2022 IEEE International Symposium on Information Theory (ISIT) | 978-1-6654-2159-1/22/$31.00 ©2022 IEEE | DOI: 10.1109/ISIT50566.2022.9834411

One-Shot Point-to-Point Channel Simulation

Michael X. Cao*†, Navneeth Ramakrishnan†‡, Mario Berta‡§¶, Marco Tomamichel*†
*Department of Electrical and Computer Engineering, National University of Singapore, Singapore
†Centre for Quantum Technologies, National University of Singapore, Singapore
‡Department of Computing, Imperial College London, London, UK
§Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, USA
¶AWS Center for Quantum Computing, Pasadena, USA

***Abstract*—We study the problem of one-shot channel simulation of DMCs with unlimited shared randomness. For any fixed tolerance measured in total variational distance, we propose an achievability bound and a converse bound on the size of the code to simulate the channel. The achievability bound utilizes the convex split lemma, whereas the converse bound is the result of the relationships between smoothed max-divergences and the max-mutual information. The achievability proof does not rely on a "universal state" (compared with some previous related works), and provides a tighter bound. Using the two bounds, we also provide an alternative proof to the reverse Shannon theorem.**

I. Introduction

Channel simulation has been a fundamental topic in both classical and quantum information theory (*e.g.*, see [1], [2]). It has mostly been studied in asymptotic setups in the context of the reverse Shannon theorem [3], [4]. The theorems state that to simulate a discrete memoryless channel (DMC) W (or a quantum channel $\mathcal{N}$) with unlimited shared randomness (or shared entanglement), one needs $C(W)$ (or $C_{\mathrm{E}}(\mathcal{N})$) perfect channels per channel simulated asymptotically. In one-shot cases, quantum channel simulation has also been studied (see *e.g.*, [5]). However, we are unaware of similar one-shot studies in classical setups.

In this paper, we consider the problem of channel simulation with unlimited shared randomness in one-shot setups. Given a DMC W and a shared random variable S distributed on $\mathcal{S}$, the task of simulating W within a tolerance of ϵ in total variational distance (TVD) is to find a set of encoders $\mathcal{E}_s \in \mathcal{P}(\{1,\dots,M\}|\mathcal{X})$ and decoders $\mathcal{D}_s \in \mathcal{P}(\mathcal{Y}|\{1,\dots,M\})$ for each $s \in \mathcal{S}$ such that, *for every input distribution* p_{X}, the induced joint distribution

$$\tilde{p}_{\mathsf{XY}}(x,y) \triangleq \sum_{s\in\mathcal{S}} p_{\mathsf{S}}(s)\cdot p_{\mathsf{X}}(x)\cdot \sum_{m\in\{1,\dots,M\}} \mathcal{E}_s(m|x)\cdot\mathcal{D}_s(y|m) \quad (1)$$

is ϵ-close to the original joint distribution $p_{\mathsf{XY}}(x,y) \triangleq p_{\mathsf{X}}(x)\cdot W_{\mathsf{Y}|\mathsf{X}}(y|x)$ in TVD. In this case, we call the 3-tuple $(\{\mathcal{E}_s\}_{s\in\mathcal{S}}, \{\mathcal{D}_s\}_{s\in\mathcal{S}}, p_{\mathsf{S}})$ a size-M ϵ-simulation code for W (see Fig. 1). On the other hand, an integer M is said to be *attainable* if there exists a size-M ϵ-simulation code. As a direct observation, if M is attainable, so is any larger integer.

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore and the Ministry of Education, Singapore under the Research Centres of Excellence programme. MC and MT are also supported by NUS startup grants (R-263-000-E32-133 and R-263-000-E32-731).

In this paper, we are interested in characterizing the minimal attainable size of ϵ-simulation codes for a given DMC.

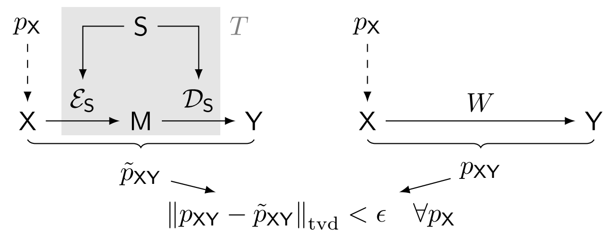

Fig. 1. The task of channel simulation

Channel simulation is also related to the task of *strong coordination*, where the latter has been studied mostly in asymptotic setups for various network topologies [6], [7]. Channel simulation can be viewed as a *universal* strong coordination of $p_{\mathsf{Y}|\mathsf{X}}$ for all p_{X} with the help of unlimited shared randomness. We have seen some recent studies in finite-length strong coordination without the help of shared randomness [8].

In the following, we propose, for any given $\epsilon \in (0,1)$, a pair of upper (achievability) and lower (converse) bounds on the minimal attainable size of ϵ-simulation codes for DMC W. Both bounds are expressed in terms of the partial smoothed max-divergence. The achievability bound is inspired by the convex split lemmas [9], [10], whereas the converse bound is a result of a series of inequalities among (partial) smoothed max-divergences and the max-mutual information. Using the pair of bounds, we provide an alternative proof to the reverse Shannon theorem, namely to show that the first-order error exponent for simulating n-fold channels $W^{\otimes n}$ equals the capacity of the channel $C(W)$. Both achievability and converse proofs use the relationship between the partial smoothed max-divergence and the spectrum divergence. Some open problems and on-going work are discussed at the end of the paper. Please note that the proofs for most lemmas have been deferred till the end of the paper.

Some of the definitions/notations used in this paper

- Throughout this paper, bold letters are used to denote vectors. In particular, we use lower and upper scripts to indicate the starting and ending indices of the components of the vector, *e.g.*, $\boldsymbol{x}_1^n \equiv (x_1,\dots,x_n)$.
- Throughout this paper, random variables are denoted by capitalized sans-serif letters, *e.g.*, X reads "random variable X".

- $\mathcal{P}(\mathcal{X})$ denotes the set of all pmfs/pdfs on set $\mathcal{X}$, and $\mathcal{P}(\mathcal{Y}|\mathcal{X})$ denotes the set of all conditional pmfs/pdfs on set $\mathcal{Y}$ conditioning on $\mathcal{X}$.

Let p and q be two distributions on a same alphabet $\mathcal{X}$, and let random variable X be distributed according to p:

- $D(p\|q) \triangleq \mathbb{E}\left(\log \frac{p(\mathsf{X})}{q(\mathsf{X})}\right) = \sum_{x\in\mathcal{X}} p(x)\cdot\log\frac{p(x)}{q(x)}$ stands for the *divergence* of p w.r.t. q.
- $V(p\|q) \triangleq \mathrm{Var}\left(\log\frac{p(\mathsf{X})}{q(\mathsf{X})}\right)$ stands for the *divergence variance* of p w.r.t. q.
- The *truncated spectrum divergence* of p w.r.t. q is defined as

$$D_{s+}^{\epsilon}(p\|q) \triangleq \inf\left\{a \geqslant 0 : P\left[\log\frac{p(\mathsf{X})}{q(\mathsf{X})} > a\right] < \epsilon\right\}. \quad (2)$$

- The *max-divergence* of p w.r.t. q is defined as

$$D_{\max}(p\|q) \triangleq \sup_{x\in\mathcal{X}}\left\{\log\frac{p(x)}{q(x)}\right\}. \quad (3)$$

- The ϵ*-smoothed max-divergence* of p w.r.t. q is defined as

$$D_{\max}^{\epsilon}(p\|q) \triangleq \inf_{\tilde{p}\in\mathcal{P}(\mathcal{X}):\|\tilde{p}-p\|_{\mathrm{tvd}}\leqslant\epsilon} D_{\max}(\tilde{p}\|q). \quad (4)$$

Let XY be joint random variables distributed on $\mathcal{X}\times\mathcal{Y}$ according to some joint pmf p_{XY}.

- Let q_{XY} be another distribution on $\mathcal{X}\times\mathcal{Y}$. The *partial* ϵ*-smoothed max-divergence* of p w.r.t. q is defined as

$$D_{\max}^{\epsilon,\mathsf{X}}(p_{\mathsf{XY}}\|q_{\mathsf{XY}}) \triangleq \inf_{\substack{\tilde{p}_{\mathsf{XY}}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})\\ \|\tilde{p}_{\mathsf{XY}}-p_{\mathsf{XY}}\|_{\mathrm{tvd}}\leqslant\epsilon\\ \tilde{p}_{\mathsf{X}}=p_{\mathsf{X}}}} D_{\max}(\tilde{p}\|q). \quad (5)$$

- The *max-mutual information* between X and $\mathcal{Y}$ is defined as

$$I_{\max}(\mathsf{X}:\mathsf{Y}) \triangleq \inf_{q_{\mathsf{Y}}\in\mathcal{P}(\mathcal{Y})} D_{\max}(p_{\mathsf{XY}}\|p_{\mathsf{X}}\times q_{\mathsf{Y}}). \quad (6)$$

II. One-Shot Bounds on Channel Simulation

In this section, we present an achievability bound and a converse bound for the attainable sizes of ϵ-simulation codes for DMC W.

A. Achievability Bound

We propose an achievability bound based on the following lemma. The lemma is adopted from the convex-split lemmas (based on $D_{\max}$ and D_s, respectively) in [9] and [10, Fact 6].

Lemma 1. *Let (X,Y) be jointly distributed over $\mathcal{X}\times\mathcal{Y}$, let q be some probability distribution over $\mathcal{Y}$, and let $\epsilon > \delta \in (0,1)$. Let M be a nonnegative integer such that*

$$\log M \geqslant D_{\max}^{\epsilon-\delta,\mathsf{X}}(p_{\mathsf{XY}}\|p_{\mathsf{X}}\times q_{\mathsf{Y}}) + \log\frac{2}{\delta^2}, \quad (7)$$

where $D_{\max}^{\epsilon-\delta,\mathsf{X}}$ has been defined in (5). Let J be uniformly distributed on $\{1,\ldots,M\}$, and let the joint random variables $(\mathsf{J},\mathsf{X},\mathsf{Y}_1,\ldots,\mathsf{Y}_M)$ be distributed according to

$$p_{\mathsf{X},\mathbf{Y}_1^M|\mathsf{J}}(x,\boldsymbol{y}_1^M|j) = p_{\mathsf{XY}}(x,y_j)\cdot\prod_{i\neq j} q(y_i). \quad (8)$$

Then, it holds that

$$\|p_{\mathsf{X},\mathsf{Y}_1,\ldots,\mathsf{Y}_M} - p_{\mathsf{X}}\times q_{\mathsf{Y}_1}\times\cdots\times q_{\mathsf{Y}_M}\|_{\mathrm{tvd}} \leqslant \epsilon. \quad (9)$$

Theorem 2. *Let W be a DMC from $\mathcal{X}$ to $\mathcal{Y}$, and let $\epsilon\in(0,1)$. A size-M ϵ-simulation code for DMC W exists if*

$$\log M > \inf_{q_{\mathsf{Y}}}\sup_{p_{\mathsf{X}}} D_{\max}^{\epsilon-\delta,\mathsf{X}}\big(p_{\mathsf{X}}\cdot W_{\mathsf{Y}|\mathsf{X}}\big\|p_{\mathsf{X}}\times q_{\mathsf{Y}}\big) + \log\frac{2}{\delta^2}, \quad (10)$$

for some $\delta\in(0,\epsilon)$.

Proof. We prove by presenting a protocol to simulate W:

0) Let q_{Y} be a pmf on $\mathcal{Y}$ such that

$$\log M > \sup_{p_{\mathsf{X}}} D_{\max}^{\epsilon-\delta,\mathsf{X}}\big(p_{\mathsf{X}}\cdot W_{\mathsf{Y}|\mathsf{X}}\big\|p_{\mathsf{X}}\times q_{\mathsf{Y}}\big) + \log\frac{2}{\delta^2}.$$

1) Let Alice and Bob share some shared i.i.d. random variables $(\mathsf{Y}_1,\ldots,\mathsf{Y}_M)$ where $\mathsf{Y}_k\sim q$ for each k.
2) Upon receiving input $\mathsf{X}=x$, Alice generates a random variable J (distributed on $\{1,\ldots,M\}$) according to the following conditional pmf

$$\tilde{p}_{\mathsf{J}|\mathsf{X},\mathsf{Y}_1^M}(j|x,\boldsymbol{y}_1^M) \triangleq \frac{W(y_j|x)\cdot\prod_{\ell\neq j} q(y_\ell)}{\sum_{i=1}^M W(y_i|x)\cdot\prod_{\ell\neq i} q(y_\ell)}.$$

3) Alice sends J losslessly to Bob using $\log M$-bits.
4) Upon receiving J, Bob outputs Y_{J}.

It suffices to show that the joint pmf of the random variables XY_{J} generated by the above protocol is ϵ-close (in TVD) to $p_{\mathsf{XY}} \triangleq p_{\mathsf{X}}\cdot W_{\mathsf{Y}|\mathsf{X}}$ for all input distribution p_{X}.

Let $\tilde{p}$ denote the (marginal or conditional, depending on the subscript) pmfs of the random variables $\mathsf{J},\mathsf{X},\mathsf{Y}_1,\ldots,\mathsf{Y}_M$ as in the above protocol. Define the joint pmf $p_{\mathsf{J},\mathsf{X},\mathsf{Y}_1,\ldots,\mathsf{Y}_M}$ as

$$p_{\mathsf{J},\mathsf{X},\mathsf{Y}_1,\ldots,\mathsf{Y}_M}(j,x,y_1,\ldots,y_M) \triangleq \frac{1}{M}\cdot p_{\mathsf{X},\mathbf{Y}_1^M|\mathsf{J}}(x,\boldsymbol{y}_1^M|j),$$

where $p_{\mathsf{X},\mathbf{Y}_1^M|\mathsf{J}}$ has been defined in (8). For any input distribution $p_{\mathsf{X}}=\tilde{p}_{\mathsf{X}}$, as a result of the protocol, we have

$$\begin{aligned}&\tilde{p}_{\mathsf{J},\mathsf{X},\mathsf{Y}_1,\ldots,\mathsf{Y}_M}(j,x,y_1,\ldots,y_M)\\ &= p_{\mathsf{X}}(x)\cdot\prod_{\ell=1}^M q(y_\ell)\cdot\frac{W(y_j|x)\cdot\prod_{\ell\neq j} q(y_\ell)}{\sum_{i=1}^M W(y_i|x)\cdot\prod_{\ell\neq i} q(y_\ell)}.\end{aligned}$$

By Lemma 1 and (10), it holds that

$$\begin{aligned}&\left\|\tilde{p}_{\mathsf{X},\mathbf{Y}_1^M} - p_{\mathsf{X},\mathbf{Y}_1^M}\right\|_{\mathrm{tvd}}\\ &= \left\|p_{\mathsf{X}}\cdot\prod_{\ell=1}^M q_{\mathsf{Y}_\ell} - \frac{1}{M}\sum_{i=1}^M p_{\mathsf{X}}\cdot W_{\mathsf{Y}_i|\mathsf{X}}\cdot\prod_{\ell\neq i} q_{\mathsf{Y}_\ell}\right\|_{\mathrm{tvd}} \leqslant \epsilon.\end{aligned}$$

Since $\tilde{p}_{\mathsf{J}|\mathsf{X},\mathbf{Y}_1^M} = p_{\mathsf{J}|\mathsf{X},\mathbf{Y}_1^M}$ (as part of the construction), we have

$$\left\|\tilde{p}_{\mathsf{J},\mathsf{X},\mathbf{Y}_1^M} - p_{\mathsf{J},\mathsf{X},\mathbf{Y}_1^M}\right\|_{\mathrm{tvd}} = \left\|\tilde{p}_{\mathsf{X},\mathbf{Y}_1^M} - p_{\mathsf{X},\mathbf{Y}_1^M}\right\|_{\mathrm{tvd}} \leqslant \epsilon.$$

Using the data processing inequality of TVD, we have

$$\epsilon \geqslant \|\tilde{p}_{\mathsf{XY_J}} - p_{\mathsf{XY_J}}\|_{\mathrm{tvd}} = \left\|\tilde{p}_{\mathsf{XY_J}} - p_{\mathsf{X}}\cdot W_{\mathsf{Y}|\mathsf{X}}\right\|_{\mathrm{tvd}}.$$

Since the above holds for all input distributions p_{X}, we have finished the proof. □

Notice that in the above protocol, we picked *a* q_{Y} such that $\log M \geqslant D_{\max}^{\epsilon-\delta,\mathsf{X}}\big(p_{\mathsf{X}} \cdot W_{\mathsf{Y}|\mathsf{X}} \big\| p_{\mathsf{X}} \times q_{\mathsf{Y}}\big) + \log \frac{2}{\delta^2}$ for *all* p_{X}. This is particularly important since this allows the protocol to guarantee the ϵ-closeness independently from the input distribution p_{X}. This differs from the approach in some of the related works in quantum setups (*e.g.*, see [5]), in which the quantum counterpart of q_{Y} is optimized first, and then one applies the post selection technique to find a series of "universal" input states. (Here, the "universal" input states are a source such that if a protocol can simulate the channel for *this* input source then it can handle *all* the other input states with a diminishing additional cost in asymptotic setups.)

B. Converse Bound

We now present a converse bound for the attainable sizes of ϵ-simulation codes for DMC W. The bound is a result of a series of relationships among $I_{\max}$, $D_{\max}$, and $D_{\max}^{\epsilon,\mathsf{X}}$.

Theorem 3. *Let W be a DMC from $\mathcal{X}$ to $\mathcal{Y}$, and let $\epsilon \in (0,1)$. For any size-M ϵ-simulation code for W, it holds that*

$$\log M \geqslant \inf_{q_{\mathsf{Y}}} \sup_{p_{\mathsf{X}}} D_{\max}^{\epsilon,\mathsf{X}}(p_{\mathsf{X}} \cdot W_{\mathsf{Y}|\mathsf{X}} \| p_{\mathsf{X}} \times q_{\mathsf{Y}}). \tag{11}$$

Proof. Suppose there exists some size-M ϵ-simulation code for W. Let the random variable S denote the shared randomness, and M denote the codeword transmitted (see Fig. 1). Then, for any input source $\mathsf{X} \sim p_{\mathsf{X}}$, we have a Markov chain $\mathsf{X} - \mathsf{MS} - \mathsf{Y}$ where

- the distribution of XY, denoted by $\tilde{p}_{\mathsf{XY}}$, is ϵ-close (in TVD) to $p_{\mathsf{XY}} \triangleq p_{\mathsf{X}} \cdot W_{\mathsf{Y}|\mathsf{X}}$;
- the marginal distribution $\sum_y \tilde{p}_{\mathsf{XY}}(x,y) = p_{\mathsf{X}}(x)\ \forall x$;
- M is distributed over $\{1,\ldots,M\}$.

In this case, we have the following series of inequalities.

1) By the non-lockability of $I_{\max}$ (*e.g.*, see [11, Corollary A.14]), we have
$$\log M = I_{\max}(\mathsf{X};\mathsf{S}) + \log M \geqslant I_{\max}(\mathsf{X};\mathsf{MS}).$$
2) By the data-processing inequality of $I_{\max}$, we have
$$I_{\max}(\mathsf{X};\mathsf{MS}) \geqslant I_{\max}(\mathsf{X};\mathsf{Y})$$
3) By the definition of the smoothed max-divergence, and noting that $\tilde{p}_{\mathsf{X}} = p_{\mathsf{X}}$, we have

$$\begin{aligned} I_{\max}(\mathsf{X};\mathsf{Y}) &\triangleq \inf_{q_{\mathsf{Y}}} D_{\max}(\tilde{p}_{\mathsf{XY}} \| p_{\mathsf{X}} \times q_{\mathsf{Y}}) \\ = \inf_{q_{\mathsf{Y}}} \max_{x,y} \log \frac{\tilde{p}_{\mathsf{Y}|\mathsf{X}}(y|x)}{q_{\mathsf{Y}}(y)} &= \inf_{q_{\mathsf{Y}}} \sup_{p_{\mathsf{X}}} D_{\max}(\tilde{p}_{\mathsf{XY}} \| p_{\mathsf{X}} \times q_{\mathsf{Y}}) \\ \geqslant \inf_{q_{\mathsf{Y}}} \sup_{p_{\mathsf{X}}} D_{\max}^{\epsilon,\mathsf{X}}(p_{\mathsf{XY}} \| p_{\mathsf{X}} \times q_{\mathsf{Y}}). \end{aligned}$$

Thus, we have shown (11). □

Notice that the partial ϵ-smoothed max-divergence is always lower bounded by the ϵ-smoothed max-divergence. Thus, one easily obtains the following looser converse bound from the above theorem:

$$\log M \geqslant \inf_{q_{\mathsf{Y}}} \sup_{p_{\mathsf{X}}} D_{\max}^{\epsilon}(p_{\mathsf{X}} \cdot W_{\mathsf{Y}|\mathsf{X}} \| p_{\mathsf{X}} \times q_{\mathsf{Y}}). \tag{12}$$

One can further relax the bound by swapping the inf and sup, and get

$$\log M \geqslant \sup_{p_{\mathsf{X}}} \inf_{q_{\mathsf{Y}}} D_{\max}^{\epsilon}(p_{\mathsf{X}} \cdot W_{\mathsf{Y}|\mathsf{X}} \| p_{\mathsf{X}} \times q_{\mathsf{Y}}). \tag{13}$$

III. Asymptotic Analysis of the Achievability Bound

In this section and the next, we consider the problem of simulating n copies of W, *i.e.*, $W^{\otimes n}$, for n being large (especially when $n \to \infty$). Here, we focus on the behavior of the achievability bound as n grows. In particular, we would like to use the one-shot achievability bound developed in the previous section to show the following achievability bound for the n-fold channel $W^{\otimes n}$.

Theorem 4. *Let W be a DMC. For every $\epsilon \in (0,1)$, there exists a sequence of 2^{nR}-size ϵ-simulation codes for $W^{\otimes n}$ for sufficiently large n as long as $R > C(W)$.*

We use the following lemma for the proof of Theorem 4.

Lemma 5. *Let $p_{\mathsf{Y}|\mathsf{X}}, q_{\mathsf{Y}|\mathsf{X}} \in \mathcal{P}(\mathcal{Y}|\mathcal{X})$, and $p_{\mathsf{X}} \in \mathcal{P}(\mathcal{X})$. For all $\epsilon \in (0,1)$, it holds that,*

$$D_{s+}^{\epsilon}\big(p_{\mathsf{Y}|\mathsf{X}} \cdot p_{\mathsf{X}} \big\| q_{\mathsf{Y}|\mathsf{X}} \cdot p_{\mathsf{X}}\big) \leqslant \sup_{x\in\mathcal{X}} D_{s+}^{\epsilon}\big(p_{\mathsf{Y}|\mathsf{X}}(\cdot|x) \big\| q_{\mathsf{Y}|\mathsf{X}}(\cdot|x)\big). \tag{14}$$

Proof of Theorem 4. Let $\epsilon \in (0,1)$ to be fixed. Let $M_n^\star$ denote the minimum integer such that an $M_n^\star$-size ϵ-simulation code exists for $W^{\otimes n}$. It suffices to prove

$$\limsup_{n\to\infty} \frac{1}{n} \log M_n^\star \leqslant C(W).$$

To show the above, first note that (for any $\delta \in (0,\epsilon)$)

$$\begin{aligned} &\frac{1}{n}\log M_n^\star \\ &\overset{a)}{\leqslant} \inf_{q_{\mathbf{Y}_1^n}} \sup_{p_{\mathbf{X}_1^n}} \frac{1}{n} D_{\max}^{\epsilon-\delta,\mathsf{X}}\Big(p_{\mathbf{X}_1^n} \cdot W_{\mathsf{Y}|\mathsf{X}}^{\otimes n} \Big\| p_{\mathbf{X}_1^n} \times q_{\mathbf{Y}_1^n}\Big) + \frac{1}{n}\log\frac{2}{\delta^2} \\ &\overset{b)}{\leqslant} \inf_{q_{\mathbf{Y}_1^n}} \sup_{p_{\mathbf{X}_1^n}} \frac{1}{n} D_{s+}^{\epsilon-\delta}\Big(p_{\mathbf{X}_1^n} \cdot W_{\mathsf{Y}|\mathsf{X}}^{\otimes n} \Big\| p_{\mathbf{X}_1^n} \times q_{\mathbf{Y}_1^n}\Big) + \frac{1}{n}\log\frac{4}{\delta^2} \\ &\overset{c)}{\leqslant} \inf_{q_{\mathbf{Y}_1^n}} \sup_{\boldsymbol{x}_1^n} \frac{1}{n} D_{s+}^{\epsilon-\delta}\Big(W_{\mathsf{Y}|\mathsf{X}}^{\otimes n}(\cdot|\boldsymbol{x}_1^n) \Big\| q_{\mathbf{Y}_1^n}\Big) + \frac{1}{n}\log\frac{4}{\delta^2}, \end{aligned}$$

where a) is based on Theorem 2, b) uses [12, Theorem 9], and c) uses Lemma 5. We pick $q_{\mathbf{Y}_1^n}$ as

$$q_{\mathbf{Y}_1^n}(\boldsymbol{y}^n) \triangleq \sum_{\lambda\in\Lambda_n} \frac{1}{|\Lambda_n|} \prod_{i=1}^n \sum_{\tilde{x}} W(y_i|\tilde{x}) p_\lambda(\tilde{x}),$$

where Λ_n is the set of all *types* of length-n sequences in $\mathcal{X}$, *i.e.*,

$$\Lambda_n \triangleq \left\{ \lambda \subset \mathcal{X}^n : \sum_{i=1}^n \delta_{x,x_i} = \sum_{i=1}^n \delta_{x,\tilde{x}_i} \quad \begin{matrix} \forall \boldsymbol{x}_1^n, \tilde{\boldsymbol{x}}_1^n \in \lambda \\ \forall x \in \mathcal{X} \end{matrix} \right\},$$

and for each $\lambda \in \Lambda_n$, p_λ is a pmf on $\mathcal{X}$ such that $p_\lambda(\tilde{x})$ equals the frequency of $\tilde{x}$ in sequences $\boldsymbol{x} \in \lambda$. In this case, we have

$\frac{1}{n}\log M_n^\star$ (continue on the next page)

$$\leqslant \sup_{\boldsymbol{x}_1^n} \frac{1}{n} D_{s+}^{\epsilon-\delta}\left(W_{\mathsf{Y}|\mathsf{X}}^{\otimes n}(\cdot|\boldsymbol{x}_1^n)\middle\| \ldots \right.$$

$$\left. \sum_{\lambda\in\Lambda_n} \frac{1}{|\Lambda_n|}\left(\sum_{\tilde{x}} W_{\mathsf{Y}|\mathsf{X}}(\cdot|\tilde{x})\cdot p_\lambda(\tilde{x})\right)^{\otimes n}\right) + \frac{1}{n}\log\frac{4}{\delta^2}$$

$$\overset{\text{d)}}{\leqslant} \sup_{\boldsymbol{x}_1^n} \frac{1}{n} D_{s+}^{\epsilon-\delta}\left(W_{\mathsf{Y}|\mathsf{X}}^{\otimes n}(\cdot|\boldsymbol{x}_1^n)\middle\|\left(\sum_{\tilde{x}} W_{\mathsf{Y}|\mathsf{X}}(\cdot|\tilde{x})\cdot f_{\boldsymbol{x}_1^n}(\tilde{x})\right)^{\otimes n}\right) + \frac{1}{n}\log|\Lambda_n| + \frac{1}{n}\log\frac{4}{\delta^2}$$

$$\overset{\text{e)}}{\leqslant} \sup_{\boldsymbol{x}_1^n}\left\{\frac{1}{n}\sum_{i=1}^n D\left(W_{\mathsf{Y}|\mathsf{X}}(\cdot|x_i)\middle\|\sum_{\tilde{x}} W_{\mathsf{Y}|\mathsf{X}}(\cdot|\tilde{x})\cdot f_{\boldsymbol{x}_1^n}(\tilde{x})\right) + \frac{1}{\sqrt{n(\epsilon-\delta)}}\cdot\sqrt{\frac{1}{n}\sum_{i=1}^n V\left(W_{\mathsf{Y}|\mathsf{X}}(\cdot|x_i)\middle\|\sum_{\tilde{x}} W_{\mathsf{Y}|\mathsf{X}}(\cdot|\tilde{x})\cdot f_{\boldsymbol{x}_1^n}(\tilde{x})\right)}\right\} + \frac{1}{n}\log|\Lambda_n| + \frac{1}{n}\log\frac{4}{\delta^2},$$

where $f_{\boldsymbol{x}_1^n}$ denotes the empirical distribution of the sequence $\boldsymbol{x}_1^n$, and we have used Lemma 3 and Lemma 5 from [13] for d) and e), respectively. By counting, the above can be rewritten as

$$\frac{1}{n}\log M_n^\star \leqslant \frac{1}{n}\log\frac{4|\Lambda_n|}{\delta^2} + \sup_{\boldsymbol{x}_1^n}\left\{ I(\mathsf{X};\mathsf{Y})_{f_{\boldsymbol{x}_1^n}\cdot W_{\mathsf{Y}|\mathsf{X}}} + \frac{1}{\sqrt{n(\epsilon-\delta)}} V\left(f_{\boldsymbol{x}_1^n}\cdot W_{\mathsf{Y}|\mathsf{X}}\middle\|\sum_{\tilde{x}} f_{\boldsymbol{x}_1^n}(\tilde{x})\cdot W_{\mathsf{Y}|\mathsf{X}}(\cdot|\tilde{x})\right)\right\}.$$

Notice that, in the RHS of the above inequality, both first and last term tend to 0 as $n\to\infty$, thus

$$\limsup_{n\to\infty}\frac{1}{n}\log M_n^\star \leqslant \limsup_{n\to\infty}\sup_{\boldsymbol{x}_1^n} I(\mathsf{X};\mathsf{Y})_{f_{\boldsymbol{x}_1^n}\cdot W_{\mathsf{Y}|\mathsf{X}}} \leqslant \sup_{p_\mathsf{X}} I(\mathsf{X};\mathsf{Y})_{p_\mathsf{X}\cdot W_{\mathsf{Y}|\mathsf{X}}} = C(W),$$

which finishes the proof. □

IV. Asymptotic Analysis of the Converse Bound

We continue to consider the problem of simulating $W^{\otimes n}$ for large n. In particular, we would like to use the one-shot converse bound developed in Section II-B to show the following converse bound for the n-fold channel $W^{\otimes n}$.

Theorem 6. *Let W be a DMC. For any $\epsilon\in(0,1)$, if there exists a sequence of size-2^{nR} ϵ-simulation code for $W^{\otimes n}$ for all n sufficiently large, then $R\geqslant C(W)$.*

We shall need the following lemmas.

Lemma 7 ([12, Lemma 10])**.** *Suppose $p_{\mathsf{XY}}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$, and $q_\mathsf{Y}\in\mathcal{P}(\mathcal{Y})$. For any $\epsilon\in(0,1)$, and $\delta\in(0,1-\epsilon)$, we have*

$$\inf_{q_\mathsf{Y}} D_{s+}^{\epsilon}(p_{\mathsf{XY}}\|p_\mathsf{X}\times q_\mathsf{Y}) \geqslant D_{s+}^{\epsilon+\delta}(p_{\mathsf{XY}}\|p_\mathsf{X}\times p_\mathsf{Y}) + \log\delta. \quad (15)$$

Lemma 8 ([12, Eq. (48)])**.** *Suppose $p_{\mathsf{XY}}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$, and $q_\mathsf{Y}\in\mathcal{P}(\mathcal{Y})$. For any $\epsilon,\delta\in(0,1)$, we have*

$$D_{\max}^{\epsilon,\mathsf{X}}(p_{\mathsf{XY}}\|p_\mathsf{X}\times q_\mathsf{Y}) \geqslant D_{s+}^{\frac{\epsilon}{1-\delta}}(p_{\mathsf{XY}}\|p_\mathsf{X}\times q_\mathsf{Y}) + \log\delta. \quad (16)$$

Proof of Theorem 6. Suppose there exists a size $M_n \triangleq 2^{nR}$ ϵ-simulation code for $W^{\otimes n}$. For n large enough, we have the following chain of inequalities

$$R = \frac{1}{n}\log M_n \overset{\text{f)}}{\geqslant} \frac{1}{n}\sup_{p_{\mathsf{X}_1^n}}\inf_{q_{\mathsf{Y}_1^n}} D_{\max}^{\epsilon}(p_{\mathsf{X}_1^n}\cdot W_{\mathsf{Y}|\mathsf{X}}^{\otimes n}\|p_{\mathsf{X}_1^n}\times q_{\mathsf{Y}_1^n})$$

$$\overset{\text{g)}}{\geqslant} \frac{1}{n}\sup_{p_{\mathsf{X}_1^n}}\inf_{q_{\mathsf{Y}_1^n}} D_{s+}^{\frac{\epsilon}{1-\delta}}(p_{\mathsf{X}_1^n}\cdot W_{\mathsf{Y}|\mathsf{X}}^{\otimes n}\|p_{\mathsf{X}_1^n}\times q_{\mathsf{Y}_1^n}) + \frac{\log\delta}{n}$$

$$\overset{\text{h)}}{\geqslant} \frac{1}{n}\sup_{p_{\mathsf{X}_1^n}} D_{s+}^{\frac{\epsilon}{1-\delta}+\gamma}(p_{\mathsf{X}_1^n}\cdot W_{\mathsf{Y}|\mathsf{X}}^{\otimes n}\|p_{\mathsf{X}_1^n}\times p_{\mathsf{Y}_1^n}) + \frac{\log\delta+\log\gamma}{n} \quad (*)$$

for all $\delta\in(0,1)$ and $\gamma\in(0,1-\frac{\epsilon}{1-\delta})$, where f) uses the relaxed bound (13), and g) uses Lemma 8, and h) uses Lemma 7.

By restricting $p_{\mathsf{X}_1^n}$ to i.i.d. pmfs, one can further lower bound $(*)$ by

$$R \geqslant (*) \geqslant \frac{1}{n}\sup_{p_\mathsf{X}} D_{s+}^{\frac{\epsilon}{1-\delta}+\gamma}((p_\mathsf{X}\cdot W_{\mathsf{Y}|\mathsf{X}})^{\otimes n}\|(p_\mathsf{X}\times p_\mathsf{Y})^{\otimes n}) + \frac{\log\delta+\log\gamma}{n} \quad (**)$$

for all $\epsilon,\delta,\gamma>0$ such that $\frac{\epsilon}{1-\delta}+\gamma<1$. For any fixed ϵ,δ,γ, it is clear that the second line in the above equation tends to 0 as $n\to\infty$; and the first line tends to $\sup_{p_\mathsf{X}} D(p_\mathsf{X}\cdot W_{\mathsf{Y}|\mathsf{X}}\|p_\mathsf{X}\times p_\mathsf{Y}) = C(W)$ as a famous result of the information spectrum method [14]. Since the above inequalities hold for all n sufficiently large, it must hold that

$$R \geqslant \lim_{n\to\infty}(**) = C(W). \qquad \square$$

V. Conclusion

In the paper, we have proposed and proved an achievability bound and a converse bound for the attainable sizes of ϵ-simulation codes for DMCs. Using these two bounds, we have provided an alternative proof to the reverse Shannon theorem. For future works, the higher order terms of the error exponents could be obtained along the same line of work. Despite that we have assumed unlimited supply of shared randomness in our setup, the protocol in our achievability proof uses finite amount of such randomness. Although optimal amount of shared randomness has already been quantified in asymptotic setups, it is an interesting yet open problem to study the trade-offs between the simulation code length and the amount of shared randomness needed.

VI. Deferred Proofs

A. Proof of Lemma 1

Proof. Let $p_{\mathsf{X}'\mathsf{Y}'}\in\mathcal{P}(\mathcal{X}\times\mathcal{Y})$ be an optimal distribution of the minimization problem in the definition of $D_{\max}^{\epsilon-\delta,\mathsf{X}}$ (see (5)), namely

- $\|p_{\mathsf{X}'\mathsf{Y}'}-p_{\mathsf{XY}}\|_{\text{tvd}}\leqslant\epsilon-\delta$,
- $\log\frac{p_{\mathsf{X}'\mathsf{Y}'}(x,y)}{p(x)\cdot q(y)}\leqslant D_{\max}^{\epsilon-\delta}(p_{\mathsf{XY}}\|p_\mathsf{X}\times q_\mathsf{Y})$ for all x,y,
- $p_{\mathsf{X}'}=p_\mathsf{X}$.

Now, construct random variables $(\mathsf{J}', \mathsf{X}', \mathsf{Y}_1, \ldots, \mathsf{Y}_M)$ from $(\mathsf{X}', \mathsf{Y}')$ in the same fashion as the construction of $(\mathsf{J}, \mathsf{X}, \mathsf{Y}_1, \ldots, \mathsf{Y}_M)$ from (X, Y). Then,

$$
\begin{aligned}
&D\big(p_{\mathsf{X}'\mathsf{Y}'_1,\ldots,\mathsf{Y}'_M} \big\| p_{\mathsf{X}'} \times q_{\mathsf{Y}_1} \times \cdots \times q_{\mathsf{Y}_M}\big) \\
&\overset{a)}{=} \frac{1}{M} \sum_{j=1}^{M} \Bigg\{ D\Big(p_{\mathsf{X}'\mathsf{Y}'_j} \Big\| p_{\mathsf{X}'} \times q_{\mathsf{Y}_j}\Big) \\
&\qquad - D\Bigg(p_{\mathsf{X}'\mathsf{Y}'_j} \times \prod_{i \neq j} q_{\mathsf{Y}'_i} \Bigg\| p_{\mathsf{X}'\mathsf{Y}'_1,\ldots,\mathsf{Y}'_M}\Bigg)\Bigg\} \\
&\overset{b)}{\leqslant} \frac{1}{M} \sum_{j=1}^{M} \Bigg\{ D\Big(p_{\mathsf{X}'\mathsf{Y}'_j} \Big\| p_{\mathsf{X}'} \times q_{\mathsf{Y}_j}\Big) \\
&\qquad - D\Big(p_{\mathsf{X}'\mathsf{Y}'_j} \Big\| \frac{1}{M} p_{\mathsf{X}'\mathsf{Y}'_j} + (1 - \frac{1}{M}) p_{\mathsf{X}'} \times q_{\mathsf{Y}'_j}\Big)\Bigg\} \\
&= \sum_{x,y} p_{\mathsf{X}'\mathsf{Y}'}(x,y) \cdot \log \frac{\frac{1}{M} \cdot p_{\mathsf{X}'\mathsf{Y}'}(x,y) + (1 - \frac{1}{M}) \cdot p_{\mathsf{X}'}(x) \cdot q_{\mathsf{Y}'}(y)}{p_{\mathsf{X}'}(x) \cdot q_{\mathsf{Y}}(y)} \\
&\leqslant \sum_{x,y} p_{\mathsf{X}'\mathsf{Y}'}(x,y) \cdot \log \left(1 + 2^{-R} \cdot \frac{p_{\mathsf{X}'\mathsf{Y}'}(x,y)}{p_{\mathsf{X}'}(x) \cdot q_{\mathsf{Y}}(y)}\right) \\
&\overset{c)}{\leqslant} 2^{-R + D_{\max}^{\epsilon-\delta,\mathsf{X}}(p_{\mathsf{XY}} \| p_{\mathsf{X}} \cdot q_{\mathsf{Y}})} \leqslant \frac{\delta^2}{2}
\end{aligned}
$$

where in step a) we used the fact that $D(\sum_i \lambda_i p_i \| q) = \sum_i \lambda_i \left(D(p_i \| q) - D(p_i \| \sum_i \lambda_i p_i)\right)$ for $\lambda_i > 0$, and in step b) we applied monotonicity of the information divergence, and step c) is based on the fact that $\log(1+x) \leqslant x$ for all $x > 0$. By Pinsker's inequality [15, Theorem 2.33], we have

$$
\begin{aligned}
&\big\|p_{\mathsf{X}'\mathsf{Y}'_1,\ldots,\mathsf{Y}'_M} - p_{\mathsf{X}} \times q_{\mathsf{Y}_1} \times \cdots \times q_{\mathsf{Y}_M}\big\|_{\mathrm{tvd}} \\
&= \big\|p_{\mathsf{X}'\mathsf{Y}'_1,\ldots,\mathsf{Y}'_M} - p_{\mathsf{X}'} \times q_{\mathsf{Y}_1} \times \cdots \times q_{\mathsf{Y}_M}\big\|_{\mathrm{tvd}} \\
&\leqslant \sqrt{\frac{1}{2} D\big(p_{\mathsf{X}'\mathsf{Y}'_1,\ldots,\mathsf{Y}'_M} \big\| p_{\mathsf{X}'} \times q_{\mathsf{Y}_1} \times \cdots \times q_{\mathsf{Y}_M}\big)} \leqslant \delta.
\end{aligned}
$$

By triangular inequality for TVD, we have

$$
\begin{aligned}
&\|p_{\mathsf{X},\mathsf{Y}_1,\ldots,\mathsf{Y}_M} - p_{\mathsf{X}} \times q_{\mathsf{Y}_1} \times \cdots \times q_{\mathsf{Y}_M}\|_{\mathrm{tvd}} \\
&\leqslant \big\|p_{\mathsf{X},\mathsf{Y}_1,\ldots,\mathsf{Y}_M} - p_{\mathsf{X}',\mathsf{Y}'_1,\ldots,\mathsf{Y}'_M}\big\|_{\mathrm{tvd}} + \\
&\quad \big\|p_{\mathsf{X}',\mathsf{Y}'_1,\ldots,\mathsf{Y}'_M} - p_{\mathsf{X}} \times q_{\mathsf{Y}_1} \times \cdots \times q_{\mathsf{Y}_M}\big\|_{\mathrm{tvd}} \\
&\leqslant \|p_{\mathsf{X}'\mathsf{Y}'} - p_{\mathsf{XY}}\|_{\mathrm{tvd}} + \delta \leqslant \epsilon.
\end{aligned}
$$

□

B. *Proof of Lemma 5*

Proof. Denote $a \triangleq \sup_{x \in \mathcal{X}} D_{s+}^{\epsilon}\big(p_{\mathsf{Y}|\mathsf{X}}(\cdot|x) \big\| q_{\mathsf{Y}|\mathsf{X}}(\cdot|x)\big)$, *i.e.*, $a \geqslant D_{s+}^{\epsilon}\big(p_{\mathsf{Y}|\mathsf{X}}(\cdot|x) \big\| q_{\mathsf{Y}|\mathsf{X}}(\cdot|x)\big)$ for all $x \in \mathcal{X}$. By definition of D_{s+}, we know that for all $\delta > 0$

$$
P_{\mathsf{Y}_x \sim p_{\mathsf{Y}|\mathsf{X}}(\cdot|x)}\left[\log \frac{p_{\mathsf{Y}|\mathsf{X}}(\mathsf{Y}_x|x)}{q_{\mathsf{Y}|\mathsf{X}}(\mathsf{Y}_x|x)} > a + \delta\right] < \epsilon \quad \forall x \in \mathcal{X}.
$$

Define subsets $\mathcal{A}^a \triangleq \left\{(x,y) \in \mathcal{X} \times \mathcal{Y} \middle| \log \frac{p_{\mathsf{Y}|\mathsf{X}}(y|x)}{q_{\mathsf{Y}|\mathsf{X}}(y|x)} > a + \delta\right\}$, and $A_x^a \triangleq \left\{y \in \mathcal{Y} \middle| \log \frac{p_{\mathsf{Y}|\mathsf{X}}(y|x)}{q_{\mathsf{Y}|\mathsf{X}}(y|x)} > a + \delta\right\}$ for each $x \in \mathcal{X}$. We have

$$
\begin{aligned}
p_{\mathsf{XY}}(\mathcal{A}^a) &= \sum_{x \in \mathcal{X}} p_{\mathsf{X}}(x) \cdot p_{\mathsf{Y}|\mathsf{X}=x}(\mathcal{A}_x^a) \\
&= \sum_{x \in \mathcal{X}} p_{\mathsf{X}}(x) \cdot P_{\mathsf{Y}_x \sim p_{\mathsf{Y}|\mathsf{X}}(\cdot|x)}\left[\log \frac{p_{\mathsf{Y}|\mathsf{X}}(\mathsf{Y}_x|x)}{q_{\mathsf{Y}|\mathsf{X}}(\mathsf{Y}_x|x)} > a + \delta\right] < \epsilon.
\end{aligned}
$$

Thus, by definition, $D_{s+}^{\epsilon}\big(p_{\mathsf{Y}|\mathsf{X}} \cdot p_{\mathsf{X}} \big\| q_{\mathsf{Y}|\mathsf{X}} \cdot p_{\mathsf{X}}\big) \leqslant a + \delta$. Since this holds for all $\delta > 0$, it must also hold when $\delta = 0$, which finishes the proof. □

C. *Proof of Lemma 7 [12, Lemma 10]*

Proof. Let $a^\star = \inf_{q_{\mathsf{Y}}} D_{s+}^{\epsilon}(p_{\mathsf{XY}} \| p_{\mathsf{X}} \times q_{\mathsf{Y}}) \geqslant 0$. There must exists some pmf $q_{\mathsf{Y}}^\star$ such that $p_{\mathsf{XY}}(\mathcal{A}) \leqslant \epsilon$ where

$$
\mathcal{A} \triangleq \left\{(x,y) \in \mathcal{X} \times \mathcal{Y} : \log \frac{p_{\mathsf{XY}}(x,y)}{p_{\mathsf{X}}(x) \cdot q_{\mathsf{Y}}^\star(y)} > a^\star\right\}.
$$

Define the set

$$
\mathcal{B} \triangleq \left\{(x,y) \in \mathcal{X} \times \mathcal{Y} : q_{\mathsf{Y}}^\star(y) > \frac{1}{\delta} p_{\mathsf{Y}}(y)\right\}.
$$

We have

$$
p_{\mathsf{XY}}(\mathcal{B}) = \sum_{y : q_{\mathsf{Y}}^\star(y) > \frac{1}{\delta} p_{\mathsf{Y}}(y)} p_{\mathsf{Y}}(y) < \sum_{y} \delta \cdot q_{\mathsf{Y}}^\star(y) = \delta.
$$

Notice that for all $(x,y) \notin \mathcal{A} \cup \mathcal{B}$, we have

$$
p_{\mathsf{XY}}(x,y) \leqslant \frac{2^{a^\star}}{\delta} \cdot p_{\mathsf{X}}(x) \cdot p_{\mathsf{Y}}(y).
$$

Thus,

$$
\begin{aligned}
&p_{\mathsf{XY}}\left(\left\{(x,y) : \log \frac{p_{\mathsf{XY}}(x,y)}{p_{\mathsf{X}}(x) \cdot p_{\mathsf{Y}}(y)} > a^\star - \log \delta\right\}\right) \\
&\leqslant p_{\mathsf{XY}}(\mathcal{A} \cup \mathcal{B}) \leqslant p_{\mathsf{XY}}(\mathcal{A}) + p_{\mathsf{XY}}(\mathcal{B}) < \epsilon + \delta.
\end{aligned}
$$

Since $a^\star - \log \delta \geqslant 0$, we know $D_{s+}^{\epsilon+\delta}(p_{\mathsf{XY}} \| p_{\mathsf{X}} \times p_{\mathsf{Y}}) \leqslant a^\star - \log \delta$. □

D. *Proof of Lemma 8 [12]*

Proof. Let $p_{\mathsf{XY}}^\star \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ such that $\|p_{\mathsf{XY}}^\star - p_{\mathsf{XY}}\|_{\mathrm{tvd}} \leqslant \epsilon$, $p_{\mathsf{X}}^\star = p_{\mathsf{X}}$, and $D_{\max}^{\epsilon,\mathsf{X}}(p_{\mathsf{XY}} \| p_{\mathsf{X}} \times q_{\mathsf{Y}}) = D_{\max}^{\epsilon}(p_{\mathsf{XY}}^\star \| p_{\mathsf{X}} \times q_{\mathsf{Y}})$. Define the set

$$
\mathcal{A} \triangleq \left\{(x,y) \in \mathcal{X} \times \mathcal{Y} : \frac{p_{\mathsf{XY}}(x,y)}{p_{\mathsf{XY}}^\star(x,y)} \geqslant \frac{1}{\delta}\right\}.
$$

Noting that

$$
\epsilon \geqslant p_{\mathsf{XY}}(\mathcal{A}) - p_{\mathsf{XY}}^\star(\mathcal{A}) \geqslant (1 - \delta) \cdot p_{\mathsf{XY}}(\mathcal{A}),
$$

we have

$$
p_{\mathsf{XY}}(\mathcal{A}) \leqslant \frac{\epsilon}{1 - \delta}.
$$

On the other hand, for all $(x,y) \notin \mathcal{A}$, we have

$$
\begin{aligned}
p_{\mathsf{XY}}(x,y) &< \frac{1}{\delta} \cdot p_{\mathsf{XY}}^\star(x,y) \\
&\leqslant \frac{1}{\delta} \cdot 2^{D_{\max}^{\epsilon,\mathsf{X}}(p_{\mathsf{XY}} \| p_{\mathsf{X}} \times q_{\mathsf{Y}})} \cdot p_{\mathsf{X}}(x) \cdot q_{\mathsf{Y}}(y).
\end{aligned}
$$

Thus, by picking $a \triangleq D_{\max}^{\epsilon,\mathsf{X}}(p_{\mathsf{XY}} \| p_{\mathsf{X}} \times q_{\mathsf{Y}}) - \log \delta \geqslant 0$, we have

$$
p_{\mathsf{XY}}\left(\left\{(x,y) : \log \frac{p_{\mathsf{XY}}(x,y)}{p_{\mathsf{X}}(x) \cdot p_{\mathsf{Y}}(y)} > a\right\}\right) \leqslant p_{\mathsf{XY}}(\mathcal{A}) \leqslant \frac{\epsilon}{1 - \delta},
$$

which implies $D_{s+}^{\frac{\epsilon}{1-\delta}}(p_{\mathsf{XY}} \| p_{\mathsf{X}} \times q_{\mathsf{Y}}) \leqslant a$. □

REFERENCES

[1] F. Haddadpour, M. H. Yassaee, S. Beigi, A. Gohari, and M. R. Aref, "Simulation of a channel with another channel," *IEEE Transactions on Information Theory*, vol. 63, no. 5, pp. 2659–2677, 2016.

[2] K. Fang, X. Wang, M. Tomamichel, and M. Berta, "Quantum channel simulation and the channel's smooth max-information," *IEEE Transactions on Information Theory*, vol. 66, no. 4, pp. 2129–2140, 2019.

[3] C. H. Bennett, P. W. Shor, J. A. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse shannon theorem," *IEEE Transactions on Information Theory*, vol. 48, no. 10, pp. 2637–2655, 2002.

[4] C. H. Bennett, I. Devetak, A. W. Harrow, P. W. Shor, and A. Winter, "The quantum reverse shannon theorem and resource tradeoffs for simulating quantum channels," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2926–2959, 2014.

[5] M. Berta, M. Christandl, and R. Renner, "The quantum reverse shannon theorem based on one-shot information theory," *Communications in Mathematical Physics*, vol. 306, no. 3, pp. 579–615, 2011.

[6] P. W. Cuff, H. H. Permuter, and T. M. Cover, "Coordination capacity," *IEEE Transactions on Information Theory*, vol. 56, no. 9, pp. 4181–4206, 2010.

[7] P. Cuff, "Distributed channel synthesis," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7071–7096, 2013.

[8] G. Cervia, T. J. Oechtering, and M. Skoglund, "(ϵ, n) fixed-length strong coordination capacity," in *2021 IEEE Information Theory Workshop (ITW)*, 2021, pp. 1–6.

[9] A. Anshu, V. K. Devabathini, and R. Jain, "Quantum communication using coherent rejection sampling," *Physical review letters*, vol. 119, no. 12, p. 120506, 2017.

[10] A. Anshu, R. Jain, and N. A. Warsi, "A unified approach to source and message compression," *arXiv preprint arXiv:1707.03619*, 2017.

[11] M. Berta, "Quantum side information: Uncertainty relations, extractors, channel simulations," *arXiv preprint arXiv:1310.4581*, 2013.

[12] A. Anshu, M. Berta, R. Jain, and M. Tomamichel, "Partially smoothed information measures," *IEEE Transactions on Information Theory*, vol. 66, no. 8, pp. 5022–5036, 2020.

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

[14] S. Verdú *et al.*, "A general formula for channel capacity," *IEEE Transactions on Information Theory*, vol. 40, no. 4, pp. 1147–1157, 1994.

[15] R. W. Yeung, *Information theory and network coding*. Springer Science & Business Media, 2008.