Minimax Converse for Identification via Channels

Shun Watanabe, *Senior Member, IEEE*

Abstract— A minimax converse for the identification via channels is derived. By this converse, a general formula for the identification capacity, which coincides with the transmission capacity, is proved without the assumption of the strong converse property. Furthermore, the optimal second-order coding rate of the identification via channels is characterized when the type I error probability is non-vanishing and the type II error probability is vanishing. Our converse is built upon the so-called partial channel resolvability approach; however, the minimax argument enables us to circumvent a flaw reported in the literature.

Index Terms— Channel resolvability, general formula identification code, information spectrum, minimax, second-order rate.

I. INTRODUCTION

THE identification is one of typical functions such that randomization significantly reduces the amount of communication necessary to compute those functions; e.g., see [26]. Inspired by the work by Ja'Ja' [25], Ahlswede and Dueck studied the problem of identification via noisy channels in [4], [5]; they have shown that, with randomization, messages of doubly exponential size of the block-length can be identified, and the optimal coefficient is given by Shannon's transmission capacity. Since then, the problem of identification in the context of information theory has been studied extensively in the literature [7], [12], [13], [18]–[20], [29], [35], [36], [45], [11], [27], [33], [34]; see [2] and [3] for a thorough review.

In many cases, the difficulties of identification problems arise in proving converse coding theorems. Initially, the so-called soft converse was proved in [5]; the converse coding theorem was only proved under the assumption that the identification error probabilities converge to zero exponentially fast in the block-length. Later, Han and Verdú proved the strong converse coding theorem of the identification via channels in [18]; see also [1] for an alternative proof. The crucial step of the proof in [18] is that we replace general stochastic encoders with stochastic encoders having specific forms, termed "M-types." In [19], Han and Verdú further studied this step as a separate problem, which they termed the channel resolvability, by introducing the information spectrum approach.

The information spectrum approach provides effective tools to investigate coding problems for general non-ergodic sources/channels; see [17] for a thorough treatment. For the channel resolvability, the optimal rate is upper bounded by the spectral sup-mutual information rate maximized over input processes. On the other hand, the identification capacity of general channels can be lower bounded by the spectral inf-mutual information rate maximized over input processes. When those upper bound and lower bound coincide, which is termed the strong converse property, it was shown in [19] that the identification capacity and the optimal rate of the channel resolvability coincide with the transmission capacity of the same channels. Later, it was proved in [20] that, without the assumption of the strong converse property, the optimal rate of the channel resolvability is characterized by the spectral sup-mutual information rate maximized over input processes.

In an attempt to determine the identification capacity without the assumption of the strong converse property, Steinberg introduced the partial channel resolvability [35]. In the partial channel resolvability, we consider a truncated channel so that the tail probability of information spectrum is not accumulated twice in the argument of relating the channel resolvability to the identification code. It should be noted that, in the modern terminology, considering the partial response is essentially equivalent to the technique termed "smoothing" [32]. For instance, the channel resolvability for smoothed channels has been effectively used to derive second-order bounds on coding problems with side-information [44].

Using the partial channel resolvability, it was claimed in [35] that the identification capacity of general channels coincides with the transmission capacity of the same channels. However, there is a flaw in the proof of [35, Lemma 2], which has been reported in [20, Remark 2]. Thus, without the assumption of the strong converse property, the identification capacity of general channels has been an open problem so far. The main purpose of this paper is to provide a remedy to the result claimed in [35]. In fact, our converse is built upon the partial channel resolvability; however, in order to circumvent the aforementioned flaw, we leverage the minimax argument described below.

In the past few decades, the argument based on the hypothesis testing has been successfully used to derive a converse bound on transmission codes of general channels [28], [24], [31], which is termed the meta converse.[1] Particularly, a useful feature of the meta converse bound is that we can choose an auxiliary output distribution; thus, the expression of the converse bound involves the minimum over the output distribution and the maximum over the input distribution. For the asymptotic analysis of discrete memoryless channels, the Shannon capacity is recovered from the minimax

[1]For a detailed historical perspective on the meta converse, see [22].

expression by the Topsoe identity [41]. In fact, the flexibility of choosing the output distribution has been effectively used to derive finer asymptotic results: the second-order coding rate [21], [31] and the third-order coding rate [40]; see also [38]. Also, Polyanskiy proved that the order of minimax in the meta converse bound can be interchanged under certain regularity conditions [30].

In this paper, we derive a minimax converse bound for the identification via channels. To that end, we utilize a modified version [23] of the so-called soft covering lemma reported in [16], [20], [29]; the modified bound on the channel resolvability involves an auxiliary output distribution. The main contribution of this paper is to apply the flexibility of choosing the auxiliary output distribution to the argument connecting the channel resolvability and the identification code.[2] The key difference between our argument and the argument in [35, Lemma 2] is as follows: in our argument, we consider a truncated channel induced from a fixed auxiliary output distribution; on the other hand, truncated channels are constructed from output distributions that depend on input distributions in [35, Lemma 2]. In the former case, we can bound the number of messages of an identification code by the number of M-types without causing any trouble; this enables us to circumvent the flaw reported in [20, Remark 2]. See Remark 2 of Section IV for more detail.

By using the minimax converse bound, we derive the identification capacity of general channels; it turns out that the identification capacity coincides with the transmission capacity without the assumption of the strong converse property. In the derivation of this result, we invoke the aforementioned result in [30] to interchange the order of the minimum over the output distribution and the maximum over the input distribution. Furthermore, we also derive the optimal second-order coding rate of the identification via channels when the type I error probability is non-vanishing and the type II error probability is vanishing.

It should be noted that the coincidence of the transmission capacity and the identification capacity is not that obvious statement. Typically, we prove the achievability of the identification capacity by using the fact that the identification capacity is lower bounded by the so-called common randomness capacity. However, the common randomness capacity and the identification capacity need not coincide in general; for more detail, see [2].

As an additional remark, we note that the general channel in the sense of [17] does not cover some classes of channels, such as the compound channel and the arbitrary varying channel. For the identification problem via those classes of channels, see [8]–[10] and references therein.

Although we connect the identification problem to the hypothesis testing as a converse proof technique, see also [6] for an alternative connection between the identification problem and the hypothesis testing.

Notation: Throughout the paper, random variables (eg. X) and their realizations (eg. x) are denoted by capital

and lower case letters, respectively. All random variables take values in some finite alphabets which are denoted by the respective calligraphic letters (eg. $\mathcal{X}$). The probability distribution of random variable X is denoted by P_X. Similarly, $X^n = (X_1, \ldots, X_n)$ and $x^n = (x_1, \ldots, x_n)$ denote, respectively, a random vector and its realization in the nth Cartesian product $\mathcal{X}^n$. For a finite set $\mathcal{S}$, the cardinality of $\mathcal{S}$ is denoted by $|\mathcal{S}|$. For a subset $\mathcal{T} \subseteq \mathcal{S}$, the complement $\mathcal{S} \backslash \mathcal{T}$ is denoted by $\mathcal{T}^c$. The set of all distributions on $\mathcal{X}$ is denoted by $\mathcal{P}(\mathcal{X})$. The indicator function is denoted by $\mathbf{1}[\cdot]$. Information theoretic quantities are denoted in the same manner as [14], [15], [17]. All information quantities and rates are evaluated with respect to the natural logarithm. For given sub-distributions P and Q that are not necessarily normalized, the variational distance is denoted by $d(P, Q) := \frac{1}{2} \sum_x |P(x) - Q(x)|$.

II. PROBLEM FORMULATION OF IDENTIFICATION VIA CHANNELS

In this section, we describe the problem formulation of the identification via channels, and review basic results. We start with the problem formulation for the single-shot regime. Given a channel W from $\mathcal{X}$ to $\mathcal{Y}$, the sender tries to transmit one of N messages; then the receiver shall identify if message $i \in \{1, \ldots, N\}$ was transmitted or not. The encoder is given by stochastic mappings $P_1, \ldots, P_N \in \mathcal{P}(\mathcal{X})$, and the decoder is given by acceptance regions $\mathcal{D}_1, \ldots, \mathcal{D}_N \subset \mathcal{Y}$ for each message. Note that, unlike the standard transmission code, the acceptance regions of an identification code need not be disjoint. In other words, if the receiver is intended to identify message i, there is no need to distinguish messages other than i.

For a given identification code $\{(P_i, \mathcal{D}_i)\}_{i=1}^N$, the first type error probability is given by

$$\mathrm{P_I} := \max_{1 \leq i \leq N} P_i W(\mathcal{D}_i^c),$$

and the second type error probability is given by

$$\mathrm{P_{II}} := \max_{1 \leq i \neq j \leq N} P_i W(\mathcal{D}_j),$$

where

$$P_i W(y) = \sum_{x \in \mathcal{X}} P_i(x) W(y|x)$$

is the output distribution of the channel W corresponding to the input distribution P_i. For given error probabilities $0 \leq \varepsilon, \delta < 1$ with $\varepsilon + \delta < 1$, an identification code $\{(P_i, \mathcal{D}_i)\}_{i=1}^N$ is called (N, ε, δ)-ID code for channel W if $\mathrm{P_I} \leq \varepsilon$ and $\mathrm{P_{II}} \leq \delta$ are satisfied. Then, the optimal code size of identification via channel W is defined by

$$N^\star(\varepsilon, \delta | W) := \sup \{N : (N, \varepsilon, \delta)\text{-ID code for } W \text{ exists}\}.$$

When we consider the block coding over n uses W^n of a channel, it is known that the optimal code size $N^\star(\varepsilon, \delta | W^n)$ grows doubly exponentially in the block length n. For a discrete memoryless channel, it has been known that the identification capacity

$$C_{\mathrm{ID}}(\varepsilon, \delta | W) := \liminf_{n \to \infty} \frac{1}{n} \log \log N^\star(\varepsilon, \delta | W^n)$$

---

[2]Recently, the flexibility of choosing the auxiliary output distribution was used in a different manner to derive the identification capacity of the covert communication [46].

coincide with the transmission capacity [5], [18], i.e.,

$$C_{\mathrm{ID}}(\varepsilon, \delta | W) = C(W) := \max_{P_X} I(X \wedge Y)$$

as long as $\varepsilon + \delta < 1$, where $I(X \wedge Y)$ is the mutual information between (X, Y) distributed according to $P_{XY}(x, y) = P_X(x)W(y|x)$. It should be noted that the identification capacity is infinite when $\varepsilon + \delta \geq 1$ [18].

III. HYPOTHESIS TESTING

In this section, we summarize known facts on the hypothesis testing and the meta converse that are needed in the rest of the paper. Consider a binary hypothesis testing with a null hypothesis $Z \sim P_Z$ and an alternative hypothesis $Z \sim Q_Z$, where P_Z and Q_Z are distribution on the same alphabet $\mathcal{Z}$. Upon observing $Z = z$, we shall decide whether the value was generated by the distribution P_Z or the distribution Q_Z. Most general test can be described by a channel T from $\mathcal{Z}$ to $\{0, 1\}$, where 0 indicates the null hypothesis and 1 indicates the alternative hypothesis. When $z \in \mathcal{Z}$ is observed, the test T chooses the null hypothesis with probability $T(0|z)$ and the alternative hypothesis with probability $T(1|z) = 1 - T(0|z)$. Then, the type I error probability of the test is defined by

$$\mathrm{P_I}[T] := \sum_z P_Z(z) T(1|z),$$

and the type II error probability of the test is defined by

$$\mathrm{P_{II}}[T] := \sum_z Q_Z(z) T(0|z).$$

For a given $0 \leq \varepsilon < 1$, denote by $\beta_\varepsilon(P_Z, Q_Z)$ the optimal type II error probability under the condition that the type I error probability is less than ε, i.e.,

$$\beta_\varepsilon(P_Z, Q_Z) := \inf_{\substack{T: \\ \mathrm{P_I}[T] \leq \varepsilon}} \mathrm{P_{II}}[T].$$

In fact, since $\beta_\varepsilon(P_Z, Q_Z)$ can be described as a linear programming when $\mathcal{Z}$ is finite, the infimum can be attained.

For a threshold parameter $\gamma \in \mathbb{R}$, the test given by

$$T(0|z) = \mathbf{1}\left[\log \frac{P_Z(z)}{Q_Z(z)} > \gamma \right]$$

is termed the likelihood ratio test, also known as the Neyman-Pearson test. For given $0 \leq \varepsilon < 1$, let

$$D_{\mathsf{s}}^\varepsilon(P_Z \| Q_Z) := \sup \left\{ \gamma \in \mathbb{R} : \Pr\left(\log \frac{P_Z(Z)}{Q_Z(Z)} \leq \gamma \right) \leq \varepsilon \right\}, \tag{1}$$

where the probability is with respect to $Z \sim P_Z$. Note that the quantity is the supremum of thresholds such that the type I error probability of the likelihood ratio test is less than ε, and it is referred to as ε-information spectrum divergence [39]. This quantity and the optimal type II error probability defined above have the following relationship (eg. see [38, Lemma 2.4]); it can be understood as a variant of the Neyman-Pearson lemma claiming that the likelihood ratio test is essentially optimal.

Lemma 1: For a given $0 \leq \varepsilon < 1$, it holds that

$$D_{\mathsf{s}}^\varepsilon(P_Z \| Q_Z) \leq -\log \beta_\varepsilon(P_Z, Q_Z)$$
$$\leq D_{\mathsf{s}}^{\varepsilon + \zeta}(P_Z \| Q_Z) + \log(1/\zeta)$$

for any $0 < \zeta < 1 - \varepsilon$.

This lemma enables us to use the two quantities almost interchangeably.

As we have mentioned in Section I, in the past few decades, the hypothesis testing has become a useful tool to derive a converse bound on transmission codes over a channel W from $\mathcal{X}$ to $\mathcal{Y}$. For such an application, we consider the hypothesis testing between the null hypothesis

$$P \times W(x, y) := P(x) W(y|x)$$

and the alternative hypothesis

$$P \times Q(x, y) := P(x) Q(y),$$

where $P \in \mathcal{P}(\mathcal{X})$ and $Q \in \mathcal{P}(\mathcal{Y})$ are given input/output distributions. More specifically, the optimal coding rate of transmission codes is bounded in terms of

$$\inf_{P \in \mathcal{P}(\mathcal{X})} \sup_{Q \in \mathcal{P}(\mathcal{Y})} \beta_\varepsilon(P \times W, P \times Q). \tag{2}$$

It can be easily verified from the definition that $\beta_\varepsilon(P \times W, P \times Q)$ is concave with respect to the output distribution $Q \in \mathcal{P}(\mathcal{Y})$. On the other hand, it was proved in [30] that $\beta_\varepsilon(P \times W, P \times Q)$ is convex with respect to the input distribution $P \in \mathcal{P}(\mathcal{X})$. Thus, $\beta_\varepsilon(P \times W, P \times Q)$ is a convex-concave function on $\mathcal{P}(\mathcal{X}) \times \mathcal{P}(\mathcal{Y})$, and regularity conditions on the saddle-point property were discussed in [30]; particularly, since $\mathcal{P}(\mathcal{X})$ and $\mathcal{P}(\mathcal{Y})$ are compact for finite alphabets $\mathcal{X}$ and $\mathcal{Y}$, the following saddle-point property follows from the classic min-max theorem.

Lemma 2 ([30]): For a given $0 \leq \varepsilon < 1$, the optimal value in (2) is attainable and

$$\min_{P \in \mathcal{P}(\mathcal{X})} \max_{Q \in \mathcal{P}(\mathcal{Y})} \beta_\varepsilon(P \times W, P \times Q)$$
$$= \max_{Q \in \mathcal{P}(\mathcal{Y})} \min_{P \in \mathcal{P}(\mathcal{X})} \beta_\varepsilon(P \times W, P \times Q).$$

When we evaluate asymptotic behavior of coding rates for a DMC, it is more convenient to use the ε-information spectrum divergence. Particularly, we will use the following symbol-wise relaxation (eg. see [40]):

$$D_{\mathsf{s}}^\varepsilon(P \times W \| P \times Q) \leq \max_{x \in \mathcal{X}} D_{\mathsf{s}}^\varepsilon(W(\cdot|x) \| Q) \tag{3}$$

for any $Q \in \mathcal{P}(\mathcal{Y})$.

IV. MAIN RESULT: MINIMAX CONVERSE FOR IDENTIFICATION VIA CHANNELS

In this section, we present our main result, i.e., the minimax converse bound on the identification via channels. To that end, we first explain the problem of channel resolvability.

For an integer M, a distribution $P \in \mathcal{P}(\mathcal{X})$ is said to be an M-type if $P(x)$ is an integer multiple of $1/M$ for every $x \in \mathcal{X}$. Then, (N, ε, δ)-ID code $\{(P_i, \mathcal{D}_i)\}_{i=1}^N$ is said to be M-canonical if P_i is an M-type for every $1 \leq i \leq N$. For M-canonical (N, ε, δ)-ID code with $\varepsilon + \delta < 1$, it is not difficult

to see that all P_is are distinct; in fact, if there exist i and j such that $P_i = P_j$, then

$$1 - \varepsilon \le P_i W(\mathcal{D}_i) = P_j W(\mathcal{D}_i) \le \delta,$$

which contradict $\varepsilon + \delta < 1$. Since the number of M-types on $\mathcal{X}$ is at most $|\mathcal{X}|^M$, we must have $N \le |\mathcal{X}|^M$ for M-canonical ID code.

In [19], among other motivations, the channel resolvability was introduced as a tool handle general ID codes by relating their analysis to that of M-canonical codes. In the channel resolvability problem, we shall approximate the output distribution PW of an arbitrarily given input distribution $P \in \mathcal{P}(\mathcal{X})$ by the output distribution $\tilde{P}W$ of an M-type $\tilde{P}$ so that

$$d(\tilde{P}W, PW) \le \zeta$$

is satisfied for a prescribed approximation error ζ. If such an approximation is realized, then we can replace each P_i with an M-type $\tilde{P}_i$, and use the above mentioned counting argument for M-canonical codes.

In an attempt to derive a tighter converse bound than that in [19], the partial channel resolvability was introduced in [35]. For a given subset $\mathcal{S} \subset \mathcal{X} \times \mathcal{Y}$, let us introduce the truncated channel

$$W^{\mathcal{S}}(y|x) := W(y|x)\mathbf{1}[(x, y) \in \mathcal{S}]$$

and the truncated output distribution

$$PW^{\mathcal{S}}(y) := \sum_{x \in \mathcal{X}} P(x)W(y|x)\mathbf{1}[(x, y) \in \mathcal{S}].$$

Note that $PW^{\mathcal{S}}$ is a sub-distribution, i.e., it may not add up to 1, and it is referred to as the partial response of the input distribution P. It can be immediately verified that

$$d(PW^{\mathcal{S}}, PW) = \frac{P \times W(\mathcal{S}^c)}{2}. \tag{4}$$

In the partial channel resolvability problem, we shall approximate the partial response $PW^{\mathcal{S}}$ of an arbitrarily given input distribution $P \in \mathcal{P}(\mathcal{X})$ by the partial response $\tilde{P}W^{\mathcal{S}}$ of an M-type $\tilde{P}$ so that

$$d(\tilde{P}W^{\mathcal{S}}, PW^{\mathcal{S}}) \le \zeta$$

is satisfied for a prescribed approximation error ζ.

A standard approach of constructing the (partial) channel resolvability code is to randomly generate M symbols $x_1, \ldots, x_M$ according to distribution P. The performance analysis of such a random code construction is referred to as the soft covering lemma [16]. The following lemma is a variant of the soft covering lemma, and it can be derived in almost the same manner as [16], [20], [29] with a simple modification. Even though the modified version is available in the literature [23], [46], we provide a proof here for completeness.

Lemma 3: For arbitrarily given $Q \in \mathcal{P}(\mathcal{Y})$ and $\gamma \in \mathbb{R}$, let

$$\mathcal{S} = \mathcal{S}(Q, \gamma) := \left\{ (x, y) \in \mathcal{X} \times \mathcal{Y} : \log \frac{W(y|x)}{Q(y)} \le \gamma \right\}. \tag{5}$$

Then, for a given input distribution $P \in \mathcal{P}(\mathcal{X})$, there exists an M-type $\tilde{P}$ such that

$$d(\tilde{P}W^{\mathcal{S}}, PW^{\mathcal{S}}) \le \frac{1}{2}\sqrt{\frac{e^{\gamma}}{M}} \tag{6}$$

Proof: Let $\mathcal{C} = \{X_1, \ldots, X_M\}$ be a codebook such that each X_i is randomly generated with distribution P. Then, we define M-type $\tilde{P} = \tilde{P}_{\mathcal{C}}$ by

$$\tilde{P}(x) = \frac{1}{M} \sum_{i=1}^{M} \mathbf{1}[X_i = x].$$

We shall evaluate the approximation error averaged over the random generation of the codebook $\mathcal{C}$. By Jensen's inequality and the convexity of $t \to t^2$, we have

$$\mathbb{E}_{\mathcal{C}}\left[d(\tilde{P}W^{\mathcal{S}}, PW^{\mathcal{S}})\right]^2 \le \mathbb{E}_{\mathcal{C}}\left[d(\tilde{P}W^{\mathcal{S}}, PW^{\mathcal{S}})^2\right]. \tag{7}$$

Then, we have

$$4\mathbb{E}\left[d(\tilde{P}W^{\mathcal{S}}, PW^{\mathcal{S}})^2\right]$$

$$= \mathbb{E}_{\mathcal{C}}\left[\left(\sum_y |\tilde{P}W^{\mathcal{S}}(y) - PW^{\mathcal{S}}(y)|\right)^2\right]$$

$$= \mathbb{E}_{\mathcal{C}}\left[\left(\sum_y \sqrt{Q(y)}\sqrt{Q(y)}\left|\frac{\tilde{P}W^{\mathcal{S}}(y) - PW^{\mathcal{S}}(y)}{Q(y)}\right|\right)^2\right]$$

$$\le \mathbb{E}_{\mathcal{C}}\left[\sum_y Q(y)\left|\frac{\tilde{P}W^{\mathcal{S}}(y) - PW^{\mathcal{S}}(y)}{Q(y)}\right|^2\right], \tag{8}$$

where the summation y is taken over $\mathrm{supp}(Q)$,[3] and the last inequality follows from the Cauchy-Schwarz inequality. Denoting $Y \sim Q$, we can rewrite the above formula as

$$\mathbb{E}_{\mathcal{C}}\left[\sum_y Q(y)\left|\frac{\tilde{P}W^{\mathcal{S}}(y) - PW^{\mathcal{S}}(y)}{Q(y)}\right|^2\right]$$

$$= \mathbb{E}_Y \mathbb{E}_{\mathcal{C}}\left[\left(\frac{\tilde{P}W^{\mathcal{S}}(Y)}{Q(Y)} - \frac{PW^{\mathcal{S}}(Y)}{Q(Y)}\right)^2\right]$$

$$= \mathbb{E}_Y \mathbb{E}_{\mathcal{C}}\left[\left(\sum_{i=1}^{M} \frac{1}{M}\frac{W^{\mathcal{S}}(Y|X_i)}{Q(Y)} - \frac{PW^{\mathcal{S}}(Y)}{Q(Y)}\right)^2\right]$$

$$= \mathbb{E}_Y \mathbb{E}_{\mathcal{C}}\left[\frac{1}{M^2}\sum_{i=1}^{M}\left(\frac{W^{\mathcal{S}}(Y|X_i)}{Q(Y)}\right)^2\right.$$

$$+ \sum_{\substack{i,j=1:\\i \ne j}}^{M} \frac{1}{M^2}\frac{W^{\mathcal{S}}(Y|X_i)}{Q(Y)}\frac{W^{\mathcal{S}}(Y|X_j)}{Q(Y)}$$

$$\left. - \sum_{i=1}^{M} \frac{2}{M}\frac{W^{\mathcal{S}}(Y|X_i)}{Q(Y)}\frac{PW^{\mathcal{S}}(Y)}{Q(Y)} + \left(\frac{PW^{\mathcal{S}}(Y)}{Q(Y)}\right)^2\right]. \tag{9}$$

Furthermore, by noting that, for $i \ne j$,

$$\mathbb{E}_{\mathcal{C}}\left[\frac{W^{\mathcal{S}}(Y|X_i)}{Q(Y)}\right] = \frac{PW^{\mathcal{S}}(Y)}{Q(Y)}$$

---

[3]Note that $\tilde{P}W^{\mathcal{S}}(y) = PW^{\mathcal{S}}(y) = 0$ whenever $Q(y) = 0$ from the definition of $\mathcal{S}$.

and

$$\mathbb{E}_{\mathcal{C}}\left[\frac{W^{\mathcal{S}}(Y|X_i)}{Q(Y)}\frac{W^{\mathcal{S}}(Y|X_j)}{Q(Y)}\right]$$
$$= \mathbb{E}_{X_i}\left[\frac{W^{\mathcal{S}}(Y|X_i)}{Q(Y)}\right]\mathbb{E}_{X_j}\left[\frac{W^{\mathcal{S}}(Y|X_j)}{Q(Y)}\right]$$
$$= \left(\frac{PW^{\mathcal{S}}(Y)}{Q(Y)}\right)^2,$$

we can rewrite (9) as

$$\frac{1}{M}\mathbb{E}_Y\mathbb{E}_X\left[\left(\frac{W^{\mathcal{S}}(Y|X)}{Q(Y)}\right)^2 - \left(\frac{PW^{\mathcal{S}}(Y)}{Q(Y)}\right)^2\right]$$
$$\leq \frac{1}{M}\mathbb{E}_Y\mathbb{E}_X\left[\left(\frac{W^{\mathcal{S}}(Y|X)}{Q(Y)}\right)^2\right]$$
$$= \frac{1}{M}\sum_{x,y}P(x)\frac{W(y|x)^2}{Q(y)}\mathbf{1}[(x,y)\in\mathcal{S}]$$
$$\leq \frac{1}{M}\sum_{x,y}P(x)W(y|x)e^{\gamma}\mathbf{1}[(x,y)\in\mathcal{S}]$$
$$\leq \frac{e^{\gamma}}{M}, \tag{10}$$

where $X \sim P$. By combining (7), (8), (9), and (10), we have

$$\mathbb{E}_{\mathcal{C}}\left[d(\tilde{P}W^{\mathcal{S}}, PW^{\mathcal{S}})\right] \leq \frac{1}{2}\sqrt{\frac{e^{\gamma}}{M}},$$

which implies the existence of an M-type $\tilde{P}$ satisfying (6). ∎

The difference between Lemma 3 and the standard soft covering lemmas is that we can arbitrarily choose an auxiliary output distribution $Q \in \mathcal{P}(\mathcal{Y})$ instead of the output distribution PW that corresponds to the input distribution P. A similar usage of the auxiliary distribution has been known in the context of a related problem, the privacy amplification [32].

The main innovation of this paper is that we use the above mentioned flexibility of choosing the auxiliary output distribution to derive a novel converse bound on the ID code.

Theorem 1: For arbitrarily given $Q \in \mathcal{P}(\mathcal{Y})$ and $\gamma \in \mathbb{R}$, let $\mathcal{S} = \mathcal{S}(Q,\gamma)$ be defined as in (5). Then, for an arbitrary integer M, any (N,ε,δ)-ID code with $N > |\mathcal{X}|^M$ must satisfy

$$\varepsilon + \delta \geq \inf_{P\in\mathcal{P}(\mathcal{X})} P \times W(\mathcal{S}) - \sqrt{\frac{e^{\gamma}}{M}}. \tag{11}$$

Proof: For an arbitrarily given (N,ε,δ)-ID code $\{(P_i,\mathcal{D}_i)\}_{i=1}^N$, we have

$$d(P_iW, P_jW) \geq P_iW(\mathcal{D}_i) - P_jW(\mathcal{D}_i)$$
$$\geq 1 - \varepsilon - \delta \tag{12}$$

for every $i \neq j$. By applying Lemma 3 for each P_i, we can find M-type $\tilde{P}_i$ such that

$$d(\tilde{P}_iW^{\mathcal{S}}, P_iW^{\mathcal{S}}) \leq \frac{1}{2}\sqrt{\frac{e^{\gamma}}{M}}. \tag{13}$$

Since the number of distinct M-types is upper bonded by $|\mathcal{X}|^M$ and since $N > |\mathcal{X}|^M$ by assumption, there must

exist a pair i and j such that $\tilde{P}_i = \tilde{P}_j$. For such a pair, by applying the triangular inequality twice, we have

$$d(P_iW, P_jW)$$
$$\leq d(P_iW, \tilde{P}_iW^{\mathcal{S}}) + d(\tilde{P}_iW^{\mathcal{S}}, \tilde{P}_jW^{\mathcal{S}}) + d(\tilde{P}_jW^{\mathcal{S}}, P_jW)$$
$$= d(P_iW, \tilde{P}_iW^{\mathcal{S}}) + d(\tilde{P}_jW^{\mathcal{S}}, P_jW)$$
$$\leq d(P_iW, P_iW^{\mathcal{S}}) + d(P_iW^{\mathcal{S}}, \tilde{P}_iW^{\mathcal{S}})$$
$$\quad + d(\tilde{P}_jW^{\mathcal{S}}, P_jW^{\mathcal{S}}) + d(P_jW^{\mathcal{S}}, P_jW)$$
$$\leq \frac{P_i \times W(\mathcal{S}^c) + P_j \times W(\mathcal{S}^c)}{2} + \sqrt{\frac{e^{\gamma}}{M}}$$
$$\leq \sup_{P\in\mathcal{P}(\mathcal{X})} P \times W(\mathcal{S}^c) + \sqrt{\frac{e^{\gamma}}{M}}, \tag{14}$$

where the second last inequality follows from (13) and (4). Then, (14) together with (12) imply (11). ∎

Remark 1: Without using the partial channel resolvability, it can be proved that any (N,ε,δ)-ID code with $N > |\mathcal{X}|^M$ must satisfy[4]

$$\varepsilon + \delta \geq \inf_{P\in\mathcal{P}(\mathcal{X})}\left[1 - 2P \times W(\mathcal{T}_P^c)\right] - \sqrt{\frac{e^{\gamma}}{M}}, \tag{15}$$

where

$$\mathcal{T}_P = \mathcal{T}(P,\gamma) := \left\{(x,y) \in \mathcal{X}\times\mathcal{Y} : \log\frac{W(y|x)}{PW(y)} \leq \gamma\right\}. \tag{16}$$

The factor 2 of the first term in (15) has prevented us from deriving a general formula of the ID-capacity without the strong converse property.

Remark 2: The proof of Theorem 1 is inspired in part from the argument in [35, Lemma 2], which has a flaw reported in [20, Remark 2]. A crucial difference between our argument and that in [35, Lemma 2] is that the set $\mathcal{S}$ for a fixed Q is used to construct the truncated channel $W^{\mathcal{S}}$ in our argument, while the set $\mathcal{T}_{P_i}$ defined by (16) is used to construct the truncated channel $W^{\mathcal{T}_{P_i}}$ for each i in [35, Lemma 2]. In the former case, $\tilde{P}_i = \tilde{P}_j$ implies $\tilde{P}_iW^{\mathcal{S}} = \tilde{P}_jW^{\mathcal{S}}$, and the size N of the ID code is bounded by the number $|\mathcal{X}|^M$ of M-types eventually.[5] On the other hand, in the latter case, we cannot conclude that $\tilde{P}_1,\ldots,\tilde{P}_N$ are all distinct since $\tilde{P}_i = \tilde{P}_j$ does not necessarily imply $\tilde{P}_iW^{\mathcal{T}_{P_i}} = \tilde{P}_jW^{\mathcal{T}_{P_j}}$; thus, the size N of the ID code cannot be bounded by the number of M-types. Instead, it was attempted in [35, Lemma 2] to bound N by the number of some alternative measures induced by M-types, which has a flaw [20, Remark 2].

Corollary 1: For $0 \leq \varepsilon, \delta < 1$ with $\varepsilon + \delta < 1$ and arbitrary $0 < \eta < 1 - \varepsilon - \delta$, we have

$$\log\log N^{\star}(\varepsilon,\delta|W)$$
$$\leq \inf_{Q\in\mathcal{P}(\mathcal{Y})}\sup_{P\in\mathcal{P}(\mathcal{X})} D_{\mathsf{s}}^{\varepsilon+\delta+\eta}(P \times W \| P \times Q)$$
$$\quad + \log\log|\mathcal{X}| + 2\log(1/\eta) + 2. \tag{17}$$

---

[4]For instance, see Eq. (17) and Lemma 3 of [20].
[5]More precisely, we have used the contraposition of this claim.

Proof: For arbitrary (N, ε, δ)-ID code, set[6]

$$M = \left\lfloor \frac{\log(N-1)}{\log|\mathcal{X}|} \right\rfloor$$
$$\geq \frac{\log(N-1)}{e \log|\mathcal{X}|}$$
$$\geq \frac{\log N}{e^2 \log|\mathcal{X}|}$$

so that $N > |\mathcal{X}|^M$, and set

$$\gamma = 2\log\eta + \log\log N - \log\log|\mathcal{X}| - 2 \tag{18}$$

so that

$$\sqrt{\frac{e^\gamma}{M}} = \sqrt{\frac{\eta^2}{M} \frac{\log N}{e^2 \log|\mathcal{X}|}}$$
$$\leq \eta.$$

Then, since $N > |\mathcal{X}|^M$, we can apply Theorem 1 and obtain

$$\inf_{P \in \mathcal{P}(\mathcal{X})} P \times W(\mathcal{S}) \leq \varepsilon + \delta + \eta, \tag{19}$$

where $\mathcal{S} = \mathcal{S}(Q, \gamma)$ is defined as in (5) for arbitrarily fixed $Q \in \mathcal{P}(\mathcal{Y})$. In fact, since the lefthand side of (19) is linear with respect to P and $\mathcal{P}(\mathcal{X})$ is a compact set, the infimum in (19) can be attained for some $P \in \mathcal{P}(\mathcal{X})$. Since $D_{\mathsf{s}}^{\varepsilon+\delta+\eta}(P \times W \| P \times Q)$ is defined as the supremum of threshold $\tilde{\gamma}$ satisfying $P \times W(\mathcal{S}(Q, \tilde{\gamma})) \leq \varepsilon + \delta + \eta$ (cf. (1)), the bound (19) implies that $D_{\mathsf{s}}^{\varepsilon+\delta+\eta}(P \times W \| P \times Q) \geq \gamma$ for some $P \in \mathcal{P}(\mathcal{X})$, which together with (18) imply

$$\log\log N \leq \sup_{P \in \mathcal{P}(\mathcal{X})} D_{\mathsf{s}}^{\varepsilon+\delta+\eta}(P \times W \| P \times Q)$$
$$+ \log\log|\mathcal{X}| + 2\log(1/\eta) + 2.$$

Since this bound holds for arbitrary (N, ε, δ)-ID code and $Q \in \mathcal{P}(\mathcal{Y})$, we have the claim of the corollary. ∎

From Lemma 1 and Lemma 2, Corollary 1 implies the following corollary.

Corollary 2: For $0 \leq \varepsilon, \delta < 1$ with $\varepsilon + \delta < 1$ and arbitrary $0 < \eta < 1 - \varepsilon - \delta$, we have

$$\log\log N^\star(\varepsilon, \delta | W)$$
$$\leq \min_{Q \in \mathcal{P}(\mathcal{Y})} \max_{P \in \mathcal{P}(\mathcal{X})} -\log\beta_{\varepsilon+\delta+\eta}(P \times W, P \times Q)$$
$$+ \log\log|\mathcal{X}| + 2\log(1/\eta) + 2 \tag{20}$$
$$= \max_{P \in \mathcal{P}(\mathcal{X})} \min_{Q \in \mathcal{P}(\mathcal{Y})} -\log\beta_{\varepsilon+\delta+\eta}(P \times W, P \times Q)$$
$$+ \log\log|\mathcal{X}| + 2\log(1/\eta) + 2. \tag{21}$$

Up to some residual terms, the upper bounds on the doubly exponential rate of the optimal ID code in Corollary 1 and Corollary 2 have the same form as the upper bounds on the rate of the optimal transmission code reported in the literature [31]. In the next section, we will discuss asymptotic behaviors of those bounds.

---

[6]Since $D_{\mathsf{s}}^{\varepsilon+\delta+\eta}(P \times W \| P \times Q) + \log(1/\eta) \geq -\log\beta_{\varepsilon+\delta}(P \times W, P \times Q) \geq 0$ by Lemma 1, (17) trivially holds if $N \leq |\mathcal{X}|$. Thus, we only consider the case with $N > |\mathcal{X}|$, which implies $M \geq 1$. Also, (17) trivially holds if N is too small, say $\log N < 2$. Thus, we only consider the case with $\log N \geq 2$; in this case, note that $\log(N-1) \geq \log(N/2) = \log N - 1 \geq (\log N)/2$.

V. CAPACITY FOR GENERAL CHANNELS

In this section, we derive the identification capacity of general channels. Let $\boldsymbol{W} = \{W^n\}_{n=1}^\infty$ be a sequence of general channels from $\mathcal{X}^n$ to $\mathcal{Y}^n$, where $\mathcal{X}$ and $\mathcal{Y}$ are finite alphabets; the channel $\boldsymbol{W}$ may not be stationary nor ergodic. For each integer n, an $(N_n, \varepsilon_n, \delta_n)$-ID code for channel W^n is defined exactly in the same manner as in Section II. We are interested in characterizing the doubly exponential optimal growth rate of the message size N_n.

Definition 1: For given $0 \leq \varepsilon, \delta < 1$, a rate R is said to be (ε, δ)-achievable ID rate for general channel $\boldsymbol{W}$ if there exists a sequence of $(N_n, \varepsilon_n, \delta_n)$-ID codes satisfying

$$\limsup_{n\to\infty} \varepsilon_n \leq \varepsilon, \tag{22}$$
$$\limsup_{n\to\infty} \delta_n \leq \delta, \tag{23}$$

and

$$\liminf_{n\to\infty} \frac{1}{n} \log\log N_n \geq R. \tag{24}$$

Then, the supremum of (ε, δ)-achievable ID rates for $\boldsymbol{W}$ is termed the (ε, δ)-ID capacity, and is denoted by $C_{\mathrm{ID}}(\varepsilon, \delta | \boldsymbol{W})$. Particularly, for $(\varepsilon, \delta) = (0, 0)$, it is termed the ID capacity and denoted by $C_{\mathrm{ID}}(\boldsymbol{W})$.

For a sequence $\boldsymbol{X} = \{X^n\}_{n=1}^\infty$ of input processes, denote by $\boldsymbol{Y} = \{Y^n\}_{n=1}^\infty$ the corresponding output processes via $\boldsymbol{W} = \{W^n\}_{n=1}^\infty$, i.e., $P_{Y^n} = P_{X^n} W^n$ for each n. Then, for $0 \leq \varepsilon < 1$, let

$$\underline{I}^\varepsilon(\boldsymbol{X} \wedge \boldsymbol{Y})$$
$$:= \sup\left\{ a : \limsup_{n\to\infty} \Pr\left(\frac{1}{n} \log\frac{W^n(Y^n | X^n)}{P_{Y^n}(Y^n)} \leq a \right) \leq \varepsilon \right\}. \tag{25}$$

be the ε-spectral inf-mutual information rate. Particularly, when $\varepsilon = 0$, we just denote $\underline{I}(\boldsymbol{X} \wedge \boldsymbol{Y})$.

In [20], the following lower bound on the (ε, δ)-ID capacity was derived.

Proposition 1: For $0 \leq \varepsilon, \delta < 1$ and a sequence $\boldsymbol{W} = \{W^n\}_{n=1}^\infty$ of general channels, we have

$$C_{\mathrm{ID}}(\varepsilon, \delta | \boldsymbol{W}) \geq \sup_{\boldsymbol{X}} \underline{I}^\varepsilon(\boldsymbol{X} \wedge \boldsymbol{Y}), \tag{26}$$

where the supremum is taken over all sequences of input processes $\boldsymbol{X}$.

Note that the right-side of (26) does not depend on δ. Before [20], it had been known that $C_{\mathrm{ID}}(\varepsilon, \varepsilon | \boldsymbol{W})$ can be lower bounded by the right-side of (26); see [17].

On the other hand, from Corollary 2, we can derive the following upper bound on the (ε, δ)-ID capacity.

Theorem 2: For $0 \leq \varepsilon, \delta < 1$ with $\varepsilon + \delta < 1$ and a sequence $\boldsymbol{W} = \{W^n\}_{n=1}^\infty$ of general channels, we have

$$C_{\mathrm{ID}}(\varepsilon, \delta | \boldsymbol{W}) \leq \sup_{\boldsymbol{X}} \underline{I}^{\varepsilon+\delta}(\boldsymbol{X} \wedge \boldsymbol{Y}).$$

Proof: Suppose that R is (ε, δ)-achievable ID rate, i.e., there exists a sequence of $(N_n, \varepsilon_n, \delta_n)$-ID codes

satisfying (22), (23), and (24). By Corollary 2, we have

$$\frac{1}{n} \log \log N_n \leq$$

$$\max_{P_{X^n}} \min_{Q_{Y^n}} -\frac{1}{n} \log \beta_{\varepsilon_n + \delta_n + \eta_n}(P_{X^n} \times W^n, P_{X^n} \times Q_{Y^n}) + \Delta_n \tag{27}$$

for $\eta_n = 1/n$,[7] where[8]

$$\Delta_n = \frac{1}{n}\big(\log n + \log \log |\mathcal{X}| + 2\log(1/\eta_n) + 2\big).$$

Let $\hat{X} = \{\hat{X}^n\}$ be a sequence of input processes that attain the maximum in (27) for each n, and let $\hat{Y} = \{\hat{Y}^n\}$ be the corresponding output process. Then, we have

$$\frac{1}{n} \log \log N_n$$
$$\leq -\frac{1}{n} \log \beta_{\varepsilon_n + \delta_n + \eta_n}(P_{\hat{X}^n} \times W^n, P_{\hat{X}^n} \times P_{\hat{Y}^n}) + \Delta_n. \tag{28}$$

Furthermore, by applying the righthand inequality of Lemma 1, we have

$$\frac{1}{n} \log \log N_n$$
$$\leq -\frac{1}{n} \log D_{\mathsf{s}}^{\varepsilon_n + \delta_n + 2\eta_n}(P_{\hat{X}^n} \times W^n \| P_{\hat{X}^n} \times P_{\hat{Y}^n})$$
$$\quad + \Delta_n + \frac{1}{n} \log(1/\eta_n) \tag{29}$$

for sufficiently large n.

For arbitrary $\tau > 0$, let $\xi = \underline{I}^{\varepsilon+\delta}(\hat{X} \wedge \hat{Y}) + \tau$. Then, from the definition in (25), there exists $\nu > 0$ such that

$$\Pr\left(\frac{1}{n} \log \frac{W^n(\hat{Y}^n|\hat{X}^n)}{P_{\hat{Y}^n}(\hat{Y}^n)} \leq \xi\right) \geq \varepsilon + \delta + \nu$$

for infinitely many n. Then, for those n's, since $\limsup_{n\to\infty} \varepsilon_n + \delta_n + 2\eta_n \leq \varepsilon + \delta$ from the assumption that $(N_n, \varepsilon_n, \delta_n)$-ID codes satisfy (22) and (23), we have

$$D_{\mathsf{s}}^{\varepsilon_n + \delta_n + 2\eta_n}(P_{\hat{X}^n} \times W^n \| P_{\hat{X}^n} \times P_{\hat{Y}^n}) \leq \xi \tag{30}$$

provided that n is sufficiently large. Thus, by combining (28), (29) and (30), we have

$$R \leq \liminf_{n\to\infty} \frac{1}{n} \log \log N_n$$
$$\leq \xi$$
$$\leq \sup_{X} \underline{I}^{\varepsilon+\delta}(X \wedge Y) + \tau.$$

Since τ is arbitrary, any (ε, δ)-achievable ID rate R must satisfy $R \leq \sup_{X} \underline{I}^{\varepsilon+\delta}(X \wedge Y)$, which implies the claim of the theorem. ∎

When the requirement of the type-II error probability is $\delta = 0$, we can completely characterize the ID capacity from Proposition 1 and Theorem 2 as follows.

[7]Since $\varepsilon + \delta < 1$, we have $\eta_n < 1 - \varepsilon_n - \delta_n$ for sufficiently large n.
[8]We apply Corollary 2 to channel W^n from product alphabet $\mathcal{X}^n$; the extra factor $\log n$ in Δ_n appears since $\log \log |\mathcal{X}^n| = \log n + \log \log |\mathcal{X}|$.

Corollary 3: For $0 \leq \varepsilon < 1$ and a sequence $\boldsymbol{W} = \{W^n\}_{n=1}^{\infty}$ of general channels, we have

$$C_{\mathrm{ID}}(\varepsilon, 0|\boldsymbol{W}) = \sup_{X} \underline{I}^{\varepsilon}(X \wedge Y).$$

Particularly, for $\varepsilon = 0$, we have

$$C_{\mathrm{ID}}(\boldsymbol{W}) = \sup_{X} \underline{I}(X \wedge Y). \tag{31}$$

Note that (31) coincides with the general formula of the transmission capacity [42]. Thus, the ID capacity and the transmission capacity coincide for general channels. Previously, the coincidence of the ID capacity and the transmission capacity was known only for channels satisfying the strong converse property [17]; it should be emphasized that (31) holds without the assumption of the strong converse property.

VI. SECOND-ORDER CODING RATE

In this section, we consider the second-order coding rate of the identification via discrete memoryless channels (DMCs) W^n. As we have mentioned at the end of Section II, the optimal code size $N^{\star}(\varepsilon, \delta|W^n)$ behaves like

$$\log \log N^{\star}(\varepsilon, \delta|W^n) = nC(W) + o(n),$$

where $C(W)$ is the transmission capacity of channel W. In this section, we are interested in characterizing $L_{\mathrm{ID}}(\varepsilon, \delta|W)$ in the expansion

$$\log \log N^{\star}(\varepsilon, \delta|W^n) = nC(W) + \sqrt{n} L_{\mathrm{ID}}(\varepsilon, \delta|W) + o(\sqrt{n})$$

for fixed $0 < \varepsilon < 1$ and vanishing $\delta \to 0$.

Definition 2: For given $0 < \varepsilon, \delta < 1$ and DMC W, the second-order ID rate L is defined to be (ε, δ)-achievable if there exists a sequence of $(N_n, \varepsilon_n, \delta_n)$-ID codes for W^n satisfying

$$\limsup_{n\to\infty} \varepsilon_n \leq \varepsilon, \tag{32}$$
$$\limsup_{n\to\infty} \delta_n \leq \delta, \tag{33}$$

and

$$\liminf_{n\to\infty} \frac{1}{\sqrt{n}}\big(\log \log N_n - nC(W)\big) \geq L. \tag{34}$$

Then, the supremum of (ε, δ)-achievable second-order ID rates is termed the second-order (ε, δ)-ID capacity, and is denoted by $L_{\mathrm{ID}}(\varepsilon, \delta|W)$. Particularly,

$$L_{\mathrm{ID}}(\varepsilon|W) := \lim_{\delta\to 0} L_{\mathrm{ID}}(\varepsilon, \delta|W)$$

is termed the second-order ε-ID capacity.

In order to characterize the second-order rate, we need to introduce certain information quantities. Let

$$\Pi(W) := \big\{P_X \in \mathcal{P}(\mathcal{X}) : I(X \wedge Y) = C(W)\big\}$$

be the set of all capacity achieving input distributions. Even though capacity achieving input distributions may not be unique in general, it is known that the capacity achieving output distribution P_Y^* is unique.

For a given output distribution Q_Y, let

$$V(W\|Q_Y|P_X) :=$$

$$\sum_x P_X(x) \sum_y W(y|x) \left(\log \frac{W(y|x)}{Q_Y(y)} - D(W(\cdot|x)\|Q_Y) \right)^2$$

be the conditional variance of the log-likelihood ratio between $W(\cdot|x)$ and Q_Y, where $D(\cdot\|\cdot)$ is the KL-divergence. Then, we define the minimum and the maximum of conditional information variances as

$$V_{\min}(W) := \min_{P_X \in \Pi(W)} V(W\|P_Y^*|P_X),$$
$$V_{\max}(W) = \max_{P_X \in \Pi(W)} V(W\|P_Y^*|P_X).$$

Using these quantities, ε-dispersion of channel W is defined as

$$V_\varepsilon(W) := \begin{cases} V_{\min}(W) & \text{if } \varepsilon < \frac{1}{2} \\ V_{\max}(W) & \text{if } \varepsilon \geq \frac{1}{2} \end{cases}.$$

For a given input distribution P_X and corresponding output distribution $P_Y = P_X W$, let

$$U(P_X, W)$$
$$:= \sum_{x,y} P_X(x) W(y|x) \left(\log \frac{W(y|x)}{P_Y(y)} - I(X \wedge Y) \right)^2$$

be the unconditional information variance. Then, we define the minimum and the maximum of unconditional information variances as

$$U_{\min}(W) := \min_{P_X \in \Pi(W)} U(P_X, W),$$
$$U_{\max}(W) := \max_{P_X \in \Pi(W)} U(P_X, W).$$

Even though the unconditional information variance $U(P_X, W)$ can be strictly larger than the conditional information variance $V(W\|P_X W|P_X)$ in general, for capacity achieving input distributions, these quantities coincide. Thus, the quantity

$$U_\varepsilon(W) := \begin{cases} U_{\min}(W) & \text{if } \varepsilon < \frac{1}{2} \\ U_{\max}(W) & \text{if } \varepsilon \geq \frac{1}{2} \end{cases}.$$

coincides with the ε-dispersion $V_\varepsilon(W)$ defined above [38].

Now, we are ready to present the characterization of the second-order ε-ID capacity.

Theorem 3: For given DMC W and $0 < \varepsilon < 1$, if $V_\varepsilon(W) > 0$, then the second-order ε-ID capacity is given by

$$L_{\text{ID}}(\varepsilon|W) = \sqrt{V_\varepsilon(W)} \Phi^{-1}(\varepsilon), \tag{35}$$

where $\Phi^{-1}(\cdot)$ is the inverse function of the cumulative distribution function

$$\Phi(a) = \int_\infty^a \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

of the Gaussian distribution.

Note that the characterization of the second-order ε-ID capacity in (35) coincides with the second-order ε-transmission capacity [21], [31], [37].

A. Proof of Achievability

The achievability part of Theorem 3 is a straightforward consequence of the achievability bound derived in [20].

Lemma 4 ([20, Theorem 1]): For given channel W and input distribution P_X, let P_Y be the corresponding output distribution. Assume that real numbers $a, a', b, b', \tau, \kappa > 0$ satisfy

$$\kappa \log \left(\frac{1}{\tau} - 1 \right) > \log 2 + 1, \ 0 < \tau < 1/3, \ 0 < \kappa < 1 \tag{36}$$

and

$$1 > \frac{1}{a} + \frac{1}{a'}, \ c := 1 - \frac{1}{b} - \frac{1}{b'} > 0. \tag{37}$$

Then, for any integer $M > 0$ and for any real number $K > 0$, there exists an (N, ε, δ)-ID code such that

$$\varepsilon \leq ab \Pr \left(\log \frac{W(Y|X)}{P_Y(Y)} \leq \log K \right),$$
$$\delta \leq \kappa + a'b' \frac{1}{K} \left\lceil \frac{M}{c} \right\rceil,$$
$$N = \left\lfloor \frac{e^{\tau M}}{Me} \right\rfloor$$

provided that[9]

$$ab \Pr \left(\log \frac{W(Y|X)}{P_Y(Y)} \leq \log K \right) + a'b' \frac{1}{K} \left\lceil \frac{M}{c} \right\rceil < 1,$$

where $(X, Y) \sim P_X \times W$.

Now, we go back to the proof of achievability. For a given $0 < \varepsilon < 1$, fix a capacity achieving input distribution P_X that attains $U_\varepsilon(W)$; then, let P_Y be the corresponding output distribution of channel W. By setting $a = b = 1 + \frac{2}{n}$, $a' = b' = (n + 2)$, $\tau = \frac{1}{n+2}$, and $\kappa = \frac{1+\log 2}{\log n}$, we can verify that the conditions in (36) and (37) are satisfied for $n \geq 2$. For $R > 0$, we apply Lemma 4 by setting $K = e^{nR}$ and $M = \lceil e^{nR}/(n+2)^4 \rceil$; then, there exist a constant $F > 0$ and a sequence of $(N_n, \varepsilon_n, \delta_n)$-ID codes such that

$$\frac{1}{n} \log \log N_n \geq R - \frac{F}{n} \log n$$

and

$$\varepsilon_n \leq \left(1 + \frac{2}{n} \right)^2 \Pr \left(\frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P_Y^n(Y^n)} \leq R \right),$$
$$\delta_n \leq \frac{1 + \log 2}{\log n} + \frac{2}{n+2}$$

provided that

$$\left(1 + \frac{2}{n} \right)^2 \Pr \left(\frac{1}{n} \log \frac{W^n(Y^n|X^n)}{P_Y^n(Y^n)} \leq R \right) + \frac{2}{n+2} < 1, \tag{38}$$

where $(X^n, Y^n) \sim P_X^n \times W^n$. Here, set

$$R = C(W) + \sqrt{\frac{U_\varepsilon(W)}{n}} \Phi^{-1}(\varepsilon).$$

---

[9]In [20, Eq. (3)], there is a typo that α (a in our notation) is missing.

Then, by applying the central limit theorem, we have

$$\lim_{n\to\infty} \Pr\left(\frac{1}{n}\log\frac{W^n(Y^n|X^n)}{P_Y^n(Y^n)} \leq R\right) = \varepsilon.$$

Thus, the condition in (38) is satisfied for sufficiently large n, and there exists a sequence of $(N_n, \varepsilon_n, \delta_n)$-ID codes satisfying (32)-(34) for $L = \sqrt{U_\varepsilon(W)}\Phi^{-1}(\varepsilon)$ and an arbitrary $\delta > 0$. Thus, we have

$$L_{\mathrm{ID}}(\varepsilon|W) \geq \sqrt{U_\varepsilon(W)}\Phi^{-1}(\varepsilon)$$
$$= \sqrt{V_\varepsilon(W)}\Phi^{-1}(\varepsilon),$$

which completes the proof of the achievability part of Theorem 3. ∎

B. Proof of Converse

By Corollary 1 and the symbol-wise relaxation (3), we have

$$\log\log N^\star(\varepsilon, \delta|W^n)$$
$$\leq \inf_{Q_n\in\mathcal{P}(\mathcal{Y}^n)} \max_{x^n\in\mathcal{X}^n} D_{\mathsf{s}}^{\varepsilon+\delta+\eta}(W^n(\cdot|x^n)\|Q_n)$$
$$+ \log\log|\mathcal{X}^n| + 2\log(1/\eta) + 2. \tag{39}$$

Since the terms other than the first one in (39) are $o(\sqrt{n})$, the remaining task is to evaluate the first term of (39) for an appropriate choice of the output distribution Q_n. For the purpose of deriving the second-order rate, it suffices to choose a mixture of the capacity achieving output distribution and output distributions induced from types on $\mathcal{X}^n$ [21]. Although it is more than necessary to derive the second-order rate, we refer to a stronger result that is derived by a more sophisticated choice of the output distribution [40].

Lemma 5 ([40, Proposition 8]): Suppose that $V_{\varepsilon+\delta} > 0$. For $\eta = 1/\sqrt{n}$, there exists a constant F such that

$$\inf_{Q_n\in\mathcal{P}(\mathcal{Y}^n)} \max_{x^n\in\mathcal{X}^n} D_{\mathsf{s}}^{\varepsilon+\delta+\eta}(W^n(\cdot|x^n)\|Q_n)$$
$$\leq nC(W) + \sqrt{nV_{\varepsilon+\delta}}\Phi^{-1}(\varepsilon+\delta) + F$$

for sufficiently large n.

By (39) and Lemma 5, we have

$$L_{\mathrm{ID}}(\varepsilon, \delta|W) \leq \sqrt{V_{\varepsilon+\delta}}\Phi^{-1}(\varepsilon+\delta).$$

Finally, by taking the limit of $\delta \to 0$, we have the converse part of Theorem 3. ∎

VII. DISCUSSION

In this paper, we have derived a minimax converse bound for the identification via channel. By using this converse bound, we have derived the general formula for the identification capacity without the assumption of the strong converse property; the problem has been unsolved for a long time. Our converse is built upon the partial channel resolvability introduced in [35]. When we derive the converse bound for the identification code using the channel resolvability, a crucial observation is the counting argument in which the number of messages of the identification code is bounded by the number of M-types. Even though the partial channel resolvability approach have had a potential to improve the bound based on the channel resolvability, there was a difficulty that the counting argument does not work for the partial channel resolvability, at least without an additional trick; cf. Remark 2. We have overcome this difficulty by utilizing the auxiliary output distribution, an idea that has become popular in the past decade. As a future direction, it is tempting to apply the auxiliary output distribution approach to other identification problems.

One of open problems of the identification code is to determine the general (ε, δ)-ID capacity and the second-order (ε, δ)-ID capacity for positive δ. At this point, it is not clear whether we need to improve the achievability bound or the converse bound; perhaps both of them. It is also interesting to revisit the second-order capacity of channel resolvability, which has been partially solved in [43].

REFERENCES

[1] R. Ahlswede, "On concepts of performance parameters for channels," in *General Theory of Information Transfer and Combinatorics* (Lecture Notes in Computer Science), vol. 4123. Berlin, Germany: Springer-Verlag, 2006, pp. 639–663.
[2] R. Ahlswede, "General theory of information transfer: Updated," *Discrete Appl. Math.*, vol. 156, no. 9, pp. 1348–1388, May 2008.
[3] A. Ahlswede, I. Althöfer, C. Deppe, and U. Tamm, Eds., *Identification and Other Probabilistic Models*. Cham, Switzerland: Springer, 2021.
[4] R. Ahlswede and G. Dueck, "Identification in the presence of feedback—A discovery of new capacity formulas," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 30–36, Jan. 1989.
[5] R. Ahlswede and G. Dueck, "Identification via channels," *IEEE Trans. Inf. Theory*, vol. 35, no. 1, pp. 15–29, Jan. 1989.
[6] R. Ahlswede and E. Haroutunian, "On logarithmically asymptotically optimal testing of hypotheses and identification," in *General Theory of Information Transfer and Combinatorics* (Lecture Notes in Computer Science), vol. 4123. Berlin, Germany: Springer-Verlag, 2006, pp. 553–571.
[7] R. Ahlswede and A. Winter, "Strong converse for identification via quantum channels," *IEEE Trans. Inf. Theory*, vol. 48, no. 3, pp. 569–579, Mar. 2002.
[8] H. Boche and C. Deppe, "Secure identification for wiretap channels; robustness, super-additivity and continuity," *IEEE Trans. Inf. Forensics Security*, vol. 13, no. 7, pp. 1641–1655, Jul. 2018.
[9] H. Boche and C. Deppe, "Secure identification under passive eavesdroppers and active jamming attacks," *IEEE Trans. Inf. Forensics Security*, vol. 14, no. 2, pp. 472–485, Feb. 2019.
[10] H. Boche, C. Deppe, and A. Winter, "Secure and robust identification via classical-quantum channels," *IEEE Trans. Inf. Theory*, vol. 65, no. 10, pp. 6734–6749, Oct. 2019.
[11] H. Boche, R. F. Schaefer, and H. V. Poor, "Identification capacity of channels with feedback: Discontinuity behavior, super-activation, and Turing computability," *IEEE Trans. Inf. Theory*, vol. 66, no. 10, pp. 6184–6199, Oct. 2020.
[12] A. Bracher and A. Lapidoth, "Identification via the broadcast channel," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 3480–3501, Jun. 2017.
[13] M. V. Burnashev, "On identification capacity of infinite alphabets or continuous-time channels," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2407–2414, Nov. 2000.
[14] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed. Hoboken, NJ, USA: Wiley, 2006.
[15] I. Csiszár and J. Körner, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[16] P. Cuff, "Distributed channel synthesis," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7071–7096, Nov. 2013.

[17] T. S. Han, *Information-Spectrum Methods in Information Theory*. Berlin, Germany: Springer-Verlag, 2003.

[18] T. S. Han and S. Verdú, "New results in the theory of identification via channels," *IEEE Trans. Inf. Theory*, vol. 38, no. 1, pp. 14–25, Jan. 1992.

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

[20] M. Hayashi, "General nonasymptotic and asymptotic formulas in channel resolvability and identification capacity and their application to the wiretap channel," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1562–1575, Apr. 2006.

[21] M. Hayashi, "Information spectrum approach to second-order coding rate in channel coding," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 4947–4966, Nov. 2009.

[22] M. Hayashi, "Finite-block-length analysis in classical and quantum information theory," *Proc. Jpn. Acad. B*, vol. 93, no. 3, pp. 99–124, 2017.

[23] M. Hayashi, *Quantum Information Theory: Mathematical Foundation*, 2nd ed. Berlin, Germany: Springer-Verlag, 2017.

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

[25] J. J. Ja, "Identification is easier than decoding," in *Proc. 26th Annu. Symp. Found. Comput. Sci. (FOCS)*, 1985, pp. 43–50.

[26] E. Kushilevitz and N. Nisan, *Communication Complexity*. Cambridge, U.K.: Cambridge Univ. Press, 1997.

[27] W. Labidi, H. Boche, C. Deppe, and M. Wiese, "Identification over the Gaussian channel in the presence of feedback," 2021, *arXiv:2102.01198*. [Online]. Available: http://arxiv.org/abs/2102.01198

[28] H. Nagaoka, "Strong converse theorems in quantum information theory," in *Asymptotic Theory of Quantum Statistical Inference: Selected Papers*, M. Hayashi, Ed. Tokyo, Japan: World Scientific, 2005, ch. 4, p. 33.

[29] Y. Oohama, "Converse coding theorems for identification via channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 2, pp. 744–759, Feb. 2013.

[30] Y. Polyanskiy, "Saddle point in the minimax converse for channel coding," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 2576–2595, May 2013.

[31] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[32] R. Renner, "Security of quantum key distribution," Ph.D. dissertation, Diploma Phys., ETH, Zürich, Switzerland, Feb. 2005.

[33] M. J. Salariseddigh, U. Pereg, H. Boche, and C. Deppe, "Deterministic identification over fading channels," in *Proc. IEEE Inf. Theory Workshop*, Apr. 2021, pp. 1–5. [Online]. Available: https://arxiv.org/abs/2010.10010

[34] M. J. Salariseddigh, U. Pereg, H. Boche, and C. Deppe, "Deterministic identification over channels with power constraints," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2021, pp. 1–6. [Online]. Available: https://arxiv.org/abs/2010.04239

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

[36] Y. Steinberg and N. Merhav, "Identification in the presence of side information with application to watermarking," *IEEE Trans. Inf. Theory*, vol. 47, no. 4, pp. 1410–1422, May 2001.

[37] V. Strassen, "Asymptotische abschätzungen in Shannons informationstheorie," in *Proc. Trans. 3rd Prague Conf. Inf. Theory*, 1962, pp. 689–723.

[38] V. Y. F. Tan, *Foundations and Trends in Communications and Information Theory*, vol. 11, nos. 1–2. Norwell, MA, USA: Now Publishers, 2014.

[39] M. Tomamichel and M. Hayashi, "A hierarchy of information quantities for finite block length analysis of quantum tasks," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7693–7710, Nov. 2013.

[40] M. Tomamichel and V. Y. F. Tan, "A tight upper bound for the third-order asymptotics for most discrete memoryless channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7041–7051, Nov. 2013.

[41] F. Topsøe, "An information theoretical identity and a problem involving capacity," *Stud. Sci. Math. Hung.*, vol. 2, pp. 291–292, Jan. 1967.

[42] 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.

[43] S. Watanabe and M. Hayashi, "Strong converse and second-order asymptotics of channel resolvability," in *Proc. IEEE Int. Symp. Inf. Theory*, Honolulu, HI, USA, Jun. 2014, pp. 1882–1886.

[44] S. Watanabe, S. Kuzuoka, and V. Y. F. Tan, "Nonasymptotic and second-order achievability bounds for coding with side-information," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1574–1605, Apr. 2015.

[45] H. Yamamoto and M. Ueda, "Multiple object identification coding," *IEEE Trans. Inf. Theory*, vol. 61, no. 8, pp. 4269–4276, Aug. 2015.

[46] Q. Zhang and V. Y. F. Tan, "Covert identification over binary-input discrete memoryless channels," 2020, *arXiv:2007.13333*. [Online]. Available: http://arxiv.org/abs/2007.13333

Shun Watanabe (Senior Member, IEEE) received the B.E., M.E., and Ph.D. degrees from the Tokyo Institute of Technology in 2005, 2007, and 2009, respectively. From April 2009 to February 2015, he was an Assistant Professor with the Department of Information Science and Intelligent Systems, Tokushima University. From April 2013 to March 2015, he was a Visiting Assistant Professor with the Institute for Systems Research, University of Maryland, College Park, USA. From March 2016 to April 2016, he was a Visiting Fellow with the Institute of Henri Poincaré. Since February 2015, he has been an Associate Professor with the Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology. His current research interests include the areas of information theory, cryptography, quantum information, and information geometry. From 2016 to 2020, he served as an Associate Editor for the IEEE TRANSACTIONS ON INFORMATION THEORY.