambda \tag{157}$$

respectively. So,

$$\begin{aligned}
&H(B_n|Z_n)\\
=&\sum_{m=1}^{M} P_{Z_n}(m) H(B_n|Z_n = m)\\
\geq& -\sum_{m=1}^{M} P_{Z_n}(m) \sum_{j \in \mathcal{B}^{(m)}} P_{B_n|Z_n}(j|m) \log P_{B_n|Z_n}(j|m)\\
\geq& -\sum_{m=1}^{M} P_{Z_n}(m) \sum_{j \in \mathcal{B}^{(m)}} P_{B_n|Z_n}(j|m) \log J^{-1}2^{n\delta}\\
=&(\log J - n\delta) \sum_{m=1}^{M} P_{Z_n}(m) P_{B_n|Z_n}(B_n \in \mathcal{B}^{(m)}|m)\\
\geq&(\log J - n\delta)(1 - \lambda)
\end{aligned} \tag{158}$$

where the second inequality holds by (156) and the last inequality follows from (157). Or equivalently

$$\frac{1}{n} \log J \leq \frac{\frac{1}{n} H(B_n|Z_n)}{1 - \lambda} + \delta. \tag{159}$$

Since $H(B_n) \le \log J$, (159) implies that for a function θ such that $\theta(\delta, \lambda) \to 0$ as $\delta, \lambda \to 0$,

$$\frac{1}{n}\log J - \theta(\delta,\lambda) < \frac{1}{n}H(B_n|Z_n) \le \frac{1}{n}H(B_n) \le \frac{1}{n}\log J. \tag{160}$$

which says that B_n and Z_n are "nearly independent". Moreover because Z_n is independent of V^n, by Fano's inequality,

$$\begin{aligned} R_1 - \varepsilon < \frac{1}{n}\log M &= \frac{1}{n}H(Z_n) \\ &= \frac{1}{n}H(Z_n|V^n) \\ &\le \frac{1}{n}H(B_n, Z_n|V^n) \\ &\le \frac{1}{n}[H(B_n, Z_n|V^n) - H(B_n, Z_n|Y^n)] + \lambda \log JM + \frac{1}{n}h(\lambda) \\ &= \frac{1}{n}[I(B_n, Z_n; Y^n) - I(B_n, Z_n; V^n)] + \lambda\frac{1}{n}\log JM + \frac{1}{n}h(\lambda) \end{aligned} \tag{161}$$

where the second inequality follows from Fano's inequality. Since B_n is a function of V^n and Z_n, we have also

$$H(B_n, Z_n|V^n) \le H(V^n, Z_n|V^n) = H(Z_n), \tag{162}$$

which and (160) are followed by

$$\begin{aligned} R_2 - \varepsilon < \frac{1}{n}\log J &< \frac{1}{n}H(B_n|Z_n) + \theta(\delta,\lambda) \\ &= \frac{1}{n}[H(B_n, Z_n) - H(Z_n)] + \theta(\delta,\lambda) \\ &\le \frac{1}{n}[H(B_n, Z_n) - H(B_n, Z_n|V^n)] + \theta(\delta,\lambda) \\ &= \frac{1}{n}I(B_n, Z_n; V^n) + \theta(\delta,\lambda). \end{aligned} \tag{163}$$

So far we have had a non-single-letter characterization of the capacity region (161) and (163). In the rest part of the proof we shall reduce it to a single letter one.

First we substitute A^n, B^n, and C in (123) by V^n, Y^n, and (B_n, Z_n) respectively and obtain that

$$\begin{aligned} &H(V^n|B_n, Z_n) - H(Y^n|B_n, Z_n) \\ =&\sum_{t=1}^{n}[H(V_t|V_{t+1}, V_{t+2}, \ldots, V_n, Y^{t-1}, B_n, Z_n) - H(Y_t|V_{t+1}, V_{t+2}, \ldots, V_n, Y^{t-1}, B_n, Z_n)]. \end{aligned} \tag{164}$$

Next we note that $H(V^n) = \sum_{t=1}^n H(V_t)$ because the source is memoryless and $H(Y^n) = \sum_{t=1} H(Y_t|Y^{t-1})$. Therefore, we have

$$\begin{aligned}
&I(B_n, Z_n; Y^n) - I(B_n, Z_n; V^n)\\
=&H(Y^n) - H(V^n) + [H(V^n|B_n, Z_n) - H(Y^n|B_n, Z_n)]\\
=&\sum_{t=1}^n H(Y_t|Y^{t-1}) - \sum_{t=1}^n H(V_t) + \sum_{t=1}^n [H(V_t|V_{t+1}, V_{t+2}, \dots, V_n, Y^{t-1}, B_n, Z_n)\\
&- H(Y_t|V_{t+1}, V_{t+2}, \dots, V_n, Y^{t-1}, B_n, Z_n)]\\
=&\sum_{t=1}^n [H(Y_t|Y^{t-1}) - H(Y_t|V_{t+1}, V_{t+2}, \dots, V_n, Y^{t-1}, B_n, Z_n)]\\
&- \sum_{t=1}^n [H(V_t) - H(V_t|V_{t+1}, V_{t+2}, \dots, V_n, Y^{t-1}, B_n, Z_n)]\\
=&\sum_{t=1}^n I(V_{t+1}, V_{t+2}, \dots, V_n, B_n, Z_n; Y_t|Y^{t-1})\\
&- \sum_{t=1}^n I(V_{t+1}, V_{t+2}, \dots, V_n, Y^{t-1}, B_n, Z_n; V_t)\\
\leq&\sum_{t=1}^n I(V_{t+1}, V_{t+2}, \dots, V_n, Y^{t-1}, B_n, Z_n; Y_t)\\
&- \sum_{t=1}^n I(V_{t+1}, V_{t+2}, \dots, V_n, Y^{t-1}, B_n, Z_n; V_t). \qquad (165)
\end{aligned}$$

Moreover,

$$\begin{aligned}
&I(B_n, Z_n; V^n)\\
=&\sum_{t=1}^n I(B_n, Z_n; V_t|V_{t+1}, V_{t+2}, \dots, V_n)\\
\leq&\sum_{t=1}^n I(V_{t+1}, V_{t+2}, \dots, V_n, B_n, Z_n; V_t)\\
\leq&\sum_{t=1}^n I(V_{t+1}, V_{t+2}, \dots, V_n, Y^{t-1}, B_n, Z_n; V_t). \qquad (166)
\end{aligned}$$

So we may let I be a random variable taking values in $\{1, 2, \dots, n\}$ uniformly and $U' = (V_{I+1}, V_{I+2}, \dots, V_n, Y^{I-1}, B_n, Z_n)$ and conclude by (163), (164), (165), (166)

$$\begin{aligned}
R_1 - \varepsilon &\leq I(U'; Y_I|I) - I(U'; V_I|I) + \lambda \log JM + \frac{1}{n} h(\lambda)\\
&\leq I(U', I; Y_J) - I(U', I; V_I) + I(I; V_I) + \lambda \log JM + \frac{1}{n} h(\lambda), \qquad (167)
\end{aligned}$$

and

$$R_2 - \varepsilon \leq I(U'; V_I | I) \leq I(U', I; V_I) + \theta(\delta, \lambda). \tag{168}$$

Let $U = (U', I), V' = V_I, X = X_I$ and $Y = Y_I$. Then $P_{V'} = P_V$, $(V'U, X, Y)$ forms a Markov chain and (168) can be re-written as

$$R_2 \leq I(U; V') + \theta(\delta, \lambda),$$

and

$$EP(v', x') < D.$$

Further that $I(I; V_I) = 0$ (as the source is stationary) and (167) are followed by

$$R_1 \leq I(U; Y) - I(U; V') + \lambda \log JM + \frac{1}{n} h(\lambda) + \varepsilon.$$

Finally $|\mathcal{U}|$ is bounded by the support Lemma in the standard way.

References

[1] R. Ahlswede, Channels with arbitrarily varying channel probability functions in the presence of noiseless feedback, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 25, 239-252, 1973.

[2] R. Ahlswede, Elimination of correlation in random codes for arbitrarily varying channels, *Z. Wahrsch. Verw. Gebiete*, 44, no. 2, 159-175, 1978.

[3] R. Ahlswede, Coloring hypergraphs: a new approach to multi-user source coding. I. *J. Combin. Inform. System Sci.* 4, no. 1, 76-115, 1979.

[4] R. Ahlswede, Coloring hypergraphs: a new approach to multi-user source coding II, *J. Combin. Inform. System Sci.* 5, no. 3, 220-268, 1980.

[5] R. Ahlswede, General theory of information trannsfer, Preprint 97-118, SFB 343 Diskrete Strukturen in der Mathematik, 1997, this volume pages.

[6] R. Ahlswede and V. B. Balakirsky, Identification by means of a random process, (Russian), *Problemy Peredachi Informatsii*, 32, no. 1, 144-160, 1996, translation in *Problems Inform. Transmission*, 32, no. 1, 123-138, 1996.

[7] R. Ahlswede and N. Cai, Information and control: matching channels. *IEEE Trans. Inform. Theory*, 44, no. 2, 542–563, 1998.

[8] R. Ahlswede and N. Cai, The AVC with noiseless feedback and maximal error probability: a capacity formula with a trichotomy, *Numbers, Information, and Complexity* (Festschrift for Rudolf Ahlswede), eds: I. Althöfer, N. Cai, G. Dueck, L. Khachatrian, M. Pinsker, A. Sarkozy, I. Wegener, and Z. Zhang, Kluwer, 151-176, 2000.

[9] R. Ahlswede and I. Csiszár, Common randomness in information theory and cryptography I, Secret sharing, *IEEE Trans. Inform. Theory*, 39, no. 4, 1121-1132, 1993.

[10] R. Ahlswede and I. Csiszár, Common randomness in information theory and cryptograpy II, CR capacity, *IEEE Trans. Inform. Theory*, 44, no. 1, 225-240, 1998.

[11] R. Ahlswede and G. Dueck, Identification via channels, *IEEE Trans. Inform. Theory*, 35, no. 1, 15-29, 1989.

[12] R. Ahlswede and G. Dueck, Identification in the presence of feedback - a discovery of new capacity formulas. *IEEE Trans. Inform. Theory*, 35, no. 1, 30-36, 1989.

[13] R. Ahlswede and J. Körner, Source coding with side information and a converse for degraded broadcast channels. *IEEE Trans. Information Theory*, IT-21, no. 6, 629-637, 1975.

[14] R. Ahlswede and C. Kleinewächter, Pathological examples with different common randomness and (second order) identification capacities, this volume pages.

[15] R. Ahlswede and Z. Zhang, New directions in the theory of identification via channels, *IEEE Trans. Inform. Theory*, 41, no. 4, 1040-1050, 1995.

[16] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, Capactiy of the watermark channel: how many bits can be hidden within a digital image?, Proc. SPIE, SPIE, vol. 3657, 437-448, San Jose, CA, Jan, 1999.

[17] M. Burnashav, On identification capacity of infinite alphabets or continuous time channel, *IEEE Trans. Inf. Theory*, IT–46, 2407–2414, 2000.

[18] N. Cai and L. Y. Lam, On identification secret sharing schemes, *Information and Computation*, to appear.

[19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 1991.

[20] I. J. Cox, M. L. Miller, and A. Mckellips, Watermarking as communications with side information, *Proc. of IEEE*, V. 87, no 7, 1127-1141, 1999.

[21] I. Csiszár, Almost independence and secrecy capacity, *Probl. Inform. Trans.*, vol. 32, pp. 40-47, 1996.

[22] I. Csiszár and J. Körner, *Information Theory: Coding theorems for Discrete memoryless Systems*, Academic, 1981.

[23] I. Csiszár and P. Narayan, Common randomness and secret key generation with a helper, *IEEE Trans. Inform. Theory*, V. 46, no 2, 344-366, 2000.

[24] S.I. Gelfand and M.S. Pinsker, Coding for channels with random parameters, Problems of Control and Inform. Theory, V. 9, 19–31, 1980.

[25] T. S. Han and S. Verdú, New results in the theory of identification via channels, *IEEE Trans. Inform. Theory*, V. 38, 14-25, 1992.

[26] U. M. Maurer, Secret key argreenent by public discussion from common information, *IEEE Trans. Inform. Theory*, V. 39, 733-742, 1993.

[27] N. Merhav, On random coding error exponents of watermarking codes,,*IEEE Trans. Inform. Theory*, V. 46

[28] P. Moulin and J. A. O'Sulliovan, Information-theoretic analysis of information hiding, preprint, 1999.

[29] J. A. O'Sullivan, P. Moulin, and J. M. Ettinger, Information theoretic analysis of steganography, *Proc. ISIT '98*, 297, 1998.

[30] S. D. Servetto, C. I. Podilchuk, and K. Ramchandran, Capacity issues in digital image watermarking, *Proc. ICIP '98*, 1998.

[31] Y. Steinberg, New converses in the theory of identification via channels, *IEEE Trans. Inform. Theory*, V. 44, 984-998, 1998.

[32] Y. Steinberg and N. Merhav, Indentification in the presence of side infromation with application to watermarking, IEEE Trans. Inform. Theory, V. 47, 1410–1422, 2001.

[33] S. Venkatesan and V. Anantharam, The common randomness capacity of a pair of independent dicrete memoryless channels, *IEEE Trans. Inform. Theory*, 215-224, 1998.

[34] S. Venkatesan and V. Anantharam, The common randomness capacity of network of discret mamoryless channels, *IEEE Trans. Inform. Theory*, 367-387, 2000.

[35] R.W. Yeung, A First Course in Information Theory, Kluwer Academic, 2002.