Strong Converse Exponent for Degraded Broadcast Channels at Rates outside the Capacity Region

Yasutada Oohama
University of Electro-Communications, Tokyo, Japan
Email: oohama@uec.ac.jp

***Abstract*—We consider the discrete memoryless degraded broadcast channels. We prove that the error probability of decoding tends to one exponentially for rates outside the capacity region and derive an explicit lower bound of this exponent function. We shall demonstrate that the information spectrum approach is quite useful for investigating this problem.**

I. The Capacity Region of the Degraded Broadcast Channels

Let $\mathcal{X}, \mathcal{Y}$, $\mathcal{Z}$ be finite sets. The broadcast channel we study in this paper is defined by a discrete memoryless channel specified with the following stochastic matrix:

$$W \stackrel{\triangle}{=} \{W(y,z|x)\}_{(x,y,z)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}}. \quad (1)$$

Here the set $\mathcal{X}$ stands for a set of channel input. The sets $\mathcal{Y}$ and $\mathcal{Z}$ stand for sets of two channel outputs. Let X^n be a random variable taking values in $\mathcal{X}^n$. We write an element of $\mathcal{X}^n$ as $x^n = x_1 x_2 \cdots x_n$. Suppose that X^n has a probability distribution on $\mathcal{X}^n$ denoted by $p_{X^n} = \{p_{X^n}(x^n)\}_{x^n\in\mathcal{X}^n}$. Similar notations are adopted for other random variables. Let $Y^n \in \mathcal{Y}^n$ and $Z^n \in \mathcal{Y}^n$ be random variables obtained as the channel output by connecting X^n to the input of channel. We write a conditional distribution of (Y^n, Z^n) on given X^n as

$$W^n = \{W^n(y^n, z^n|x^n)\}_{(x^n,y^n,z^n)\in\mathcal{X}^n\times\mathcal{Y}^n\times\mathcal{Z}^n}.$$

In this paper we deal with the case where the components $W(z,y|x)$ of W satisfy the following conditions:

$$W(y,z|x) = W_1(y|x)W_2(z|y). \quad (2)$$

In this case we say that the broadcast channel W is *degraded*. The degraded broadcast channel (DBC) is specified by (W_1, W_2). Transmission of messages via the degraded BC is shown in Fig. 1. Let K_n and L_n be uniformly distributed random variables taking values in message sets $\mathcal{K}_n$ and $\mathcal{L}_n$, respectively. The random variable K_n is a message sent to the receiver 1. The random variable L_n is a message sent to the receiver 2. A sender transforms K_n and L_n into a transmitted sequence X^n using an encoder function $\varphi^{(n)}$ and sends it to the receivers 1 and 2. In this paper we assume that the encoder function $\varphi^{(n)}$ is a stochastic encoder. In this case, $\varphi^{(n)}$ is a stochastic matrix given by

$$\varphi^{(n)} = \{\varphi^{(n)}(x^n|k,l)\}_{(k,l,x^n)\in\mathcal{K}_n\times\mathcal{L}_n\times\mathcal{X}^n},$$

where $\varphi^{(n)}(x^n|k,l)$ is a conditional probability of $x^n \in \mathcal{X}^n$ given message pair $(k,l) \in \mathcal{K}_n \times \mathcal{L}_n$. The joint probability

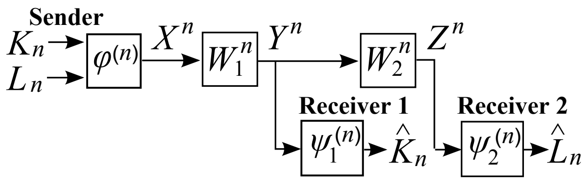

Fig. 1. Transmission of messages via the degraded BC.

mass function on $\mathcal{K}_n \times \mathcal{L}_n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$ is given by

$$\begin{aligned}&\Pr\{(K_n, L_n, X^n, Y^n, Z^n) = (k, l, x^n, y^n, z^n)\}\\ &= \frac{\varphi^{(n)}(x^n|k,l)}{|\mathcal{K}_n||\mathcal{L}_n|}\prod_{t=1}^{n} W_1(y_t|x_t)\, W_2(z_t|y_t),\end{aligned}$$

where $|\mathcal{K}_n|$ is a cardinality of the set $\mathcal{K}_n$. The decoding functions at the receiver 1 and the receiver 2, respectively, are denoted by $\psi_1^{(n)}$ and $\psi_2^{(n)}$. Those functions are formally defined by $\psi_1^{(n)} : \mathcal{Y}^n \to \mathcal{K}_n, \psi_2^{(n)} : \mathcal{Z}^n \to \mathcal{L}_n$. The average error probabilities of decoding at the receivers 1 and 2 are defined by

$$\begin{aligned}\mathrm{P}_{\mathrm{e},1}^{(n)} &= \mathrm{P}_{\mathrm{e}}^{(n)}(\varphi^{(n)}, \psi_1^{(n)}) \stackrel{\triangle}{=} \Pr\{\psi_1^{(n)}(Y^n) \neq K_n\},\\ \mathrm{P}_{\mathrm{e},2}^{(n)} &= \mathrm{P}_{\mathrm{e}}^{(n)}(\varphi^{(n)}, \psi_2^{(n)}) \stackrel{\triangle}{=} \Pr\{\psi_2^{(n)}(Z^n) \neq L_n\}.\end{aligned}$$

Furthermore, we set

$$\begin{aligned}\mathrm{P}_{\mathrm{e}}^{(n)} &= \mathrm{P}_{\mathrm{e}}^{(n)}(\varphi^{(n)}, \psi_1^{(n)}, \psi_2^{(n)})\\ &\stackrel{\triangle}{=} \Pr\{\psi_1^{(n)}(Y^n) \neq K_n \text{ or } \psi_2^{(n)}(Z^n) \neq L_n\}.\end{aligned}$$

It is obvious that we have the following relation.

$$\mathrm{P}_{\mathrm{e}}^{(n)} \leq \mathrm{P}_{\mathrm{e},1}^{(n)} + \mathrm{P}_{\mathrm{e},2}^{(n)}. \quad (3)$$

For $k \in \mathcal{K}_n$ and $l \in \mathcal{L}_n$, set $\mathcal{D}_1(k) \stackrel{\triangle}{=} \{y^n : \psi_1^{(n)}(y^n) = k\}, \mathcal{D}_2(l) \stackrel{\triangle}{=} \{z^n : \psi_2^{(n)}(z^n) = l\}$. The families of sets $\{\mathcal{D}_1(k)\}_{k\in\mathcal{K}_n}$ and $\{\mathcal{D}_2(l)\}_{l\in\mathcal{L}_n}$ are called the decoding regions. Using the decoding region, $\mathrm{P}_{\mathrm{e}}^{(n)}$ can be written as

$$\begin{aligned}\mathrm{P}_{\mathrm{e}}^{(n)} = \frac{1}{|\mathcal{K}_n||\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\ \sum_{\substack{(x^n,y^n,z^n)\in\mathcal{X}^n\times\mathcal{Y}^n\times\mathcal{Z}^n:\\ y^n\in\mathcal{D}_1^c(k) \text{ or } z^n\in\mathcal{D}_2^c(l)}} 1\\ \times\varphi^{(n)}(x^n|k,l)W_1^n(y^n|x^n)W_2^n(z^n|y^n).\end{aligned}$$

Set

$$\mathrm{P}_{\mathrm{c}}^{(n)} = \mathrm{P}_{\mathrm{c}}^{(n)}(\varphi^{(n)}, \psi_1^{(n)}, \psi_2^{(n)}) \stackrel{\triangle}{=} 1 - \mathrm{P}_{\mathrm{e}}^{(n)}(\varphi^{(n)}, \psi_1^{(n)}, \psi_2^{(n)}).$$

The quantity $\mathrm{P}_{\mathrm{c}}^{(n)}$ is called the average correct probability of decoding. For given $(\varepsilon_1,\varepsilon_2)\in(0,1)^2$, a pair (R_1,R_2) is $(\varepsilon_1,\varepsilon_2)$-*achievable* if there exists a sequence of triples $\{(\varphi^{(n)}, \psi_1^{(n)},\psi_2^{(n)})\}_{n=1}^{\infty}$ such that

$$\begin{aligned}&\mathrm{P}_{\mathrm{e},i}^{(n)}(\varphi^{(n)},\psi_i^{(n)})\leq\varepsilon_i, i=1,2,\\&\liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{K}_n|\geq R_1,\liminf_{n\to\infty}\ \frac{1}{n}\log|\mathcal{L}_n|\geq R_2.\end{aligned}$$

The set that consists of all $(\varepsilon_1,\varepsilon_2)$-achievable rate pair is denoted by $\mathcal{C}_{\mathrm{DBC}}(\varepsilon_1,\varepsilon_2|W_1,W_2)$, which is called the capacity region of the DBC. We can define another capacity region based on the error probability $\mathrm{P}_{\mathrm{e}}^{(n)}(\varphi^{(n)},\ \psi_1^{(n)},\ \psi_2^{(n)})$. For given $\varepsilon\in(0,1)$, a pair (R_1,R_2) is ε-*achievable* if there exists a sequence of triples $\{(\varphi^{(n)},\ \psi_1^{(n)},\psi_2^{(n)})\}_{n=1}^{\infty}$ such that

$$\begin{aligned}&\mathrm{P}_{\mathrm{e}}^{(n)}(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)})\leq\varepsilon,\\&\liminf_{n\to\infty}\frac{1}{n}\log|\mathcal{K}_n|\geq R_1,\liminf_{n\to\infty}\ \frac{1}{n}\log|\mathcal{L}_n|\geq R_2.\end{aligned}$$

The set that consists of all ε-achievable rate pair is denoted by $\mathcal{C}_{\mathrm{DBC}}(\varepsilon|W_1,W_2)$. It is obvious that for $0<\varepsilon_1+\varepsilon_2\leq 1$, we have

$$\mathcal{C}_{\mathrm{DBC}}(\varepsilon_1,\varepsilon_2|W_1,W_2)\subseteq\mathcal{C}_{\mathrm{DBC}}(\varepsilon_1+\varepsilon_2|W_1,W_2).$$

We set

$$\mathcal{C}_{\mathrm{DBC}}(W_1,W_2)\stackrel{\triangle}{=}\bigcap_{\varepsilon\in(0,1)}\mathcal{C}_{\mathrm{DBC}}(\varepsilon|W_1,W_2),$$

which is called the capacity region of the DBC. The two maximum error probabilities of decoding are defined by as follows:

$$\begin{aligned}\mathrm{P}_{\mathrm{e,m,1}}^{(n)}&=\mathrm{P}_{\mathrm{e,m,1}}^{(n)}(\varphi^{(n)},\psi_1^{(n)})\\&\stackrel{\triangle}{=}\max_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\Pr\{\psi_1^{(n)}(Y^n)\neq k|K_n=k\},\\\mathrm{P}_{\mathrm{e,m,2}}^{(n)}&=\mathrm{P}_{\mathrm{e,m,2}}^{(n)}(\varphi^{(n)},\psi_2^{(n)})\\&\stackrel{\triangle}{=}\max_{l\in\mathcal{L}_n}\Pr\{\psi_2^{(n)}(Z^n)\neq l|L_n=l\}.\end{aligned}$$

Based on those quantities, we define the maximum capacity region $\mathcal{C}_{\mathrm{m,DBC}}(\varepsilon_1,\varepsilon_2|W_1,W_2)$ in a manner quite similar to the definition of $\mathcal{C}_{\mathrm{DBC}}(\varepsilon_1,\varepsilon_2|W_1,W_2)$. To describe previous works on $\mathcal{C}_{\mathrm{DBC}}(W_1,W_2)$ and $\mathcal{C}_{\mathrm{m,DBC}}(\varepsilon_1,\varepsilon_2|W_1,W_2)$, we introduce an auxiliary random variable U taking values in a finite set $\mathcal{U}$. We assume that the joint distribution of (U,X,Y,Z) is

$$p_{UXYZ}(u,x,y,z)=p_U(u)p_{X|U}(x|u)W_1(y|x)W_2(z|y).$$

The above condition is equivalent to $U\leftrightarrow X\leftrightarrow Y\leftrightarrow Z$. Define the set of probability distribution $p=p_{UXYZ}$ of $(U, X, Y, Z)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ by

$$\begin{aligned}&\mathcal{P}(W_1,W_2)\stackrel{\triangle}{=}\{p:|\mathcal{U}|\leq|\mathcal{X}|+1,\\&\quad p_{Y|X}=W_1,p_{Z|Y}=W_2,U\leftrightarrow X\leftrightarrow Y\leftrightarrow Z\}.\end{aligned}$$

Set

$$\begin{aligned}\mathcal{C}(p)&\stackrel{\triangle}{=}\{(R_1,R_2):R_1,R_2\geq 0\,,\\&\qquad R_1\leq I_p(X;Y|U),R_2\leq I_p(U;Z)\}.\\\mathcal{C}(W_1,W_2)&=\bigcup_{p\in\mathcal{P}(W_1,W_2)}\mathcal{C}(p).\end{aligned}$$

We can show that the above functions and sets satisfy the following property.

Property 1:

a) The region $\mathcal{C}(W_1,W_2)$ is a closed convex set of The region $\mathcal{C}(W_1,W_2)$ is a closed convex subset of $\mathbb{R}_+^2$, where

$$\mathbb{R}_+^2\stackrel{\triangle}{=}\{(R_1,R_2):R_1\geq 0,R_2\geq 0\}.$$

b) The region $\mathcal{C}(W_1,W_2)$ can be expressed with a family of supporting hyperplanes. To describe this result we define the set of probability distribution $p=p_{UXYZ}$ of $(U, X, Y, Z)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ by

$$\begin{aligned}&\mathcal{P}_{\mathrm{sh}}(W_1,W_2)\stackrel{\triangle}{=}\{p:|\mathcal{U}|\leq|\mathcal{X}|,\\&\quad p_{Y|X}=W_1,p_{Z|Y}=W_2,U\leftrightarrow X\leftrightarrow Y\leftrightarrow Z\}.\end{aligned}$$

We set

$$\begin{aligned}&C^{(\mu)}(W_1,W_2)\\&\stackrel{\triangle}{=}\max_{p\in\mathcal{P}_{\mathrm{sh}}(W_1,W_2)}\{\mu I_p(X;Y|U)+I_p(U;Z)\}\,,\\&\mathcal{C}_{\mathrm{sh}}(W_1,W_2)\\&=\bigcap_{\mu>0}\{(R_1,R_2):\mu R_1+R_2\leq C^{(\mu)}(W_1,W_2)\}.\end{aligned}$$

Then we have the following

$$\mathcal{C}(W_1,W_2)=\mathcal{C}_{\mathrm{sh}}(W_1,W_2).$$

Property 1 is a well known result. We omit the proof of this property. The broadcast channel was posed and investigated by Cover [1]. Bergmans [2] proved that $\mathcal{C}(W_1,W_2)$ serves as an inner bound of $\mathcal{C}_{\mathrm{DBC}}(W_1,W_2)$. Gallager [3], Ahlswede and Körner [4], proved that the inner bound $\mathcal{C}(W_1,W_2)$ is tight, thereby establishing the following theorem.

Theorem 1 (Gallager [3],Ahlswede and Körner [4]): For any DBC (W_1,W_2), we have

$$\mathcal{C}_{\mathrm{DBC}}(W_1,W_2)=\mathcal{C}(W_1,W_2).$$

The strong converse theorem was proved by Ahlswede *et al.* [5]. Their result is the following:

Theorem 2 (Ahlswede et al. [5]): For each fixed $(\varepsilon_1,\varepsilon_2)\in(0,1)^2$ and any DBC (W_1,W_2), we have

$$\mathcal{C}_{\mathrm{m,DBC}}(\varepsilon_1,\varepsilon_2|W_1,W_2)=\mathcal{C}_{\mathrm{DBC}}(W_1,W_2).$$

Their method used to prove the strong converse theorem was extended to the method called the image size characterization by Csiszár and Körner [6].

To examine an asymptotic behavior of $\mathrm{P}_{\rm c}^{(n)}$ for rates outside the capacity region $\mathcal{C}(W_1,W_2)$, we define the following quantity.

$$
\begin{aligned}
&G^{(n)}(R_1,R_2|W_1,W_2)\\
&\stackrel{\triangle}{=} \min_{\substack{(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)}):\\ (1/n)\log|\mathcal{K}_n|\geq R_1,\\ (1/n)\log|\mathcal{L}_n|\geq R_2}} \left(-\frac{1}{n}\right)\log \mathrm{P}_{\rm c}^{(n)}(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)}),\\
&G(R_1,R_2|W_1,W_2) \stackrel{\triangle}{=} \lim_{n\to\infty} G^{(n)}(R_1,R_2|W_1,W_2).
\end{aligned}
$$

Our main aim is to find an explicit In this paper we derive an explicit lower bound of $G(R_1,R_2|(W_1,W_2)$ that is positive if and only if $(R_1,R_2)\notin\mathcal{C}(W_1,W_2)$.

II. Main Result

In this section we state our main result. Define

$$
\begin{aligned}
&\omega_q^{(\mu)}(x,y,z|u)\\
&\stackrel{\triangle}{=} \mu\log\frac{q_{Y|X}(y|x)}{q_{Y|U}(y|u)}+\log\frac{q_{Z|U}(z|u)}{q_Z(z)},\\
&\Lambda_q^{(\mu,\lambda)}(XYZ|U)\\
&\stackrel{\triangle}{=} \sum_{(u,x,y,z)\in\mathcal{U}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}} q_{UX}(u,x)q_{Y|X}(y|x)q_{Z|Y}(z|y)\\
&\quad\times\exp\left\{\lambda\omega_q^{(\mu)}(x,y,z|u)\right\},\\
&\Omega_q^{(\mu,\lambda)}(XYZ|U)\stackrel{\triangle}{=}\log\Lambda_q^{(\mu,\lambda)}(XYZ|U),\\
&\Omega^{(\mu,\lambda)}(W_1,W_2)\stackrel{\triangle}{=}\max_{q\in\mathcal{P}_{\rm sh}(W_1,W_2)}\Omega_q^{(\mu,\lambda)}(XYZ|U),\\
&F^{(\mu,\lambda)}(\mu R_1+R_2|W_1,W_2)\\
&\stackrel{\triangle}{=}\frac{\lambda(\mu R_1+R_2)-\Omega^{(\mu,\lambda)}(W_1,W_2)}{1+2\lambda+\lambda\mu},\\
&F(R_1,R_2|W_1,W_2)\\
&\stackrel{\triangle}{=}\sup_{\mu,\lambda>0}F^{(\mu,\lambda)}(\mu R_1+R_2|W_1,W_2).
\end{aligned}
$$

We can show that the above functions and sets satisfy the following property.

Property 2:

a) For each $q\in\mathcal{P}(W_1,W_2)$, $\Omega_q^{(\mu,\lambda)}(XYZ|U)$ is a monotone increasing and convex function of $\lambda>0$.

b) For every $q\in\mathcal{P}_{\rm sh}(W_1,W_2)$, we have

$$
\begin{aligned}
&\lim_{\lambda\to+0}\frac{\Omega_q^{(\mu,\lambda)}(XYZ|U)}{\lambda}\\
&=\mu I_q(X;Y|U)+I_q(U;Z).
\end{aligned}
$$

c) If $(R_1,R_2)\notin\mathcal{C}(W_1,W_2)$, then we have $F(R_1,R_2|W_1,W_2)>0$.

Proof of Property 2 is given in Appendix B. Our main result is the following.

Theorem 3: For any degraded BC (W_1,W_2), we have

$$
G(R_1,R_2|W_1,W_2)\geq F(R_1,R_2|W_1,W_2). \tag{4}
$$

Proof of this theorem will be given in Section III. It follows from Theorem 3 and Property 2 part c) that if (R_1,R_2) is outside the capacity region, then the error probability of decoding goes to one exponentially and its exponent is not below $F(R_1,R_2|W_1,W_2)$. From this theorem we immediately obtain the following corollary, which partially recovers the strong converse theorem by Ahlswede *et al.* [5].

Corollary 1: For each pair $(\varepsilon_1,\varepsilon_2)\in(0,1)^2$ satisfying $\varepsilon_1+\varepsilon_2<1$, we have

$$
\begin{aligned}
&\mathcal{C}_{\rm m,DBC}(\varepsilon_1,\varepsilon_2|W_1,W_2)\\
&=\mathcal{C}_{\rm DBC}(\varepsilon_1,\varepsilon_2|W_1,W_2)=\mathcal{C}_{\rm DBC}(\varepsilon_1+\varepsilon_2|W_1,W_2)\\
&=\mathcal{C}_{\rm DBC}(W_1,W_2)=\mathcal{C}(W_1,W_2).
\end{aligned}
$$

In particular, for each $\varepsilon\in(0,1/2)$, we have

$$
\mathcal{C}_{\rm DBC}(\varepsilon,\varepsilon|W)=\mathcal{C}_{\rm DBC}(2\varepsilon|W_1,W_2)=\mathcal{C}(W_1,W_2).
$$

The exponent function at rates outside the channel capacity was derived by Arimoto [7] and Dueck and Körner [8]. The techniques used by them are not useful to prove Theorem 3. Some novel techniques based on the information spectrum method introduced by Han [9] are necessary to prove this theorem.

III. Proof of the Results

We first prove the following lemma.

Lemma 1: For any $\eta>0$ and for any $(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)})$ satisfying $(1/n)\log|\mathcal{K}_n|\geq R_1$, $(1/n)\log|\mathcal{L}_n|\geq R_2$. we have

$$
\begin{aligned}
&\mathrm{P}_{\rm c}^{(n)}(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)})\leq p_{L_nX^nY^nZ^n}\Biggl\{\\
&R_1\leq\frac{1}{n}\log\frac{W_1^n(Y^n|X^n)W_2^n(Z^n|Y^n)}{q_{Y^nZ^n|L_n}(Y^n,Z^n|L_n)}+\eta \qquad (5)\\
&R_2\leq\frac{1}{n}\log\frac{p_{Z^n|L_n}(Z^n|L_n)}{\tilde{q}_{Z^n}(Z^n)}+\eta\Biggr\}+2{\rm e}^{-n\eta}. \qquad (6)
\end{aligned}
$$

In (5), we can choose any conditional distribution $q_{Y^nZ^n|L_n}$ on $\mathcal{Y}^n\times\mathcal{Z}^n$ given $L_n\in\mathcal{L}_n$. In (6) we can choose any probability distribution $\tilde{q}_{Z^n}$ on $\mathcal{Z}^n$.

Proof of this lemma is given in Appendix C.

For $t=1,2,\cdots,n$, set

$$
\begin{aligned}
&\mathcal{U}_t\stackrel{\triangle}{=}\mathcal{L}_n\times\mathcal{Y}^{t-1}\times\mathcal{Z}^{t-1},\mathcal{V}_t\stackrel{\triangle}{=}\mathcal{L}_n\times\mathcal{Z}^{t-1},\\
&U_t\stackrel{\triangle}{=}(L_n,Y^{t-1},Z^{t-1})\in\mathcal{U}_t,V_t\stackrel{\triangle}{=}(L_n,Z^{t-1})\in\mathcal{V}_t,\\
&u_t\stackrel{\triangle}{=}(l,y^{t-1},z^{t-1})\in\mathcal{U}_t,v_t\stackrel{\triangle}{=}(l,z^{t-1})\in\mathcal{V}_t.
\end{aligned}
$$

For each $t=1,2\cdots,l$, let κ_t be a natural projection from $\mathcal{U}_t$ onto $\mathcal{V}_t$. Using κ_t, we have $V_t=\kappa_t(U_t)$, $t=1,2,\cdots,n$. For each $t=1,2,\cdots,n$, let $\mathcal{Q}(\mathcal{U}_t\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z})$ be a set of all probability distributions on

$$
\mathcal{U}_t\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}=\mathcal{L}_n\times\mathcal{X}\times\mathcal{Y}^t\times\mathcal{Z}^t.
$$

For $t=1,2,\cdots,n$, we simply write $\mathcal{Q}_t=\mathcal{Q}(\mathcal{U}_t\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z})$. Similarly, for $t=1,2,\cdots,n$, we simply write $q_t=q_{U_tX_tY_tZ_t}\in\mathcal{Q}_t$. Set

$$
\begin{aligned}
&\mathcal{Q}^n\stackrel{\triangle}{=}\prod_{t=1}^n\mathcal{Q}_t=\prod_{t=1}^n\mathcal{Q}(\mathcal{U}_t\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}),\\
&q^n\stackrel{\triangle}{=}\{q_t\}_{t=1}^n\in\mathcal{Q}^n.
\end{aligned}
$$

By Lemma 1 and some computations we have the following lemma.

Lemma 2: For any $\eta > 0$, for any $(\varphi^{(n)}, \psi_1^{(n)}, \psi_2^{(n)})$ satisfying $(1/n)\log|\mathcal{K}_n| \geq R_1$, $(1/n)\log|\mathcal{L}_n| \geq R_2$, and for any $q^n \in \mathcal{Q}^n$, we have

$$\begin{aligned}
&\mathrm{P}_{\rm c}^{(n)}(\varphi^{(n)}, \psi_1^{(n)}, \psi_2^{(n)}) \leq p_{L_n X^n Y^n Z^n}\Bigg\{ \\
&R_1 \leq \frac{1}{n}\sum_{t=1}^{n} \log \frac{W_1(Y_t|X_t)}{q_{Y_t|U_t}(Y_t|U_t)} + \eta, \\
&R_2 \leq \frac{1}{n}\sum_{t=1}^{n} \log \frac{p_{Z_t|V_t}(Z_t|V_t)}{q_{Z_t}(Z_t)} + \eta \Bigg\} + 2\mathrm{e}^{-n\eta}, \qquad (7)
\end{aligned}$$

where for each $t = 1, 2, \cdots, n$, the conditional probability distribution $q_{Y_t|U_t}$ and the probability distribution q_{Z_t} appearing in the first term in the right members of (7) are chosen so that they are induced by the joint distribution $q_t = q_{U_t X_t Y_t Z_t} \in \mathcal{Q}_t$.

Proof of this lemma is given in Appendix D.

To evaluate an upper bound of (7) in Lemma 2. We use the following lemma, which is well known as the Cramèr's bound in the large deviation principle.

Lemma 3: For any real valued random variable Z and any $\theta > 0$, we have

$$\Pr\{Z \geq a\} \leq \exp\left[-\left(\lambda a - \log \mathrm{E}[\exp(\theta Z)]\right)\right].$$

Here we define a quantity which serves as an exponential upper bound of $\mathrm{P}_{\rm c}^{(n)}(\varphi^{(n)}, \psi_1^{(n)}, \psi_2^{(n)})$. Let $\mathcal{P}^{(n)}(W_1, W_2)$ be a set of all probability distributions $p_{L_n X^n Y^n Z^n}$ on $\mathcal{L}_n \times \mathcal{X}^n \times \mathcal{Y}^n \times \mathcal{Z}^n$ having the form:

$$\begin{aligned}
&p_{L^n X^n Y^n Z^n}(l, x^n, y^n, z^n) \\
&= p_{L^n}(l) \prod_{t=1}^{n} p_{X_t|L_n X^{t-1}}(x_t|l, x^{t-1}) W_1(y_t|x_t) W_2(z_t|y_t).
\end{aligned}$$

For simplicity of notation we use the notation $p^{(n)}$ for $p_{L_n X^n Y^n Z^n} \in \mathcal{P}^{(n)}(W_1, W_2)$. We assume that $p_{U_t X_t Y_t Z_t} = p_{L_n X_t Y^t Z_t}$ is a marginal distribution of $p^{(n)}$. For $t = 1, 2, \cdots, n$, we simply write $p_t = p_{U_t X_t Y_t Z_t}$. For $p^{(n)} \in \mathcal{P}^{(n)}(W_1, W_2)$ and $q^n \in \mathcal{Q}^n$, we define

$$\begin{aligned}
&\Omega^{(\mu,\theta)}_{p^{(n)}||q^n}(X^n Y^n Z^n|L_n) \\
&\triangleq \log \mathrm{E}_{p^{(n)}}\left[\prod_{t=1}^{n} \frac{W_1^{\theta\mu}(Y_t|X_t) p^{\theta}_{Z_t|V_t}(Z_t|V_t)}{q^{\theta\mu}_{Y_t|U_t}(Y_t|U_t) q^{\theta}_{Z_t}(Z_t)}\right],
\end{aligned}$$

where for each $t = 1, 2, \cdots, n$, the conditional probability distribution $q_{Y_t|U_t}$ and the probability distribution q_{Z_t} appearing in the definition of $\Omega^{(\mu,\theta)}_{p^{(n)}||q^n}(X^n Y^n Z^n|L_n)$ are chosen so that they are induced by the joint distribution $q_t = q_{U_t X_t Y_t Z_t} \in \mathcal{Q}_t$.

Here we give a remark on an essential difference between $p^{(n)} \in \mathcal{P}^{(n)}(W_1, W_2)$ and $q^n \in \mathcal{Q}^n$. For the former the n probability distributions p_t, $t = 1, 2, \cdots, n$, are consistent with $p^{(n)}$, since all of them are marginal distributions of $p^{(n)}$. On the other hand, for the latter, q^n is just *a sequence* of n probability distributions. Hence, we may not have the consistency between the n elements q_t, $t = 1, 2, \cdots, n$, of q^n.

By Lemmas 2 and 3, we have the following proposition.

Proposition 1: For any μ, $\theta > 0$, any $q^n \in \mathcal{Q}^n$, and any $(\varphi^{(n)}, \psi_1^{(n)}, \psi_2^{(n)})$ satisfying

$$\frac{1}{n}\log|\mathcal{K}_n| \geq R_1, \frac{1}{n}\log|\mathcal{L}_n| \geq R_2, \qquad (8)$$

we have

$$\begin{aligned}
&\mathrm{P}_{\rm c}^{(n)}(\varphi^{(n)}, \psi_1^{(n)}, \psi_2^{(n)}) \\
&\leq 3\exp\left\{-n\frac{\theta(\mu R_1 + R_2) - \frac{1}{n}\Omega^{(\mu,\theta)}_{p^{(n)}||q^n}(X^n Y^n Z^n|L_n)}{1 + \theta + \theta\mu}\right\}.
\end{aligned}$$

Proof: Under the condition (8), we have the following chain of inequalities:

$$\begin{aligned}
&\mathrm{P}_{\rm c}^{(n)}(\varphi^{(n)}, \psi_1^{(n)}, \psi_2^{(n)}) \overset{\rm (a)}{\leq} p_{L_n X^n Y^n Z^n}\Bigg\{ \\
&R_1 \leq \frac{1}{n}\sum_{t=1}^{n} \log \frac{W_1(Y_t|X_t)}{q_{Y_t|U_t}(Y_t|U_t)} + \eta, \\
&R_2 \leq \frac{1}{n}\sum_{t=1}^{n} \log \frac{p_{Z_t|V_t}(Z_t|V_t)}{q_{Z_t}(Z_t)} + \eta \Bigg\} + 3\mathrm{e}^{-n\eta} \\
&\leq p_{L_n X^n Y^n Z^n}\Bigg\{\mu R_1 + R_2 - (\mu + 1)\eta \\
&\quad \leq \frac{1}{n}\sum_{t=1}^{n} \log\left[\frac{W_1(Y_t|X_t) p_{Z_t|V_t}(Z_t|V_t)}{q^{\mu}_{Y_t|U_t}(Y_t|U_t) q^{\mu}_{Z_t}(Z_t)}\right]\Bigg\} + 3\mathrm{e}^{-n\eta} \\
&\overset{\rm (b)}{\leq} \exp\Bigg[n\Bigg\{-\theta(\mu R_1 + R_2) + \theta(\mu + 1)\eta \\
&\quad + \frac{1}{n}\Omega^{(\mu,\theta)}_{p^{(n)}||q^n}(X^n Y^n Z^n|L_n)\Bigg\}\Bigg] + 3\mathrm{e}^{-n\eta}. \qquad (9)
\end{aligned}$$

Step (a) follows from Lemma 2. Step (b) follows from Lemma 3. We choose η so that

$$\begin{aligned}
-\eta = &-\theta(\mu R_1 + R_2) + \theta(\mu + 1)\eta \\
&+ \frac{1}{n}\Omega^{(\mu,\theta)}_{p^{(n)}||q^n}(X^n Y^n Z^n|L_n). \qquad (10)
\end{aligned}$$

Solving (10) with respect to η, we have

$$\eta = \frac{\theta(\mu R_1 + R_2) - \frac{1}{n}\Omega^{(\mu,\theta)}_{p^{(n)}||q^n}(X^n Y^n Z^n|L_n)}{1 + \theta + \theta\mu}.$$

For this choice of η and (9), we have

$$\begin{aligned}
&\mathrm{P}_{\rm c}^{(n)} \leq 3\mathrm{e}^{-n\eta} \\
&= 3\exp\left\{-n\frac{\theta(\mu R_1 + R_2) - \frac{1}{n}\Omega^{(\mu,\theta)}_{p^{(n)}||q^n}(X^n Y^n Z^n|L_n)}{1 + \theta + \theta\mu}\right\},
\end{aligned}$$

completing the proof. ∎

Set

$$\begin{aligned}
&\overline{\Omega}^{(\mu,\theta)}(W_1, W_2) \\
&\triangleq \sup_{n \geq 1} \max_{p^{(n)} \in \mathcal{P}^{(n)}(W_1, W_2)} \min_{q^n \in \mathcal{Q}^n} \frac{1}{n}\Omega^{(\mu,\theta)}_{p^{(n)}||q^n}(X^n Y^n Z^n|L_n).
\end{aligned}$$

By the above definition of $G^{(n)}(R_1,\, R_2|W_1,W_2)$ and Proposition 1, we have

$$\begin{aligned}&G^{(n)}(R_1,R_2|W_1,W_2)\\&\geq \frac{\theta(\mu R_1+R_2)-\overline{\Omega}^{(\mu,\theta)}(W_1,W_2)}{1+\theta+\theta\mu}-\frac{1}{n}\log 3.\end{aligned} \tag{11}$$

Then from (11), we obtain the following corollary.

Corollary 2: For any $\theta > 0, \mu > 0$, we have

$$G(R_1,R_2|W_1,W_2) \geq \frac{\theta(\mu R_1+R_2)-\overline{\Omega}^{(\mu,\theta)}(W_1,W_2)}{1+\theta+\theta\mu}.$$

We shall call $\overline{\Omega}^{(\mu,\theta)}(W_1,W_2)$ the communication potential. The above corollary implies that the analysis of $\overline{\Omega}^{(\mu,\theta)}(W_1,W_2)$ leads to an establishment of a strong converse theorem for the degraded BC.

The following proposition is a mathematical core to prove our main result.

Proposition 2: For $\theta\in(0,1)$, set

$$\lambda=\frac{\theta}{1-\theta}\Leftrightarrow\theta=\frac{\lambda}{1+\lambda}. \tag{12}$$

Then, for any $\theta\in(0,1)$, we have

$$\overline{\Omega}^{(\mu,\theta)}(W_1,W_2)\leq\frac{1}{1+\lambda}\Omega^{(\mu,\lambda)}(W_1,W_2).$$

Proof of this proposition is in Appendix E. The proof is not so simple. We must introduce a new method for the proof.

Proof of Theorem 3: For $\theta\in(0,1)$, set

$$\lambda=\frac{\theta}{1-\theta}\Leftrightarrow\theta=\frac{\lambda}{1+\lambda}. \tag{13}$$

Then we have the following:

$$\begin{aligned}&G(R_1,R_2|W_1,W_2)\\&\stackrel{\rm (a)}{\geq}\frac{\theta(\mu R_1+R_2)-\overline{\Omega}^{(\mu,\theta)}(W_1,W_2)}{1+\theta(1+\mu)}\\&\stackrel{\rm (b)}{\geq}\frac{\frac{\lambda}{1+\lambda}(\mu R_1+R_2)-\frac{1}{1+\lambda}\Omega^{(\mu,\lambda)}(W_1,W_2)}{1+\frac{\lambda}{1+\lambda}(1+\mu)}\\&=\frac{\lambda(\mu R_1+R_2)-\Omega^{(\mu,\lambda)}(W_1,W_2)}{1+\lambda+\lambda(1+\mu)}\\&=F^{(\mu,\lambda)}(\mu R_1+R_2|W_1,W_2).\end{aligned}$$

Step (a) follows from Corollary 2. Step (b) follows from Proposition 2 and (13). Since (14) holds for any $\lambda,\mu>0$, we have (4) in Theorem 3. ■

IV. Concluding Remarks

For the DBC, we have derived an explicit lower bound of the optimal exponent function on the correct probability of decoding for rates outside the capacity region. Our method for the DBC can also be applied to the derivation of an explicit lower bound of the optimal exponent function outside the capacity region for the asymmetric broadcast channels(ABCs)(or said the broadcast channels with degraded message sets) investigated by [6], [10]-[12]. In fact the author [13] succeeded deriving an explicit lower bound of the exponent function that is positive for rates outside the capacity region of the ABC. In the case of ABC, some additional techniques are also needed.

References

[1] T. M. Cover, "Broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-18, no.1, pp. 2–13, Jan. 1972.

[2] P. P. Bergmans, "Random coding theorems for broadcast channels with degraded components," *IEEE Trans. Inform. Theory*, vol. IT-19, pp. 197-207, Mar. 1973.

[3] R. G. Gallager, "Capacity and coding for degraded broadcast channels," *Problemy Peredachi Informatsii*, vol. 10, pp. 3-14, July-Sept. 1974.

[4] R. F. Ahlswede and J. Körner, "Source coding with side information and a converse for degraded broadcast channels," *IEEE Trans. Inform. Theory*, vol. IT-21, pp. 629-637, Nov. 1975.

[5] R. Ahlswede, P. Gàs, and J. Körner, "Bounds on conditional probabilities with applications in multi-user communication," *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, vol. 34, pp. 157-177, 1976.

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

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

[8] G. Dueck and J. Körner, "Reliability function of a discrete memoryless channel at rates above capacity," *IEEE Trans. Inform. Theory*, vol. IT-25, no. 1, pp. 82–85, Jan. 1979.

[9] T. S. Han, *Information-Spectrum Methods in Information Theory.* Springer-Verlag, Berlin, New York, 2002. The Japanese edition was published by Baifukan-publisher, Tokyo, 1998.

[10] J. Körner and K. Marton, "General broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. IT-23, no. 1, pp. 60-64, Jan 1977.

[11] J. Körner and A. Sgarro, "Universally attainable error exponents for broadcast channels with degraded message sets," *IEEE Trans. Inform. Theory*, vol. IT-26, no. 6, pp. 670-679, Nov. 1980.

[12] Y. Kaspi and N. Merhav, "Error exponents for broadcast channels with degraded Message sets," *IEEE Trans. Inform. Theory*, vol. 57, no. 1, pp.101-123, Jan. 2011.

[13] Y. Oohama, "New converse for asymmetric broadcast channels," *Proceedings of the 36th Symposium on Information Theory and its Applications(SITA2013)*, pp. 273–278, Ito, Shizuoka, Japan, Nov. 26-29, 2013.

Appendix

A. Cardinality Bound on Auxiliary Random Variables

We have the following lemma.

Lemma 4: For each integer $n\geq 2$, we define

$$\begin{aligned}&\hat{\Omega}_n^{(\mu,\lambda)}(W_1,W_2)\\&\triangleq\max_{\substack{q=q_{UXYZ}:U\leftrightarrow X\leftrightarrow Y\leftrightarrow Z,\\ q_{Y|X}=W_1,q_{Z|Y}=W_2,\\ |\mathcal{U}|\leq|\mathcal{L}_n||\mathcal{Y}|^{n-1}}}\\&\Omega^{(\mu,\lambda)}(W_1,W_2)\\&\triangleq\max_{\substack{q=q_{UXYZ}:U\leftrightarrow X\leftrightarrow Y\leftrightarrow Z,\\ q_{Y|X}=W_1,q_{Z|Y}=W_2,\\ |\mathcal{U}|\leq|\mathcal{X}|}}\Omega_q^{(\mu,\lambda)}(XYZ|U).\end{aligned}$$

Then we have

$$\hat{\Omega}^{(\mu,\lambda)}(W_1,W_2)=\Omega^{(\mu,\lambda)}(W_1,W_2).$$

Proof: We bound the cardinality $|\mathcal{U}|$ of U to show that the bound $|\mathcal{U}|\leq|\mathcal{X}|$ is sufficient to describe $\hat{\Omega}_n^{(\mu,\lambda)}(W_1,W_2)$ and

$\tilde{\Omega}_n^{(\mu,\lambda)}$ (W_1, W_2). Observe that

$$q_X(x) = \sum_{u\in\mathcal{U}} q_U(u) q_{X|U}(x|u), \tag{14}$$

$$\Lambda_q^{(\mu,\lambda)}(XYZ|U) = \sum_{u\in\mathcal{U}} q_U(u)\zeta^{(\mu,\lambda)}(q_{X|U}(\cdot|u)), \tag{15}$$

where

$$\begin{aligned}&\zeta^{(\mu,\lambda)}(q_{X|U}(\cdot|u))\\ \triangleq &\sum_{(x,y,z)\in\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}} q_{X|U}(x|u)W_1(y|x)W_2(z|y)\\ &\times \exp\left\{\lambda\omega_q^{(\mu)}(x,y,z|u)\right\}\end{aligned}$$

For the quantities $q_Z(\cdot)$ contained in the forms of $\zeta^{(\mu,\lambda)}$ $(q_{X|U}(\cdot|u))$, $u \in \mathcal{U}$, we regard them as constants under (14). For each $u \in \mathcal{U}$, $\zeta^{(\mu,\lambda)}(q_{X|U}(\cdot|u))$ are continuous functions of $q_{X|U}(\cdot|u)$. Then by the support lemma,

$$|\mathcal{U}| \le |\mathcal{X}| - 1 + 1 = |\mathcal{X}|$$

is sufficient to express $|\mathcal{X}| - 1$ values of (14) and one value of (15). ∎

B. *Proof of Property 2*

In this appendix we prove Property 2.

Proof of Property 2: We first prove part a) and b). For simplicity of notations, set

$$\begin{aligned}&\underline{a} \triangleq (u,x,y,z), \underline{A} \triangleq (U,X,Y,Z), \underline{\mathcal{A}} \triangleq \mathcal{U}\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z},\\ &\omega_q^{(\mu)}(x,y,z|u) \triangleq \rho(\underline{a}), \Omega_q^{(\mu,\lambda)}(XYZ|U) \triangleq \xi(\lambda).\end{aligned}$$

Then we have

$$\Omega_q^{(\mu,\lambda)}(XYZ|U) = \xi(\lambda) = \log\left[\sum_{\underline{a}\in\underline{\mathcal{A}}} q_{\underline{A}}(\underline{a}) \mathrm{e}^{\lambda\rho(\underline{a})}\right].$$

By simple computations we have

$$\xi'(\lambda) = \mathrm{e}^{-\xi(\lambda)}\left[\sum_{\underline{a}} q_{\underline{A}}(\underline{a})\rho(\underline{a})\mathrm{e}^{\lambda\rho(\underline{a})}\right], \tag{16}$$

$$\begin{aligned}&\xi''(\lambda) = \mathrm{e}^{-2\xi(\lambda)}\\ &\times\left[\sum_{\underline{a},\underline{b}\in\underline{\mathcal{A}}} q_{\underline{A}}(\underline{a})q_{\underline{A}}(\underline{b})\frac{\{\rho(\underline{a})-\rho(\underline{b})\}^2}{2}\mathrm{e}^{\lambda\{\rho(\underline{a})+\rho(\underline{b})\}}\right].\end{aligned} \tag{17}$$

From (17), it is obvious that $\xi''(\lambda)$ is nonnegative. Hence $\Omega_q^{(\mu,\lambda)}(XYZ|U)$ is a convex function of λ. It follows from (16) that for each $q \in \mathcal{P}(W_1, W_2)$, we have

$$\begin{aligned}\xi'(0) &= \sum_{\underline{a}} q_{\underline{A}}(\underline{a})\rho(\underline{a})\\ &= \mu I_q(X;Y|U) + I_q(U;Z) \ge 0.\end{aligned} \tag{18}$$

Hence we have the part b). Since $\xi'(0) \ge 0$ and $\xi''(\lambda) \ge 0$, we have $\xi'(\lambda) \ge 0$ for $\lambda > 0$. Hence for each $q \in \mathcal{P}(W_1, W_2)$, $\Omega_q^{(\mu,\lambda)}(XYZ|U)$ is monotone increasing for $\lambda > 0$. Next we prove the part c). We assume that $(R_1, R_2) \notin \mathcal{C}(W_1, W_2)$, then by Property 1 part b), there exist $\mu^* > 0$ and $\epsilon > 0$, such that

$$\mu^* R_1 + R_2 \ge C^{(\mu^*)}(W_1, W_2) + \epsilon. \tag{19}$$

Set

$$\zeta(\lambda) \triangleq \xi(\lambda) - \lambda\left[I_q(X;Y|U) + I_q(U;Z) + \frac{\epsilon}{2}\right].$$

Then we have the following:

$$\zeta(0) = 0, \zeta'(0) = -\frac{\epsilon}{2}, \zeta''(\lambda) = \xi''(\lambda) \ge 0. \tag{20}$$

It follows from (20) that there exists $\nu(\epsilon) > 0$ such that we have $\zeta(\lambda) \le 0$ for $\lambda \in (0, \nu(\epsilon)]$. Hence for any $\lambda \in (0, \nu(\epsilon)]$, for any $\mu \ge 0$, and for every $q \in \mathcal{P}_{\rm sh}(W_1, W_2)$, we have

$$\Omega_q^{(\mu,\lambda)}(UXYZ) \le \lambda\left(\mu I_q(X;Y|U) + I_q(U;Z) + \frac{\epsilon}{2}\right). \tag{21}$$

From (21), we have that for any $\lambda \in (0, \nu(\epsilon)]$ and for any $\mu > 0$,

$$\begin{aligned}&\Omega^{(\mu,\lambda)}(W_1, W_2)\\ &= \max_{q\in\mathcal{P}_{\rm sh}(W_1,W_2)} \Omega_q^{(\mu,\lambda)}(UXYZ)\\ &\le \lambda\left[\max_{q\in\mathcal{P}_{\rm sh}(W_1,W_2)}\{\mu I_q(X;Y|U) + I_q(U;Z)\} + \frac{\epsilon}{2}\right]\\ &= \lambda\left[\max_{q\in\mathcal{P}_{\rm sh}(W_1,W_2)}\{\mu I_q(X;Y|U) + I_q(U;Z)\} + \frac{\epsilon}{2}\right]\\ &= \lambda\left[C^{(\mu)}(W_1, W_2) + \frac{\epsilon}{2}\right].\end{aligned} \tag{22}$$

Under (19) and (22), we have the following chain of inequalities:

$$\begin{aligned}&F(R_1, R_2|W_1, W_2)\\ &= \sup_{\lambda>0}\sup_{\mu>0} F^{(\mu,\lambda)}(\mu R_1 + R_2|W_1, W_2)\\ &\ge \sup_{\lambda\in(0,\nu(\epsilon)]} F^{(\mu^*,\lambda)}(\mu^* R_1 + R_2|W_1, W_2)\\ &= \sup_{\lambda\in(0,\nu(\epsilon)]} \frac{\lambda(\mu^* R_1 + R_2) - \Omega^{(\mu^*,\lambda)}(W_1, W_2)}{1 + 2\lambda + \lambda\mu^*}\\ &\overset{\rm (a)}{\ge} \sup_{\lambda\in(0,\nu(\epsilon)]} \lambda\frac{\mu^* R_1 + R_2 - C^{(\mu^*)}(W_1, W_2) - \frac{\epsilon}{2}}{1 + 2\lambda + \lambda\mu^*}\\ &\overset{\rm (b)}{\ge} \sup_{\lambda\in(0,\nu(\epsilon)]} \frac{1}{2}\cdot\frac{\lambda\epsilon}{1 + 2\lambda + \lambda\mu^*}\\ &= \frac{1}{2}\cdot\frac{\nu(\epsilon)\epsilon}{1 + 2\nu(\epsilon) + \nu(\epsilon)\mu^*} > 0.\end{aligned}$$

Step (a) follows from (22). Step (b) follows from (19). ∎

C. *Proof of Lemma 1*

In this appendix we prove Lemma 1.

Proof of Lemma 1: For $l \in \mathcal{L}_n$, set

$$\begin{aligned}\mathcal{A}_1(l) &\triangleq \{(x^n, y^n, z^n) : W_2^n(z^n|y^n)W_1^n(y^n|x^n)\\ &\qquad \ge |\mathcal{K}_n|\mathrm{e}^{-n\eta} q_{Y^nZ^n|L_n}(y^n, z^n|l)\},\\ \mathcal{A}_2(l) &\triangleq \{(x^n, y^n, z^n) : p_{Z^n|L_n}(z^n|l) \ge |\mathcal{L}_n|\mathrm{e}^{-n\eta}\tilde{q}_{Z^n}(z^n)\},\\ \mathcal{A}(l) &\triangleq \mathcal{A}_1(l) \cap \mathcal{A}_2(l).\end{aligned}$$

Then we have the following:

$$
\begin{aligned}
\mathrm{P}_{\mathrm{c}}^{(n)} &= \frac{1}{|\mathcal{K}_n||\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{\substack{(x^n,y^n,z^n)\in\mathcal{A}(l),\\ y^n\in\mathcal{D}_1(k),z^n\in\mathcal{D}_2(l)}} 1\\
&\quad\times\varphi^{(n)}(x^n|k,l)W_1^n(y^n|x^n)W_2^n(z^n|y^n)\\
&\quad+\frac{1}{|\mathcal{K}_n||\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{\substack{(x^n,y^n,z^n)\in\mathcal{A}^c(l):\\ y^n\in\mathcal{D}_1(k),z^n\in\mathcal{D}_2(l)}} 1\\
&\quad\times\varphi^{(n)}(x^n|k,l)W_1^n(y^n|x^n)W_2^n(z^n|y^n)\\
&\le \sum_{i=0,1,2}\Delta_i,
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta_0 &\triangleq \frac{1}{|\mathcal{K}_n||\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{(x^n,y^n,z^n)\in\mathcal{A}(l)}\\
&\quad\times p_{X^nY^nZ^n|K_n,L_n}(x^n,y^n,z^n|k,l),\\
\Delta_i &\triangleq \frac{1}{|\mathcal{K}_n||\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{\substack{(x^n,y^n,z^n)\in\mathcal{A}_i^c(l),\\ y^n\in\mathcal{D}_1(k),z^n\in\mathcal{D}_2(l)}} 1\\
&\quad\times p_{X^nY^nZ^n|K_n,L_n}(x^n,y^n,z^n|k,l)\\
&\quad\text{for } i=1,2.
\end{aligned}
$$

By definition we have

$$
\begin{aligned}
&\Delta_0\\
&= p_{L_nX^nY^nZ^n}\Biggl\{\\
&\frac{1}{n}\log|\mathcal{K}_n| \le \frac{1}{n}\log\frac{W_1^n(Y^n|X^n)}{q_{Y^n|L_n}(Y^n|L_n)}+\eta,\\
&\frac{1}{n}\log|\mathcal{L}_n| \le \frac{1}{n}\log\frac{p_{Z^n|L_n}(Z^n|L_n)}{\tilde{q}_{Z^n}(Z^n)}+\eta\Biggr\}. \qquad (23)
\end{aligned}
$$

From (23), it follows that if $(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)})$ satisfies $(1/n)\log|\mathcal{K}_n|\ge R_1$, $(1/n)\log|\mathcal{L}_n|\ge R_2$, then the quantity $\tilde{\Delta}_0$ is upper bounded by the first term in the right members of (6) in Lemma 1. Hence it suffices to show $\tilde{\Delta}_i \le \mathrm{e}^{-n\eta}, i=1,2$ to prove Lemma 1. We first prove $\tilde{\Delta}_1\le \mathrm{e}^{-n\eta}$. We have the following chain of inequalities:

$$
\begin{aligned}
\Delta_1 &= \frac{1}{|\mathcal{K}_n||\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{\substack{(x^n,y^n,z^n):\\ y^n\in\mathcal{D}_1(k),z^n\in\mathcal{D}_2(l)\\ W_1^n(y^n|x^n)W_2(z^n|y^n)\\ <\mathrm{e}^{-n\eta}|\mathcal{K}_n|\\ \times q_{Y^nZ^n|L_n}(y^n,z^n|l)}} 1\\
&\quad\times\varphi^{(n)}(x^n|k,l)W_1^n(y^n|x^n)W_2^n(z^n|y^n)\\
&\le \frac{\mathrm{e}^{-n\eta}}{|\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{\substack{(x^n,y^n,z^n):\\ y^n\in\mathcal{D}_1(k),z^n\in\mathcal{D}_2(l)}} 1\\
&\quad\times\varphi^{(n)}(x^n|k,l)q_{Y^nZ^n|L_n}(y^n,z^n|l)\\
&= \frac{\mathrm{e}^{-n\eta}}{|\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n} q_{Y^nZ^n|L_n}\left(\mathcal{D}_1(k)\times\mathcal{D}_2(l)\middle|\, l\right)\\
&\le \frac{\mathrm{e}^{-n\eta}}{|\mathcal{L}_n|}\sum_{l\in\mathcal{L}_n}\sum_{k\in\mathcal{K}_n} q_{Y^n|L_n}\left(\mathcal{D}_1(k)\middle|\, l\right)\\
&= \frac{\mathrm{e}^{-n\eta}}{|\mathcal{L}_n|}\sum_{l\in\mathcal{L}_n} q_{Y^n|L_n}\left(\bigcup_{k\in\mathcal{K}_n}\mathcal{D}_1(k)\middle|\, l\right)\\
&\le \frac{\mathrm{e}^{-n\eta}}{|\mathcal{L}_n|}\sum_{l\in\mathcal{L}_n} 1 = \mathrm{e}^{-n\eta}.
\end{aligned}
$$

Next we prove $\Delta_2\le \mathrm{e}^{-n\eta}$. We have the following chain of inequalities:

$$
\begin{aligned}
\Delta_2 &= \frac{1}{|\mathcal{L}_n|}\sum_{(k,l)\in\mathcal{K}_n\times\mathcal{L}_n}\sum_{\substack{(x^n,y^n,z^n):\\ y^n\in\mathcal{D}_1(k),z^n\in\mathcal{D}_2(l)\\ p_{Z^n|L_n}(z^n|l)<\mathrm{e}^{-n\eta}\\ \times|\mathcal{L}_n|\tilde{q}_{Z^n}(z^n)}} 1\\
&\quad\times p_{K_nX^nY^nZ^n|L_n}(k,x^n,y^n,z^n|l)\\
&\le \frac{1}{|\mathcal{L}_n|}\sum_{l\in\mathcal{L}_n}\sum_{\substack{z^n\in\mathcal{D}_2(l),\\ p_{Z^n|L_n}(z^n|l)<\mathrm{e}^{-n\eta}\\ \times|\mathcal{L}_n|\tilde{q}_{Z^n}(z^n)}}\sum_{k\in\mathcal{K}_n}\sum_{(x^n,y^n)\in\mathcal{X}^n\times\mathcal{Y}^n} 1\\
&\quad\times p_{K_nX^nY^nZ^n|L_n}(k,x^n,y^n,z^n|l)\\
&\le \frac{1}{|\mathcal{L}_n|}\sum_{l\in\mathcal{L}_n}\sum_{\substack{z^n\in\mathcal{D}_2(l),\\ p_{Z^n|L_n}(z^n|l)<\mathrm{e}^{-n\eta}\\ \times|\mathcal{L}_n|\tilde{q}_{Z^n}(z^n)}} p_{Z^n|L_n}(z^n|l)\\
&\le \mathrm{e}^{-n\eta}\sum_{l\in\mathcal{L}_n}\sum_{z^n\in\mathcal{D}_2(l)}\tilde{q}_{Z^n}(z^n)\\
&= \mathrm{e}^{-n\eta}\sum_{l\in\mathcal{L}_n}\tilde{q}_{Z^n}\left(\mathcal{D}_2(l)\right)\\
&= \mathrm{e}^{-n\eta}\tilde{q}_{Z^n}\left(\bigcup_{l\in\mathcal{L}_n}\mathcal{D}_2(l)\right)\le \mathrm{e}^{-n\eta}.
\end{aligned}
$$

Thus Lemma 1 is proved ∎

D. *Proof of Lemma 2*

From Lemma 1, we have the following lemma

Lemma 5: For any $\eta>0$ and for any $(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)})$ satisfying $(1/n)\log|\mathcal{K}_n|\ge R_1$, $(1/n)\log|\mathcal{L}_n|\ge R_2$, we have

$$
\begin{aligned}
&\mathrm{P}_{\mathrm{c}}^{(n)}(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)})\le p_{L_nX^nY^nZ^n}\Biggl\{\\
&R_1\le\frac{1}{n}\sum_{t=1}^{n}\log\frac{W_1(Y_t|X_t)}{q_{Y_t|L_nY^{t-1}}(Y_t|L_n,Y^{t-1},Z^{t-1})}+\eta,\\
&R_2\le\frac{1}{n}\sum_{t=1}^{n}\log\frac{p_{Z_t|L_nZ^{t-1}}(Z_t|L_n,Z^{t-1})}{\tilde{q}_{Z_t}(Z_t)}+\eta\Biggr\}+2\mathrm{e}^{-n\eta}.
\end{aligned}
$$

Proof: In (5) in Lemma 1, we choose $q_{Z^nY^n|L_n}$

$$
\begin{aligned}
&q_{Y^nZ^n|L_n}(y^n,z^n|l)\\
&=\prod_{t=1}^{n}\left\{q_{Y_t|L_nY^{t-1}Z^{t-1}}(y_t|l,y^{t-1},z^{t-1})\right.\\
&\qquad\left.\times\, q_{Z_t|L_nY^tZ^{t-1}}(z_t|l,y^t,z^{t-1})\right\}\\
&=\prod_{t=1}^{n}\{q_{Y_t|L_nY^{t-1}Z^{t-1}}(y_t|l,y^{t-1},z^{t-1})W_2(z_t|y_t)\}.
\end{aligned}
$$

In (6) in Lemma 1, we choose $\tilde{q}_{Z^n}$ having the form

$$\tilde{q}_{Z^n}(Z^n)=\prod_{t=1}^{n}\tilde{q}_{Z_t}(Z_t).$$

Then from the bound (6) in Lemma 1, we obtain

$$\begin{aligned}
&\mathrm{P}_{\mathrm{c}}^{(n)}(\varphi^{(n)},\psi_1^{(n)},\psi_2^{(n)})\le p_{L_nX^nY^nZ^n}\Bigg\{\\
&R_1\le\frac{1}{n}\sum_{t=1}^{n}\log\frac{W_1(Y_t|X_t)}{q_{Y_t|L_nY^{t-1}Z^{t-1}}(Y_t|L_n,Y^{t-1},Z^{t-1})}+\eta,\\
&R_2\le\frac{1}{n}\sum_{t=1}^{n}\log\frac{p_{Z_t|L_nZ^{t-1}}(Z_t|L_n,Z^{t-1})}{\tilde{q}_{Z_t}(Z_t)}+\eta\Bigg\}+2\mathrm{e}^{-n\eta},
\end{aligned}$$

completing the proof. ■

From Lemma 5, we immediately obtain Lemma 2.

E. Upper Bound of $\overline{\Omega}^{(\mu,\theta)}(W_1,W_2)$

In this appendix we derive an explicit upper bound of $\overline{\Omega}^{(\mu,\theta)}(W_1,W_2)$ to prove Proposition 2. For each $t=1,2,\cdots,n$, define the function of $(u_t,x_t,y_t,z_t)\in\mathcal{U}_t\times\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$ by

$$f_{p_t||q_t,\kappa_t}^{(\mu,\theta)}(x_t,y_t,z_t|u_t)\triangleq\frac{W_1^{\theta\mu}(y_t|x_t)p_{Z_t|U_t}^{\theta}(z_t|u_t)}{q_{Y_t|U_t}^{\theta\mu}(y_t|u_t)\tilde{q}_{Z_t}^{\theta}(z_t)}.$$

For each $t=1,2,\cdots,n$, we define the probability distribution

$$\begin{aligned}
&p_{L_nX^tY^tZ^t}^{(\mu,\theta;q^t,\kappa^t)}\\
&\triangleq\left\{p_{L_nX^tY^tZ^t}^{(\mu,\theta;q^t,\kappa^t)}(l,x^t,y^t,z^t)\right\}_{(l,x^t,y^t,z^t)\in\mathcal{L}_n\times\mathcal{X}^t\times\mathcal{Y}^t\times\mathcal{Z}^t}
\end{aligned}$$

by

$$\begin{aligned}
&p_{L_nX^tY^tZ^t}^{(\mu,\theta;q^t,\kappa^t)}(l,x^t,y^t,z^t)\\
&\triangleq C_t^{-1}p_{L_n}(l)p_{X^t|L_n}(x^t|l)\prod_{i=1}^{t}\{W_1(y_i|x_i)W_2(z_i|y_i)\\
&\quad\times f_{p_i||q_i,\kappa_i}^{(\mu,\theta)}(x_i,y_i,z_i|u_i)\},
\end{aligned}$$

where

$$\begin{aligned}
C_t\triangleq\sum_{l,x^t,y^t,z^t}&p_{L_n}(l)p_{X^t|L_n}(x^t|l)\prod_{i=1}^{t}\{W_1(y_i|x_i)W_2(z_i|y_i)\\
&\times f_{p_i||q_i,\kappa_i}^{(\mu,\theta)}(x_i,y_i,z_i|u_i)\},
\end{aligned}$$

are constants for normalization. For each $t=1,2,\cdots,n$, set

$$\Phi_{t,q^t,\kappa^t}^{(\mu,\theta)}\triangleq C_tC_{t-1}^{-1},\tag{24}$$

where we define $C_0=1$. Then we have the following lemma.

Lemma 6:

$$\Omega_{p^{(n)}||q^n}^{(\mu,\theta)}(X^nY^nZ^n|L_n)=\sum_{t=1}^{n}\log\Phi_{t,q^t}^{(\mu,\theta)}.\tag{25}$$

Proof: From (24) we have

$$\log\Phi_{t,q^t,\kappa^t}^{(\mu,\theta)}=\log C_t-\log C_{t-1}.\tag{26}$$

Furthermore, by definition we have

$$\Omega_{p^{(n)}||q^n}^{(\mu,\theta)}(X^nY^nZ^n|L_n)=\log C_n,C_0=1.\tag{27}$$

From (26) and (27), (25) is obvious. ■

The following lemma is useful for the computation of $\Phi_{t,q^t,\kappa^t}^{(\mu,\theta)}$ for $t=1,2,\cdots,n$.

Lemma 7: For each $t=1,2,\cdots,n$, and for any $(l,x^t,y^t,z^t)\in\mathcal{L}_n\times\mathcal{X}^t\times\mathcal{Y}^t\times\mathcal{Z}^t$, we have

$$\begin{aligned}
&p_{L_nX^tY^tZ^t}^{(\mu,\theta;q^t,\kappa^t)}(l,x^t,y^t,z^t)\\
&=(\Phi_t^{(\mu,\theta;q^t,\kappa^t)})^{-1}p_{L_nX^{t-1}Y^{t-1}Z^{t-1}}^{(\mu,\theta;q^{t-1},\kappa^{t-1})}(l,x^{t-1},y^{t-1},z^{t-1})\\
&\quad\times p_{X_t|L_n,X^{t-1}}(x_t|l,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\
&\quad\times f_{p_t||q_t,\kappa_t}^{(\mu,\theta)}(x_t,y_t,z_t|u_t).
\end{aligned}\tag{28}$$

Furthermore, we have

$$\begin{aligned}
&\Phi_{t,q^t,\kappa^t}^{(\mu,\theta)}\\
&=\sum_{l,x^t,y^t,z^t}p_{L_nX^{t-1}Y^{t-1}Z^{t-1}}^{(\mu,\theta;q^{t-1},\kappa^{t-1})}(l,x^{t-1},y^{t-1},z^{t-1})\\
&\qquad\times p_{X_t|L_n,X^{t-1}}(x_t|l,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\
&\qquad\times f_{p_t||q_t,\kappa_t}^{(\mu,\theta)}(x_t,y_t,z_t|u_t).
\end{aligned}\tag{29}$$

Proof of Lemma 7: By the definition of $p_{L_nX^tY^tZ^t}^{(\mu,\theta;q^t,\kappa^t)}$ (l,x^t,y^t,z^t), $t=1,2,\cdots,n$, we have

$$\begin{aligned}
&p_{L_nX^tY^tZ^t}^{(\mu,\theta;q^t,\kappa^t)}(l,x^t,y^t,z^t)\\
&=C_t^{-1}p_{L_n}(l)p_{X^t|L_n}(x^t|l)\prod_{i=1}^{t}\{W_1(y_i|x_i)W_2(z_i|y_i)\\
&\quad\times f_{p_i||q_i,\kappa_i}^{(\mu,\theta)}(x_i,y_i,z_i|u_{i-1})\}.
\end{aligned}\tag{30}$$

Then we have the following chain of equalities:

$$\begin{aligned}
&p_{L_nX^tY^tZ^t}^{(\mu,\theta;q^t,\kappa^t)}(l,x^t,y^t,z^t)\\
&\overset{\text{(a)}}{=}C_t^{-1}p_{L_n}(l)p_{X^t|L_n}(x^t|l)\prod_{i=1}^{t}\{W_1(y_i|x_i)W_2(z_i|y_i)\\
&\quad\times f_{p_i||q_i,\kappa_i}^{(\mu,\theta)}(x_i,y_i,z_i|u_i)\}\\
&=C_t^{-1}p_{L_n}(l)p_{X^{t-1}|L_n}(x^{t-1}|l)\prod_{i=1}^{t}\{W_1(y_i|x_i)W_2(z_i|y_i)\\
&\quad\times f_{p_i||q_i,\kappa_i}^{(\mu,\theta)}(x_i,y_i,z_i|u_i)\}\\
&\quad\times p_{X_t|L_nX^{t-1}}(x_t|l,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\
&\quad\times f_{p_t||q_t,\kappa_t}^{(\mu,\theta)}(x_t,y_t,z_t|u_t)\\
&\overset{\text{(b)}}{=}C_t^{-1}C_{t-1}p_{L_nX^{t-1}Y^{t-1}Z^{t-1}}^{(\mu,\theta;q^{t-1},\kappa^{t-1})}(l,x^{t-1},y^{t-1},z^{t-1})\\
&\quad\times p_{X_t|L_nX^{t-1}}(x_t|l,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\
&\quad\times f_{p_t||q_t,\kappa_t}^{(\mu,\theta)}(x_t,y_t,z_t|u_t)\\
&=(\Phi_{t,q^t,\kappa^t}^{(\mu,\theta)})^{-1}p_{L_nX^{t-1}Y^{t-1}Z^{t-1}}^{(\mu,\theta;q^{t-1},\kappa^{t-1})}(l,x^{t-1},y^{t-1},z^{t-1})\\
&\quad\times p_{X_t|L_nX^{t-1}}(x_t|l,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\
&\quad\times f_{p_t||q_t,\kappa_t}^{(\mu,\theta)}(x_t,y_t,z_t|u_t).
\end{aligned}\tag{31}$$

Steps (a) and (b) follow from (30). From (31), we have

$$\begin{aligned}
&\Phi^{(\mu,\theta)}_{t,q^t,\kappa^t} p^{(\mu,\mu;q^t,\kappa^t)}_{L_nX^tY^tZ^t}(l,x^t,y^t,z^t) \qquad (32)\\
&= p^{(\mu,\theta;q^{t-1},\kappa^{t-1})}_{L_nX^{t-1}Y^{t-1}Z^{t-1}}(l,x^{t-1},y^{t-1},z^{t-1})\\
&\quad \times p_{X_t|L_nX^{t-1}}(x_t|l,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\
&\quad \times f^{(\mu,\theta)}_{p_t||q_t,\kappa_t}(x_t,y_t,z_t|u_t). \qquad (33)
\end{aligned}$$

Taking summations of (32) and (33) with respect to l, x^t, y^t, z^t, we obtain

$$\begin{aligned}
&\Phi^{(\mu,\theta)}_{t,q^t,\kappa^t}\\
&= \sum_{l,x^t,y^t,z^t} p^{(\mu,\theta;q^{t-1},\kappa^{t-1})}_{L_nX^{t-1}Y^{t-1}Z^{t-1}}(l,x^{t-1},y^{t-1},z^{t-1})\\
&\quad \times p_{X_t|L_n,X^{t-1}}(x_t|l,x^{t-1})W_1(y_t|x_t)W_2(z_t|y_t)\\
&\quad \times f^{(\mu,\theta)}_{p_t||q_t,\kappa_t}(x_t,y_t,z_t|u_t),
\end{aligned}$$

completing the proof. ■

We set

$$\begin{aligned}
&p^{(\mu,\theta;q^{t-1},\kappa^{t-1})}_{U_tX_t}(u_t,x_t) = p^{(\mu,\theta;q^{t-1},\kappa^{t-1})}_{L_nX_tY^{t-1}Z^{t-1}}(l,x_t,y^{t-1},z^{t-1})\\
&\triangleq \sum_{x^{t-1}} p^{(\mu,\theta;q^{t-1},\kappa^{t-1})}_{L_nX^{t-1}Y^{t-1}Z^{t-1}}(l,x^{t-1},y^{t-1},z^{t-1})\\
&\quad \times p_{X_t|L_nX^{t-1}}(x_t|l,x^{t-1}).
\end{aligned}$$

Then by (29) in Lemma 7 and the definition of $f^{(\mu,\theta)}_{p_t||q_t,\kappa_t}$ $(x_t,y_t,z_t\,|u_t)$, we have

$$\begin{aligned}
&\Phi^{(\mu,\theta)}_{t,q^t,\kappa^t}\\
&= \sum_{u_t,x_t,y_t,z_t} p^{(\mu,\theta;q^{t-1},\kappa^{t-1})}_{U_tX_t}(u_t,x_t)W_1(y_t|x_t)W_2(z_t|y_t)\\
&\quad \times \frac{W_1^{\theta\mu}(y_t|x_t)p^{\theta}_{Z_t|V_t}(z_t|v_t)}{q^{\theta\mu}_{Y_t|U_t}(y_t|u_t)q^{\theta}_{Z_t}(z_t)}. \qquad (34)
\end{aligned}$$

Proof of Proposition 2 is as follows.

Proof of Proposition 2: Set

$$\begin{aligned}
\hat{\mathcal{P}}_n(W_1,W_2) &\triangleq \{q : |\mathcal{U}| \le |\mathcal{L}_n||\mathcal{Y}|^{n-1},\\
&\qquad q_{Y|X}=W_1, q_{Z|Y}=W_2, U\leftrightarrow X\leftrightarrow Y\leftrightarrow Z\},\\
\hat{\Omega}^{(\mu,\lambda)}_n(W_1,W_2) &\triangleq \max_{q\in\hat{\mathcal{P}}_n(W_1,W_2)} \Omega^{(\mu,\lambda)}_q(XYZ|U).
\end{aligned}$$

We choose $q_t = q_{U_tX_tY_tZ_t}$ so that

$$\begin{aligned}
&q_{U_tX_tY_tZ_t}(u_t,x_t,y_t,z_t)\\
&= p^{(\mu,\theta;q^{t-1},\kappa^{t-1})}_{U_tX_t}(u_t,x_t)W_1(y_1|x_t)W_2(z_t|y_t).
\end{aligned}$$

It is obvious that $q_t \in \hat{\mathcal{P}}_n(W_1,W_2)$ for $t=1,2,\cdots,n$. By (34) and the above choice of q_t, we have

$$\begin{aligned}
&\Phi^{(\mu,\theta)}_{t,q^t,\kappa^t}\\
&= \sum_{u_t,x_t,y_t,z_t} q_{U_t}(u_t)q_{X_t|U_t}(x_t|u_t)W_1(y_t|x_t)W_2(z_t|y_t)\\
&\quad \times \left\{\frac{W_1^{\mu}(y_t|x_t)}{q^{\mu}_{Y_t|U_t}(y_t|u_t)}\frac{p_{Z_t|V_t}(z_t|v_t)}{q_{Z_t}(z_t)}\right\}^{\theta}\\
&= \mathrm{E}_{q_t}\left[\left\{\frac{W_1^{\mu}(Y_t|X_t)}{q^{\mu}_{Y_t|U_t}(Y_t|U_t)}\frac{p_{Z_t|U_t}(Z_t|V_t)}{q_{Z_t}(Z_t)}\right\}^{\theta}\right]\\
&= \mathrm{E}_{q_t}\left[\left\{\frac{W_1^{\mu}(Y_t|X_t)}{q^{\mu}_{Y_t|U_t}(Y_t|U_t)}\frac{q_{Z_t|U_t}(Z_t|U_t)}{q_{Z_t}(Z_t)}\frac{p_{Z_t|V_t}(Z_t|V_t)}{q_{Z_t|U_t}(Z_t|U_t)}\right\}^{\theta}\right]\\
&\overset{\text{(a)}}{\le} \left(\mathrm{E}_{q_t}\left[\left\{\frac{W_1^{\mu}(Y_t|X_t)}{q^{\mu}_{Y_t|U_t}(Y_t|U_t)}\frac{q_{Z_t|U_t}(Z_t|U_t)}{q_{Z_t}(Z_t)}\right\}^{\frac{\theta}{1-\theta}}\right]\right)^{1-\theta}\\
&\quad \times \left(\mathrm{E}_{q_t}\left\{\frac{p_{Z_t|V_t}(Z_t|V_t)}{q_{Z_t|U_t}(Z_t|U_t)}\right\}\right)^{\theta}\\
&= \exp\left\{(1-\theta)\Omega^{(\mu,\frac{\theta}{1-\theta})}_{q_t}(X_tY_tZ_t|U_t)\right\}\\
&\overset{\text{(b)}}{=} \exp\left\{\frac{1}{1+\lambda}\Omega^{(\mu,\lambda)}_{q_t}(X_tY_tZ_t|U_t)\right\}\\
&\overset{\text{(c)}}{\le} \exp\left\{\frac{1}{1+\lambda}\hat{\Omega}^{(\mu,\lambda)}_{n}(W_1,W_2)\right\}\\
&\overset{\text{(d)}}{=} \exp\left\{\frac{1}{1+\lambda}\Omega^{(\mu,\lambda)}(W_1,W_2)\right\}. \qquad (35)
\end{aligned}$$

Step (a) follows from Hölder's inequality. Step (b) follows from (12). Step (c) follows from $q_t \in \hat{\mathcal{P}}_n(W_1,W_2)$ and the definition of $\hat{\Omega}^{(\mu,\lambda)}_n$ (W_1,W_2). Step (d) follows from Lemma 4 in Appendix A. To prove this lemma we bound the cardinality $|\mathcal{U}|$ appearing in the definition of $\hat{\Omega}^{(\mu,\lambda)}_n(W_1,W_2)$ to show that the bound $|\mathcal{U}| \le |\mathcal{X}|$ is sufficient to describe $\hat{\Omega}^{(\mu,\lambda)}_n(W_1,W_2)$. Hence we have the following:

$$\begin{aligned}
&\min_{q^n\in\mathcal{Q}^n} \frac{1}{n}\Omega^{(\mu,\theta)}_{p^{(n)}||q^n}(X^nY^nZ^n|L_n)\\
&\le \frac{1}{n}\Omega^{(\mu,\theta)}_{p^{(n)}||q^n}(X^nY^nZ^n|L_n) \overset{\text{(a)}}{=} \frac{1}{n}\sum_{t=1}^{n}\log\Phi^{(\mu,\theta)}_{t,q^t,\kappa^t}\\
&\overset{\text{(b)}}{\le} \frac{1}{1+\lambda}\Omega^{(\mu,\lambda)}(W_1,W_2). \qquad (36)
\end{aligned}$$

Step (a) follows from (25) in Lemma 6. Step (b) follows from (35). Since (36) holds for any $n \ge 1$ and any $p^{(n)} \in \mathcal{P}^{(n)}$ (W_1,W_2), we have

$$\overline{\Omega}^{(\mu,\theta)}(W_1,W_2) \le \frac{1}{1+\gamma}\Omega^{(\mu,\lambda)}(W_1,W_2).$$

Thus, Proposition 2 is proved. ■