General Nonasymptotic and Asymptotic Formulas in Channel Resolvability and Identification Capacity and Their Application to the Wiretap Channel

Masahito Hayashi

***Abstract*—Several nonasymptotic formulas are established in channel resolvability and identification capacity, and they are applied to the wiretap channel. By using these formulas, the ϵ capacities of the above three problems are considered in the most general setting, where no structural assumptions such as the stationary memoryless property are made on a channel. As a result, we solve an open problem proposed by Han and Verdú. Moreover, we obtain lower bounds of the exponents of error probability and the wiretapper's information in the wiretap channel.**

***Index Terms*—Channel resolvability, identification code, information spectrum, nonasymptotic setting, wiretap channel.**

I. Introduction

In 1989, Ahlswede and Dueck [1] proposed the identification code as a new framework for communication system using noisy channels. However, the upper bound of the rate of the reliable identification codes was not solved in their paper. In 1993, for analysis of the converse part of this problem, Han and Verdú [2] proposed the channel resolvability problem, in which we approximate the output distribution to a desired output distribution by using a uniform input distribution with smaller support. In particular, the capacity of this problem is defined as the rate of the maximal number of the size of support for every desired output distribution. In order to discuss the channel resolvability problem, they introduced the concepts of "general sequence of channels" and the "information spectrum method." They gave the relation between identification code and channel resolvability, and succeeded in proving the converse part of the capacity of identification code for the discrete memoryless channel. In this method, it is essential that the performances of several problems be characterized by using the probability distribution of the random variable with a form of "likelihood" function in this method. This insight is very useful for obtaining the overview of information theory [3]. In particular, it gives a useful insight into quantum information theory [12], [11], [10]. Therefore, Han and Verdú's paper [2] is undoubtably the landmark of information spectrum.

However, while Han and Verdú's paper gives the capacity of channel resolvability for general sequence of channels [2], their proof of the converse part contains mistakes that is recognized in Section 6.3. of Han [3]. They proved the achievability of channel resolvability with the asymptotic zero error setting for a general sequence of channels. Concerning the converse part, their proof is valid for the asymptotic ϵ error setting when the general sequence of channels has a strong converse property. However, their proof is not valid in the general channel even in the asymptotic zero error setting.

In this paper, we give several useful nonasymptotic formulas for identification code and channel resolvability, which are divided into two parts. One is the direct part of the identification code. The existence of a good identification code is proved in Theorem 1. This construction is much improved from Ahlswede and Dueck's construction. The other is the direct part of channel resolvability. The existence of a good approximation regarding the output statistics is proved in the two criteria, variational distance and Kullback–Leibler (K-L) divergence as in Theorem 2. In this discussion, we derived upper bounds of the average of the variational distance and K-L divergence between the output distribution of a given distribution p and the output distribution of the input uniform distribution on M elements of the input signal space, when the M elements are randomly chosen with the distribution p (Lemma 2). Combining Han and Verdú's relation between identification code and channel resolvability, we derived the capacity of the channel resolvability for general sequence of channels with the asymptotic zero error setting, which was conjectured by Han and Verdú [2] ((26) and (27) of Theorem 4). This discussion is valid even though the strong converse property does not hold.

As another application, we give an upper bound of the capacity of the channel resolvability for a general sequence of channels with the asymptotic ϵ error. As a byproduct, we show that there exists a sequence of codes whose second error probability goes to 0 in any general sequence of channels, and only the first error probability is asymptotically related to the probability distribution of the random variable with the form of "likelihood" ((24) and (25) of Theorem 4). We also derived several lower bounds of exponent of channel resolvability in the stationary memoryless setting with respective error criteria (Theorem 6).

Moreover, we apply our nonasymptotic formulas for channel resolvability to wiretap channel, in which there are two receivers, i.e., the eavesdropper and the normal receiver. Wyner [4] introduced this wiretap channel, and proved that its capacity is greater then the difference between the normal receiver's information and the eavesdropper's information. Csiszár and Narayan [16] showed that the capacity does not depend on

Manuscript received July 11, 2004; revised December 20, 2005. The material in this paper was presented in part at the 2004 IEEE International Symposium on Information Theory and Its Applications, Parma, Italy, October 2004.

The author is with the Quantum Computation and Information Project, ERATO, JST, 5-28-3, Hongo, Bunkyo-ku, Tokyo 113-0033, Japan (e-mail: masahito@qci.jst.go.jp).

Communicated by K. Kobayashi, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2006.871040

the following two conditions for eavesdropper's information: i) The eavesdropper's information must be less than $n\epsilon$ for given $\epsilon > 0$, where n is the number of transmissions. ii) The eavesdropper's information must go to 0 exponentially. However, there are no results giving an explicit lower bound of the optimal exponents of wiretapper's information.

Indeed, this problem is closely related to the channel resolvability as follows. In Wyner's proof [4], in the asymptotic independent and identically distributed (i.i.d.) setting with a large enough number M, he essentially showed that when M elements of the input signal space are randomly chosen with a given distribution p, the output distribution of the distribution p can be approximated with a high probability by the output distribution of the input uniform distribution on the above M elements of the input signal space. This idea is also applied in Devetak [5] and Winter, Nascimento, and Imai [6]. Using the same idea in the nonasymptotic setting, we can apply our formulas of channel resolvability to wiretap channel, and derive a good nonasymptotic formula for wiretap channel (Theorem 3). As a result, we obtain the capacity of general sequence of wiretap channel (Theorem 5) and lower bounds of the exponents of error probability and the wiretapper's information in the stationary memoryless setting (Theorem 7). We can expect that these results will be applied to evaluations of the security of channels.

Finally, we should remark that our nonasymptotic resolvability formula regarding variational distance can be regarded as essentially the same results as Oohama [9]'s formula, where he treated the partial resolvability. Furthermore, he also derived a lower bound of exponent of channel resolvability by type method [8].

II. Identification Code in Nonasymptotic Setting

Let $W : x \mapsto W_x$ be an arbitrary channel with the input alphabet $\mathcal{X}$ and the output alphabet $\mathcal{Y}$. The identification channel code for the channel W is defined in the following way. First, let $\mathcal{N} = \{1, \ldots, N\}$ be a set of messages to be transmitted, and denote by $\mathcal{P}(\mathcal{X})$ the set of all probability distribution over $\mathcal{X}$. A transmitter prepares N probability distributions $Q_1, \ldots, Q_N \in \mathcal{P}(\mathcal{X})$. If the transmitter wants to send a message $i \in \mathcal{N}$, an encoder generates an input sequence $x_i \in \mathcal{X}$ randomly subject to the probability distribution Q_i. In this case, the output signal y obeys the distribution W_{Q_i}, where the output distribution W_p of a given input distribution p is defined as

$$W_p(y) \stackrel{\mathrm{def}}{=} \sum_x p(x) W_x(y).$$

On the other hand, at the decoder side an N-tuple of decoders is prepared. For every $i = 1, \ldots, N$, the ith decoder judges that $i \in \mathcal{N}$ is transmitted if a channel output y belongs to $\mathcal{D}_i$, where $\{\mathcal{D}_1, \ldots, \mathcal{D}_N\}$ are N subsets of $\mathcal{Y}$ in advance. The ith decoder judges that a message different from $i \in \mathcal{N}$ if $y \notin \mathcal{D}_i$. Here, $\mathcal{D}_i$ is called the *decoding region*, of the message i. It is not required that $\mathcal{D}_1, \ldots, \mathcal{D}_N$ be disjoint. In the identification coding problem, the ith decoder is only interested in transmission of the corresponding message i. Thus, we call the tuple of

$$\Phi \stackrel{\mathrm{def}}{=} (N, \{Q_1, \ldots Q_N\}, \{\mathcal{D}_1, \ldots, \mathcal{D}_N\})$$

an identification code of channel W. The performance of this code can be characterized by the following three quantities. One is the size N of the message sent and is denoted by $|\Phi|$, and the others are the maximum values of the two-type error probabilities given as

$$\mu(\Phi) \stackrel{\mathrm{def}}{=} \max_i W_{Q_i}(\mathcal{D}_i^c), \qquad \lambda(\Phi) \stackrel{\mathrm{def}}{=} \max_{i \neq j} W_{Q_j}(\mathcal{D}_i)$$

where $\mathcal{D}_i^c$ is the complement set of $\mathcal{D}_i$. Concerning this problem, as discussed in the following theorem, the "likelihood" function

$$\frac{W_x}{W_p}(y) \stackrel{\mathrm{def}}{=} \frac{W_x(y)}{W_p(y)}$$

suitablely characterizes the performance of good identification codes.

Theorem 1: Assume that real numbers $\alpha, \alpha', \beta, \beta', \tau, \kappa > 0$ satisfy

$$\kappa \log \left(\frac{1}{\tau} - 1 \right) > \log 2 + 1, \quad 1/3 > \tau > 0, \quad 1 > \kappa > 0 \tag{1}$$

$$1 > \frac{1}{\alpha} + \frac{1}{\alpha'}, \quad \gamma \stackrel{\mathrm{def}}{=} 1 - \frac{1}{\beta} - \frac{1}{\beta'} > 0. \tag{2}$$

Then, for any integer $M > 0$, any real number $C > 0$, any channel W, and any probability distribution $p \in \mathcal{P}(\mathcal{X})$, there exists an identification code Φ such that

$$\mu(\Phi) \le \alpha\beta \mathrm{E}_{p,x} W_x \left\{ y \left| \frac{W_x}{W_p}(y) \le C \right. \right\}$$
$$\lambda(\Phi) \le \kappa + \alpha'\beta' \frac{1}{C} \left\lceil \frac{M}{\gamma} \right\rceil, \quad |\Phi| = \left\lfloor \frac{e^{\tau M}}{Me} \right\rfloor$$

if

$$\beta \mathrm{E}_{p,x} W_x \left\{ y \left| \frac{W_x}{W_p}(y) \le C \right. \right\} + \alpha'\beta' \frac{1}{C} \left\lceil \frac{M}{\gamma} \right\rceil < 1 \tag{3}$$

where $\mathrm{E}_{p,x}$ denotes the expectation concerning random variable x obeying the probability distribution p.

In the following, we omit x or p in the notation $\mathrm{E}_{p,x}$, and abbreviate the set $\{y \,|\, \frac{W_x}{W_p}(y) \le C\}$ as $\{\frac{W_x}{W_p}(y) \le C\}$. We also denote the probability that the random variable X belongs to the set $\mathcal{D}$, by $\mathrm{P}_X(\mathcal{D})$ or $\mathrm{P}_X\mathcal{D}$. If we do not need to take note of the random variable X, we simplify it to $\mathrm{P}(\mathcal{D})$ or $\mathrm{P}\mathcal{D}$. This theorem is proven by using the following lemma.

Lemma 1: (Ahlswede and Dueck [1]) Let $\mathcal{M}$ be an arbitrary finite set of the size $M = |\mathcal{M}|$. Choose constants τ and κ satisfying the condition (1). Then there exist $N (\stackrel{\mathrm{def}}{=} \lfloor \frac{e^{\tau M}}{Me} \rfloor)$ subsets $A_1, \ldots, A_N \subset \mathcal{M}$ satisfying

$$|A_i| = \lfloor \tau M \rfloor, \quad |A_i \cap A_j| < \kappa \lfloor \tau M \rfloor (i \neq j). \tag{4}$$

Proof of Theorem 1: In this proof, the subset

$$\mathcal{U}_x \stackrel{\mathrm{def}}{=} \left\{ \frac{W_x}{W_p}(y) > C \right\}$$

plays an important role. First, we assume the existence of M distinct elements $x_1, \ldots, x_M$ of $\mathcal{X}$ satisfying

$$W_{x_i}(\mathcal{U}_{x_i}^c) \le \alpha\beta \mathrm{E}_p W_x \left\{ \frac{W_x}{W_p}(y) \le C \right\} \tag{5}$$

$$W_{x_i}\left(\bigcup_{j \ne i} \mathcal{U}_{x_j} \right) \le \alpha'\beta' \frac{1}{C} \left\lceil \frac{M}{\gamma} \right\rceil. \tag{6}$$

From Lemma 1, we can choose $N \stackrel{\mathrm{def}}{=} \lfloor \frac{e^{\tau M}}{Me} \rfloor$ subsets $A_i, \ldots, A_N$ of the set $\{x_1, \ldots, x_M\}$ satisfying (4). Let Q_i be the uniform distribution on the subset A_i whose cardinality is $\lfloor \tau M \rfloor$, that is, Q_i is defined as

$$Q_i(x) \stackrel{\mathrm{def}}{=} \frac{1}{|A_i|} \sum_{x' \in A_i} 1_{x'}(x) \tag{7}$$

where $1_{x'} = 1_{x'}(x)$ is an indicator function taking value 1 if $x = x'$ and 0 otherwise. Defining the subset $\mathcal{D}_i$ as $\mathcal{D}_i \stackrel{\mathrm{def}}{=} \cup_{x \in A_i} \mathcal{U}_x$, we evaluate

$$\begin{aligned}
W_{Q_i}(\mathcal{D}_i) &= \sum_{x \in A_i} \frac{1}{|A_i|} W_x(\mathcal{D}_i) \ge \sum_{x \in A_i} \frac{1}{|A_i|} W_x(\mathcal{U}_x) \\
&\ge 1 - \alpha\beta \mathrm{E}_p W_x \left\{ \frac{W_x}{W_p}(y) \le C \right\} \\
W_{Q_i}(\mathcal{D}_j) &= \sum_{x \in A_i} \frac{1}{|A_i|} W_x(\mathcal{D}_j) \\
&= \sum_{x \in A_i \cap A_j} \frac{1}{|A_i|} W_x(\mathcal{D}_j) + \sum_{x \in A_i \cap A_j^c} \frac{1}{|A_i|} W_x(\mathcal{D}_j) \\
&\le \frac{|A_i \cap A_j|}{|A_i|} + \sum_{x \in A_i \cap A_j^c} \frac{1}{|A_i|} W_x\left(\bigcup_{x' \ne x} \mathcal{U}_{x'} \right) \\
&\le \kappa + \alpha'\beta' \frac{1}{C} \left\lceil \frac{M}{\gamma} \right\rceil.
\end{aligned}$$

Therefore, we obtain the desired argument.

Next, we prove the existence of M elements and M subsets satisfying (5) and (6) by a random coding method. Let M' be $\lceil \frac{M}{\gamma} \rceil$, and $X = (X_1, \ldots, X_{M'})$ be M independent and identical random variables subject to the probability distribution $p \in \mathcal{P}(\mathcal{X})$, then we have

$$W_p(\mathcal{U}_x) \le \frac{1}{C} W_x(\mathcal{U}_x) \le \frac{1}{C}.$$

Using this inequality, we obtain

$$\begin{aligned}
\mathrm{E}_X \frac{1}{M'} \sum_{i=1}^{M'} W_{X_i}\left(\bigcup_{j \ne i} \mathcal{U}_{X_j} \right) &\le \mathrm{E}_X \sum_{j=1}^{M'} \frac{1}{M'} \sum_{i \ne j} W_{X_i}(\mathcal{U}_{X_j}) \\
&= \sum_{j=1}^{M'} \mathrm{E}_{X_j} \frac{M'-1}{M'} W_p(\mathcal{U}_{X_j}) \\
&\le \sum_{j=1}^{M'} \mathrm{E}_{X_j} \frac{M'-1}{M'C} \le \frac{M'-1}{C}.
\end{aligned}$$

Further

$$\begin{aligned}
\mathrm{E}_X \frac{1}{M'} \sum_{i=1}^{M'} W_{X_i}(\mathcal{U}_{X_i}^c) &= \sum_{i=1}^{M'} \mathrm{E}_{X_i} W_{X_i} \frac{(\mathcal{U}_{X_i}^c)}{M'} \\
&= \mathrm{E}_p W_x(\mathcal{U}_x^c).
\end{aligned}$$

Using the Markov inequality $\mathrm{P}_X\{X > \alpha \mathrm{E}X\} < \frac{1}{\alpha}$, i.e., $\mathrm{P}_X\{X \le \alpha \mathrm{E}X\} > 1 - \frac{1}{\alpha}$, we can show that

$$\begin{aligned}
\mathrm{P}_X \left\{ \frac{1}{M'} \sum_{i=1}^{M'} W_{X_i}\left(\bigcup_{j \ne i} \mathcal{U}_{X_j} \right) \le \alpha' \frac{M'-1}{C} \right\} &> 1 - \frac{1}{\alpha'} \\
\mathrm{P}_X \left\{ \frac{1}{M'} \sum_{i=1}^{M'} W_{X_i}(\mathcal{U}_{X_i}^c) \le \alpha \mathrm{E}_p W_x(\mathcal{U}_x^c) \right\} &> 1 - \frac{1}{\alpha}.
\end{aligned}$$

Since $(1 - \frac{1}{\alpha'}) + (1 - \frac{1}{\alpha}) > 1$, there exist M elements $x_1, \ldots, x_{M'}$ such that

$$\begin{aligned}
\frac{1}{M'} \sum_{i=1}^{M'} W_{x_i}\left(\bigcup_{j \ne i} \mathcal{U}_{x_j} \right) &\le \alpha' \frac{M'-1}{C} \\
\frac{1}{M'} \sum_{i=1}^{M'} W_{x_i}(\mathcal{U}_{x_i}^c) &\le \alpha \mathrm{E}_p W_c(\mathcal{U}_x^c).
\end{aligned}$$

In the following, the above M elements $x_1, \ldots, x_{M'}$ are fixed, and we only focus on the random variable i subject to the uniform distribution on the set $\{1, \ldots, M'\}$. Combining Markov inequality and the preceding inequalities, we have

$$\begin{aligned}
\mathrm{P}_i \left\{ W_{x_i}\left(\bigcup_{j \ne i} \mathcal{U}_{x_j} \right) \le \beta'\alpha' \frac{M'-1}{C} \right\} &> 1 - \frac{1}{\beta'} \\
\mathrm{P}_i \left\{ W_{x_i}(\mathcal{U}_{x_i}^c) \le \beta\alpha \mathrm{E}_p W_x(\mathcal{U}_x^c) \right\} &> 1 - \frac{1}{\beta}.
\end{aligned}$$

Hence, we obtain

$$\begin{aligned}
\mathrm{P}_i \left\{ \begin{array}{l} W_{x_i}\left(\bigcup_{j \ne i} \mathcal{U}_{x_j} \right) \le \beta'\alpha' \frac{M'-1}{C}, \\ W_{x_i}(\mathcal{U}_{x_i}^c) \le \beta\alpha \mathrm{E}_p W_x(\mathcal{U}_x^c) \end{array} \right\} \\
> \left(1 - \frac{1}{\beta}\right) + \left(1 - \frac{1}{\beta'}\right) - 1 = \gamma
\end{aligned}$$

which yields

$$\left| \left\{ i \left| \begin{array}{l} W_{x_i}\left(\bigcup_{j \ne i} \mathcal{U}_{x_j} \right) \le \beta'\alpha' \frac{M'-1}{C}, \\ W_{x_i}(\mathcal{U}_{x_i}^c) \le \beta\alpha \mathrm{E}_p W_x(\mathcal{U}_x^c) \end{array} \right. \right\} \right| > \lfloor \gamma M' \rfloor.$$

Since $\lfloor \gamma M' \rfloor = \lfloor \gamma \lceil \frac{M}{\gamma} \rceil \rfloor \ge M$, there exist M elements of $\mathcal{X}$ satisfying (5) and (6). Here, one may think that these M elements may not be distinct. However, if $x_i = x_{i'} (i \ne i')$, the relation $W_{x_i}(\mathcal{U}_{x_i}^c) + W_{x_i}(\bigcup_{j \ne i} \mathcal{U}_{x_j}) \ge 1$ holds. From condition (3), this contradicts (5) and (6). Hence, we obtain the desired bound. □

III. Channel Resolvability in Nonasymptotic Setting

In the channel resolvability, we choose M elements $x_1,\ldots,x_M$ in the input set $\mathcal{X}$ for every probability distribution $p \in \mathcal{P}(\mathcal{X})$, such that the output distribution of the input distribution

$$\sum_{i=1}^{M} \frac{1}{M} 1_{x_i}$$

close enough to the output distribution of p through the channel W. In particular, we call the distribution with the preceding form an M-type. In this setting, our purpose is to disenable the receiver of the given channel W to distinguish whether the sender generates the input signal based on "the given distribution p" or "the M-type $\sum_{i=1}^{M} \frac{1}{M} 1_{x_i}$ with a smaller number M." This kind of indistinguishability cannot be applied to any realistic model, directly, but it can be technically related to wiretap channel. In particular, we prove Lemma 2 in this section as the technically essential part, but this lemma is also the technically essential part for the direct part of wiretap channel.

In the following, we call the pair of the integer M and the M elements $x_1,\ldots,x_M$ of $\mathcal{X}$, a resolvability code Ψ with the size $|\Psi| \stackrel{\rm def}{=} M$. The performance of a resolvability code Ψ is characterized by its size $|\Psi|$ and the variational distance

$$\epsilon(\Psi, W_p) \stackrel{\rm def}{=} d\left(\sum_{i=1}^{M} \frac{1}{M} W_{x_i}, W_p\right)$$

where the variational distance $d(p,q)$ defined by

$$d(p,q) = \sum_y |p(y) - q(y)|$$

which equals the l_1 norm $\|p-q\|_1$. Another characterization of its performance is given by K-L divergence

$$D(\Psi, W_p) \stackrel{\rm def}{=} D\left(\sum_{i=1}^{M} \frac{1}{M}\, W_{x_i} \| W_p\right), \tag{8}$$

where $D(p\,\|\,q) \stackrel{\rm def}{=} \sum_y p(y) \log \frac{p(y)}{q(y)}$.

Theorem 2: For any integer $M > 0$, any real number $C > 0$, and any probability distribution $p \in \mathcal{P}(\mathcal{X})$, there exists a resolvability code Ψ such that $|\Psi| = M$ and

$$\begin{aligned} \epsilon(\Psi, W_p) &\le 2\delta_{p,W,C} + \sqrt{\frac{\delta'_{p,W,C}}{M}} \\ \delta'_{p,W,C} &\le C \end{aligned} \tag{9}$$

where

$$\delta_{p,W,C} \stackrel{\rm def}{=} \mathrm{E}_p W_x \left\{ \frac{W_x}{W_p}(y) > C \right\}$$

and

$$\delta'_{p,W,C} \stackrel{\rm def}{=} \mathrm{E}_p \frac{W_x^2}{W_p} \left\{ \frac{W_x}{W_p}(y) \le C \right\}.$$

If the cardinality $|\mathcal{Y}|$ is finite, for any $0 > t \ge -1/2$, there exists a resolvability code Ψ' such that $|\Psi'| = M$ and either of

$$D(\Psi', W_p) \le \frac{\log(1 + M^t e^{\phi(t\,|\,W,p)})}{-t} \tag{10}$$

$$D(\Psi', W_p) \le \eta(\delta_{p,W,C}) + \delta_{p,W,C} \log |\mathcal{Y}| + \frac{\delta'_{p,W,C}}{M} \tag{11}$$

holds, where

$$\eta(x) \stackrel{\rm def}{=} -x \log x$$

and

$$\phi(t\,|\,W,p) \stackrel{\rm def}{=} \log \sum_y (\mathrm{E}_p W_x^{1/(1+t)}(y))^{1+t}.$$

Remark 1: The partial resolvability version of inequality (9) has been obtained by Oohama [9]. Inequality (9) can be regarded as the essentially same result as Oohama's inequality.

Proof: In the following, the indicator functions I_x and I_x^c on the sets $\mathcal{U}_x = \{\frac{W_x}{W_p}(y) > C\}$ and their compliment sets $\mathcal{U}_x^c$ play important roles. In our proof of Theorem 2, we use the random coding method, i.e., we consider the M independent and identical random variables $X = (X_1,\ldots,X_M)$ subject to p. Using the notations

$$\begin{aligned} &W_x^\alpha(y) \stackrel{\rm def}{=} W_x(y) I_x^c(y), \quad W_x^\beta(y) \stackrel{\rm def}{=} W_x(y) I_x(y) \\ &W_p^\alpha(y) \stackrel{\rm def}{=} \mathrm{E}_p W_x^\alpha(y), \quad W_p^\beta(y) \stackrel{\rm def}{=} \mathrm{E}_p W_x^\beta(y) \\ &W_X^\alpha(y) \stackrel{\rm def}{=} \frac{1}{M} \sum_{i=1}^{M} W_{X_i}^\alpha(y), \quad W_X^\beta(y) \stackrel{\rm def}{=} \frac{1}{M} \sum_{i=1}^{M} W_{X_i}^\beta(y) \\ &W_X^M(y) \stackrel{\rm def}{=} W_X^\alpha(y) + W_X^\beta(y) = \frac{1}{M} \sum_{i=1}^{M} W_{X_i}(y) \end{aligned}$$

we have the following lemma.

Lemma 2: The M random variables $X = (X_1,\ldots,X_M)$ satisfy the following inequality:

$$\mathrm{E}_X \left\| W_X^M - W_p \right\|_1 \le 2\delta_{p,W,C} + \sqrt{\frac{\delta'_{p,W,C}}{M}} \tag{12}$$

$$\mathrm{E}_X D(W_X^M \| W_p) \le \frac{\log(1 + M^t e^{\phi(t\,|\,W,p)})}{-t} \tag{13}$$

for $0 > t \ge -\frac{1}{2}$. If the cardinality $\mathcal{Y}$ is finite, the inequality

$$\mathrm{E}_X D(W_X^M \| W_p) \le \eta(\delta_{p,W,C}) + \delta_{p,W,C} \log |\mathcal{Y}| + \frac{\delta'_{p,W,C}}{M} \tag{14}$$

holds.

Since there exists a resolvability code Ψ with the size M such that

$$\epsilon(\Psi, W_p) \le \mathrm{E}_X \left\| W_X^M - W_p \right\|_1$$

the inequality (12) guarantees the existence of a resolvability code Ψ satisfying (9). On the other hand, the relation

$$\frac{W_x^\alpha(y)}{W_p(y)} = \frac{W_x(y)}{W_p(y)} I_x^c(y) \le C$$

holds. Thus,

$$\delta'_{p,W,C} = \sum_y W_x(y) \frac{W_x^\alpha(y)}{W_p(y)} \le C.$$

Similarly, since there exists a resolvability code Ψ with the size M such that

$$D(\Psi', W_p) \le \mathrm{E}_X D(W_X^M \,\|\, W_p)$$

the inequalities (13) and (14) guarantees the existence of a resolvability code Ψ satisfying (10) and (11). □

Proof of Lemma 2: First, we show (12). Since

$$\delta_{p,W,C} = \mathrm{E}_{p,x} W_x(\mathcal{U}_x) = \mathrm{E}_p \|W_x^\beta\| = \|W_p^\beta\|$$

we can evaluate

$$\begin{aligned}
&\mathrm{E}_X \left\|W_X^M - W_p\right\|_1 \\
&= \mathrm{E}_X \left\|W_X^\alpha - W_p^\alpha + W_X^\beta - W_p^\beta\right\|_1 \\
&\le \mathrm{E}_X \left\|W_X^\alpha - W_p^\alpha\right\|_1 + \sum_{i=1}^M \frac{1}{M} \mathrm{E}_X \left\|W_{X_i}^\beta\right\|_1 + \left\|W_p^\beta\right\|_1 \\
&= \mathrm{E}_X \left\|W_X^\alpha - W_p^\alpha\right\|_1 + 2\mathrm{E}_{p,x} W_x(\mathcal{U}_x).
\end{aligned}$$

Next, we focus on the Schwarz inequality regarding the random variable

$$l_X(y) \stackrel{\mathrm{def}}{=} \frac{W_X^\alpha}{W_p}(y)$$

and the sign function

$$\tilde{l}_X(y) \stackrel{\mathrm{def}}{=} \frac{l_X(y)}{|l_X(y)|}$$

(we can check that $\tilde{l}_X^2 = 1$), then we obtain

$$\begin{aligned}
(\|W_p l_X\|_1)^2 = (\mathrm{E}_{W_p} |l_X(y)|)^2 &= (\mathrm{E}_{W_p} l_X(y) \tilde{l}_X(y))^2 \\
&\le \mathrm{E}_{W_p} l_X^2(y) \mathrm{E}_{W_p} \tilde{l}_X^2(y) = \mathrm{E}_{W_p} l_X^2(y).
\end{aligned}$$

Thus, the Jensen inequality yields that

$$\left(\mathrm{E}_X \left\|W_X^\alpha - W_p^\alpha\right\|_1\right)^2 \le \mathrm{E}_X \left\|W_X^\alpha - W_p^\alpha\right\|_1^2 \le \mathrm{E}_X \mathrm{E}_{W_p} l_X^2(y).$$

Since

$$\mathrm{E}_x \frac{W_x^\alpha(y)}{W_p(y)} = \frac{W_p^\alpha(y)}{W_p(y)}$$

we have

$$\begin{aligned}
\mathrm{E}_X \mathrm{E}_{W_p} l_X^2 &= \mathrm{E}_{W_p} \mathrm{E}_X l_X^2(y) \\
&= \mathrm{E}_{W_p} \mathrm{E}_X \frac{1}{M^2} \sum_{i=1}^M \left(\frac{W_{X_i}^\alpha(y)}{W_p(y)} - \frac{W_p^\alpha(y)}{W_p(y)}\right)^2 \\
&= \mathrm{E}_{W_p} \frac{1}{M} \mathrm{E}_x \left(\left(\frac{W_x^\alpha(y)}{W_p(y)}\right)^2 - \left(\frac{W_p^\alpha(y)}{W_p(y)}\right)^2\right) \\
&\le \mathrm{E}_x \frac{1}{M} \mathrm{E}_{W_p} \left(\frac{W_x^\alpha(y)}{W_p(y)}\right)^2 = \frac{\delta'_{p,W,C}}{M}.
\end{aligned}$$

Therefore, we obtain

$$\mathrm{E}_X \left\|\sum_{i=1}^M \frac{1}{M} W_{X_i} - W_p\right\|_1 \le 2\delta_{p,W,C} + \sqrt{\frac{\delta'_{p,W,C}}{M}}.$$

Hence, we obtain (12).

Next, we show (14). Since

$$\frac{W_X^M(y)}{W_p(y)} \le \frac{1}{W_p^\beta(y)}$$

by using the inequality $\log x \le x - 1$, we can evaluate

$$\begin{aligned}
&\mathrm{E}_X D(W_X^M \,\|\, W_p) \\
&= \mathrm{E}_X \sum_y \left(W_X^\alpha(y) \log \frac{W_X^M(y)}{W_p(y)} + \sum_y W_X^\beta(y) \log \frac{W_X^M(y)}{W_p(y)}\right) \\
&\le \mathrm{E}_X \sum_y \left(W_X^\alpha(y) \left(\frac{W_X^M(y)}{W_p(y)} - 1\right) + W_X^\beta(y) \log \frac{1}{W_p^\beta(y)}\right) \\
&= \sum_y \mathrm{E}_X W_X^\alpha(y) \left(\frac{W_X^M(y)}{W_p(y)} - 1\right) + \sum_y W_p^\beta(y) \log \frac{1}{W_p^\beta(y)}.
\end{aligned}$$

Regarding the first term, we can calculate

$$\begin{aligned}
&\sum_y \mathrm{E}_X W_X^\alpha(y) \left(\frac{W_X^M(y)}{W_p(y)} - 1\right) \\
&= \sum_y \mathrm{E}_X \frac{1}{M^2} \sum_{i,j} W_{X_i}^\alpha(y) \left(\frac{W_{X_j}(y)}{W_p(y)} - 1\right) \\
&= \sum_y \frac{1}{M} \mathrm{E}_{p,x} W_x^\alpha(y) \left(\frac{W_x(y)}{W_p(y)} - 1\right) \\
&\le \sum_y \frac{1}{M} \mathrm{E}_{p,x} \frac{W_x^\alpha(y)}{W_p(y)} W_x(y) = \frac{\delta'_{p,W,C}}{M}
\end{aligned}$$

where we use the relation

$$\mathrm{E}_X W_{X_i}^\alpha(y) \left(\frac{W_{X_j}(y)}{W_p(y)} - 1\right) = 0, \qquad \text{for } i \ne j.$$

Concerning the second term, letting $K \stackrel{\mathrm{def}}{=} \sum_y W_p^\beta(y)$, we have

$$\begin{aligned}
&\sum_y W_p^\beta(y) \log \frac{1}{W_p^\beta(y)} \\
&= -K \log K - K \sum_y \frac{W_p^\beta(y)}{K} \log \frac{W_p^\beta(y)}{K} \\
&\le \eta\left(\sum_y W_p^\beta(y)\right) + \sum_y W_p^\beta(y) \log |\mathcal{Y}|
\end{aligned}$$

because $\log |\mathcal{Y}|$ is the maximal entropy of the distribution on the probability space $\mathcal{Y}$. Since $\sum_y W_p^\beta(y) = \delta_{p,W,C}$, we obtain (14).

Finally, we prove (13) by a different method. The quantity $\mathrm{E}_X D(W_X^M \,\|\, W_p)$ can be regarded as the mutual information of channel $X \mapsto W_X^M$ with the input probability $p^M(X)$ which equals the M-fold i.i.d. of p. We can check that the function $t \mapsto \phi(t \,|\, W^M, p^M)$ satisfies the following property:

$$\begin{aligned}
\phi(0 \,|\, W^M, p^M) &= 0 \\
\left.\frac{d\phi(t \,|\, W^M, p^M)}{dt}\right|_{t=0} &= -\mathrm{E}_X D(W_X^M \| W_p), \\
\frac{d^2\phi(t|W^M, p^M)}{dt^2} &\ge 0.
\end{aligned}$$

Hence, its convexity guarantees the inequality

$$t\mathrm{E}_X D(W_X^M\|W_p) \le \phi(t|W^M,p^M)$$

which implies the inequality

$$\mathrm{E}_X D(W_X^M \,\|\, W_p) \le \frac{\phi(t\,|\,W^M,p^M)}{-t} \tag{15}$$

for $0 > t \ge -\frac{1}{2}$.

Let $1+s=\frac{1}{1+t}$, then $1 \ge s > 0$ and $t=\frac{-s}{1+s}$. Since $x \mapsto x^s$ is concave

$$\mathrm{E}_X \left(\sum_{j\neq i} W_{X_j}(y)\right)^s \le \left[\mathrm{E}_X \sum_{j\neq i} W_{X_j}(y)\right]^s = (M-1)^s W_p^s(y). \tag{16}$$

Using (16) and the relation $(x+y)^s \le x^s + y^s$ for two positive real numbers x, y, we obtain

$$\begin{aligned}
e^{\phi(t\,|\,W^M,p^M)} &= \sum_y \left(\mathrm{E}_X (W_X^M)^{1+s}(y)\right)^{\frac{1}{1+s}} \\
&= \frac{1}{M}\sum_y \left(\mathrm{E}_X \sum_{i=1}^M W_{X_i}(y) \left(W_{X_i}(y) + \sum_{j\neq i} W_{X_j}(y)\right)^s\right)^{\frac{1}{1+s}} \\
&\le \frac{1}{M}\sum_y \left(\mathrm{E}_X \sum_{i=1}^M W_{X_i}(y) \left(W_{X_i}^s(y) + \left(\sum_{j\neq i} W_{X_j}(y)\right)^s\right)\right)^{\frac{1}{1+s}} \\
&= \frac{1}{M}\sum_y \left(\sum_{i=1}^M \mathrm{E}_X W_{X_i}^{1+s}(y) + \sum_{i=1}^M \mathrm{E}_X W_{X_i}(y) \left(\sum_{j\neq i} W_{X_j}(y)\right)^s\right)^{\frac{1}{1+s}} \\
&\le \sum_y \frac{1}{M}\left(\sum_{i=1}^M \mathrm{E}_X W_{X_i}^{1+s}(y) \sum_{i=1}^M (M-1)^s W_p^{1+s}(y)\right)^{\frac{1}{1+s}} \\
&= \frac{1}{M}\sum_y \left(M\mathrm{E}_x W_x^{1+s}(y) M(M-1)^s W_p^{1+s}(y)\right)^{\frac{1}{1+s}} \\
&\le \frac{1}{M}\sum_y \left(M\mathrm{E}_x W_x^{1+s}(y)\right)^{\frac{1}{1+s}} \\
&\quad + \left(M(M-1)^s W_p^{1+s}(y)\right)^{\frac{1}{1+s}} \\
&= \sum_y \frac{\left(\mathrm{E}_x W_x^{1+s}(y)\right)^{\frac{1}{1+s}}}{M^{\frac{s}{1+s}}} + \left(\frac{M-1}{M}\right)^{\frac{s}{1+s}} W_p(y) \\
&\le 1 + \frac{1}{M^{\frac{s}{1+s}}}\sum_y \left(\mathrm{E}_x W_x^{1+s}(y)\right)^{\frac{1}{1+s}} \\
&= 1 + M^t e^{\phi(t\,|\,W,p)}.
\end{aligned}$$

Since $-t$ is positive, the desired inequality (13) follows from (15) and the above inequality. □

Next, we proceed to the relation with identification codes. In order to discuss this relation, we focus on channel resolvability of the worst input case, and define the following values:

$$\epsilon(M,W) \stackrel{\mathrm{def}}{=} \max_{p\in\mathcal{P}(\mathcal{X})} \min_{\Psi:|\Psi|\le M} \epsilon(\Psi, W_p)$$
$$D(M,W) \stackrel{\mathrm{def}}{=} \max_{p\in\mathcal{P}(\mathcal{X})} \min_{\Psi:|\Psi|\le M} D(\Psi, W_p)$$

which satisfies

$$\epsilon(M,W) \le 2\max_p \mathrm{E}_p W_x\left\{\frac{W_x}{W_p}(y) > C\right\} + \sqrt{\frac{C}{M}} \tag{17}$$

for any real number $C > 0$.

Lemma 3: (Han and Verdú [2]) If the cardinality $|\mathcal{X}|$ is finite, and if an identification code Φ and an integer M satisfy

$$1 - \mu(\Phi) - \lambda(\Phi) > \epsilon(M,W)$$

then

$$|\mathcal{X}|^M \ge |\Phi|. \tag{18}$$

Proof: Let the identification code Φ be a triplet

$$(N, \{Q_1,\ldots,Q_N\}, \{\mathcal{D}_1,\ldots,\mathcal{D}_N\}),$$

then there exist NM-types $Q'_1,\ldots Q'_N$ such that

$$d(W_{Q_i}, W_{Q'_i}) \le \epsilon(M,W).$$

Since the inequalities

$$\begin{aligned}
&2\epsilon(M,W) + d(W_{Q'_i}, W_{Q'_j}) \\
&\quad \ge d(W_{Q_i}, W_{Q'_i}) + d(W_{Q_j}, W_{Q'_j}) + d(W_{Q'_i}, W_{Q'_j}) \\
&\quad \ge d(W_{Q_i}, W_{Q_j}) \ge 2(W_{Q_i}(\mathcal{D}_i) - W_{Q_j}(\mathcal{D}_i)) \\
&\quad \ge 2(1-\mu(\Phi)-\lambda(\Phi))
\end{aligned}$$

hold for any $i \neq j$, we can show

$$d(W_{Q'_i}, W_{Q'_j}) > 0$$

which implies that Q'_i is different from Q'_j. However, the total number of M-types is less than $|\mathcal{X}|^M$. Therefore, we obtain (18). □

IV. Wiretap Channel in Nonasymptotic Setting

Next, we discuss the message transmission with the wiretapper who has less information than the main receiver. This problem is formulated as follows. Let $\mathcal{Y}$ be the probability space of the main receiver, and $\mathcal{Z}$ be the space of the wiretapper, then the main channel from the transmitter to the main receiver is described by $W^B : x \mapsto W_x^B$, and the wiretapper channel from the transmitter to the the wiretapper is described by $W^E : x \mapsto W_x^E$. In this setting, the transmitter chooses M distributions $Q_1,\ldots,Q_M$ on $\mathcal{X}$, and he generates $x \in \mathcal{X}$ subject to Q_i when he wants to send the message $i \in \{1,\ldots,M\}$. The normal receiver prepares M disjoint subsets $\mathcal{D}_1,\ldots,\mathcal{D}_M$

of $\mathcal{Y}$ and judges that a message is i if y belongs to $\mathcal{D}_i$. Therefore, the triplet $(M, \{Q_1, \ldots, Q_M\}, \{\mathcal{D}_1, \ldots, \mathcal{D}_M\})$ is called a code, and is described by Φ. Its performance is given by the following quantities. One is the size M, which is denoted by $|\Phi|$. The second one is the average error probability $\epsilon_B(\Phi)$

$$\epsilon_B(\Phi) \stackrel{\mathrm{def}}{=} \frac{1}{M}\sum_{i=1}^{M} W_{Q_i}^B(\mathcal{D}_i^c)$$

and the third one is the wiretapper's information regarding the transmitted message $I_E(\Phi)$

$$I_E(\Phi) \stackrel{\mathrm{def}}{=} \sum_i \frac{1}{M} D(W_{Q_i}^E \| W_\Phi^E), \qquad W_\Phi^E \stackrel{\mathrm{def}}{=} \sum_i \frac{1}{M} W_{Q_i}^E.$$

A different measure of the wiretapper's information is given by the average variational distance $d_E(\Phi)$

$$d_E(\Phi) \stackrel{\mathrm{def}}{=} \frac{1}{M(M-1)} \sum_{i \neq j} d(W_{Q_i}^E, W_{Q_j}^E).$$

Theorem 3: There exists a code Φ for any integers L, M, any real numbers $C, C' > 0$, and any probability distribution p on $\mathcal{X}$ such that

$$|\Phi| = M$$

$$\epsilon_B(\Phi) \le 3 \min_{0 \le s \le 1} (ML)^s \sum_y \left(\mathrm{E}_p (W_x^B(y))^{1/(1+s)} \right)^{1+s} \tag{19}$$

$$\epsilon_B(\Phi) \le 3 \left(\mathrm{E}_p W_x \left\{ \frac{W_x^B}{W_p^B}(y) \le C' \right\} + \frac{ML}{C'} \right) \tag{20}$$

$$I_E(\Phi) \le 3 \Bigg(\eta(\delta_{p,W^E,C}) + \delta_{p,W^E,C} \log |\mathcal{Z}| + \frac{\delta'_{p,W^E,C}}{L} \Bigg) \tag{21}$$

$$I_E(\Phi) \le 3 \min_{0 > t \ge -1/2} \frac{\log(1 + L^t e^{\phi(t\,|\,W^E,p)})}{t} \tag{22}$$

$$d_E(\Phi) \le 6 \left(2\delta_{p,W^E,C} + \sqrt{\frac{\delta'_{p,W^E,C}}{L}} \right). \tag{23}$$

Proof: We prove Theorem 3 by a random coding method. Let $X = (X_{l,m})$ be LM independent and identical random variables subject to the distribution p on $\mathcal{X}$ for integers $l = 1, \ldots, L$ and $m = 1, \ldots, M$, and $\mathcal{D}'_{l,m}(X)$ be the maximum-likelihood decoder of the code $X_{l,m}$, then we can evaluate as follows by Gallager upper bound [7]:

$$\mathrm{E}_X \frac{1}{ML} \sum_{l,m} W_{X_{m,l}}^B (\mathcal{D}'_{l,m}(X)^c) \le \min_{0 \le s \le 1} (ML)^s \sum_y \left(\mathrm{E}_p \left(W_x^B(y) \right)^{1/(1+s)} \right)^{1+s}.$$

Since the maximum-likelihood decoder is better than the code

$$\mathcal{D}''_{l,m}(X) = \left\{ \frac{W_x^B}{W_p^B}(y) > C' \right\} \setminus \bigcup_{(l',m') \neq (l,m)} \left\{ \frac{W_x^B}{W_p^B}(y) > C' \right\}$$

we have another evaluation as

$$\begin{aligned}
&\mathrm{E}_X \frac{1}{ML} \sum_{l,m} W_{X_{m,l}}^B (\mathcal{D}'_{l,m}(X)^c) \\
&\le \mathrm{E}_X \frac{1}{ML} \sum_{l,m} W_{X_{m,l}}^B (\mathcal{D}''_{l,m}(X)^c) \\
&\le \mathrm{E}_X \frac{1}{ML} \sum_{l,m} W_{X_{m,l}}^B \left\{ \frac{W_{X_{m,l}}^B}{W_p^B}(y) \le C' \right\} \\
&\quad + \mathrm{E}_X \frac{1}{ML} \sum_{l,m} W_{X_{m,l}}^B \sum_{(l',m') \neq (l,m)} \left\{ \frac{W_{X_{m',l'}}^B}{W_p^B}(y) \le C' \right\} \\
&\le \mathrm{E}_{p,x} W_x^B \left\{ \frac{W_x^B}{W_p^B}(y) \le C' \right\} \\
&\quad + W_p^B (ML - 1) \mathrm{E}_{p,x} \left\{ \frac{W_x^B}{W_p^B}(y) \le C' \right\} \\
&\le \mathrm{E}_{p,x} W_x^B \left\{ \frac{W_x^B}{W_p^B}(y) \le C' \right\} + \frac{ML}{C'}.
\end{aligned}$$

Let $Q_m(X)$ be the uniform distribution on $\{X_{1,m}, \ldots, X_{L,m}\}$, $\mathcal{D}_m(X)$ be $\cup_l \mathcal{D}'_{l,m}(X)$, and $\Phi(X)$ be the code $(M, \{Q_m(X)\}, \{\mathcal{D}_m(X)\})$, then $\mathrm{E}_X \epsilon_B(\Phi(X))$ is less than the right-hand sides (RHSs) of (19) and (20) because the average error probability of $\Phi(X)$ is less than the one of the code $(ML, \{X_{m,l}\}, \{\mathcal{D}'_{l,m}(X)\})$.

Since

$$\sum_{m=1}^{M} \frac{1}{M} D\left(W_{Q_m(X)}^E \middle\| W_{\Phi(X)}^E \right) + D\left(W_{\Phi(X)}^E \middle\| W_p^E \right) = \sum_{m=1}^{M} \frac{1}{M} D\left(W_{Q_m(X)}^E \middle\| W_p^E \right)$$

we obtain

$$\begin{aligned}
\mathrm{E}_X I_E(\Phi(X)) &= \mathrm{E}_X \sum_{m=1}^{M} \frac{1}{M} D\left(W_{Q_m(X)}^E \middle\| W_{\Phi(X)}^E \right) \\
&\le \mathrm{E}_X \sum_{m=1}^{M} \frac{1}{M} D\left(W_{Q_m(X)}^E \middle\| W_p^E \right) \\
&\le \eta(\delta_{p,W^E,C}) + \delta_{p,W^E,C} \log |\mathcal{Z}| + \frac{\delta'_{p,W,C}}{L}
\end{aligned}$$

where the last inequality follows from Lemma 2. Similarly, we can show

$$\mathrm{E}_X I_E(\Phi(X)) \le \frac{\log(1 + L^t e^{\phi(t\,|\,W^E,p)})}{t}.$$

Regarding $d_E(\Phi(X))$, we can calculate

$$\begin{aligned}
&\mathrm{E}_X \frac{1}{M(M-1)} \sum_{i \neq j} d\left(W_{Q_i(X)}^E, W_{Q_j(X)}^E \right) \\
&\le \mathrm{E}_X \frac{1}{M(M-1)} \sum_{i \neq j} d\left(W_{Q_i(X)}^E, W_p^E \right) + d(W_{Q_j(X)}^E, W_p^E) \\
&= 2\mathrm{E}_X d\left(W_{Q_1(X)}^E, W_p^E \right) \\
&\le 2 \left(2\delta_{p,W,C} + \sqrt{\frac{\delta'_{p,W,C}}{L}} \right).
\end{aligned}$$

Using Markov inequality, we obtain

$$\mathrm{P}_X\{\epsilon_B(\Phi(X)) \le 3\mathrm{E}\epsilon_B(\Phi(X))\}^c < \frac{1}{3}$$
$$\mathrm{P}_X\{I_E(\Phi(X)) \le 3\mathrm{E}I_E(\Phi(X))\}^c < \frac{1}{3}$$
$$\mathrm{P}_X\{d_E(\Phi(X)) \le 3\mathrm{E}d_E(\Phi(X))\}^c < \frac{1}{3}.$$

Therefore, there exists a code Φ satisfying desired conditions. □

V. General Asymptotic Setting

A. Identification Code and Channel Resolvability

Next, we focus on an arbitrary sequence of channels $\boldsymbol{W} = \{W^n\}_{n=1}^\infty$, in which W^n is an arbitrary channel from $\mathcal{X}^n$ to $\mathcal{Y}^n$. In this setting, two-types of (μ,λ)-identification capacities are defined by

$$D(\mu,\lambda\,|\,\boldsymbol{W}) \stackrel{\mathrm{def}}{=} \sup_{\{\Phi_n\}} \left\{ \underline{\lim}\frac{1}{n}\log\log|\Phi_n| \,\middle|\, \overline{\lim}\mu(\Phi_n) < \mu, \overline{\lim}\lambda(\Phi_n) \le \lambda \right\}$$
$$D^\dagger(\mu,\lambda\,|\,\boldsymbol{W}) \stackrel{\mathrm{def}}{=} \sup_{\{\Phi_n\}} \left\{ \underline{\lim}\frac{1}{n}\log\log|\Phi_n| \,\middle|\, \underline{\lim}\mu(\Phi_n) < \mu, \overline{\lim}\lambda(\Phi_n) \le \lambda \right\}.$$

However, in the case of $\mu = 0$, we replace

$$\overline{\lim}\mu(\Phi_n) < \mu \qquad (\underline{\lim}\mu(\Phi_n) < \mu)$$

by

$$\overline{\lim}\mu(\Phi_n) = 0 \qquad (\underline{\lim}\mu(\Phi_n) = 0)$$

at the above two definitions. On the other hand, two-types ϵ-resolvability capacities are defined by

$$S(\epsilon|\boldsymbol{W}) \stackrel{\mathrm{def}}{=} \sup\{R\,|\,\overline{\lim}\epsilon(e^{nR}, W^n) \le \epsilon\}$$
$$S^\dagger(\epsilon|\boldsymbol{W}) \stackrel{\mathrm{def}}{=} \sup\{R\,|\,\underline{\lim}\epsilon(e^{nR}, W^n) \le \epsilon\}$$

where in the case of $\epsilon = 2$, we replace $\le\epsilon$ by <2 in the above two definitions.

In the information spectrum method, the following quantities are defined for arbitrary sequence $\boldsymbol{p} = \{p^n\}_{n=1}^\infty$ of input probability distributions:

$$\bar{I}(\epsilon|\boldsymbol{p},\boldsymbol{W}) \stackrel{\mathrm{def}}{=} \inf\left\{ a \,\middle|\, \overline{\lim}\mathrm{E}_{p^n} W_x^n \left\{ \frac{1}{n}\log\frac{W_x^n}{W_{p^n}^n}(y) > a \right\} \le \epsilon \right\}$$
$$\underline{I}(\epsilon|\boldsymbol{p},\boldsymbol{W}) \stackrel{\mathrm{def}}{=} \inf\left\{ a \,\middle|\, \underline{\lim}\mathrm{E}_{p^n} W_x^n \left\{ \frac{1}{n}\log\frac{W_x^n}{W_{p^n}^n}(y) > a \right\} \le \epsilon \right\}$$

where the case of $\epsilon = 1$, we replace $\le$ by $<$ at the above definitions. These quantities have another expression as

$$\bar{I}(\epsilon\,|\,\boldsymbol{p},\boldsymbol{W}) = \sup\left\{ a \,\middle|\, \underline{\lim}\mathrm{E}_{p^n} W_x^n \left\{ \frac{1}{n}\log\frac{W_x^n}{W_{p^n}^n}(y) \le a \right\} < 1-\epsilon \right\}$$
$$\underline{I}(\epsilon|\boldsymbol{p},\boldsymbol{W}) = \sup\left\{ a \,\middle|\, \overline{\lim}\mathrm{E}_{p^n} W_x^n \left\{ \frac{1}{n}\log\frac{W_x^n}{W_{p^n}^n}(y) \le a \right\} < 1-\epsilon \right\}.$$

Theorem 4: Assume that $|\mathcal{X}^n| = d^n$, then the above quantities satisfy the following relations:

$$\sup_{\boldsymbol{p}} \underline{I}(\epsilon\,|\,\boldsymbol{p},\boldsymbol{W}) \le D(1-\epsilon,0\,|\,\boldsymbol{W}) \le S^\dagger(\epsilon|\boldsymbol{W}) \le \sup_{\boldsymbol{p}} \underline{I}\left(\frac{\epsilon}{2}|\boldsymbol{p},\boldsymbol{W}\right) \tag{24}$$
$$\sup_{\boldsymbol{p}} \bar{I}(\epsilon|\boldsymbol{p},\boldsymbol{W}) \le D^\dagger(1-\epsilon,0|\boldsymbol{W}) \le S(\epsilon\,|\,\boldsymbol{W}) \le \sup_{\boldsymbol{p}} \bar{I}\left(\frac{\epsilon}{2}|\boldsymbol{p},\boldsymbol{W}\right) \tag{25}$$

for any real number $0 \le \epsilon < 1$. However, the first inequalities in (24) and (25) hold for $0 \le \epsilon \le 1$, and the third ones hold for $0 \le \epsilon \le 2$. In particular, we obtain

$$\sup_{\boldsymbol{p}} \underline{I}(0|\boldsymbol{p},\boldsymbol{W}) = D(1,0|\boldsymbol{W}) = S^\dagger(0|\boldsymbol{W}) \tag{26}$$
$$\sup_{\boldsymbol{p}} \bar{I}(0\,|\,\boldsymbol{p},\boldsymbol{W}) = D^\dagger(1,0|\boldsymbol{W}) = S(0\,|\,\boldsymbol{W}) \tag{27}$$

which is desired in Han and Verdú [2] and Han [3].[1]

This theorem indicates the existence of a code satisfying the following: The second error probability λ is asymptotically independent for the behavior of the distribution of the random variable of likelihood and always goes to 0, and only the second error probability μ asymptotically depends on it.

Remark 2: Steinberg [14] claims the inequalities

$$\sup_{\boldsymbol{p}} \underline{I}(\epsilon|\boldsymbol{p},\boldsymbol{W}) \ge D(\lambda_1,\lambda_2|\boldsymbol{W})$$
$$\sup_{\boldsymbol{p}} \bar{I}(\epsilon\,|\,\boldsymbol{p},\boldsymbol{W}) \ge D^\dagger(\lambda_1,\lambda_2\,|\,\boldsymbol{W})$$

for $\lambda_1+\lambda_2 < 1-\epsilon$. If they are proved, by combining the above inequalities and Theorem 4, we can prove the equalities of the above inequalities in the continuous case. However, it seems that his paper has a gap in counting the maximum number of different pairs of a partial response and an M'-type measure at the proof of Lemma 2, which is essential for these inequalities. That is, he estimated the total number of positive functions on $\mathcal{X}\times\mathcal{Y}$ of the form

$$f(x,y) = \frac{1}{M'}\sum_{i=1}^{M'} 1_{x_i}(x) \sum_{(x',y')\in F} 1_{(x',y')}(x,y)$$

where F is an arbitrary subset of $\mathcal{X}\times\mathcal{Y}$. The total measure of f, i.e., $\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} f(x,y)$ is not necessarily less than 1, while he indicated that it is less than 1. Hence, this total number cannot be bounded by $|\mathcal{X}|^{M'}$.

Proof: In order to prove the first inequalities, we choose an arbitrary real number $R < \sup_{\boldsymbol{p}} \underline{I}(\epsilon\,|\,\boldsymbol{p},\boldsymbol{W})$ and a sequence of input probability distributions $\boldsymbol{p}$ such that

$$R < R' \stackrel{\mathrm{def}}{=} \underline{I}(1-\mu\,|\,\boldsymbol{p},\boldsymbol{W}).$$

Substitute $M = e^{nR}, C = e^{nR'}, \alpha = \beta = 1+\frac{2}{n}, \alpha' = \beta' = \frac{1}{n+2}, \tau = \frac{1}{n+2}, \kappa = \frac{\log 2+1}{\log n}$ in Theorem 1, then the conditions

[1] Theorem 6 in Han and Verdú [2] claims that $S(0\,|\,\boldsymbol{W}) = \sup_{\boldsymbol{p}} \bar{I}(0\,|\,\boldsymbol{p},\boldsymbol{W})$ always holds for any channel $\boldsymbol{W}$ if the input alphabet is finite. However, the proof in [2] contains a mistake in part, as mentioned in Han [3, Sec. 6.3]. Therefore, it has been an open problem as to whether this inequality holds or not.

(1) and (2) are satisfied and $\gamma = \frac{1}{n+2}$. Thus, there exists an identification code Φ_n such that

$$\begin{aligned}
|\Phi_n| &= \left\lfloor \frac{e^{\frac{e^{nR}}{n+2}}}{e^{1+nR}} \right\rfloor \\
\mu(\Phi_n) &\le \left(1+\frac{2}{n}\right)^2 \mathrm{E}_{p^n} W_x^n \left\{ \frac{1}{n} \log \frac{W_x^n}{W_{p^n}^n}(y) \le R' \right\} \\
\lambda(\Phi_n) &\le \frac{\log 2 + 1}{\log n} + (n+2)^2 \frac{1}{e^{nR'}} \lceil (n+2) e^{nR} \rceil \\
&\cong \frac{\log 2 + 1}{\log n} + (n+2)^3 e^{-n(R'-R)}.
\end{aligned}$$

Therefore, we obtain

$$\begin{aligned}
&\lim \frac{1}{n} \log \log |\Phi_n| = R, \\
&\overline{\lim} \mu(\Phi_n) \le \overline{\lim} \mathrm{E}_{p^n} W_x^n \left\{ \frac{1}{n} \log \frac{W_x^n}{W_{p^n}^n}(y) \le R' \right\} < \mu \\
&\lim \lambda(\Phi_n) = 0
\end{aligned} \tag{28}$$

which implies that $D(\mu, 0 \,|\, \boldsymbol{W}) \ge R'$. Thus, we obtain the first inequality in (24) for $0 \le \epsilon < 1$. In the case of $\epsilon = 1$, we need to replace $<\mu$ by $=0$ at (28). By replacing $\overline{\lim}$ by $\underline{\lim}$ at (28), we can similarly prove $D^\dagger(\mu, 0 \,|\, \boldsymbol{W}) \ge \sup_{\boldsymbol{p}} \bar{I}(1-\mu \,|\, \boldsymbol{p}, \boldsymbol{W})$.

Next, we proceed to the second inequalities. Let R be an arbitrary real number such that $R > D(1-\epsilon, 0 \,|\, \boldsymbol{W})$. Then, there exists a sequence $\{\Phi_n\}$ of identification codes such that

$$R = \overline{\lim} \frac{1}{n} \log \log |\Phi_n|, \quad \overline{\lim} \mu(\Phi_n) < 1-\epsilon, \quad \lim \lambda(\Phi_n) = 0.$$

Therefore, we can choose an integer N large enough, such that

$$1 - \mu(\Phi_n) - \lambda(\Phi_n) \ge 1 - \overline{\lim} \mu(\Phi_n) > \epsilon.$$

Moreover, we choose a strictly increasing sequence $\{a_n\}$ of integers such that $a_1 \ge N$ and

$$1 - \overline{\lim} \mu(\Phi_{a_n}) > \epsilon(e^{a_n R'}, W^n)$$

where $R' = S^\dagger(\epsilon, \boldsymbol{W})$.

Thus, Lemma 3 yields that $(d^{a_n})^{e^{a_n R'}} \ge |\Phi_{a_n}|$, which implies that $R' \ge R$. We obtain the second inequalities in (24). We can prove the second inequalities in (25) by choosing a strictly increasing sequence $\{a_n\}$ of integers such that

$$1 - \mu(\Phi_{a_n}) - \lambda(\Phi_{a_n}) \ge 1 - \underline{\lim} \mu(\Phi_n) > \epsilon.$$

Finally, we prove the third inequalities by using another expression of $\sup_{\boldsymbol{p}} \underline{I}(\epsilon \,|\, \boldsymbol{p}, \boldsymbol{W})$:

$$\begin{aligned}
&\sup_{\boldsymbol{p}} \underline{I}(\epsilon \,|\, \boldsymbol{p}, \boldsymbol{W}) \\
&\quad = \inf \left\{ a \left| \overline{\lim} \max_{p^n} \mathrm{E}_{p^n} W_x^n \left\{ \frac{1}{n} \log \frac{W_x^n}{W_{p^n}^n}(y) > a \right\} \le \epsilon \right. \right\}.
\end{aligned}$$

Let R and R' be arbitrary real numbers such that $R > \sup_{\boldsymbol{p}} \underline{I}(\epsilon/2 \,|\, \boldsymbol{p}, \boldsymbol{W})$ and $R > R' > \sup_{\boldsymbol{p}} \underline{I}(\epsilon \,|\, \boldsymbol{p}, \boldsymbol{W})$, then the inequality (17) yields that

$$\begin{aligned}
&\epsilon(e^{nR}, W^n) \\
&\quad \le 2 \min_{p^n} \mathrm{E}_{p^n} W_x^n \left\{ \frac{1}{n} \log \frac{W_x^n}{W_{p^n}^n}(y) > R' \right\} + e^{-n(R-R')/2}.
\end{aligned}$$

Taking the limit $\underline{\lim}$, we obtain

$$\underline{\lim} \epsilon(e^{nR}, W^n) \le \epsilon \tag{29}$$

which implies $S^\dagger(2\epsilon, \boldsymbol{W}) \le R$. Thus, we obtain the third inequality in (24) for $0 \le \epsilon < 1$. In the case of $\epsilon = 2$, we need to replace $\le\epsilon$ by <1 at (29). By replacing $\underline{\lim}$ by $\overline{\lim}$ in the above, we can prove the third one in (25). □

B. Wiretap Channel

Next, we focus on a general sequence

$$(\boldsymbol{W}^B = \{W^{B,n}\}, \boldsymbol{W}^E = \{W^{E,n}\})$$

of wiretap channels, and define the following two kinds of capacities by

$$\begin{aligned}
&C_d(\boldsymbol{W}^B, \boldsymbol{W}^E) \\
&\quad \stackrel{\mathrm{def}}{=} \sup_{\{\Phi_n\}} \left\{ \underline{\lim} \frac{1}{n} \log |\Phi_n| \middle| \lim \epsilon_B(\Phi_n) = \lim d_E(\Phi_n) = 0 \right\} \\
&C_I(\boldsymbol{W}^B, \boldsymbol{W}^E) \\
&\quad \stackrel{\mathrm{def}}{=} \sup_{\{\Phi_n\}} \left\{ \underline{\lim} \frac{1}{n} \log |\Phi_n| \middle| \lim \epsilon_B(\Phi_n) = \lim \frac{I_E(\Phi_n)}{n} = 0 \right\}.
\end{aligned}$$

Lemma 4: The inequality

$$C_d(\boldsymbol{W}^B, \boldsymbol{W}^E) \ge \underline{I}(1 \,|\, \boldsymbol{p}, \boldsymbol{W}^B) - \bar{I}(0 \,|\, \boldsymbol{p}, \boldsymbol{W}^E) \tag{30}$$

holds for any sequence of input distributions $\boldsymbol{p} = \{p^n\}$. Furthermore, if $|\mathcal{Z}^n| = d^n$

$$C_I(\boldsymbol{W}^B, \boldsymbol{W}^E) \ge \underline{I}(1 \,|\, \boldsymbol{p}, \boldsymbol{W}^B) - \bar{I}(0 \,|\, \boldsymbol{p}, \boldsymbol{W}^E). \tag{31}$$

This theorem is an information spectrum version of Wyner's result [4], that will be mentioned in the next section.

Proof: Let $R' > \bar{I}(0 \,|\, \boldsymbol{p}, \boldsymbol{W}^E)$, $R < \underline{I}(1 \,|\, \boldsymbol{p}, \boldsymbol{W}^B) - R'$, and choose a real number a such that

$$0 < a < \min\{\underline{I}(1 \,|\, \boldsymbol{p}, \boldsymbol{W}^B) - (R+R'), R' - \bar{I}(0 \,|\, \boldsymbol{p}, \boldsymbol{W}^E)\}.$$

Substituting $M = e^{nR}, L = e^{nR'}, C = e^{n(R'-a)}, C' = e^{n(R+R'+a)}$, we can show that the RHS of (20) goes to 0, and that

$$\delta_{p^n, W^{E,n}, e^{n(R'-a)}} \to 0, \quad \frac{\delta'_{p^n, W^{E,n}, e^{n(R'-a)}}}{e^{nR'}} \to 0.$$

Hence, the RHS of (23) goes to 0. Concerning (21), the relations

$$\begin{aligned}
&\frac{1}{n} \left(\eta \left(\delta_{p^n, W^{E,n}, e^{n(R'-a)}} \right) + \delta_{p, W^{E,n}, e^{n(R'-a)}} \log |\mathcal{Z}^n| \right. \\
&\qquad \left. + \frac{\delta'_{p^n, W^{E,n}, e^{n(R'-a)}}}{e^{nR'}} \right) \\
&\quad = \frac{1}{n} \eta \left(\delta_{p^n, W^{E,n}, e^{n(R'-a)}} \right) + \delta_{p, W^E, e^{n(R'-a)}} \log d \\
&\qquad + \frac{1}{n} \frac{\delta'_{p^n, W^{E,n}, e^{n(R'-a)}}}{e^{nR'}} \\
&\to 0
\end{aligned}$$

hold. Therefore, we obtain (30) and (31). □

Conversely, we obtain the following lemma.

Lemma 5: Let $\boldsymbol{Q} = \{Q^n\}$ be a sequence of channels from arbitrary set $\tilde{\mathcal{X}}^n$ to the set $\mathcal{X}^n$ and $\boldsymbol{p} = \{p^n\}$ be a sequence of distributions on $\tilde{\mathcal{X}}^n$. Then, the inequalities

$$C_d(\boldsymbol{W}^B, \boldsymbol{W}^E) \le \sup_{\boldsymbol{p},\boldsymbol{Q}} \left\{ \underline{I}(1\,|\,\boldsymbol{p}, \boldsymbol{W}^B\boldsymbol{Q}) - \bar{I}(0\,|\,\boldsymbol{p}, \boldsymbol{W}^E\boldsymbol{Q}) \right\} \tag{32}$$

$$C_I(\boldsymbol{W}^B, \boldsymbol{W}^E) \le \sup_{\boldsymbol{p},\boldsymbol{Q}} \left\{ \underline{I}(1\,|\,\boldsymbol{p}, \boldsymbol{W}^B\boldsymbol{Q}) - \bar{I}(0\,|\,\boldsymbol{p}, \boldsymbol{W}^E\boldsymbol{Q}) \right\} \tag{33}$$

hold, where $\boldsymbol{W}\boldsymbol{Q} = \{W^nQ^n\}$ denotes the sequence of channels from $\tilde{\mathcal{X}}^n$ to $\mathcal{Y}^n$

$$(W^nQ^n)_{\tilde{x}}(y) \stackrel{\mathrm{def}}{=} \sum_{x \in \mathcal{X}^n} W_x^n(y) Q_{\tilde{x}}^n(x)$$

for a sequence of channels $\boldsymbol{W} = \{W^n\}$ from $\mathcal{X}^n$ to $\mathcal{Y}^n$.

Hence, applying Lemma 4 to the sequence of the channels $\boldsymbol{W}^B\boldsymbol{Q}, \boldsymbol{W}^E\boldsymbol{Q}$, we obtain the following theorem.

Theorem 5:

$$\begin{aligned} C_d(\boldsymbol{W}^B, \boldsymbol{W}^E) &= C_I(\boldsymbol{W}^B, \boldsymbol{W}^E) \\ &= \sup_{\boldsymbol{p},\boldsymbol{Q}} \left\{ \underline{I}(1\,|\,\boldsymbol{p}, \boldsymbol{W}^B\boldsymbol{Q}) - \bar{I}(0\,|\,\boldsymbol{p}, \boldsymbol{W}^E\boldsymbol{Q}) \right\}. \end{aligned}$$

Proof of Lemma 5: Let

$$\{\Phi_n = (M_n, \{Q_1^n, \ldots, Q_{M_n}^n\}, \{\mathcal{D}_1^n, \ldots, \mathcal{D}_{M_n}^n\})\}$$

be a sequence of codes of wiretap channel such that

$$R = \underline{\lim} \frac{1}{n} \log |\Phi_n|, \quad \lim \epsilon_B(\Phi_n) = 0, \quad \lim d_E(\Phi_n) = 0.$$

Hence, Verdú–Han's result [19] yields that the transmission capacity of the sequence of channel $\boldsymbol{W}^B\boldsymbol{Q}$ is less than $\underline{I}(1|\boldsymbol{p}, \boldsymbol{W}^B\boldsymbol{Q})$, which implies

$$R \le \underline{I}(1\,|\,\boldsymbol{p}, \boldsymbol{W}^B\boldsymbol{Q}).$$

Furthermore, the property $\lim d_E(\Phi_n) = 0$ implies that $S(0|\boldsymbol{W}^E\boldsymbol{Q}) = 0$. Hence, we have

$$\bar{I}(0\,|\,\boldsymbol{p}, \boldsymbol{W}^E\boldsymbol{Q}) = 0. \tag{34}$$

Thus, we obtain

$$R = \underline{I}(1\,|\,\boldsymbol{p}, \boldsymbol{W}^B\boldsymbol{Q}) - \bar{I}(0\,|\,\boldsymbol{p}, \boldsymbol{W}^E\boldsymbol{Q})$$

which implies (32).

Next, we assume that a sequence of codes of wiretap channel $\{\Phi_n = (M_n, \{Q_1^n, \ldots, Q_{M_n}^n\}, \{\mathcal{D}_1^n, \ldots, \mathcal{D}_{M_n}^n\})\}$ satisfies that

$$R = \underline{\lim} \frac{1}{n} \log |\Phi_n|, \quad \lim \epsilon_B(\Phi_n) = 0, \quad \lim \frac{I_E(\Phi_n)}{n} = 0.$$

Since the mutual information

$$I_E(\Phi_n) = \sum_{i=1}^{M_n} \frac{1}{M_n} \mathrm{E}_{(W^{E,n}Q^n)_i, y} \log \frac{(W^{E,n}Q^n)_i}{\sum_{i=1}^{M_n} \frac{1}{M_n} (W^{E,n}Q^n)_i}(y)$$

can be regarded as K-L divergence, Lemma 6 yields that

$$\sum_{i=1}^{M_n} \frac{1}{M_n} (W^{E,n}Q^n)_i \left\{ \frac{1}{n} \log \frac{(W^{E,n}Q^n)_i}{\sum_{i=1}^{M_n} \frac{1}{M_n} (W^{E,n}Q^n)_i}(y) \ge a \right\} \le \frac{I_E(\Phi_n) + \frac{1}{e}}{na} \to 0$$

for any $a > 0$. Thus, we obtain (34). Therefore, similarly to (32), we obtain (33). □

Lemma 6: Assume that p and q are two probability distributions on Ω. Then, we have

$$D(p\,\|\,q) + \frac{1}{e} \ge \alpha \cdot p\left\{ \log \frac{p(\omega)}{q(\omega)} \ge \alpha \right\}. \tag{35}$$

Proof: We focus on the two probability distributions on $\Omega_0 \stackrel{\mathrm{def}}{=} \{\log \frac{p}{q}(\omega) < \alpha\}$

$$p_0(\omega) \stackrel{\mathrm{def}}{=} \frac{p(\omega)}{p\{\Omega_0\}}, \qquad q_0(\omega) \stackrel{\mathrm{def}}{=} \frac{q(\omega)}{q\{\Omega_0\}}.$$

Hence,

$$\begin{aligned} D(p\,\|\,q) &= \sum_{\omega \in \Omega_0^c} p(\omega) \log \frac{p(\omega)}{q(\omega)} + \sum_{\omega \in \Omega} p(\omega) \log \frac{p(\omega)}{q(\omega)} \\ &\ge \alpha p\{\Omega_0^c\} + \sum_{\omega \in \Omega} p_0(\omega) \left(\log \frac{p\{\Omega_0\}}{q\{\Omega_0\}} + \log \frac{p_0(\omega)}{q_0(\omega)} \right) \\ &= \alpha p\{\Omega_0^c\} + p\{\Omega_0\} \log \frac{p\{\Omega_0\}}{q\{\Omega_0\}} + D(p_0\,\|\,q_0) \\ &\ge \alpha p\{\Omega_0^c\} + p\{\Omega_0\} \log \frac{p\{\Omega_0\}}{q\{\Omega_0\}} \\ &\ge \alpha p\{\Omega_0^c\} + p\{\Omega_0\} \log p\{\Omega_0\}. \end{aligned}$$

Finally, the convexity of the map $x \mapsto x \log x$ guarantees that $p\{\Omega_0\} \log p\{\Omega_0\} \ge -\frac{1}{e}$. We obtain (35).

VI. Exponents in Stationary Memoryless Channel

A. Channel Resolvability

Next, we proceed to the stationary memoryless channel of a given channel W as a special case.

First, we treat channel resolvability. As was shown by Han and Verdú [2] and Han [3], the information spectrum quantities of discrete memoryless channel of W is calculated as

$$\sup_{\boldsymbol{p}} \bar{I}(\epsilon\,|\,\boldsymbol{p}, \boldsymbol{W}) = \sup_{\boldsymbol{p}} \underline{I}(\epsilon\,|\,\boldsymbol{p}, \boldsymbol{W}) = \max_p I(p; W)$$

for $1 \ge \epsilon \ge 0$, where

$$I(p; W) \stackrel{\mathrm{def}}{=} \mathrm{E}_p D(W_x\,\|\,W_p).$$

Hence, Theorem 4 yields

$$S(\epsilon\,|\,\boldsymbol{W}) = S^\dagger(\epsilon\,|\,\boldsymbol{W}) = \max_p I(p; W)$$

which has been obtained by Han and Verdú [2]. Furthermore, using Theorem 2, we can discuss these problems in more detail by treating the following optimal exponents:

$$e_\epsilon(R|W,p) \stackrel{\rm def}{=} \sup_{\{\Psi_n\}} \left\{ \underline{\lim} \frac{-1}{n} \log \epsilon(\Psi_n, W_{p^n}^n) \middle| \overline{\lim} \frac{1}{n} \log |\Psi_n| \le R \right\}$$

$$e_D(R\,|\,W,p) \stackrel{\rm def}{=} \sup_{\{\Psi_n\}} \left\{ \underline{\lim} \frac{-1}{n} \log D(\Psi_n, W_{p^n}^n) \middle| \overline{\lim} \frac{1}{n} \log |\Psi_n| \le R \right\}$$

and

$$e_\epsilon(R\,|\,W) \stackrel{\rm def}{=} \underline{\lim} \frac{-1}{n} \log \epsilon(e^{nR}, W^n)$$

$$e_D(R\,|\,W) \stackrel{\rm def}{=} \underline{\lim} \frac{-1}{n} \log D(e^{nR}, W^n)$$

where p^n is the n-fold identical independent distribution of p. As is discussed by Oohama [8], by using Lemma 3, the exponent $e_\epsilon(R,W)$ gives a lower bound of strong converse exponent of identification code.

Theorem 6: Assume that the cardinality $|\mathcal{Y}|$ is finite, then

$$e_\epsilon(R|W,p) \ge \max_{1\ge s\ge 0} \left\{ \frac{-\psi(s\,|\,W,p) + sR}{1+s} \right\} \tag{36}$$

$$e_D(R|W,p) \ge \max_{0\ge t\ge -1/2} \{-\phi(t\,|\,W,p) - tR\} \tag{37}$$

$$e_\epsilon(R|W) \ge \max_{1\ge s\ge 0} \left\{ \frac{-\psi(s\,|\,W) + sR}{1+s} \right\} \tag{38}$$

$$e_D(R|W) \ge \max_{0\ge t\ge -1/2} \{-\max_p \phi(t\,|\,W,p) - tR\} \tag{39}$$

where

$$\psi(s\,|\,W,p) \stackrel{\rm def}{=} \log \mathrm{E}_p \sum_y W_x^{1+s}(y) W_p^{-s}(y)$$

and

$$\psi(s\,|\,W) \stackrel{\rm def}{=} \log \max_p \sum_y \left(\mathrm{E}_p W_x^{1+s}(y)\right)^{1-s}.$$

Using Pinsker's inequality $D(p\|q) \ge \|p-q\|^2$, we obtain two inequalities

$$\frac{1}{2} e_D(R\,|\,W,p) \le e_\epsilon(R\,|\,W,p) \text{ and } \frac{1}{2} e_D(R\,|\,W) \le e_\epsilon(R\,|\,W)$$

which implies different lower bounds of exponents

$$e_\epsilon(R|W,p) \ge \frac{1}{2} \max_{0\ge t\ge -1/2} \{-\phi(t\,|\,W,p) - tR\} \tag{40}$$

$$e_\epsilon(R|W) \ge \frac{1}{2} \max_{0\ge t\ge -1/2} \left\{ -\max_p \phi(t\,|\,W,p) - tR \right\}. \tag{41}$$

We can derive different lower bounds of $e_D(R\,|\,W,p)$ and $e_D(R\,|\,W)$ from the inequality (11). However, these bounds are smaller than the bound presented here.

Remark 3: Arimoto's strong converse exponent [15] of channel coding of transmission code equals

$$\max_{0\ge t\ge -1} \{-\max_p \phi(t\,|\,W,p) - tR\}$$

which is a bit greater than the RHS of (37) when R is sufficiently large.

Remark 4: By using inequality (9) and type method, Oohama [8] has obtained a lower bound of $e_\epsilon(R\,|\,W)$

$$\frac{1}{2} \max_{0\ge t\ge -1} \left\{ -\max_p \phi(t\,|\,W,p) - tR \right\}$$

which is a bit better than (41) when R is sufficiently large. It is interesting that his approach is in contrast to our approach to (41), which is based on (10) not on (9).

Remark 5: It is difficult to treat the exponent of the sum of two error probabilities in identification code based on Theorem 1. For this purpose, we need a modified version of Theorem 1.

The following lemma is a preparation of our proof of Theorem 6.

Lemma 7: For any $s \ge 0$ and $0 \ge t > -1$, the equalities

$$\max_{p\in\mathcal{P}(\mathcal{X}^n)} \sum_{y^n\in\mathcal{Y}^n} \left(\mathrm{E}_p \left(W_x^n(y^n)\right)^{1+s}\right)^{1-s} = \left(\max_{p\in\mathcal{P}(\mathcal{X})} \sum_y \left(\mathrm{E}_p W_x^{1+s}(y)\right)^{1-s} \right)^n \tag{42}$$

$$\max_{p\in\mathcal{P}(\mathcal{X}^n)} \sum_{y^n\in\mathcal{Y}^n} \left(\mathrm{E}_p \left(W_x^n(y^n)\right)^{\frac{1}{1+t}}\right)^{1+t} = \left(\max_{p\in\mathcal{P}(\mathcal{X})} \sum_y \left(\mathrm{E}_p W_x^{\frac{1}{1+t}}(y)\right)^{1+t} \right)^n \tag{43}$$

hold.

Proof: Since (43) has been shown by Arimoto [15], we prove only (42) by the same method. Since the function

$$f : p \mapsto \max_{p\in\mathcal{P}(\mathcal{X})} \sum_y (\mathrm{E}_p W_x^{1+s}(y))^{1-s}$$

is continuous and convex function, if and only if $f(p^*) = \max_p f(p)$, there exists a constant λ such that

$$\sum_y W_x^{1+s}(y) \left(\sum_x p^*(x) W_x^{1+s}(y) \right)^{-s} = \frac{\partial f}{\partial p(x)} \begin{cases} = \lambda, & \text{if } p^*(x) > 0, \\ \le \lambda, & \text{if } p^*(x) = 0. \end{cases}$$

Indeed, λ is calculated as

$$\sum_x p(x)\lambda = \sum_x p(x) \sum_y W_x^{1+s}(y) \left(\sum_x p^*(x) W_x^{1+s}(y) \right)^{-s} = \left(\sum_x p^*(x) W_x^{1+s}(y) \right)^{1-s}.$$

Thus, if and only if $f(p^*) = \max_p f(p)$,

$$\sum_y W_x^{1+s}(y) \left(\sum_x p^*(x) W_x^{1+s}(y) \right)^{-s} \begin{cases} = \left(\sum_x p^*(x) W_x^{1+s}(y)\right)^{1-s}, & \text{if } p^*(x) > 0 \\ \le \left(\sum_x p^*(x) W_x^{1+s}(y)\right)^{1-s}, & \text{if } p^*(x) = 0, \end{cases}$$

p^* gives the maximum. Hence, if p^* satisfies the above condition, $(p^*)^n$ also satisfies the following condition shown at the

$$\sum_{y^n}(W^n_{x^n})^{1+s}(y^n)\left(\sum_{x^n}(p^*)^n(x^n)\,(W^n_{x^n})^{1+s}\,(y^n)\right)^{-s}\begin{cases}=\left(\sum_{x^n}(p^*)^n(x^n)(W^n_{x^n})^{1+s}(y^n)\right)^{1-s}, & \text{if } (p^*)^n(x^n)>0\\ \le\left(\sum_{x^n}(p^*)^n(x^n)\,(W^n_{x^n})^{1+s}\,(y^n)\right)^{1-s}, & \text{if } (p^*)^n(x^n)=0\end{cases}$$

top of the page, which is a necessary and sufficient condition for

$$\sum_{y^n\in\mathcal{Y}^n}\left(\mathrm{E}_{(p^*)^n}\,(W^n_x(y^n))^{1+s}\right)^{1-s} = \max_{p\in\mathcal{P}(\mathcal{X}^n)}\sum_{y^n\in\mathcal{Y}^n}\left(\mathrm{E}_p\,(W^n_x(y^n))^{1+s}\right)^{1-s}.$$

It implies(42). □

Proof of Theorem 6: By inequality (10) of Theorem 2, we have

$$\begin{aligned}D(e^{nR},W_{\cdot p^n}) &\le \frac{\log(1+(e^{nR})^t e^{\phi(t\,|\,W^n,p^n)})}{-t}\\ &\le \frac{(e^{nR})^t e^{\phi(t\,|\,W^n,p^n)}}{-t}\\ &= \frac{e^{n(\phi(t\,|\,W,p)+tR)}}{-t}\end{aligned}\tag{44}$$

for $0>t\ge -1/2$, where the second inequality follows from $\log(1+x)\le x$. From (44), we obtain

$$e_D(R\,|\,W,p)\ge -\phi(t\,|\,W,p)-tR\tag{45}$$

for $0>t\ge -1/2$. Since $\phi(t\,|\,W,p)+tR$ is continuous for t, the inequality (37) holds. By inequality (10) and Lemma 7, we have

$$\begin{aligned}D(e^{nR},W^n) &\le \max_{p\in\mathcal{P}(\mathcal{X}^n)}\frac{\log(1+(e^{nR})^t e^{\phi(t\,|\,W^n,p)})}{-t}\\ &\le \frac{(e^{nR})^t\cdot\max_{p\in\mathcal{P}(\mathcal{X}^n)}e^{\phi(t\,|\,W^n,p)}}{-t}\\ &= \frac{(e^{nR})^t\cdot\max_{p\in\mathcal{P}(\mathcal{X})}e^{n\phi(t\,|\,W,p)}}{-t}\end{aligned}\tag{46}$$

for $0>t\ge -1/2$. Hence, in a manner similar to the derivation of (37) from (44), we obtain (39) from (46).

Next, we derive (36) and (38). To this end, we first derive an upper bound of

$$2\delta_{p,W,e^{R'}}+\sqrt{\frac{\delta'_{p,W,e^{R'}}}{e^R}}.$$

For any $1\ge s\ge 0$, we choose $R'\stackrel{\mathrm{def}}{=}\frac{\psi(s\,|\,W,p)+R}{1+s}$. By using Markov inequality, we can evaluate $\delta_{p,W,e^{R'}}$ and $\delta'_{p,W,e^{R'}}$ as

$$\begin{aligned}\delta_{p,W,e^{R'}} &\le \mathrm{E}_p\sum_{y\in\left\{\frac{W_x}{W_p}(y)>e^{R'}\right\}}W_x(y)\left(\frac{e^{-R'}W_x(y)}{W_p(y)}\right)^s\\ &\le \mathrm{E}_p\sum_y W_x(y)\left(\frac{W_x(y)}{W_p(y)}\right)^s e^{-sR'}\\ &= e^{\psi(s\,|\,W,p)-sR'}=e^{\frac{\psi(s\,|\,W,p)-sR}{1+s}}\end{aligned}\tag{47}$$

and

$$\begin{aligned}\delta'_{p,W,e^{R'}} &\le \mathrm{E}_p\sum_{y\in\left\{\frac{W_x}{W_p}(y)\le e^{R'}\right\}}\frac{W_x(y)^2}{W_p(y)}\left(\frac{W_p(y)}{e^{-R'}W_x(y)}\right)^{1-s}\\ &\le \mathrm{E}_p\sum_y W_x(y)\left(\frac{W_x(y)}{W_p(y)}\right)^s e^{(1-s)R'}\\ &= e^{\psi(s\,|\,W,p)+(1-s)R'}\end{aligned}\tag{48}$$

respectively. Inequality (48) yields

$$\sqrt{\frac{\delta'_{p,W,e^{R'}}}{e^R}}\le e^{\frac{\psi(s\,|\,W,p)+(1-s)R'-R}{2}}=e^{\frac{\psi(s\,|\,W,p)-sR}{1+s}}.\tag{49}$$

Combining (47) and (49), we have

$$2\delta_{p,W,e^{R'}}+\sqrt{\frac{\delta'_{p,W,e^{R'}}}{e^R}}\le 3e^{\frac{\psi(s\,|\,W,p)-sR}{1+s}}.\tag{50}$$

Hence, (50) and (9) in Theorem 2 guarantee that

$$\epsilon\left(e^{nR},W^n_{p^n}\right)\le 3e^{\frac{\psi(s\,|\,W^n,p^n)-snR}{1+s}}=3e^{n\frac{\psi(s\,|\,W,p)-sR}{1+s}}\tag{51}$$

for $1\ge s\ge 0$ because $\psi(s\,|\,W^n,p^n)=n\psi(s\,|\,W,p)$. Thus, (51) implies that

$$e_\epsilon(R|W,p)\ge\frac{-\psi(s\,|\,W,p)+sR}{1+s}\tag{52}$$

for $1\ge s\ge 0$. Taking the maximum for $1\ge s\ge 0$, we obtain (36).

We proceed to the proof of (38). By inequalities (9) and (50), we obtain

$$\begin{aligned}\epsilon(e^{nR},W^n) &\le \max_{p\in\mathcal{P}(\mathcal{X}^n)}3e^{\frac{\psi(s\,|\,W^n,p)-snR}{1+s}}\\ &= 3e^{\frac{-snR}{1+s}}\left[\max_{p\in\mathcal{P}(\mathcal{X}^n)}e^{\frac{\psi(s\,|\,W^n,p)}{1+s}}\right]^{\frac{1}{1+s}}.\end{aligned}\tag{53}$$

Now, we estimate an upper bound of

$$e^{\psi(s\,|\,W^n,p)}=\mathrm{E}_p\sum_y W_x^{1+s}(y)W_p^{-s}(y).\tag{54}$$

Since the map $x\mapsto x^{1+s}$ is convex, we have

$$W_p(y)=\mathrm{E}_pW_x(y)\ge\mathrm{E}_pW_x^{1+s}(y)$$

which imply that

$$W_p^{-s}(y)\le(\mathrm{E}_p(W_x(y))^{1+s})^{-s}.$$

Hence, the relations

$$\begin{aligned}\mathrm{E}_p\sum_y W_x^{1+s}(y)W_p^{-s}(y)&=\sum_y\mathrm{E}_pW_x^{1+s}(y)W_p^{-s}(y)\\ &\le\sum_y\left(\mathrm{E}_pW_x^{1+s}(y)\right)^{1-s}\end{aligned}\tag{55}$$

hold. Using (55) and Lemma 7, we can evaluate

$$\begin{aligned}&\max_{p\in\mathcal{P}(\mathcal{X}^n)} \mathrm{E}_p \sum_{y\in\mathcal{Y}^n} W_x^n(y)^{1+s} W_p^n(y)^{-s} \\ &\le \max_{p\in\mathcal{P}(\mathcal{X}^n)} \sum_{y\in\mathcal{Y}^n} \left(\mathrm{E}_p \left(W_x^n(y)\right)^{1+s}\right)^{1-s} \\ &= \left(\max_{p\in\mathcal{P}(\mathcal{X})} \sum_{y} \left(\mathrm{E}_p W_x^{1+s}(y)\right)^{1-s}\right)^n = e^{n(\psi(s\,|\,W))}.\end{aligned} \tag{56}$$

Combining (53) and (56), we have

$$\epsilon(e^{nR}, W^n) \le e^{n\frac{\psi(s\,|\,W)-sR}{1+s}} \tag{57}$$

for any $1 \ge s \ge 0$. In a manner similar to the derivation of (36) from (51), we can derive (38) from (57). □

B. Wiretap Channel

Next, we proceed to discrete memoryless wiretap channel. Applying Theorem 4 to this case with the input i.i.d., we obtain

$$\begin{aligned}&C(\boldsymbol{W}^B, \boldsymbol{W}^E) \\ &\stackrel{\mathrm{def}}{=} \sup_{\{\Phi_n\}} \left\{ \underline{\lim} \frac{1}{n} \log\log|\Phi_n| \,\middle|\, \lim \epsilon_B(\Phi_n) = \lim d_E(\Phi_n) = 0 \right\} \\ &\ge \sup_p \{I(p; W^B) - I(p; W^E)\}\end{aligned}$$

which has been obtained by Wyner [4]. Hence, Theorem 4 can be regarded as a general extension of Wyner's result. Moreover, using Lemma 3, we derived several explicit lower bounds of exponents.

Theorem 7: Assume that the cardinality $|\mathcal{Z}|$ is finite, then there exists a sequence $\{\Phi_n\}$ of codes for any real numbers R, R', and any probability distribution p such that

$$\lim \frac{1}{n} \log|\Phi_n| = R$$

$$\underline{\lim} \frac{-1}{n} \log \epsilon_B(\Phi_n) \ge \max_{1\ge s\ge 0} \{-\phi(s\,|\,W^B, p) - s(R+R')\} \tag{58}$$

$$\underline{\lim} \frac{-1}{n} \log I_E(\Phi_n) \ge \max_{0\ge t\ge -1/2} \{-\phi(t\,|\,W^E, p) - tR'\} \tag{59}$$

$$\underline{\lim} \frac{-1}{n} \log d_E(\Phi_n) \ge \max_{1\ge s\ge 0} \left\{ \frac{-\psi(s\,|\,W^E, p) + sR'}{1+s} \right\} \tag{60}$$

$$\underline{\lim} \frac{-1}{n} \log d_E(\Phi_n) \ge \frac{1}{2} \max_{0\ge t\ge -1/2} \left\{-\phi(t\,|\,W^E, p) - tR'\right\}. \tag{61}$$

Indeed, these exponents are very useful for evaluating error and wiretapper's information for a finite n.

Proof: The inequality (58) immediately follows from (19). By using an evaluation similar to (37), we can show (59) from (21). Furthermore, by using an evaluation similar to (36), we can show (60) from (23). □

VII. Comparison of Lower Bounds of Exponents

Finally, we compare the lower bounds (36), (38), (40), and (41) of error exponents of channel resolvability.

Theorem 8: Assume that $\Delta \stackrel{\mathrm{def}}{=} R - I(p; W)$ is sufficiently small. Then, RHSs of (36) and (40) (which are lower bounds of exponent of the variational distance) are approximately calculated as

$$\text{RHS of (36)} \quad \max_{1\ge s\ge 0} \left\{ \frac{-\psi(s\,|\,W, p) + sR}{1+s} \right\} \cong \frac{\Delta^2}{4J(p; W)}$$

$$\text{RHS of (40)} \quad \frac{1}{2} \max_{0\ge t\ge -1/2} \{-\phi(t\,|\,W, p) - tR\} \cong \frac{\Delta^2}{8J(p; W)}$$

where

$$J(p; W) \stackrel{\mathrm{def}}{=} \frac{1}{2}(\mathrm{E}_{p,x}\mathrm{E}_{W_x,y}(\log W_x(y) - \log W_p(y))^2 - I^2(p; W)).$$

Moreover, RHSs of (38) and (41) (which are lower bounds of exponent of the worst variational distance) are approximately calculated as

$$\begin{aligned}\text{RHS of (38)} \quad &\max_{s\ge 0} \left\{ \frac{-\psi(s\,|\,W) + sR}{1+s} \right\} \\ &\cong \frac{\Delta^2}{4(J(p_0; W) + \mathrm{E}_{p_0} H(W_x))} \\ \text{RHS of (41)} \quad &\frac{1}{2} \max_{0\ge t\ge -1/2} \left\{ -\max_p \phi(t\,|\,W, p) - tR \right\} \\ &\cong \frac{\Delta^2}{8J(p_0; W)},\end{aligned}$$

where $p_0 \stackrel{\mathrm{def}}{=} \mathrm{argmax}_p I(p; W)$.

Thus, when R is sufficiently close to $\max_p I(p; W)$, (36) gives a better lower bound than (40). Of course, this comparison can be applied to exponents of eavesdropper's information in wiretap channel, i.e., the comparison of RHSs of (60) and (61). On the other hand, (38) gives a better lower bound than (41), if and only if

$$\begin{aligned}&\mathrm{E}_{p_0} H(W_x) \\ &\quad \le \frac{1}{2}(\mathrm{E}_{p,x}\mathrm{E}_{W_x,y}(\log W_x(y) - \log W_p(y))^2 - I^2(p; W)).\end{aligned}$$

Therefore, although $R - \max_p I(p; W)$ is small enough, the relation between bounds (38) and (41) is not clear.

Proof: By using a Taylor expansion, we obtain the approximations

$$\begin{aligned}\psi(s\,|\,W, p) &\cong I(p; W)s + J(p; W)s^2 \\ \phi(t\,|\,W, p) &\cong -I(p; W)t + J(p; W)t^2 \\ \psi(s\,|\,W) &\cong I(p_0; W)s + (J(p_0; W) + \mathrm{E}_{p_0} H(W_x))s^2.\end{aligned}$$

Thus,

$$\begin{aligned}
&\max_{1\ge s\ge 0}\left\{\frac{-\psi(s\,|\,W,p)+sR}{1+s}\right\}\\
&\quad\cong \max_{1\ge s\ge 0}\left\{\frac{-I(p;W)s-J(p;W)s^2+(I(p;W)+\Delta)s}{1+s}\right\}\\
&\quad\cong \max_{1\ge s\ge 0}\left\{-J(p;W)s^2+\Delta s\right\}\cong\frac{\Delta^2}{4J(p;W)}\\
&\max_{0\ge t\ge -1/2}\{-\phi(t\,|\,W,p)-tR\}\\
&\quad\cong \max_{0\ge t\ge -1/2}\{I(p;W)t-J(p;W)t^2-(I(p;W)+\Delta)t\}\\
&\quad= \max_{0\ge t\ge -1/2}\left\{-J(p;W)t^2-\Delta t\right\}=\frac{\Delta^2}{4J(p;W)}\\
&\max_{s\ge 0}\left\{\frac{-\psi(s\,|\,W)+sR}{1+s}\right\}\\
&\quad\cong \max_{s\ge 0}\left\{\frac{-I(p_0;W)s-(J(p_0;W)+\mathrm{E}_{p_0}H(W_x))s^2}{1+s}\right.\\
&\qquad\qquad\left.+\frac{s(I(p_0;W)+\Delta)}{1+s}\right\}\\
&\quad\cong \max_{s\ge 0}\{-(J(p_0;W)+\mathrm{E}_{p_0}H(W_x))s^2+\Delta s\}\\
&\quad=\frac{\Delta^2}{4(J(p_0;W)+\mathrm{E}_{p_0}H(W_x))}.
\end{aligned}$$

□

VIII. Conclusion

We give several nonasymptotic formulas in identification code, channel resolvability, and wiretap channel. Using these formulas, we give the achievable rate channel resolvability for the general channel, which had been an open problem. Also, we derived several asymptotic relations among divergence rates, capacities of identification code, and ϵ capacities of channel resolvability.

From these nonasymptotic formulas, we obtained lower bounds of error exponents of channel resolvability in the stationary memoryless setting. Moreover, we derived lower bounds of error probability and wiretapper's information in the stationary memoryless setting in wiretap channel.

Concerning the quantum setting, wiretap channel has been discussed in the discrete memoryless channel case by Devetak [5], Winter *et al.* [6] and Cai and Yeung [17], and identification codes has been discussed by Ahlswede and Winter [18]. Hence, several quantum extensions of the results presented here can be expected. Some has been obtained by the author. And some of them have appeared in the author's textbook [13]. Those not already presented will appear in a forthcoming paper.

Acknowledgment

The author would like to thank Prof. Hiroshi Imai of the QCI project for support. He is grateful to Prof. Yasutada Oohama for useful discussions. He is also grateful to reviewers for their kind and useful comments.

References

[1] R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 15–29, Jan. 1989.

[2] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 752–772, May 1993.

[3] T. S. Han, *Information-Spectrum Methods in Information Theory*. Berlin, Germany: Springer-Verlag, 2003. (The original Japanese edition was published by Baifukan-Press, Tokyo, Japan, in 1998.).

[4] A. D. Wyner, "The wiretap channel," *Bell. Syst. Tech. J.*, vol. 54, pp. 1355–1387, 1975.

[5] I. Devetak, "The private classical information capacity and quantum information capacity of a quantum channel," quant-ph/0304127, 2003.

[6] A. Winter, A. C. A. Nascimento, and H. Imai, "Commitment Capacity of Discrete Memoryless Channels," e-print cs.CR/0304014, 2003.

[7] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[8] Y. Oohama, "Error probability of identification via channels at rates above capacity," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 26.

[9] ——, "Average error probability of identification via channels at rates above identification capacity," in *Proc. IEEE Int. Symp. Information Theory and Its Applications*, Xi'an, China, 2002, pp. 859–862.

[10] M. Hayashi and H. Nagaoka, "General formulas for capacity of classical-quantum channels," *IEEE Trans. Inf. Theor.*, vol. 49, no. 7, pp. 1753–1768, Jul. 2003.

[11] H. Nagaoka and M. Hayashi, "An Information-Spectrum Approach to Classical and Quantum Hypothesis Testing," quant-ph/0206185, 2002.

[12] T. Ogawa and H. Nagaoka, "A new proof of the channel coding theorem via hypothesis testing in quantum information theory," in *Proc. 2002 IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 73.

[13] M. Hayashi, *Introduction to Quantum Information Theory* (in Japanese). Tokyo, Japan: Saiensu-sha, 2004. English version will be published by Springer-Verlag in Mar. 2006.

[14] Y. Steinberg, "New converses in the theory of identification via chanels," *IEEE Trans. Inf. Theor.*, vol. 44, no. 3, pp. 984–998, May 1998.

[15] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inf. Theor.*, vol. IT-19, no. 3, pp. 357–359, May 1973.

[16] I. Csiszár and P. Narayan, "Common randomness and secret key generation with a helper," *IEEE Trans. Inf. Theor.*, vol. 46, no. 2, pp. 344–366, Mar. 2000.

[17] N. Cai and R. Yeung, "Quantum Privacy and Quantum Wiretap Channels," unpublished manuscript, 2003.

[18] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels," *IEEE Trans. Inf. Theor.*, vol. 48, no. 3, pp. 569–579, Mar. 2002.

[19] S. Verdú and T. S. Han, "A general formula for channel capacity," *IEEE Trans. Inf. Theory*, vol. 40, no. 4, pp. 1147–1157, Jul. 1994.