Tight Exponential Strong Converse for Source Coding Problem With Encoded Side Information

Daisuke Takeuchi, Graduate Student Member, IEEE, and Shun Watanabe, Senior Member, IEEE

Abstract—The source coding problem with encoded side information is considered. A lower bound on the strong converse exponent has been derived by Oohama, but its tightness has not been clarified. In this paper, we derive a tight strong converse exponent. For the special case where the side-information does not exist, we demonstrate that our tight exponent of the Wyner-Ahlswede-Körner (WAK) problem reduces to the known tight expression of that special case where Oohama’s lower bound is strictly loose. The converse part is proved by a judicious use of the change-of-measure argument, which was introduced by Gu and Effros and further developed by Tyagi and Watanabe. A key component of the methodology by Tyagi and Watanabe is the use of soft Markov constraint, which was originally introduced by Oohama, as a penalty term to prove the Markov constraint at the end. A technical innovation of this paper compared to Tyagi and Watanabe is recognizing that the soft Markov constraint is a part of the exponent, rather than a penalty term that should vanish at the end; this recognition enables us to derive the matching achievability bound. In fact, via numerical experiment, we provide evidence that the soft Markov constraint is strictly positive. Compared to Oohama’s derivation of the lower bound, which relies on the single-letterization of a certain moment-generating function, the derivation of our tight exponent only involves manipulations of the Kullback-Leibler divergence and Shannon entropies. The achievability part is derived by a careful analysis of the type argument; however, unlike the conventional analysis for the achievable rate region, we need to derive the soft Markov constraint in the analysis of the correct probability. Furthermore, we present an application of our derivation of the strong converse exponent to the privacy amplification.

Index Terms—Source coding with side information, network information theory, strong converse theorem, exponent of correct probability, change of measure argument.

I. INTRODUCTION

FOR source coding problems, the (weak) converse theorems state that, at transmission rates under the theoretical limit, there is no error-free coding scheme as the block length tends to infinity, i.e. the error probability cannot go to 0. On the other hand, the strong converse theorems state that, at transmission rates below the theoretical limit, the error probability converges to 1 as the block length tends to infinity.

One fundamental problem in network information theory is the source coding problem with encoded side information, also known as the Wyner-Ahlswede-Körner (WAK) problem (see Fig. 1). The WAK network consists of two encoders \( \varphi_1^{(n)} \), \( \varphi_2^{(n)} \) and one decoder \( \psi^{(n)} \). Both encoders are connected to the decoder by each channel, and the decoder aims to reproduce the source observed by the first encoder. Under the condition of i.i.d. sources, this problem was solved by Wyner [1], and Ahlswede and Körner [2] in the middle of the 1970s. According to their results, the information theoretical limit involves an auxiliary random variable with a Markov chain, which makes this problem difficult.

In this paper, we are interested in the strong converse theorem of WAK network. As described later, it is known that for the WAK network, the error probability exponentially goes to 1 when a rate pair is outside the achievable rate region. On the other hand, an explicit form of the exponent function has not been clear. Oohama [3] obtained a lower bound of the exponent by using the recursive method. However, it was unknown whether Oohama’s bound is tight or not. Thus, we consider the exponent for the WAK network by using another method called the change of measure argument. Furthermore, we discuss a property of the exponent through a numerical experiment, and we present an application of our derivation of the strong converse exponent to one of the important concepts of cryptography: privacy amplification.

A. Background

The study of strong converse was initiated by Wolfowitz [4]; he proved that, for a transmission rate above the capacity, the error probability converges to 1. For multi-terminal network problems, Ahlswede, Gács and Körner proved the strong converse theorems for the WAK network and degraded broadcast channel in [5]. A difficulty of proving the strong converse
of these problems is that these problems involve an auxiliary random variable and a Markov chain constraint; the authors of [5] overcame the difficulty by introducing the so-called blowing-up lemma. For a long time, the blowing-up lemma was the only technique to tackle the strong converse for these kinds of problems.

On the other hand, the study of exponential strong converse for the point-to-point channel coding problem was initiated by Arimoto [6] in 1973. After that, in 1976, Dueck and Körner [7] gave the tight strong converse exponent by using the type method. Also, the exponential strong converse has naturally been extended for some basic multi-terminal network problems. For the Slepian-Wolf problem, Oohama and Han [8] derived the optimal exponent by the type method in 1994. However, the exponential strong converse for more involved multi-terminal network problems such as the WAK network remained open for a long. By carefully analyzing the derivation of the strong converse for the WAK network in [5], Gu and Effros [9] showed that the correct probability decreases exponentially, i.e., the exponent is positive; however, they did not clarify the explicit form of the correct probability exponent in terms of the rate pair \((R_1, R_2)\).

More recently, new techniques have been developed to prove the strong converse. In the following, we review two lines of work that are directly related to this paper; some other techniques are reviewed in Section I-C. In a series of papers, Oohama made significant progress on the above mentioned open problem for the exponential strong converse. Specifically, he derived lower bounds on the strong converse exponent for the WAK network [3], the Wyner-Ziv problem [10], and the asymmetric broadcast channel [11]. To derive these results, he developed new techniques; particularly, to overcome a difficulty that stems from an auxiliary random variable and a Markov chain constraint, he introduced the idea of “soft Markov constraint”, i.e., adding a conditional mutual information as a penalty term describing the Markov chain constraint.

Another line of work is the approach based on the change of measure argument. The change of measure argument introduced by Gu and Effros [9] is a useful technique for the strong converses for various problems. A key step of this argument is to introduce a measure (distribution) via conditioning so that there is no error under the new measure; this enables us to prove strong converses using similar ways as those used for weak converses. A difficulty of using this argument for problems involving Markov chain constraint is that, under the new measure, it is not possible to derive the Markov chain constraint. By combining with the above mentioned idea of soft Markov constraint, Tyagi and Watanabe [12] enabled the change of measure argument to be applied to a lot of distributed coding problems including the WAK network (see also an early attempt in [13]). After that, various coding problems (for example, [14], [15]) have been analyzed by using the change of measure argument. More recently, Hamad, Wigger and Sarkiss [16] further developed the change of measure argument so that the Markov constraint can be proved in a different manner from [12].

In this paper, we derive a tight exponent of the correct probability for the WAK network by judiciously using the change of measure argument. Particularly, it should be emphasized that this is the first instance that a tight strong converse exponent is derived for a problem such that there are both an auxiliary random variable and a Markov chain in the rate region. Our technical contribution is explained next.

**B. Main Contribution**

In order to derive the tight strong converse exponent, we need novel ideas for both the converse part and the achievability part.

In the converse part, we apply the change of measure argument to the exponential strong converse problem for source coding with an auxiliary random variable and a Markov chain constraint. Originally, the change of measure argument has been used in order to prove strong converses. We now enable this technique to be applied to the analysis of strong converse exponents by focusing on the correct probability rather than achievable rates. As mentioned above, one of the main difficulties of the WAK network is handling of the Markov chain constraint that appears in the characterization of the achievable rate region. Based on Oohama’s idea of soft Markov constraint [10], the authors of [12] put a penalty term describing the Markov constraint; then, by controlling the multiplier of the penalty term so that the Markov constraint is satisfied at the end, they derived the strong converse. A main contribution of this paper is to recognize that the soft Markov constraint is a part of the tight exponent rather than a penalty term that must vanish at the end; namely, we fix the multiplier of the soft Markov constraint to be unity. This change of perspective enables us to derive the tight strong converse exponent. In fact, our numerical experiment suggests that the soft Markov constraint term is strictly positive.

In the achievability part, even though we use manipulations of the standard type method, proving the matching exponent is not trivial. Particularly, as mentioned in the previous paragraph, our (converse) characterization of the exponent may violate the Markov constraint. However, in the WAK problem, the second encoder’s coding can only depend on the side information. Thus, we need to derive the Markov constraint term in the analysis of the correct probability. A key idea is the recognition that the large deviation behavior of the source plus the soft Markov constraint can be interpreted (via a judicious decomposition) as divergence terms that describe the large deviation behavior of our coding scheme. For more detail, see the paragraph around (65).

Finally, our exponent is strictly tight compared to the previous bound [3] in a certain situation. In the channel coding, there are two kinds of expressions for the strong converse exponent: the one by Arimoto [6] and the other by Dueck and Körner [7], which corresponds to the parametric function and the optimization one, respectively. While the lower bound derived by Oohama [3] is closer to the expression by Arimoto, our characterization of the exponent is along the lines of the expression by Dueck and Körner. Therefore, we cannot

---

1It should be noted that the soft Markov constraint exists as a part of exponent in [10]; however, it was recognized as a penalty term in the sense that it involves a multiplier in [10].
directly compare our exponent with Oohama’s bound. By converting our exponent expressed in an optimization problem into a parametric form, it becomes possible to analytically compare with the bound in [3] for the special case of single-user source coding.

We also apply our derivation of the strong converse exponent of the WAK network to an instance of a random number generation problem, the problem of privacy amplification introduced in [17]. It is well known that there is a duality between the source coding and random number generation, which has been investigated by many authors. The relationship between the single-user source coding and the single-source random number generation was investigated by some authors [18], [19], and [20]; and the relationship between the Slepian-Wolf source coding problem and the distributed random number generation was investigated by Oohama [21]. Watanabe and Oohama [17] introduced a privacy amplification problem for a bounded storage eavesdropper. In this situation, an eavesdropper can access data that is correlated with the random seed observed by the legitimate users, and the eavesdropper’s observation is stored in a bounded-sized storage. The authors of [17] analyzed the security of that problem by using the results regarding the strong converse for the WAK network [5]. This problem is interesting as a dual problem for the WAK source coding problem, and, to analyze its security, we need an analysis that is equivalent to the strong converse for the WAK network. Also, Santos and Oohama [22] analyzed the security using the analysis method of the exponential strong converse for the WAK network by Oohama [3]. To demonstrate the utility of our derivation to random number generation problem, we apply our derivation of the tight exponent for the WAK network to a security analysis of this problem.

C. Related Work

Apart from Oohama’s method and the change of measure argument reviewed in Section I-A, there are new developments of strong converse proof techniques in the past decade. Kosut and Kliewer [23] proposed an association between edge removal and strong converses, and also indicated that the capacity is not changed if an extra edge is removed. However, their derivation of the strong converse results relies on a variant of the blowing-up lemma, and it seems that their approach does not directly provide the tight strong converse exponent for the WAK network. Another approach was proposed by Jose and Kulkarni [24], [25]. Their approach based on the linear programming (LP) analyzes and improves the converse bound for coding problems by introducing the so-called LP relaxation. However, it is not clear if this approach can handle problems involving auxiliary random variables. Liu, Handel and Verdú [26] developed a method for non-asymptotic converses by using reverse hypercontractivity of Markov semigroups. They have derived tighter strong converse bounds than those obtained via the blowing-up lemma. Furthermore, by combining the method in [26] with the type method, Liu have derived a second order (dispersion) converse bound for the WAK network in [27]. By a judicious adaptation of the method in [26], it may be possible to derive the tight strong converse exponent of the WAK network as well; however, at this point, the method in [26] and our method are incomparable.

D. Paper Organization

This paper is organized as follows. In Section II, after introducing notations, we define the problem that we treat. In Section III, we state our main result and provide some discussion of them. In Section IV, we consider the comparison with Oohama’s bound for the special case. In Section V, we give the proof of our result. In Section VI, we consider an application to privacy amplification. In Section VII, we conclude the paper.

II. Preliminaries

A. Notations

Random variables are in capital case (e.g. $X$), and their realizations are in lower case (e.g. $x$). All random variables take values in some alphabets that are in calligraphic letters (e.g. $\mathcal{X}$). We shall restrict our attention to finite alphabets only. The cardinality of $\mathcal{X}$ is denoted by $|\mathcal{X}|$. The probability distribution of a random variable $X$ is denoted by $P_X$. Random vectors and their realizations in the $n$th Cartesian product $\mathcal{X}^n$ are denoted by $X^n = (X_1, \ldots, X_n)$ and $x^n = (x_1, \ldots, x_n)$, respectively. Also, define $P(X)$ as the set of all distributions on $\mathcal{X}$. For information quantities, we use the same notations as [28], such as the entropy $H(X)$, the mutual information $I(X \wedge Y)$ and the KL-divergence $D(P_X || Q_X)$. For simplicity, we write $T_X$ for the type class of a type $P_X$. Also, we define the total variation distance $d_{\text{tv}}$ between two discrete distributions $P$ and $Q$ on a finite set $\mathcal{X}$ as

$$d_{\text{tv}}(P, Q) = \frac{1}{2} \sum_{x \in \mathcal{X}} |P(x) - Q(x)|. \quad (1)$$

In the following, the base of logarithm and exponentiation is 2.

B. WAK Network

We consider the network described in Fig. 1. For a discrete memoryless source that has a probability distribution $P_{XY}$ over a finite alphabet $\mathcal{X} \times \mathcal{Y}$, this coding system consists of two encoders

$$\varphi_1^{(n)} : \mathcal{X}^n \rightarrow \mathcal{M}_1^{(n)}, \quad \varphi_2^{(n)} : \mathcal{Y}^n \rightarrow \mathcal{M}_2^{(n)}, \quad (2)$$

and one decoder

$$\psi^{(n)} : \mathcal{M}_1^{(n)} \times \mathcal{M}_2^{(n)} \rightarrow \mathcal{X}^n. \quad (3)$$

For simplicity, we write encoders and a decoder as $\varphi_1, \varphi_2$ and $\psi$, respectively. Input sequences $X^n$ and $Y^n$ are encoded by $\varphi_1$ and $\varphi_2$ separately, and the decoder outputs the estimation $\hat{X}^n = \psi(\varphi_1(X^n), \varphi_2(Y^n))$ of $X^n$.

For i.i.d. $(X^n, Y^n) \sim P_{X^nY^n}$, the error probability of code $\Phi_n = (\varphi_1, \varphi_2, \psi)$ is defined as

$$P_e(\Phi_n | P_{X^nY^n}) := \Pr(\hat{X}^n \neq X^n). \quad (4)$$
A rate pair \((R_1, R_2)\) is called an achievable rate pair if there exists a sequence of codes such that
\[
\limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{M}_1^{(n)}| \leq R_1, \quad \limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{M}_2^{(n)}| \leq R_2,
\]
and
\[
\lim_{n \to \infty} P_e(\Phi_n|P_{X^nY^n}) = 0.
\]
Also, the achievable rate region \(\mathcal{R}_{WAK}\) is denoted by
\[
\mathcal{R}_{WAK} = \{(R_1, R_2) : (R_1, R_2) \text{ is achievable}\}.
\]
Wyner [1], Ahlswede and Körner [2] determined \(\mathcal{R}_{WAK}\) as below.

**Theorem 1:** Consider the region
\[
\mathcal{R} = \bigcup_{U \preceq Y \preceq X, |U| \leq |X||Y| + 1} \{(R_1, R_2) : R_1 \geq H(X|U), R_2 \geq I(U \wedge Y)\},
\]
where the Markov chain \(U \preceq Y \preceq X\) means
\[
P_{UXY} = P_X P_Y|X P_U|Y.
\]
Then, it holds
\[
\mathcal{R}_{WAK} = \mathcal{R}.
\]

A rate pair \((R_1, R_2)\) is called an achievable rate pair if there exists a sequence of codes such that
\[
\limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{M}_1^{(n)}| \leq R_1, \quad \limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{M}_2^{(n)}| \leq R_2,
\]
and
\[
\lim_{n \to \infty} P_e(\Phi_n|P_{X^nY^n}) = 0.
\]
Also, the achievable rate region \(\mathcal{R}_{WAK}\) is denoted by
\[
\mathcal{R}_{WAK} = \{(R_1, R_2) : (R_1, R_2) \text{ is achievable}\}.
\]
Wyner [1], Ahlswede and Körner [2] determined \(\mathcal{R}_{WAK}\) as below.

**Theorem 1:** Consider the region
\[
\mathcal{R} = \bigcup_{U \preceq Y \preceq X, |U| \leq |X||Y| + 1} \{(R_1, R_2) : R_1 \geq H(X|U), R_2 \geq I(U \wedge Y)\},
\]
where the Markov chain \(U \preceq Y \preceq X\) means
\[
P_{UXY} = P_X P_Y|X P_U|Y.
\]
Then, it holds
\[
\mathcal{R}_{WAK} = \mathcal{R}.
\]

A rate pair \((R_1, R_2)\) is called an achievable rate pair if there exists a sequence of codes such that
\[
\limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{M}_1^{(n)}| \leq R_1, \quad \limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{M}_2^{(n)}| \leq R_2.
\]
and
\[
\lim_{n \to \infty} P_e(\Phi_n|P_{X^nY^n}) = 0.
\]
Also, the achievable rate region \(\mathcal{R}_{WAK}\) is denoted by
\[
\mathcal{R}_{WAK} = \{(R_1, R_2) : (R_1, R_2) \text{ is achievable}\}.
\]
Wyner [1], Ahlswede and Körner [2] determined \(\mathcal{R}_{WAK}\) as below.

**Theorem 1:** Consider the region
\[
\mathcal{R} = \bigcup_{U \preceq Y \preceq X, |U| \leq |X||Y| + 1} \{(R_1, R_2) : R_1 \geq H(X|U), R_2 \geq I(U \wedge Y)\},
\]
where the Markov chain \(U \preceq Y \preceq X\) means
\[
P_{UXY} = P_X P_Y|X P_U|Y.
\]
Then, it holds
\[
\mathcal{R}_{WAK} = \mathcal{R}.
\]

A rate pair \((R_1, R_2)\) is called an achievable rate pair if there exists a sequence of codes such that
\[
\limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{M}_1^{(n)}| \leq R_1, \quad \limsup_{n \to \infty} \frac{1}{n} \log |\mathcal{M}_2^{(n)}| \leq R_2.
\]
and
\[
\lim_{n \to \infty} P_e(\Phi_n|P_{X^nY^n}) = 0.
\]
Also, the achievable rate region \(\mathcal{R}_{WAK}\) is denoted by
\[
\mathcal{R}_{WAK} = \{(R_1, R_2) : (R_1, R_2) \text{ is achievable}\}.
\]
Wyner [1], Ahlswede and Körner [2] determined \(\mathcal{R}_{WAK}\) as below.

**Theorem 1:** Consider the region
\[
\mathcal{R} = \bigcup_{U \preceq Y \preceq X, |U| \leq |X||Y| + 1} \{(R_1, R_2) : R_1 \geq H(X|U), R_2 \geq I(U \wedge Y)\},
\]
where the Markov chain \(U \preceq Y \preceq X\) means
\[
P_{UXY} = P_X P_Y|X P_U|Y.
\]
Then, it holds
\[
\mathcal{R}_{WAK} = \mathcal{R}.
\]
with the constraint, i.e., $q$ shows the plots of the values of (23) and the horizontal axis shows the value of $R$. The solid line indicates the value without the Markov constraint for $P_{U \hat{X} \hat{Y}}$, and the dashed line indicates the value with the constraint.

BSC($p$) with crossover probability $p$ when $X$ serves as the input. Also, fix the auxiliary alphabet as $|A| = 2$. Furthermore, for the purpose of numerical calculation, we choose $P_{U \hat{X} \hat{Y}} = \text{DSBS} (\beta)$ and

$$P_{X|\hat{U} \hat{Y}} (x|u, y) = \begin{cases} q_0 & (u = y \text{ and } x \neq y), \\ 1 - q_0 & (u = y \text{ and } x = y), \\ q_1 & (u \neq y \text{ and } x \neq y), \\ 1 - q_1 & (u \neq y \text{ and } x = y), \end{cases}$$

for $q_0, q_1 \in (0, 1/2)$. In this situation, we get

$$P_{X|\hat{U}} (x|u) = \begin{cases} (1 - \beta)(1 - q_0) + \beta q_1 & (u = x), \\ \beta(1 - q_1) + (1 - \beta)q_0 & (u \neq x). \end{cases}$$

Therefore, we can see

$$H(\hat{X}|\hat{Y}) = h((1 - \beta)(1 - q_0) + \beta q_1),$$

$$I(\hat{U} \wedge \hat{Y}) = 1 - h(\beta),$$

$$D(P_{\hat{U} \hat{X} \hat{Y}} || P_{\hat{U} | \hat{Y}} P_{XY}) = D(P_{\hat{U} \hat{Y}} || P_{\hat{Y}}) + D(P_{\hat{U} \hat{X} \hat{Y}} || P_{XY} P_{P_{\hat{U} \hat{Y}}})$$

$$= (1 - \beta)D(q_0|p) + \beta D(q_1|p),$$

where $h(\alpha)$ is a binary entropy function and $D(x|y) = x \log(x/y) + (1 - x) \log((1 - x)/(1 - y))$. Thus, we can derive an upper bound of the strong converse exponent as

$$\min_{(\beta, q_0, q_1) \in [0, 1]^3} \{ (1 - \beta)D(q_0|p) + \beta D(q_1|p) + |1 - h(\beta) - R_2| : h((1 - \beta)(1 - q_0) + \beta q_1) \leq R_1 \}. \quad (23)$$

Note that the Markov chain $\hat{U} - \hat{Y} - \hat{X}$ holds if and only if $q_0 = q_1$.

Now we compute (23) with numerical experiment. Figure 2 shows the plots of the values of (23) for $(p, R_2) = (0.1, 1 - h(0.2))$ in the range $R_1 \in [0, 1]$, where the vertical axis is the value of (23) and the horizontal axis is the value of $R_1$. Also, the solid line indicates the value without the Markov constraint for $P_{U \hat{X} \hat{Y}}$, and the dashed line indicates the value with the constraint, i.e., $q_0 = q_1$.

In Figure 2, we can see that the value is smaller when there is not the Markov constraint. We recall that

$I(\hat{U} \wedge \hat{X}|\hat{Y}) = 0$ corresponds to a Markov constraint for our exponent. Furthermore, we recall that the auxiliary random variable must satisfy the Markov constraint in the characterization of the achievable rate region. Therefore, one might expect that the optimal value of (23) is attained when the Markov constraint is satisfied, i.e., when $q_0 = q_1$. However, Figure 2 indicates the different result, which implies that the Markov constraint for $P_{U \hat{X} \hat{Y}}$ does not necessarily yield the optimal exponent.

B. Network With Non-Encoded Side Information

Let us consider a sub-problem of WAK coding (also Slepian-Wolf coding), a network with non-coded side information described in Fig. 3. This network corresponds to the WAK network such that $R_2 \geq \log |\mathcal{Y}|$.

For this network, we set

$$F_{NE}(R_1|P_{XY}) = \min_{P_{\hat{X} \hat{Y}} \in \mathcal{P} (\mathcal{X} \times \mathcal{Y})} \left\{ D(P_{\hat{X} \hat{Y}} || P_{XY}) + |H(\hat{X}|\hat{Y}) - R_1| \right\}.$$ 

Then, by using this quantity, we have the exponent. For any $P_{XY}$, we can obtain the following corollary.

**Corollary 1:** For $R_2 \geq \log |\mathcal{Y}|$, we have

$$G(R_1, R_2|P_{XY}) = F_{NE}(R_1|P_{XY}).$$ 

**Proof:** To prove Corollary 1, we shall show two inequalities:

$$G(R_1, R_2|P_{XY}) \geq F_{NE}(R_1|P_{XY}),$$

$$G(R_1, R_2|P_{XY}) \leq F_{NE}(R_1|P_{XY}).$$

For (26), let $\hat{Y} = \hat{Y}$ for $F(R_1, R_2|P_{XY})$, and this derives

$$G(R_1, R_2|P_{XY}) = F(R_1, R_2|P_{XY})$$

$$\geq \min_{P_{\hat{X} \hat{Y}} \in \mathcal{P} (\mathcal{X} \times \mathcal{Y})} \left\{ D(P_{\hat{X} \hat{Y}} || P_{XY}) + |I(\hat{U} \wedge \hat{Y}) - R_2| : R_1 \geq H(\hat{X}|\hat{Y}), |U| \leq |X||\mathcal{Y}| + 2 \right\}.$$ 

$$= \min_{P_{\hat{X} \hat{Y}} \in \mathcal{P} (\mathcal{X} \times \mathcal{Y})} \left\{ D(P_{\hat{X} \hat{Y}} || P_{XY}) : R_1 \geq H(\hat{X}|\hat{Y}) \right\}.$$ 

$$= \min_{P_{\hat{X} \hat{Y}} \in \mathcal{P} (\mathcal{X} \times \mathcal{Y})} \left\{ D(P_{\hat{X} \hat{Y}} || P_{XY}) + |H(\hat{X}|\hat{Y}) - R_1| \right\},$$

Fig. 3. Network with non-encoded side information.
where the last equality can be obtained by considering similarly to part of [28, Problem 2.6].

On the other hand, for (27), when \( R_1 \leq H(\hat{X}|\hat{Y}) \), the KL-divergence \( D(P_{\hat{U} \hat{X} \hat{Y}} \| P_{\hat{U} \hat{Y}} P_{X \hat{Y}}) \) can be evaluated as

\[
D(P_{\hat{U} \hat{X} \hat{Y}} \| P_{\hat{U} \hat{Y}} P_{X \hat{Y}}) = D(P_{\hat{U} \hat{X} \hat{Y}} \| P_{X \hat{Y}}) + I(\hat{U} \land \hat{X}|\hat{Y})
\]

(32)

Then, for any \( P_{X \hat{Y}} \), we have

\[
F_o(R_1, R_2|P_{X \hat{Y}}) = \sup_{(\mu, \alpha) \in [0,1]^2} f_o^{(\mu,\alpha)}(R_1, R_2|P_{X \hat{Y}}).
\]

(45)

Remark 1: Oohama showed that \( F_o(R_1, R_2|P_{X \hat{Y}}) \) is strictly positive if \((R_1, R_2)\) is outside the rate region \( R_W \).

Since the expression in Proposition 1 is very involved, it is extremely difficult to directly compare the bound in Proposition 1 and our tight exponent in Theorem 2. Thus, we compare the two expressions for the single-user network described in Fig. 4. It should be noted that the single-user network is a special case of the WAK network such that the side information \( Y \) is a constant random variable. Thus, if we consider a parametrized family \( \{P_{XY,\theta}\}_{\theta \in [0,1]} \) of sources such that \( Y \) is constant for \( \theta = 0 \) (eg. \( P_X|Y \) is the binary symmetric channel and \( P_Y(1) = \theta \)), then Proposition 2 below shows that our exponent is strictly tighter than that in [3] for \( \theta = 0 \); by the continuity of the two single-letter expressions, this means that our exponent is strictly tighter than that in [3] for sufficiently small \( \theta > 0 \), i.e., for the case with non-trivial side information.

A. Specialization to the Single-User Network

In this section, we consider two expressions for the single-user network (Fig. 4). This network corresponds to the WAK network that the encoder \( \varphi_2 \) sends nothing or \( |\mathcal{Y}| = 1 \). In this case, we can simplify Oohama’s bound and our exponent as Corollary 2 and Corollary 3 below.

Corollary 2: For \((\mu, \alpha) \in [0,1]^2\), define

\[
\tilde{\tau}_X^{(\mu,\alpha)}(x) = \alpha(1 - \mu) \log \frac{P_X(x)}{P_X(x)},
\]

(47)

\[
\tilde{\Omega}^{(\mu,\alpha)}(P_X) = -\log \sum_x P_X(x) \exp(-\tilde{\tau}_X^{(\mu,\alpha)}(x)),
\]

(48)

\[
\hat{F}_o(\mu, \alpha)(R_1|P_X) = \frac{\tilde{\Omega}^{(\mu,\alpha)}(P_X) - \alpha(1 - \mu) R_1}{2 + \alpha(1 - \mu)}.
\]

(49)

Then, for any \( P_X \), we have

\[
F_o(R_1, R_2|P_X) = \tilde{F}_o(R_1|P_X).
\]

(51)

Taking \( |\mathcal{Y}| = 1 \) in Corollary 1, we have the following corollary.

Corollary 3: Set

\[
\hat{F}_i(R_1|P_X) = \min_{P_{X \hat{Y}} \in \mathcal{P}(X)} \left\{ D(P_{\hat{X}} \| P_X) + \left| H(\hat{X}) - R_1 \right|^+ \right\}.
\]

(52)

Then, for any \( P_X \), we have

\[
F(R_1, R_2|P_X) = \hat{F}_i(R_1|P_X).
\]

(53)
B. Comparison

In this section, we compare two expressions in Corollary 2 and Corollary 3. If we can verify Proposition 2 in the following, our exponent is strictly tighter than that in [3] for the single-user network.

Proposition 2: For the special case described in Fig. 4 and $R_1 < H(X)$, we have

$$
\hat{F}_o(R_1|P_X) < \hat{F}(R_1|P_X).
$$

(54)

As mentioned in Introduction, the two expressions are different in characterization. Since we cannot compare them as they are, we convert our exponent using [29, Exercise 2.41], which is stated in the following:

Lemma 1: We can rewrite $\hat{F}_t(R_1|P_X)$ as

$$
\hat{F}_t(R_1|P_X) = \max_{\theta < 0} \frac{1 - \theta}{s(\theta) + \theta R_1},
$$

(55)

where the function $s(\theta)$ is denoted as

$$
s(\theta) = \log \sum_x \{P_X(x)\}^{1-\theta}.
$$

(56)

Using this lemma, we prove the main result, Proposition 2.

Proof: [Proof of Proposition 2] Putting $\theta = -\alpha(1 - \mu) \in [-1, 0]$, we first see that

$$
\hat{F}_t^{(\mu, \alpha)}(P_X) = -s(\theta),
$$

(57)

which can be derived from

$$
s(\theta) = \log \sum_x \{P_X(x)\}^{1-\theta}
= \log \left( \sum_x P_X(x) \left( \frac{1}{P_X(x)} \right)^\theta \right)
= \log \left( \sum_x P_X(x) \exp \left( \theta \log \frac{1}{P_X(x)} \right) \right)
= -\hat{F}_t^{(\mu, \alpha)}(P_X).
$$

(58)

Then, we can rewrite $\hat{F}_o(R_1|P_X)$ as

$$
\hat{F}_o(R_1|P_X) = \max_{\theta < 0} \frac{1 - \theta}{s(\theta) + \theta R_1}.
$$

(59)

Now we show that for some $\theta \in [-1, 0)$, the numerator $-s(\theta) + \theta R_1$ is positive, i.e., $-s(\theta) + \theta R_1 > 0$. Considering the condition of $R_1$: $R_1 < H(X)$, we rewrite $R_1 \leq H(X) - \delta$ for a small positive number $\delta$. Also, from the definition of derivatives, we get

$$
\frac{d}{d\theta} s(0) = \lim_{\theta \to 0} \frac{s(\theta)}{\theta} = H(X).
$$

(60)

Then, when we take $\theta_0 \in [-1, 0)$ sufficiently close to 0, there exists $\delta$ such that

$$
\frac{s(\theta_0)}{\theta_0} \geq H(X) - \frac{\delta}{2},
$$

which derives

$$
R_1 \leq H(X) - \delta \leq \frac{s(\theta_0)}{\theta_0} - \frac{\delta}{2} < \frac{s(\theta_0)}{\theta_0}.
$$

(62)

Therefore, noting that $\theta_0$ is a negative number, we showed that $-s(\theta) + \theta R_1 > 0$ for some $\theta \in [-1, 0)$, which indicates that there exists $\theta_1 \in [-1, 0)$ such that

$$
\hat{F}_o(R_1|P_X) = \frac{-s(\theta_1) + \theta_1 R_1}{2 - \theta_1}.
$$

(63)

Thus, we have

$$
\hat{F}_o(R_1|P_X) = \frac{-s(\theta_1) + \theta_1 R_1}{2 - \theta_1} < \frac{-s(\theta_1) + \theta_1 R_1}{1 - \theta_1} \leq \max_{\theta < 0} \frac{-s(\theta) + \theta R_1}{1 - \theta} = \hat{F}_t(R_1|P_X),
$$

(64)

which implies Proposition 2.

□

V. PROOF OF THE MAIN THEOREM

In this section, we shall prove Theorem 2. Before going into the detail of the proofs of Theorem 2, we shall explain the high-level strategy of the proof.

In the direct part, we follow the type method [28]. We consider code construction to derive the lower bound of the correct probability $P_c$. Encoder 1 $\varphi_1$ uses the standard random binning. Encoder 2 $\varphi_2$ uses the quantization scheme. The main difficulty is that, unlike the achievable region reviewed in Theorem 1, the auxiliary random variable $U$ in the expression of the strong converse exponent in (14) may not satisfy the Markov chain constraint. Since Encoder 2 can only observe the side information $Y^n$, it is not clear, at first glance, how the side information $Y^n$ could be encoded. To resolve this challenge, a key observation is the following decomposition of the divergence term of (14):

$$
D(P_{\hat{U}\hat{Y}}||P_{\hat{U}|\hat{Y}} P_{X|\hat{Y}} P_{\hat{Y}}) = D(P_{\hat{U}}||P_{\hat{U}}) + D(P_{X|\hat{Y}}||P_{X|\hat{Y}} P_{\hat{Y}}).
$$

(65)

Based on this observation, we consider the following code construction and analysis. First, we focus on sequences having a given marginal type $P_{\hat{U}}$ of the side information; the probability of occurrence of this type class contributes to the first term $D(P_{\hat{U}}||P_{\hat{U}})$ of (65). Then, we construct the quantization scheme of Encoder 2 only based on the marginal distribution $P_{\hat{Y}}$. In order to analyze the correct probability of random binning, we focus on sequences having conditional type $P_{X|\hat{Y}}$; the probability of occurrence of this conditional type class contributes to the second term $D(P_{X|\hat{Y}}||P_{X|\hat{Y}} P_{\hat{Y}})$ of (65). Under the constraint $R_1 \geq H(X|\hat{Y})$, the correct probability of random binning is close to 1, and we can derive the desired exponent except for the following issue. When $I(\hat{U} \wedge \hat{Y}) - R_2^+$ is active, the rate $R_2$ is not enough to send quantized codewords. In order to overcome this issue, Encoder 2 sends only part of the quantized codewords; the proportion of such a reduction contributes to the term $I(\hat{U} \wedge \hat{Y}) - R_2^+$.  

\footnote{Lemma 1 is a strong converse counterpart of the Rényi entropy expression of the error exponent (eg. see [28, Problem 2.15]). Note that $\frac{\theta}{1 - \alpha} s(1 - \alpha)$ is the Rényi entropy of order $\alpha$.}
In the converse part, we apply the change of measure argument [12]. We first define the set of sequences where no error occurs, and introduce the probability measure by conditioning the original distribution on this set. In fact, the logarithm of the inverse of the correct probability \( \log(1/P_n) \) is equal to the KL-divergence between the changed distribution and the original i.i.d. distribution. Under this changed probability measure, we bound the rates \( R_1 \) and \( R_2 \). The next step is to single-letterize the rates and the KL-divergence. Then, from the relationship between the logarithm of the inverse of the correct probability and the KL-divergence, we can derive the bound of the logarithm of the inverse of the correct probability. In the single-letterization step, we cannot derive the Markov chain under the changed measure. To resolve this issue, the authors in [12] used the idea of soft Markov constraint introduced by Oohama [10]; namely, instead of proving the Markov constraint, they added a conditional mutual information as a penalty term, and then derived the strong converse by controlling the multiplier of the penalty term so that the Markov constraint is satisfied at the end. Even though basic manipulations of our proof are similar to those in [12], a significant departure from the methodology in [12] is the recognition that the soft Markov constraint is a part of the exponent rather than a penalty term that must vanish at the end; namely, we fix the multiplier of the soft Markov constraint to be unity. This change of perspective enables us to derive the tight strong converse exponent.

### A. Direct Part

In this section, we prove the direct part. Since we are interested in a lower bound on the correct probability \( P_c \), it suffices to construct a code such that sequences \((x^n, y^n)\) in a fixed type class \( T_{X^nY^n} \) can be decoded correctly. First, we fix \( P_{\hat{X}|\hat{Y}} \) that attains the minimum in (14) for \((R_1 - 2\varepsilon, R_2)\) where \( \varepsilon \) is a small positive number. Let \( P_{\hat{X}|\hat{Y}} \) be a joint type satisfying

\[
d_{\text{var}}(P_{\hat{X}|\hat{Y}}, P_{\hat{X}|\hat{Y}}^c) \leq \frac{|U||X|||Y|}{n}.
\]

(66)

Note that, by the continuity of entropy, we see

\[
H(\hat{X}|\hat{U}) \leq H(\hat{X}|\hat{U}) + \varepsilon \\
\leq R_1 - \varepsilon,
\]

(67)

for sufficiently large \( n \). In order to prove the direct part, we use the following covering lemma.

**Lemma 2:** For a joint type \( P_{\hat{X}|\hat{Y}} \), there exists a codebook \( \mathcal{C} = \{u^n_1, \ldots, u^n_L\} \subseteq T_{\hat{X}|\hat{Y}} \) such that \( L = |\mathcal{C}| \) satisfies

\[
L \leq \exp\left\{ nI(\hat{U} \wedge \hat{Y}) + (|U||X||Y| + 4) \log(n + 1) \right\},
\]

(68)

and such that for any \( y^n \in T_{\hat{Y}} \), there exists \( u^n \in \mathcal{C} \) satisfying \((u^n, y^n) \in T_{\hat{X}|\hat{Y}} \).

**Proof:** Since the strategy of proving this lemma is almost the same as [30, Lemma 4], we omit it. \( \square \)

We describe an overview of code construction. For this coding scheme, we want to evaluate the correct probability \( P_c \). For \( \varphi_1 \), we use the standard random binning with rate \( R_1 \). For \( \varphi_2 \), we use the quantization using the test channel \( P_{\hat{U}|\hat{Y}} \). However, \( R_2 \) may be smaller than \( I(\hat{U} \wedge \hat{Y}) \) that is needed to quantize sequences in \( T_{\hat{Y}} \) via the test channel \( P_{\hat{U}|\hat{Y}} \). In that case, \( \varphi_2 \) assigns messages to a part of the quantization codebook, and assigns a prescribed constant message to the remaining. The detail of the coding scheme is described as follows.

1) **Encoding of \( \varphi_1 \):** Randomly and independently assign an index \( m_1 = F(x^n) \in \{1, 2, \ldots, 2^{nR_1}\} \) to each sequence \( x^n \in X^n \). The set of sequences with the same index \( m_1 \) forms a bin \( B(m_1) \). Observing \( x^n \in B(m_1) \), \( \varphi_1 \) sends the bin index \( m_1 \).

2) **Encoding of \( \varphi_2 \):** Consider a codebook \( C \) given by Lemma 2 and function \( f : T_{\hat{Y}} \to \{1, \ldots, L\} \) such that \((u^n_{f(y^n)}, y^n) \in T_{\hat{X}|\hat{Y}} \). Without loss of generality, we assume that \(|f^{-1}(1)| \geq |f^{-1}(2)| \geq \cdots \geq |f^{-1}(L)| \). We define a function \( \tilde{f} \) as

\[
\tilde{f}(y^n) = \begin{cases} f(y^n) & (f(y^n) \leq 2^{nR_{2}},) \\ 1 & (f(y^n) > 2^{nR_{2}}). \end{cases}
\]

(69)

Then, upon observing \( y^n \in \mathcal{Y}^n \), \( \varphi_2 \) sends \( m_2 = \tilde{f}(y^n) \) if \( y^n \in T_{\hat{Y}} \) and \( m_2 = 1 \) otherwise.

3) **Decoding:** Observing \((m_1, m_2) = (f(x^n), \tilde{f}(y^n)) \), \( \psi \) declares the estimate \( \hat{x} \in T_{\hat{X}|\hat{U}}(u^n_{m_2}) \) if it is the unique sequence satisfying \( F(\hat{x}) = n m_1 \); otherwise it declares a prescribed sequence.

4) **Analysis of the correct probability:** We bound the correct probability \( P_c \) averaged over random bin assignments. We lower bound the correct probability quite generously, and estimate the probability such that the following three events occur consecutively.

1) The event \( E_1 \) such that \( y^n \in T_{\hat{Y}} \) and \( \tilde{f}(y^n) = f(y^n) \); we denote the probability of this event by \( P_{c,1} = \Pr(E_1) \).
2) The event \( E_2 \) such that \( E_1 \) occurs and \( x^n \) is included in the conditional type class \( T_{\hat{X}|\hat{U}}(u^n_{f(y^n)}) \); we denote the conditional probability of this event conditioned that \( E_1 \) occurs by \( P_{c,2} = \Pr(E_2|E_1) \).
3) The event \( E_3 \) such that \( E_1 \land E_2 \) occurs and there exists no sequence \( \hat{x} \in T_{\hat{X}|\hat{U}}(u^n_{f(y^n)}) \) satisfying \( \hat{x} \neq x^n \) and \( F(\hat{x}) = F(x^n) \); we denote the conditional probability of this event conditioned on the first and second events by \( P_{c,3} = \Pr(E_3|E_1 \land E_2) \).

By noting that \( T_{\hat{X}|\hat{U}}(u^n_{f(y^n)}) \subseteq T_{\hat{X}|\hat{U}}(u^n_{f(y^n)}) \), we can verify that the decoder \( \psi \) outputs the correct source sequence \( x^n \) if the three events occur \( E_1, E_2, \) and \( E_3 \) consecutively. Thus, we can lower bound the correct probability as

\[
P_c \geq \prod_{i=1}^{3} P_{c,i}.
\]

(70)

In the following, we shall calculate each probability \( P_{c,i} \).

When \( L \leq 2^{nR_2} \), the second condition \( \tilde{f}(y^n) = f(y^n) \) of the event \( E_1 \) is always satisfied. On the other hand, by noting that \(|f^{-1}(i)| \) is sorted in descending order, when \( L > 2^{nR_2} \), we have

\[
\frac{1}{2^{nR_2}} \sum_{i=1}^{2^{nR_2}} |f^{-1}(i)| \geq \frac{1}{L} \sum_{i=1}^{L} |f^{-1}(i)| \geq \frac{|T_{\hat{Y}}|}{L}.
\]

(71)

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
where we used the following equality

\[ D(P_{\hat{X}|Y}||P_{\hat{Y}|Y}) + D(P_Y||P_{\hat{Y}}) \]

which is derived by the definition of the conditional KL-divergence.

Thus, by rearranging (76), we have

\[
\frac{1}{n} \log \frac{1}{P_{1}} \leq D(P_{\hat{X}|\hat{Y}}||P_{X|Y}) + D(P_{\hat{Y}}||Y) - R_2 + \delta_n^+ + \left( |U||X||Y| \right) \frac{\log(n+1) + 1}{n} \log \frac{1}{1 - 2^{-ne}}.
\]

Consequently, by taking the limit of \( n \) and by the continuity of the information quantities, we obtain

\[
G(R_1, R_2|P_{XY}) \leq F(R_1 - 2e, R_2|P_{XY}).
\]

Finally, by the continuity of \( F \) with respect to \( R_1 \), we obtain the desired result.

\[ \text{B. Converse Part} \]

In this section, we prove the converse part by using the change of measure argument. Define \( C = \{(x^n, y^n) : \psi(\varphi_1(x^n), \varphi_2(y^n)) = x^n \} \) as the set of sequences without error, and define the changed probability measure

\[
P_{\hat{X}^n \hat{Y}^n}(x^n, y^n) = \Pr \{(X^n, Y^n) = (x^n, y^n) : (X^n, Y^n) \in C \} = \frac{P_{X^n Y^n}(x^n, y^n)}{P_{X^n Y^n}(C)} \mathbb{1}[\{x^n, y^n \in C \},
\]

where \( \mathbb{1} \) is the indicator function. Then, we can verify

\[
D(P_{\hat{X}^n \hat{Y}^n}||P_{X^n Y^n}) = \log \frac{1}{P_{1}}.
\]

First, define random variables \( \hat{M}_1 = \varphi_1(\hat{X}^n) \) and \( \hat{M}_2 = \varphi_2(Y^n) \). Since \( nR_1 \geq \log |C| \geq H(\hat{M}_1) \), we can write a chain of inequalities

\[
nR_1 \geq H(\hat{M}_1) \geq H(M_1|M_2) \geq H(\hat{M}_1|M_2) - H(\hat{M}_1|\hat{X}^n) = I(M_1 \wedge \hat{X}^n|M_2) \geq H(\hat{X}^n|M_2) = H(\hat{X}^n|M_2),
\]

Here, note that \( H(\hat{M}_1|M_2, \hat{X}^n) = 0 \) since \( \hat{M}_1 \) is generated from \( X^n \); and \( H(\hat{X}^n|M_1, M_2) = 0 \) since \( \psi \) observes \( \hat{M}_1 \) and \( \hat{M}_2 \), then outputs the error-free estimation of \( \hat{X}^n \). Also, we consider \( H(X^n|M_2) \). Denoting \( X_j^{-} = (X_1, \ldots, X_{j-1}) \), we obtain

\[
H(\hat{X}^n|M_2) = \sum_{j=1}^{n} H(\hat{X}_j|\hat{M}_2, \hat{X}_j^{-}) \geq \sum_{j=1}^{n} H(\hat{X}_j|\hat{M}_2, \hat{X}_j^{-}, \hat{Y}_j) = nH(\hat{X}_j|\hat{U}_j, J).
\]
where $J \sim \text{unif}(\{1, \ldots, n\})$ is the time-sharing random variable and is assumed to be independent of all the other random variables involved. Also, we put $\tilde{U}_j = (\bar{M}_2, \tilde{X}_j, \tilde{Y}_j)$.

Next, since $I(\bar{M}_2 \wedge \tilde{X}^n|\tilde{Y}^n) = 0$, which follows from the structure of the network, we have

\[
D(P_{\tilde{X}^n\tilde{Y}^n}|P_{X^nY^n}) = D(P_{\tilde{X}^n\tilde{Y}^n}|P_{X^nY^n}) + I(\bar{M}_2 \wedge \tilde{X}^n|\tilde{Y}^n)
\]

\[
= D(P_{\tilde{X}^n\tilde{Y}^n}|P_{X^nY^n}) + H(\tilde{X}^n|\tilde{Y}^n) - H(\tilde{X}^n|\tilde{Y}^n, \bar{M}_2).
\]

(84)

By applying [12, Proposition 1], we obtain

\[
D(P_{\tilde{X}^n\tilde{Y}^n}|P_{X^nY^n}) + H(\tilde{X}^n|\tilde{Y}^n) \geq nD(P_{\tilde{X}^n\tilde{Y}^n}|P_{XY}) + H(\tilde{X}^n|\tilde{Y}^n). \tag{85}
\]

Also,

\[
H(\tilde{X}^n|\tilde{Y}^n, \bar{M}_2) = \sum_{j=1}^n H(\tilde{X}_j|\tilde{Y}^n, \bar{M}_2, \tilde{X}_j^{-})
\]

\[
\leq \sum_{j=1}^n H(\tilde{X}_j|\tilde{Y}_j, \bar{M}_2, \tilde{X}_j^{-}, \tilde{Y}_j^{-})
\]

\[
= nH(\tilde{X}^n|\tilde{Y}^n, \tilde{X}_j^{-}, \tilde{Y}_j^{-}). \tag{86}
\]

Then, we can obtain the following inequality

\[
D(P_{\tilde{X}^n\tilde{Y}^n}|P_{X^nY^n}) \geq nD(P_{\tilde{X}^n\tilde{Y}^n}|P_{XY}) + I(\tilde{X}^n|\tilde{Y}^n). \tag{87}
\]

Moreover, since $nR_2 \geq H(\bar{M}_2)$, we have

\[
nR_2 \geq H(\bar{M}_2)
\]

\[
= H(\bar{M}_2) - H(\bar{M}_2)\tilde{X}_j, \tilde{Y}^n)
\]

\[
= I(\bar{M}_2 \wedge \tilde{X}^n, \tilde{Y}^n)
\]

\[
= H(\tilde{X}^n|\tilde{Y}^n) - H(\tilde{X}^n|\tilde{Y}^n, \bar{M}_2). \tag{88}
\]

Then, we consider the following inequality

\[
D(P_{\tilde{X}^n\tilde{Y}^n}|P_{X^nY^n}) \geq D(P_{\tilde{X}^n\tilde{Y}^n}|P_{X^nY^n}) + H(\tilde{X}^n|\tilde{Y}^n) - H(\tilde{X}^n|\tilde{Y}^n, \bar{M}_2) - nR_2. \tag{89}
\]

A very similar argument in [12, Proposition 1] leads to

\[
D(P_{\tilde{X}^n\tilde{Y}^n}|P_{X^nY^n}) = nD(P_{\tilde{X}^n\tilde{Y}^n}|P_{XY}) + H(\tilde{X}^n|\tilde{Y}^n). \tag{90}
\]

Also,

\[
H(\tilde{X}^n|\tilde{Y}^n|\bar{M}_2) = \sum_{j=1}^n H(\tilde{X}_j, \tilde{Y}_j|\bar{M}_2, \tilde{X}_j^{-}, \tilde{Y}_j^{-})
\]

\[
= nH(\tilde{X}^n|\tilde{Y}^n, \tilde{X}_j^{-}, \tilde{Y}_j^{-}). \tag{91}
\]

Therefore, we obtain

\[
\frac{1}{n} D(P_{\tilde{X}^n\tilde{Y}^n}|P_{X^nY^n}) \geq D(P_{\tilde{X}^n\tilde{Y}^n}|P_{XY}) + H(\tilde{X}^n|\tilde{Y}^n) - H(\tilde{X}^n|\tilde{Y}^n, \bar{M}_2, \tilde{X}_j^{-}, \tilde{Y}_j^{-}) - R_2
\]

\[
= D(P_{\tilde{X}^n\tilde{Y}^n}|P_{XY}) + I(\tilde{U}_j, J \wedge \tilde{X}_j, \tilde{Y}_j) - R_2
\]

\[
= D(P_{\tilde{X}^n\tilde{Y}^n}|P_{XY}) + I(\tilde{U}_j, J \wedge \tilde{X}_j, \tilde{Y}_j) + I(\tilde{U}_j, J \wedge \tilde{Y}_j) - R_2. \tag{92}
\]

Combining (81), (87) and (92), we have

\[
\frac{1}{n} \log \frac{1}{p_{\tilde{X}^n\tilde{Y}^n}} \geq D(P_{\tilde{X}^n\tilde{Y}^n}|P_{XY}) + I(\tilde{U}_j, J \wedge \tilde{X}_j|\tilde{Y}_j)
\]

\[
+ |I(\tilde{U}_j, J \wedge \tilde{Y}_j) - R_2|^{+}. \tag{93}
\]

Noting (82) and (83), we have a condition that $R_1 \geq H(\tilde{X}^n|\tilde{Y}^n, J)$. As a consequence, we obtain

\[
\min_{r_1, r_2} \frac{1}{n} \log \frac{1}{p_{\tilde{X}^n\tilde{Y}^n}} \geq \min_{r_1 \geq H(\tilde{X}^n|\tilde{Y}^n)} \left\{ D(P_{\tilde{X}^n|\tilde{Y}^n}|P_{XY}) + I(\tilde{U}_j \wedge \tilde{X}|\tilde{Y}) + |I(\tilde{U}_j \wedge \tilde{Y}) - R_2|^{+} \right\}, \tag{94}
\]

where the minimization of $P_{U\tilde{X}^n\tilde{Y}^n}$ is taken under the condition $R_1 \geq H(\tilde{X}^n|\tilde{Y}^n)$. Note that we can see

\[
D(P_{\tilde{X}^n\tilde{Y}^n}|P_{XY}) + I(\tilde{U}_j \wedge \tilde{X}|\tilde{Y}) = D(P_{U\tilde{X}^n\tilde{Y}^n}|P_{U\tilde{Y}^n}P_{XY}). \tag{95}
\]

Consequently, we have

\[
G(R_1, R_2|P_{XY}) \geq F(R_1, R_2|P_{XY}), \tag{96}
\]

except the cardinality bound, which will be discussed in the next section.

C. Cardinality Bound

By the support lemma [28, Lemma 15.4], we can restrict the cardinality of $U$ to $|U| \leq |X||Y| + 2$ as follows. We set the following functions on $P(X \times Y)$:

\[
g_1(P_{\tilde{X}^n\tilde{Y}^n}) = H(\tilde{X}|\tilde{Y}), \tag{97}
\]

\[
g_2(P_{\tilde{X}^n\tilde{Y}^n}) = H(\tilde{X}), \tag{98}
\]

\[
g_3(P_{\tilde{X}^n\tilde{Y}^n}) = H(\tilde{Y}). \tag{99}
\]

Then, observe that

\[
P_{\tilde{X}^n\tilde{Y}^n}(x, y) = \sum_u P_{U|\tilde{Y}^n}(x, y|u) \tag{100}
\]

\[
H(\tilde{X}|\tilde{Y}) = \sum_u P_{U|\tilde{Y}^n}(y|u), \tag{101}
\]

\[
H(\tilde{X}|\tilde{Y}) = \sum_u P_{U|\tilde{Y}^n}(y|u), \tag{102}
\]

\[
H(\tilde{Y}|\tilde{U}) = \sum_u P_{U|\tilde{Y}^n}(y|u). \tag{103}
\]

Note that $I(\tilde{U} \wedge \tilde{X}|\tilde{Y}) = H(\tilde{X}|\tilde{Y}) - H(\tilde{X}|\tilde{Y}, \tilde{U})$ and $I(\tilde{U} \wedge \tilde{Y}) = H(\tilde{Y}) - H(\tilde{Y}|\tilde{U})$. Therefore, by the support lemma, it suffices to take $|U| \leq |X||Y| + 2$.

VI. APPLICATION TO PRIVACY AMPLIFICATION FOR BOUNDED STORAGE EAVESDROPPER

Now we consider an application of strong converse exponent to one of the concepts in cryptography: privacy amplification. The privacy amplification is a technique to distill a secret key from a random variable by a (possibly random) function so that the distilled key and eavesdropper’s random variable are statistically independent [31], [32].
Then, we have \(X^n\) that is correlated with \(Y^n\), and Eve (an eavesdropper) has a random variable \(Y^n\) store and Bob (a receiver) have a random variable \(X^n\) from the universal hash family (e.g. see [33, Chapter 7]). Usually, we choose \(f\) from the universal hash family. Then, we can calculate the variational distance as a security criterion denoted as \(\Delta = d_{\text{var}}(P_{K^nS_nF}, P_{\text{unif}} \times P_{S_n} \times P_F)\),

and we require the quantity \(\Delta\) to be small, where \(P_{\text{unif}}\) is the uniform distribution on \(K_n\) and \(F\) is a random mapping \(F : X^n \rightarrow K_n\). Usually, we choose \(F\) from the universal hash family (e.g. see [33, Chapter 7]).

**Lemma 3:** For a given distribution \(P_{X^nS_n}\) and a number \(\tau \in \mathbb{R}\), let

\[
\mathcal{T} = \left\{ (x^n, s^n) : \log \frac{1}{P_{X^n|S_n}(x^n|s^n)} \geq \tau \right\}.
\]

Then, we have

\[
\Delta \leq \Pr \left( \log \frac{1}{P_{X^nS_n}(X^n|S_n)} < \tau \right) + \frac{1}{2} \sqrt{2^{-\tau+nR'_1}}. \tag{108}
\]

One of the goals of privacy amplification is to evaluate a security criterion. We are interested in the trade-off between the key rate \(R_1\) and Eve’s storage size \(R_2\) under the condition that the key \(K_n\) and Eve’s information \(S_n\) are statistically independent. Now, by using our derivation of the strong converse exponent, we evaluate the first term on the right-hand side of (108) for the privacy amplification illustrated in Fig. 5. In the following theorem, we show the evaluation for the security criterion with our exponent.

**Theorem 3:** Setting \(\tau = nR'_1\) where \(R'_1 = R_1 + \delta\) for a positive number \(\delta\), the security criterion \(\Delta\) can be evaluated as

\[
\Delta \leq 2^{-nF(R'_1, R_2|P_{XY})} + \frac{1}{2} \sqrt{2^{-n\delta}}, \tag{109}
\]

where \(F(R'_1, R_2|P_{XY})\) is defined by (14). Recall that the function \(F(R'_1, R_2|P_{XY})\) is positive if and only if \((R'_1, R_2) \notin \mathcal{R}_{\text{WAK}}\).

**Proof:** In this proof, we evaluate \(\Pr(\log(1/P_{X^n|S_n}(X^n|S_n)) < nR'_1)\). First, let \(P'_c\) be a number such that

\[
P'_c = \Pr \left( \frac{1}{n} \log \frac{1}{P_{X^n|S_n}(X^n|S_n)} < R'_1 \right). \tag{110}
\]

Also, we define

\[
T^c = \left\{ (x^n, y^n) : \frac{1}{n} \log \frac{1}{P_{X^n|S_n}(x^n|g_n(y^n))} < R'_1 \right\}, \tag{111}
\]

\[
P_{X^n|Y^n} = \Pr((X^n, Y^n) = (x^n, y^n) | (X^n, Y^n) \in T^c) = \frac{P_{X^n|Y^n}(x^n, y^n)}{P_{X^n|Y^n}(T^c)} \mathbb{1}[(x^n, y^n) \in T^c]. \tag{112}
\]

Then, we can calculate

\[
D(P_{X^n|Y^n} \| P_{X^nY^n}) = \log \frac{1}{P'_c}, \tag{113}
\]

which corresponds to (81) of the converse part in Section V. This means that our converse proof method for the exponent can be also applied to the consideration of the privacy amplification for bounded storage eavesdropper.

Next, defining a random variable \(S_n = g_n(Y^n)\), we evaluate \(R'_1\) as

\[
nR'_1 \geq \mathbb{E} \left[ \log \frac{1}{P_{X^n|S_n}(X^n|S_n)} \right] = \mathbb{E} \left[ \log \frac{1}{P_{X^n|S_n}(X^n|S_n)} \right] + \mathbb{E} \left[ \log \frac{P_{X^nS_n}(X^n|S_n)}{P_{X^nS_n}(X^n|S_n)} \right] = H(S_n) + D(P_{X^nS_n} \| P_{X^nS_n}|P_{S_n}) \geq H(S_n), \tag{114}
\]

where \(\mathbb{E}[\cdot]\) denotes the expectation of \(P_{X^nS_n}\). Single-letterization in the similar way as in Section V yields

\[
R'_1 \geq H(X|U), \tag{115}
\]
where $\bar{U}$ is a similar random variable to the one that appears in the process of single-letterization in the converse of Section V. Also, in the same manner, we can obtain

$$R_2 \geq I(\bar{Y} \wedge \bar{U}).$$

(116)

According to the proof of the converse for our main theorem, we have

$$P_c \leq 2^{-nF(R_1', R_2; P_{XY})}.$$  

(117)

Thus, for any $P_{XY}$, we obtain

$$\Pr \left( \frac{1}{n} \log \frac{1}{P_{X^n|S_n}(X^n|S_n)} < R_1' \right) \leq 2^{-nF(R_1', R_2; P_{XY})}.$$  

(118)

As a conclusion, by Lemma 3, the security criterion for the privacy amplification for a bounded storage eavesdropper $\Delta$ can be bounded as

$$\Delta \leq 2^{-nF(R_1', R_2; P_{XY})} + \frac{1}{\sqrt{2^{-n\delta}}}.$$  

(119)

When $(R_1, R_2)$ is not in the achievable rate region for the WAK network, we can take $\delta > 0$ so that $(R_1', R_2)$ is not achievable as well. Then, Theorem 3 states that the security criterion $\Delta$ exponentially converges to 0.

VII. CONCLUSION

In this paper, we derived the tight exponent expression for the WAK problem. A key feature of our expression is that the soft Markov constraint is incorporated as a part of the exponent. Based on this idea, it is tempting to derive tight strong converse exponents for other multi-user networks, such as the Wyner-Ziv problem and degraded broadcast channels; lower bounds on the strong converse exponent for these problems have been studied in [10] and [35]. However, deriving a tight exponent requires a very precise discussion. Moreover, it is necessary to clarify how the soft Markov constraint corresponds to in the coding process to prove the achievability of those problems. These issues are challenging future work.

Apart from deriving the strong converse exponent itself, the analysis method in this paper has an implication for the security problem discussed in Section VI. In fact, this security problem is closely related to the side-channel attack in cryptography [22]. It is an important future agenda to derive the tight exponential security bound for this problem.

Furthermore, the analysis of strong converse is related to the parallel repetition theorem studied in computer science [36]. Thus, it is also an interesting direction to apply our idea to the analysis of the parallel repetition theorem.

ACKNOWLEDGMENT

The authors would like to thank Prof. Yasutada Oohama and Prof. Jun Chen for a valuable discussion. They would also like to thank Prof. Tomohiko Uyematsu for suggesting to consider the special case in Section IV.

REFERENCES

[1] A. Wyner, “On source coding with side information at the decoder,” IEEE Trans. Inf. Theory, vol. IT-21, no. 3, pp. 294–300, May 1975.
[2] R. Ahlswede and J. Korner, “Source coding with side information and a converse for degraded broadcast channels,” IEEE Trans. Inf. Theory, vol. IT-21, no. 6, pp. 629–637, Nov. 1975.
[3] Y. Oohama, “Exponential strong converse for source coding with side information at the decoder,” Entropy, vol. 20, no. 5, p. 352, May 2018.
[4] J. Wolfowitz, “Strong converse of the coding theorem for semicontinuous channels,” Illinois J. Math., vol. 3, no. 4, pp. 477–489, Dec. 1959.
[5] R. Ahlswede, P. Gács, and J. Körner, “Bounds on conditional probabilities with applications in multi-user communication,” Zeitschrift Für Wahrscheinlichkeitstheorie und Verwandte Gebiete, vol. 34, pp. 157–177, Jun. 1976.
[6] S. Arimoto, “On the converse to the coding theorem for discrete memoryless channels (Corresp.),” IEEE Trans. Inf. Theory, vol. IT-19, no. 3, pp. 357–359, May 1973.
[7] G. Dueck and J. Körner, “Reliability function of a discrete memoryless channel at rates above capacity (Corresp.),” IEEE Trans. Inf. Theory, vol. IT-25, no. 1, pp. 82–84, Jan. 1979.
[8] Y. Oohama and T. S. Han, “Universal coding for the Slepian–Wolf data compression system and the strong converse theorem,” IEEE Trans. Inf. Theory, vol. 40, no. 6, pp. 1908–1919, Nov. 1994.
[9] W. Gu and M. Effros, “A strong converse for a collection of network source coding problems,” in Proc. IEEE Int. Symp. Inf. Theory, Jun. 2009, pp. 2316–2320.
[10] Y. Oohama, “Exponential strong converse for one helper source coding problem,” Entropy, vol. 21, no. 6, p. 567, Jun. 2019.
[11] Y. Oohama, “Exponent function for asymmetric broadcast channels at rates outside the capacity region,” in Proc. Int. Symp. Inf. Theory Appl. (ISITA), Oct. 2016, pp. 537–541.
[12] H. Tyagi and S. Watanabe, “Strong converse using change of measure arguments,” IEEE Trans. Inf. Theory, vol. 66, no. 2, pp. 689–703, Feb. 2020.
[13] S. Watanabe, “A converse bound on Wyner-Ahlswede-Körner network via Gray-Wyner network,” in Proc. IEEE Info. Theory Workshop (ITW), Nov. 2017, pp. 81–85.
[14] S. Sreekumar and D. Gündüz, “Strong converse for testing against independence over a noisy channel,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun. 2020, pp. 1283–1288.
[15] M. Ahmadipour, M. Wigger, and S. Shamai (Shitz), “Strong converses for memoryless bi-static ISAC;,” in Proc. IEEE Int. Symp. Inf. Theory (ISIT), Jun. 2023, pp. 1818–1823.
[16] M. Hamad, M. Wigger, and M. Sarkiss, “Strong converses using typical changes of measures and asymptotic Markov chains;,” IEEE Trans. Inf. Theory, vol. 70, no. 3, pp. 1720–1737, Mar. 2024.
[17] S. Watanabe and Y. Oohama, “Privacy amplification theorem for bounded storage eavesdropper,” in Proc. IEEE Inf. Theory Workshop, Sep. 2012, pp. 177–181.
[18] S. Vembu and S. Verdu, “Generating random bits from an arbitrary source: Fundamental limits,” IEEE Trans. Inf. Theory, vol. 41, no. 5, pp. 1322–1332, Sep. 1995.
[19] T. S. Han, Information-Spectrum Methods in Information Theory. Cham, Switzerland: Springer, 2003.
[20] M. Hayashi, “Second-order asymptotics in fixed-length source coding and intrinsic randomness,” IEEE Trans. Inf. Theory, vol. 54, no. 10, pp. 4619–4637, Oct. 2008.
[21] Y. Oohama, “Intrinsic randomness problem in the framework of Slepian–Wolf separate coding system,” IEICE Trans. Fundam. Electron., Commun. Comput. Sci., vol. E90-A, no. 7, pp. 1406–1417, Jul. 2007.
[22] B. Santoso and Y. Oohama, “Information theoretic security for Shannon cipher system under side-channel attacks,” Entropy, vol. 21, no. 5, pp. 469, May 2019.
[23] O. Kosut and J. Kliewer, “Strong converses are just edge removal properties,” IEEE Trans. Inf. Theory, vol. 65, no. 6, pp. 3315–3339, Jun. 2019.
[24] T. S. Jose and A. A. Kulkarni, “Linear programming based finite blocklength converses for some network-like problems,” in Proc. IEEE Inf. Theory Workshop (ITW), Nov. 2017, pp. 544–548.
[25] T. S. Jose and A. A. Kulkarni, “Improved finite blocklength converses for Slepian–Wolf coding via linear programming,” IEEE Trans. Inf. Theory, vol. 65, no. 4, pp. 2423–2441, Apr. 2019.
Daisuke Takeuchi (Graduate Student Member, IEEE) received the B.E. and M.E. degrees from Tokyo University of Agriculture and Technology, Japan, in 2021 and 2023, respectively, where he is currently pursuing the Ph.D. degree. His research interests include information theory and cryptography.

Shun Watanabe (Senior Member, IEEE) received the B.E., M.E., and Ph.D. degrees from Tokyo Institute of Technology, Japan, 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, MD, 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. He is a co-author of the book Information-theoretic Cryptography (Cambridge University Press 2023). His current research interests include information theory, cryptography, quantum information, and information geometry. He served as an Associate Editor for IEEE TRANSACTIONS ON INFORMATION THEORY from 2016 to 2020 and a Board of Governors for the IEEE Information Theory Society from 2022 to 2024. He also served as the General Co-Chair for the 2021 IEEE Information Theory Workshop. From 2023 to 2024, he was an IEEE IT Society Distinguished Lecturer.