On Identification via Channels

Yasutada Oohama
Faculty of Information Science and Electrical Engineering, Kyushu University
E-mail: oohama@csce.kyushu-u.ac.jp

Abstract- In identification via channels the sum of two types of error probabilities of identification goes to one as the block length of transmission tends to infinity at rates above capacity when channels satisfy some stochastic properties. This is well known as a strong converse theorem for the identification via channels. In this paper we prove that the sum of two error probabilities tends to one exponentially and derive an explicit lower bound of this exponent function.

1 Introduction

In 1989, Ahlswede and Dueck [1] proposed a new framework of communication system using noisy channels. Their proposed framework called the identification via channels (or briefly say the ID channel) has opened a new and fertile area in the Shannon theory.

In some class of noisy channels the sum of two types of error probabilities of identification goes to one as the block length n of transmitted codes tends to infinity at rates above capacity. This is well known as a strong converse theorem for the ID channel. Han and Verdú [2] established the strong converse theorem for the ID channel in the case of stationary discrete memoryless channels (DMC). An extension of the above result to more general class of noisy channels was studied by Han and Verdú[3]. They introduced a new coding problem of approximation of output random variables through noisy channels. They call it the channel resolvability problem. They have established the direct and converse coding theorem for the channel resolvability problem. Furthermore, they derived an upper bound of the capacity for the ID channel by using some interesting relation between the direct coding theorem of the channel resolvability problem and the converse coding theorem of the ID channel. Using the fact that the upper bound coincides with the lower bound in the class of noisy channels having what they call the strong converse property, they determined the capacity and established the strong converse theorem for the ID channel in this class of noisy channels. The results of Han and Verdú [2] was sharpened by Steinberg [4]. He introduced a new notion called *partial resolvability.* Based on this notion he formulated a new resolvability problem, which is an extension of that posed by Han and Verdú[3]. By investigating this problem, he determined the capacity of the ID channel for general noisy channels with finite input and output alphabets.

In this paper we deal with the ID channel for general noisy channels. For transmission rates above capacity we derive some function which serves as a lower bound of the sum of two error probabilities. In particular, in the case of the stationary DMC, we show that the sum of two error probabilities tends to one exponentially as n goes to infinity at transmission rates above capacity, deriving an explicit form of the lower bound of this exponent. For the derivation of the result, we consider the channel resolvability problem formulated by Steinberg [4]. We first establish a stronger result on the direct coding theorem for this problem by deriving an exponential lower bound for the approximation error of channel outputs to tend to zero as n goes to infinity. Next, we derive the converse coding theorem for the ID channel based on an idea of converting the direct coding theorem for the channel resolvability problem into the converse coding theorem of the ID channel. This idea is similar to that of Han and Verdú [3] and Steinberg [4] used for deriving a relation between the converse coding theorem for the ID channel and the direct coding theorem for the channel resolvability problem.

2 Identification via Channels

Let $\mathcal{X}$ and $\mathcal{Y}$ be finite sets. Let $\mathcal{P}(\mathcal{X}^n)$ and $\mathcal{P}(\mathcal{Y}^n)$ be sets of probability distributions on $\mathcal{X}^n$ and $\mathcal{Y}^n$, respectively. A source $\mathbf{X}$ with alphabet $\mathcal{X}$ is the sequence $\{P_{X^n} : P_{X^n} \in \mathcal{P}(\mathcal{X}^n)\}_{n=1}^{\infty}$. Similarly, a noisy channel $\mathbf{W}$ with input alphabet $\mathcal{X}$ and output alphabet $\mathcal{Y}$ is a sequence of conditional distributions $\{W^n(\cdot|\cdot)\}_{n=1}^{\infty}$, where $W^n(\cdot|\cdot) = \{ W^n(\cdot|\mathbf{x}) \in \mathcal{P}(\mathcal{Y}^n) \}_{\mathbf{x}\in\mathcal{X}^n}$.

Next, for $P \in \mathcal{P}(\mathcal{X}^n)$ and $\mathbf{y} \in \mathcal{Y}^n$, set

$$PW^n(\mathbf{y}) = \sum_{\mathbf{x}\in\mathcal{X}^n} P(\mathbf{x})W^n(\mathbf{y}|\mathbf{x}), \tag{1}$$

which becomes a probability distribution on $\mathcal{Y}^n$. We denote it by $PW^n = \{PW^n(\mathbf{y}) \}_{\mathbf{y}\in\mathcal{Y}^n}$. Set $Q = PW^n$ and call Q *the response of P through noisy channel W^n (or briefly the response of P)* .

An $(n, N_n, \mu_n, \lambda_n)$ ID code for W^n is a collection $\{(P_i,\, \mathcal{D}_i),\, i = 1, 2, \cdots, N_n\}$ such that

1) $P_i \in \mathcal{P}(\mathcal{X}^n)\,, \mathcal{D}_i \subseteq \mathcal{Y}^n\,,$
2) Q_i is the response of $P_i\,,$
3) $\mu_n^{(i)} = Q_i(\mathcal{D}_i^c)\,, \mu_n = \max_{1\le i\le N_n} \mu_n^{(i)}\,,$
4) $\lambda_n^{(i)} = \max_{j\ne i} Q_j(\mathcal{D}_i)\,, \lambda_n = \max_{1\le i\le N_n} \lambda_n^{(i)}\,.$

The rate of an $(n, N_n, \mu_n, \lambda_n)$ ID code is defined by

$$r_n = \frac{1}{n}\log\log N_n\,. \tag{2}$$

Definition 1 A rate R is said to be (μ, λ)-achievable ID rate if there exists an $(n, N_n, \mu_n, \lambda_n)$ code such that

$$\left.\begin{aligned} &\limsup_{n\to\infty} \mu_n \le \mu\,, \\ &\limsup_{n\to\infty} \lambda_n \le \lambda\,, \\ &\liminf_{n\to\infty} \frac{1}{n}\log\log N_n \ge R\,. \end{aligned}\right\} \tag{3}$$

The supremum of the (μ, λ)-achievable ID rates for $\mathbf{W}$ is denoted by $C_{\mathrm{ID}}(\mu, \lambda|\mathbf{W})$, which we call the (μ, λ)-ID capacity.

To state results for the identification capacity, we prepare several quantities which are defined based on the notion of the *information spectrum* introduced by Han and Verdú [3].

Definition 2 For $n = 1, 2, \cdots$, let X^n be an arbitrary prescribed random variable taking values in $\mathcal{X}^n$. The probability mass function of X^n is $P_{X^n}(\mathbf{x})$, $\mathbf{x} \in \mathcal{X}^n$. Let $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$ denotes a sequence of those random variables. Let $\mathbf{Y} = \{Y^n\}_{n=1}^{\infty}$ be a sequence of output random variables when we use $\mathbf{X}$ as a channel input of the noisy channel $\mathbf{W}$. In this case the joint probability mass function of (X^n, Y^n) denoted by $P_{X^nY^n}(\mathbf{x}, \mathbf{y})$, $(\mathbf{x}, \mathbf{y}) \in \mathcal{X}^n \times \mathcal{Y}^n$ is equal to $P_{X^n}(\mathbf{x}) W^n(\mathbf{y}|\mathbf{x})$.

Definition 3 Given a joint distribution $P_{X^nY^n}(\mathbf{x}, \mathbf{y}) = P_{X^n}(\mathbf{x})\, W^n(\mathbf{y}|\mathbf{x})$, the information density is the function defined on $\mathcal{X}^n \times \mathcal{Y}^n$:

$$i_{X^nY^n}(\mathbf{x}; \mathbf{y}) = \log \frac{W^n(\mathbf{y}|\mathbf{x})}{P_{Y^n}(\mathbf{y})}. \tag{4}$$

Let $\{Z_n\}_{n=1}^{\infty}$ be a sequence of arbitrary real-valued random variables. We introduce the notion of the so-called *probabilistic limsup/inf* in the following.

Definition 4 (The limsup/inf in probability)

$$\text{p-}\limsup_{n\to\infty} Z_n \triangleq \inf\{\alpha : \lim_{n\to\infty} \Pr\{Z_n \geq \alpha\} = 0\}, \tag{5}$$

$$\text{p-}\liminf_{n\to\infty} Z_n \triangleq \sup\{\alpha : \lim_{n\to\infty} \Pr\{Z_n \leq \alpha\} = 0\}. \tag{6}$$

The *probabilistic limsup/inf* in the above definitions is considered as an extension of ordinary (deterministic) limsup/inf. The operation of liminf/sup has the same properties as those of the operation of liminf/sup. For the details see Han and Verdú [3] and Han [5]. Set

$$\overline{I}(\mathbf{X}; \mathbf{Y}) \triangleq \text{p-}\limsup \frac{1}{n} i_{X^nY^n}(X^n; Y^n), \tag{7}$$

$$\underline{I}(\mathbf{X}; \mathbf{Y}) \triangleq \text{p-}\liminf \frac{1}{n} i_{X^nY^n}(X^n; Y^n). \tag{8}$$

Furthermore, set

$$\overline{C}(\mathbf{W}) = \sup_{\mathbf{X}} \overline{I}(\mathbf{X}; \mathbf{Y}), \quad \underline{C}(\mathbf{W}) = \sup_{\mathbf{X}} \underline{I}(\mathbf{X}; \mathbf{Y}). \tag{9}$$

As a special case of noisy channels, we consider the case when $\mathbf{W} = \{W^n\}_{n=1}^{\infty}$ is a stationary discrete memoryless channel (DMC) given by

$$W^n(\mathbf{y}|\mathbf{x}) = \prod_{t=1}^{n} W(y_t|x_t). \tag{10}$$

The stationary DMC is specified with $W = \{W(y|x)\ \}_{(x,y)\in\mathcal{X}\times\mathcal{Y}}$. Let (X,Y) be a pair of random variables taking values in $\mathcal{X}\times\mathcal{Y}$ whose joint distribution $P_{XY} = \{P_{XY}(x,y)\ \}_{(x,y)\in\mathcal{X}\times\mathcal{Y}}$ is $P_{XY}(x,y) = P_X(x)W(y|x)$. Set

$$C(W) = \max_{P_X\in\mathcal{P}(\mathcal{X})} I(X;Y). \tag{11}$$

In the above stationary DMC, we have

$$\overline{C}(\mathbf{W}) = \underline{C}(\mathbf{W}) = C(W). \tag{12}$$

It is well known that $C(W)$ is the channel capacity of the stationary DMC.

Identification via channels was first posed and investigated by Ahlswede and Dueck [1] for the stationary DMC. They have established the direct coding theorem by proving that the channel capacity $C(W)$ of the stationary DMC serves as a lower bound of the identification capacity $C_{\mathrm{ID}}(\mu,\lambda|\mathbf{W})$. However, they could not obtain a satisfactory result on the converse coding theorem. Subsequently, Han and Verdú [2] established the following strong converse theorem for the stationary DMC.

Theorem 1 (Han and Verdú [2]) *Suppose that* $\mathbf{W}$ *is the stationary DMC given by (10). Then, if* $\mu+\lambda<1$, *we have*

$$C_{\mathrm{ID}}(\mu,\lambda|\mathbf{W}) = C(W). \tag{13}$$

The strong converse property stated in Theorem 1 implies that if

$$\liminf_{n\to\infty} r_n = \liminf_{n\to\infty}\frac{1}{n}\log\log N_n > C(W), \tag{14}$$

then the sum $\mu_n+\lambda_n$ of two types of error probabilities of (n,N_n,μ_n,λ_n) code with $\mu_n+\lambda_n<1$ necessarily converges to one as n tends to infinity. However, the rate of convergence has not been discussed so far. In this paper we shall prove that the rate of convergence for $\mu_n+\lambda_n$ to tend to one has at least an exponential order of the code length n and derive an explicit lower bound of this exponent.

The characterization of the ID capacity for general noisy channels was studied by Han and Verdú [2],[3] and Han [5] and Steinberg [4]. According to Han and Verdú [2],[3], $C_{\mathrm{ID}}(\mu,\lambda|\mathbf{W})$ has the following lower bound.

Theorem 2 (Han and Verdú [2],[3]) *For any* $\lambda\geq 0,\mu\geq 0$ *and any noisy channel* $\mathbf{W}$, *we have*

$$\underline{C}(\mathbf{W}) \leq C_{\mathrm{ID}}(\mu,\lambda|\mathbf{W}). \tag{15}$$

An upper bound of $C_{\mathrm{ID}}(\mu,\lambda|\mathbf{W})$ was studied by Han and Verdú [3] and Han [5]. The derivation of the upper bound has some close connection with the channel resolvability problem posed and investigated by Han and Verdú [3].

The results of Han and Verdú [3] was sharpened by Steinberg[4]. He generalized the notion of resolvability introduced by Han and Verdú. The generalized notion is called the *partial resolvability.* Steinberg posed a new extended channel resolvability problem based on the partial resolvability. Using a connection between the extended channel resolvability problem and the ID channel, he proved that the capacity $C_{\mathrm{ID}}(0,0|\mathbf{W})$ is equal to the transmission capacity $\underline{C}(\mathbf{W})$ of the general noisy channel. In this paper we derive a stronger result on upper bound of $1-\mu_n-\lambda_n$ for general noisy channels. The result of Steinberg immediately follows from our result.

3 Results

3.1 Definitions of Functions and their Properties

In this subsection, we define several functions to describe our results and state their basic properties.

Definition 5 Let $S$ be an arbitrary subset of $\mathcal{X}^n\times\ \mathcal{Y}^n$ and $\mathbf{1}_S(\mathbf{x},\mathbf{y})$ be an indicator function which takes value one on $S$ and zero outside $S$. Set

$$\zeta_{n,S}=\zeta_{n,S}(R,P_{X^n},W^n)=\mathsf{E}\left[2^{-n[R-\frac{1}{n}i_{X^nY^n}(X^n;Y^n)]}\mathbf{1}_S(X^n,Y^n)\right].$$

Definition 6 Set

$$T_\gamma=\{\ (\mathbf{x},\mathbf{y})\in\mathcal{X}^n\times\mathcal{Y}^n:\frac{1}{n}i_{X^nY^n}(\mathbf{x};\mathbf{y})\leq R-\gamma\,,\}\,. \tag{16}$$

Furthermore, set

$$\Omega^{(1)}_{n,\gamma}(R,P_{X^n},W^n)=\Pr\left[\frac{1}{n}i_{X^nY^n}(X^n;Y^n)>R-\gamma\right] \tag{17}$$

$$\Omega^{(2)}_{n,\gamma}(R,P_{X^n},W^n)=\zeta_{n,T_\gamma}(R,P_{X^n},W^n)$$

$$\Omega_{n,\gamma}(R,P_{X^n},W^n)=\Omega^{(1)}_{n,\gamma}(R,P_{X^n},W^n)+\sqrt{\Omega^{(2)}_{n,\gamma}(R,P_{X^n},W^n)}\,. \tag{18}$$

Finally, set

$$\Omega_{n,\gamma}(R,W^n)=\max_{P_{X^n}\in\mathcal{P}(\mathcal{X}^n)}\Omega_{n,\gamma}(R,P_{X^n},W^n)\,. \tag{19}$$

We can easily prove that $\Omega_{n,\gamma}(R,\ W^n)$ and $\Omega_{n,\gamma}(R,\ X^n,\ W^n)$ satisfies the following two properties.

Property 1

a) *For any* $0\leq\gamma<\tau$,

$$\Omega^{(1)}_{n,\gamma}(R,P_{X^n},W^n)=\Omega^{(1)}_{n,0}(R-\gamma,P_{X^n},W^n)\,,$$
$$\Omega^{(2)}_{n,\gamma}(R,P_{X^n},W^n)=2^{-n\gamma}\Omega^{(2)}_{n,0}(R-\gamma,P_{X^n},W^n)\,,$$
$$\Omega^{(2)}_{n,\gamma}(R,P_{X^n},W^n)\leq 2^{-n\gamma}\,,$$
$$\Omega^{(2)}_{n,\gamma}(R,P_{X^n},W^n)\leq 2^{-n\tau}+\Omega^{(1)}_{n,\tau}(R,P_{X^n},W^n)-\Omega^{(1)}_{n,\gamma}(R,P_{X^n},W^n)\,.$$

b) *For any* $\gamma \geq 0$ *and* $R \geq 0$, $\Omega^{(1)}_{n,\gamma}(R, P_{X^n}, W^n)$ *satisfies*

$$0 \leq \Omega^{(1)}_{n,\gamma}(R, P_{X^n}, W^n) \leq 1$$

and is a monotone decreasing function of R.

c) *For any* $\mathbf{X}$, *any* $\gamma \geq 0$ *and any* $R \geq 0$,

$$0 \leq \Omega_{n,\gamma}(R, P_{X^n}, W^n) \leq \frac{5}{4}\,.$$

Property 2

a) *For any* $\gamma \geq 0$ *and* $R \geq 0$,

$$0 \leq \Omega_{n,\gamma}(R, W^n) \leq \frac{5}{4}\,.$$

b) *If* $R < \underline{C}(\mathbf{W})$, *then, for any* $\gamma > 0$,

$$\lim_{n\to\infty} \Omega_{n,\gamma}(R, W^n) = 1\,,$$

and for $\gamma = 0$,

$$\liminf_{n\to\infty} \Omega_{n,0}(R, W^n) \geq 1\,.$$

c) *If* $R > \overline{C}(\mathbf{W})$, *then, for any* $0 \leq \gamma < R - \overline{C}(\mathbf{W})$,

$$\lim_{n\to\infty} \Omega_{n,\gamma}(R, W^n) = 0\,.$$

Next, we examine an asymptotic behavior of $\Omega_n(R,\, W^n)$ for $\underline{C}(\mathbf{W}) < R < \overline{C}(\mathbf{W})$, as n tends to infinity. To this end, for $0 \leq \alpha, \beta \leq 1$, define

$$\begin{aligned}\underline{C}_{\mathbf{X}}(\alpha|\mathbf{W}) &= \sup\left\{R : \limsup_{n\to\infty} \Pr\left[\frac{1}{n} i_{X^nY^n}(X^n;Y^n) \leq R\right] \leq \alpha\right\}\\ \overline{C}_{\mathbf{X}}(\beta|\mathbf{W}) &= \inf\left\{R : \limsup_{n\to\infty} \Pr\left[\frac{1}{n} i_{X^nY^n}(X^n;Y^n) \geq R\right] \leq \beta\right\}\end{aligned} \tag{20}$$

and set

$$\underline{C}(\alpha|\mathbf{W}) = \sup_{\mathbf{X}} \underline{C}_{\mathbf{X}}(\alpha|\mathbf{W})\,, \quad \overline{C}(\beta|\mathbf{W}) = \sup_{\mathbf{X}} \overline{C}_{\mathbf{X}}(\beta|\mathbf{W})\,. \tag{21}$$

From the definition it is obvious that

$$\left.\begin{aligned}&\underline{I}(\mathbf{X};\mathbf{Y}) = \underline{C}_{\mathbf{X}}(0|\mathbf{W}) \leq \underline{C}_{\mathbf{X}}(\alpha|\mathbf{W})\,,\\ &\underline{C}(\mathbf{W}) = \underline{C}(0|\mathbf{W}) \leq \underline{C}(\alpha|\mathbf{W})\,.\end{aligned}\right\} \tag{22}$$

$$\left.\begin{aligned}&\overline{C}_{\mathbf{X}}(\beta|\mathbf{W}) \leq \overline{C}_{\mathbf{X}}(0|\mathbf{W}) = \overline{I}(\mathbf{X};\mathbf{Y})\,,\\ &a\overline{C}(\beta|\mathbf{W}) \leq \overline{C}(0|\mathbf{W}) = \overline{C}(\mathbf{W})\,.\end{aligned}\right\} \tag{23}$$

Then, we have the following.

Property 3

a) *For any $\gamma > 0$, if*

$$\liminf_{n\to\infty} \Omega_{n,\gamma}(R, W^n) \geq 1 - \alpha\,, \tag{24}$$

then

$$R \leq \underline{C}(\alpha|\mathbf{W}) + \gamma\,. \tag{25}$$

b) *For any $\gamma \geq 0$, if*

$$\limsup_{n\to\infty} \Omega_{n,\gamma}(R, W^n) \leq \beta\,, \tag{26}$$

then

$$R \geq \overline{C}(\beta|\mathbf{W})\,. \tag{27}$$

With respect to Property 2 part c), we are interested in the rate of convergence of $\Omega_n(R,\, W^n)$ for $R > \overline{C}(\,\mathbf{W})$ when $n \to \infty$. To this end, set

$$\sigma(R, \mathbf{W}) = \liminf_{n\to\infty} \left(-\frac{1}{n}\right) \log \Omega_{n,0}(R, W^n)\,. \tag{28}$$

The function $\sigma(R,\mathbf{W})$ is a nonnegative function of $R \geq 0$. Since by Property 3 part b), $\sigma(R,\mathbf{W})$ vanishes if $R < \overline{C}(\,\mathbf{W})$, the condition $R \geq \overline{C}(\mathbf{W})$ is a necessary condition for $\sigma(R,\mathbf{W}) > 0$.

If the channel $\mathbf{W} = \{W^n\,\}_{n=1}^{\infty}$ is the stationary DMC, specified with W, we can derive an explicit lower bound of $\sigma(R,\mathbf{W})$. To state the result, for nonnegative λ, define

$$\begin{aligned} F(R,W) &= \min_{P\in\mathcal{P}(\mathcal{X})} \min_{\substack{V\in\mathcal{P}(\mathcal{Y}|\mathcal{X}):\\ I(P;V)\geq R}} D(V||W|P)\,, \\ F_\lambda(R,W) &= \min_{P\in\mathcal{P}(\mathcal{X})} \min_{V\in\mathcal{P}(\mathcal{Y}|\mathcal{X})} \left\{\lambda[R - I(P;V)]^+ + D(V||W|P)\right\}\,, \end{aligned} \tag{29}$$

where $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ is a set of all noisy channels with input $\mathcal{X}$ and output $\mathcal{Y}$. By an elementary computation we can show that $F(R,W)$ and $F_\lambda(R,W)$ satisfy the following properties.

Property 4

a): *$F(R,W)$ and $F_\lambda(R,W)$ are monotone increasing and convex downward function of R and are positive if and only if $R > C(W)$.*

b): *Let (P^*,V^*) be a joint probability distribution on $\mathcal{P}(\mathcal{X} \times \mathcal{Y})$ that attains the minimization stated in the definition of $F(R,W)$. Let $(P_\lambda,V_\lambda) = \{P_\lambda(x)\ V_\lambda(y|x)\ \}_{(x,y)\in\mathcal{X}\times\mathcal{Y}}$ be a joint probability distribution that attains the minimum of*

$$\lambda(R - I(P;V)) + D(V||W|P)\,.$$

Set $R_\lambda = I(P_\lambda;V_\lambda)$. Then, we have

$$F_\lambda(R,W) = \begin{cases} D(V^*||W|P^*) & \text{for } C(W) \leq R \leq R_\lambda\,, \\ \lambda(R - I(P_\lambda;V_\lambda)) + D(V_\lambda||W|P_\lambda) & \text{for } R > R_\lambda\,. \end{cases} \tag{30}$$

Furthermore, $F_\lambda(R,W)$ has the following alternative form:

$$F_\lambda(R,W) = \min_{\tilde{R}\geq 0}\left\{\lambda[R-\tilde{R}]^+ + F(\tilde{R},W)\right\}. \tag{31}$$

c): *For $R, R' > 0$*

$$|F_\lambda(R,W) - F_\lambda(R',W)| \leq \lambda|R-R'|. \tag{32}$$

Then, we have the following.

Lemma 1 For any $R \geq 0$ and any stationary DMC specified with W, we have

$$\sigma(R,\mathbf{W}) \geq \frac{1}{2}F_1(R,W). \tag{33}$$

3.2 Statement of the Results

The main result in this paper is the following.

Theorem 3 *For any $(n, N_n, \mu_n, \lambda_n)$ code with $\mu_n + \lambda_n < 1$, if the rate $r_n = \frac{1}{n}\log\log N_n$ satisfies*

$$r_n \geq R_n + \frac{\log n}{n} + \frac{1}{n}\log\log|\mathcal{X}|, \tag{34}$$

then, for any $\gamma \geq 0$, the sum $\mu_n + \lambda_n$ of two error probabilities satisfies the following:

$$1 - \mu_n - \lambda_n \leq \Omega_{n,\gamma}(R, W^n). \tag{35}$$

From Theorems 2 and 3, the following capacity formula for the identification capacity due to Steinberg [4] can be obtained as a simple corollary.

Corollary 1 (Steinberg [4]) *For any nonnegative numbers μ, λ and α that satisfy $0 \leq \mu + \lambda \leq \alpha < 1$, we have*

$$\underline{C}(\mathbf{W}) \leq C_{\mathrm{ID}}(\mu,\lambda|\mathbf{W}) \leq \underline{C}(\alpha|\mathbf{W}). \tag{36}$$

In particular, by letting α be zero, we obtain

$$C_{\mathrm{ID}}(0,0|\mathbf{W}) = \underline{C}(\mathbf{W}). \tag{37}$$

It immediately follows from Theorem 3 and the second part of Property 2 that the following strong converse result for the identification channel holds.

Corollary 2 *If*

$$\liminf_{n\to\infty} r_n > R > \overline{C}(\mathbf{W}), \tag{38}$$

then, the sum of two types of error probabilities $\mu_n + \lambda_n$ converges to one as n tends to infinity. In particular, if

$$\underline{C}(\mathbf{W}) = \overline{C}(\mathbf{W}), \tag{39}$$

then $\mu_n + \lambda_n$ converges to one as $n \to \infty$ at rates above the identification capacity. This implies that the strong converse property holds with respect to the sum of two types of error probabilities.

Next, we discuss the speed of the convergence for the sum of two types of error probabilities $\mu_n + \lambda_n$ to tend to one. We consider the case when in addition to (39), $R > \underline{C}(\mathbf{W})$ is a necessary and sufficient condition for $\sigma(R, \mathbf{W}) > 0$. The following is a direct consequence of Theorem 3 and Lemma 1.

Corollary 3 *For any* $(n, N_n,\ \mu_n,\ \lambda_n)$ *code with* $\mu_n+\ \lambda_n\ <\ 1$, *if the code rate* r_n *satisfy* $\liminf_{n\to\infty} r_n > R$, *then the sum* $\mu_n + \lambda_n$ *of two error probabilities satisfies the following:*

$$\liminf_{n\to\infty} \left(-\frac{1}{n}\right) \log\left(1 - \mu_n - \lambda_n\right) \geq \sigma(R, \mathbf{W}) . \tag{40}$$

In particular, if $\mathbf{W}$ *is the stationary DMC specified with* $W = W^1$, *we have*

$$\liminf_{n\to\infty} \left(-\frac{1}{n}\right) \log\left(1 - \mu_n - \lambda_n\right) \geq \sigma(R, \mathbf{W}) \geq (1/2)F_1(R, W) . \tag{41}$$

It follows from Corollary 3 that when $R > \overline{C}(\mathbf{W})$, for any sequence of codes $\{(n, N_n,\ \mu_n,\ \lambda_n)\ \}_{n=1}^{\infty}$satisfying $\mu_n + \lambda_n < 1$ and $\lim_{n\to\infty} r_n > R$, the sum of two types of error probabilities $\mu_n + \lambda_n$ goes to one exponentially and this exponent is lower bounded by $\sigma(R, \mathbf{W})$.

We can expect that for a fairly general class of noisy channels the condition $R > \overline{C}(\mathbf{W})$ is a necessary and sufficient condition for $\sigma(R,\ \mathbf{W}) > 0$. In particular, if $\mathbf{W}$ is the stationary DMC, $\sigma(R, \mathbf{W})$ has the explicit lower bound given by $(1/2)F_1(R, W)$, which is positive if and only if $R > C(W)$. It is interesting to note that the exponent function $F_1(R, W)$ has the same form as what appears as a reliability function for the DMC at rate above capacity in Arimoto [6] and Dueck and Körner [7].

4 Proof of the Results

4.1 Channel Resolvability Problem

Let $\mathbf{W} = \{W^n\ \}_{n=1}^{\infty}$ be an arbitrarily prescribed noisy channel. For a given $\mathbf{X} = \{X^n\}_{n=1}^{\infty}$, let $\mathbf{Y} = \{Y^n\ \}_{n=1}^{\infty}$ be a channel output when we use $\mathbf{X}$ as a channel input of the noisy channel $\mathbf{W}$. Let U_{M_n} be the uniform random variable taking values in $\mathcal{U}_{M_n} = \{1, 2, \cdots, M_n\}$. By the map $\tilde{\varphi}_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$, the uniform random variable U_{M_n} is transformed into the random variable $\tilde{X}^n = \tilde{\varphi}_n(U_{M_n})$.

Definition 7 **(M_n-type)** Let $\tilde{\mathcal{P}}_{M_n}(\mathcal{X}^n)$ be a set of all probability distributions on $\mathcal{X}^n$ that can be created by the transformation of U_{M_n}. Elements of $\tilde{\mathcal{P}}_{M_n}(\mathcal{X}^n)$ are called M_n-type. Clearly, every random variable $\tilde{X}^n = \tilde{\varphi}_n(\ U_{M_n}\ )$ created by some transformation map $\tilde{\varphi}_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ and U_{M_n} has M_n-type. Let $\tilde{\mathbf{X}} = \{\tilde{X}^n\}_{n=1}^{\infty}$ and let $\tilde{\mathbf{Y}} = \{\tilde{Y}^n\}_{n=1}^{\infty}$ be a channel output when we use $\tilde{\mathbf{X}}$ as a channel input of the noisy channel $\mathbf{W}$. We denote the distributions of $\tilde{X}^n$ and $\tilde{Y}^n$ by $\tilde{P} = \{\tilde{P}(\mathbf{x})\}_{\mathbf{x}\in\mathcal{X}^n}$ and $\tilde{Q} = \{\tilde{Q}(\mathbf{y})\}_{\mathbf{y}\in\mathcal{Y}^n}$, respectively.

Definition 8 (Partial response) Let S be a subset of $\mathcal{X}^n \times \mathcal{Y}^n$.

Define a measure on $\mathcal{Y}^n$ by

$$Q_S(\mathbf{y}) = \sum_{\mathbf{x} \in \mathcal{X}^n} W^n(\mathbf{y}|\mathbf{x}) P_{X^n}(\mathbf{x}) 1_S(\mathbf{x}, \mathbf{y}) \tag{42}$$

We call the measure Q_S *the partial response of P on S through noisy channel* W^n.

By definition of the partial response, it is obvious that

$$Q = Q_S + Q_{S^c}\,. \tag{43}$$

Note that Q_S is no longer a probability measure.

Definition 9 Let $\Phi_n(R)$ be a set of maps $\tilde{\varphi}_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ that satisfy the rate constraint $\frac{1}{n} \log M_n \leq R$. and let $\tilde{Q}_{T_\gamma}$ be a partial response of $\tilde{P}$ on $\tilde{Q}_{T_\gamma}$. Let S be an arbitrary prescribed subset of $\mathcal{X}^n \times \mathcal{Y}^n$. For $\varphi_n \in \Phi_n(R)$ let $\tilde{P} = P_{\varphi_n(U_{M_n})}$ and let $\tilde{Q}_S$ be a partial response of $P_{\tilde{X}^n}$ on S.

We consider the situation that $\tilde{Q}_S$ is used as an approximation of Q. In this situation we are interested in the asymptotic behavior of the approximation error $d(Q, \tilde{Q}_S)$ measured by the variational distance. We shall derive an explicit upper bound of $d(Q, \tilde{Q}_S)$. This result is a mathematical core of the converse coding theorem for the ID channel.

Lemma 2 *Let $\Phi_n(R)$ be a set of maps $\tilde{\varphi}_n : \mathcal{U}_{M_n} \to \mathcal{X}^n$ that satisfy the rate constraint $\frac{1}{n} \log M_n \leq R$. Then, for any n, any $P \in \mathcal{P}(\mathcal{X}^n)$, and its response $Q = PW^n$, there exists $\tilde{\varphi}_n \in \Phi_n(R)$ such that the variational distance between Q and the partial response $\tilde{Q}_S$ of M_n-type $\tilde{P} = P_{\tilde{\varphi}_n(U_{M_n})}$ on S satisfies the following:*

$$d(Q, \tilde{Q}_S) \leq \mathbf{E}\left[1_{S^c}(X^n, Y^n)\right] + \sqrt{\zeta_{n,S}}\,. \tag{44}$$

Proof of Theorem 3: Let $P_i \in \mathcal{P}(\mathcal{X}^n), i \in \mathcal{N}_n$, be codewords of $(n, N_n, \mu_n, \lambda_n)$ code of the ID channel and $D_i \subseteq \mathcal{Y}^n$, $i \in \mathcal{N}_n$ be decoding regions corresponding to the codewords. For $P_i \in \mathcal{P}(\mathcal{X}^n), i \in \mathcal{N}_n$, let the response $P_i W^n$ of P_i be denoted by Q_i. Then, for any $j \neq k$, we have

$$d(Q_j, Q_k) \geq 2\left[Q_j(D_j) - Q_k(D_j)\right] \geq 2\left(1 - \mu_n - \lambda_n\right)\,. \tag{45}$$

We denote the right member of (44) by η_n. By Lemma 2, for any $P_i \in \mathcal{P}(\mathcal{X}^n), i \in \mathcal{N}_n$ and its response Q_i, their exists $\tilde{P}_i \in \tilde{\mathcal{P}}_{M_n}(\mathcal{X}^n)$ and its partial response $\tilde{Q}_{i,T_\gamma}$ on T_γ such that $d(Q_i, \tilde{Q}_{i,T_\gamma}) \leq \eta_n$. Set $2^{nR} = M_n$. Note that the cardinality of $\tilde{\mathcal{P}}_{M_n}(\mathcal{X}^n)$ does not exceed $|\mathcal{X}|^{n2^{nR}}$. Then, if $N_n \geq |\mathcal{X}|^{n2^{nR}}$ or equivalent to

$$\frac{1}{n} \log\log N_n \geq R + \frac{\log n}{n} + \frac{1}{n} \log\log|\mathcal{X}|\,, \tag{46}$$

there exists a pair $(j, k), j \neq k \in \mathcal{N}_n$ such that $\tilde{P}_j = \tilde{P}_k$. For the above pair of integers, we have

$$d(Q_j, Q_k) \leq d(Q_j, \tilde{Q}_{j,T_\gamma}) + d(Q_k, \tilde{Q}_{j,T_\gamma}) = d(Q_j, \tilde{Q}_{j,T_\gamma}) + d(Q_k, \tilde{Q}_{k,T_\gamma}) \leq 2\eta_n\,, \tag{47}$$

which together with (45) yields that $1 - \mu_n - \lambda_n \leq \eta_n$. This completes the proof of Theorem 3. □

Proof of Corollary 1: We assume that $\mu + \lambda \leq \alpha$ and R is (μ, λ)-achievable. Then, there exists $(n, N_n, \mu_n, \lambda_n)$ code such that

$$\liminf_{n\to\infty} \frac{1}{n} \log\log N_n \geq R, \tag{48}$$

$$\left.\begin{aligned} \limsup_{n\to\infty} \mu_n \leq \mu \\ \limsup_{n\to\infty} \lambda_n \leq \lambda \end{aligned}\right\} . \tag{49}$$

Since

$$\lim_{n\to\infty} \left[\frac{\log n}{n} + \frac{1}{n}\log\log|\mathcal{X}|\right] = 0, \tag{50}$$

for any $\delta > 0$, there exists $n_1 = n_1(\delta, |\mathcal{X}|)$ such that for any $n \geq n_1$

$$\frac{\log n}{n} + \frac{1}{n}\log\log|\mathcal{X}| \leq \frac{\delta}{2}. \tag{51}$$

On the other hand, by virtue of (48), there exists $n_2 = n_2(\delta)$ such that for any $n \geq n_2$

$$\frac{1}{n}\log\log N_n \geq R - \frac{\delta}{2}. \tag{52}$$

Set $n_0 = n_0(\delta, |\mathcal{X}|) = \max\{n_1, n_2\}$. Then, for any $n \geq n_0$, we have

$$\frac{1}{n}\log\log N_n \geq R - \delta + \frac{\log n}{n} + \frac{1}{n}\log\log|\mathcal{X}|. \tag{53}$$

Applying Theorem 3 with respect to $R - \delta$, for $n \geq n_0$, we have

$$1 - \mu_n - \lambda_n \leq \Omega_{n,\gamma}(R - \delta, W^n). \tag{54}$$

Taking the limit of both sides of (54) and using (49), we obtain

$$\liminf_{n\to\infty} \Omega_{n,\gamma}(R - \delta, W^n) \geq 1 - \limsup_{n\to\infty}(\mu_n + \lambda_n) \geq 1 - (\mu + \lambda) \geq 1 - \alpha, \tag{55}$$

which together with Property 3 yields that

$$R \leq \underline{C}(\alpha|\mathbf{W}) + \delta + \gamma. \tag{56}$$

Since $\gamma > 0$ and $\delta > 0$ can be taken arbitrary small, we have $R \leq \underline{C}(\alpha|\mathbf{W})$. □

References

[1] R. Ahlswede and Gunter Dueck, "Identification via channels" *IEEE Trans. Inform. Theory*, vol. 35, pp. 15-29, Jan. 1989.

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

[3] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory,* vol. 39, pp. 752-722, May 1993.

[4] Y. Steinberg, "New converses in the theory of identification via channels," *IEEE Trans. Inform. Theory,* vol. 44, pp. 984-998, May 1998.

[5] T. S. Han, *Information-Spectrum Methods in Information Theory. (in Japanese)* Baifukan, Tokyo, 1998.

[6] S. Arimoto, "On the converse to the coding theorem for discrete memoryless channels," *IEEE Trans. Inform. Theory,* vol. IT-19, pp. 357-359, May 1973.

[7] G. Dueck and J. Körner, "Reliability function of a discrete memoryless channel at rates above capacity," *IEEE Trans. Inform. Theory,* vol. IT-25, pp. 82-85, Jan. 1979.

[8] I. Csiszár and J. Körner, *Information Theory : Coding Theorems for Discrete Memoryless Systems.* Academic Press, New York, 1981.