STRONG ASYMPTOTIC COMPOSITION THEOREMS FOR SIBSON MUTUAL INFORMATION

Benjamin Wu
Electrical and Computer Engineering
Cornell University
Ithaca, NY 14850
bhw49@cornell.edu

Aaron B. Wagner
Electrical and Computer Engineering
Cornell University
Ithaca, NY 14850
wagner@cornell.edu

Ibrahim Issa
Dept. of Electrical and Computer Engineering
American University of Beirut
Beirut, Lebanon
ii19@aub.edu.lb

G. Edward Suh
Electrical and Computer Engineering
Cornell University
Ithaca, NY 14850
suh@ece.cornell.edu

November 23, 2021

ABSTRACT

We characterize the growth of the Sibson and Arimoto mutual informations and α-maximal leakage, of any order that is at least unity, between a random variable and a growing set of noisy, conditionally independent and identically-distributed observations of the random variable. Each of these measures increases exponentially fast to a limit that is order- and measure-dependent, with an exponent that is order- and measure-independent.

1 Introduction

In the context of information leakage, composition theorems characterize how leakage increases as a result of multiple, independent, noisy observations of the sensitive data. Equivalently, they characterize how security (or privacy) degrades under the "composition" of multiple observations (or queries). In practice, attacks are often sequential in nature, whether the application is side channels in computer security [1–3] or database privacy [4–6]. Thus composition theorems are practically relevant. They also raise theoretical questions that are interesting in their own right.

Various composition theorems for differential privacy and its variants have been established (e.g., [4–6]). For the information-theoretic metrics of mutual information and maximal leakage [7–10] (throughout we assume discrete alphabets and base-2 logarithms)

$$I(X;Y) = \sum_{x,y} P(x,y) \log \frac{P(x,y)}{P(x)P(y)} \tag{1}$$

$$\mathcal{L}(X \to Y) = \log \sum_y \max_{x:P(x)>0} P(y|x), \tag{2}$$

and α-maximal leakage [11], less is known. While some results are available in the case that $P(y|x)$ is not known [12], here we assume it is known. For the metrics in (1)-(2) it is straightforward to show the "weak" composition theorem that if $Y_1, \ldots, Y_n$ are conditionally independent given X, then

$$I(X;Y^n) \leq \sum_{i=1}^n I(X;Y_i)$$

$$\mathcal{L}(X \to Y^n) \le \sum_{i=1}^{n} \mathcal{L}(X \to Y_i).$$

These bounds are indeed weak in that if $Y_1, \ldots, Y_n$ are conditionally i.i.d. given X, then as $n \to \infty$, the right-hand sides generally tend to infinity while the left-hand sides remain bounded. A "strong" (asymptotic) composition theorem would identify the limit and characterize the speed of convergence.

We prove such a result for both mutual information and maximal leakage. The limits are readily identified as the entropy and $\log$-support size, respectively, of a minimal sufficient statistic of Y given X. In both cases, the speed of convergence to the limit is exponential, and the exponent is the same. Specifically, it is the minimum Chernoff information among all pairs of distinct distributions $Q_{Y|X}(\cdot|x)$ and $Q_{Y|X}(\cdot|x')$.

Mutual information and maximal leakage are both instances of Sibson mutual information [10, 13, 14], the former being order 1 and the latter being order ∞. The striking fact that the exponents governing the convergence to the limit are the same at these two extreme points suggests that Sibson mutual information of all orders satisfies a strong asymptotic composition theorem, with the convergence rate (but not the limit) being independent of the order. Meanwhile, Shannon mutual information can also be viewed as Arimoto mutual information of order 1 [15], and α-maximal leakage is equivalently expressed as a maximization of Sibson or Arimoto mutual information of order α over $P(X)$ for $\alpha > 1$; for $\alpha = 1$, it equals Shannon mutual information [11], as opposed to the Shannon capacity. Due to the intimate interrelation between these measures, it is reasonable to suspect that similar strong asymptotic composition theorems obtain for them all. Indeed, we prove strong composition theorems for Sibson mutual information, Arimoto mutual information, and α-maximal leakage, for all orders of at least unity. In particular, we find that they all approach their respective limits at the same α-independent exponential rate, namely the minimum Chernoff information mentioned earlier.

The composition theorems proven here are different in nature from those in the differential privacy literature. Here we assume that the relevant probability distributions are known, and we characterize the growth of leakage with repeated looks from those distributions. We also assume that $Y_1, \ldots, Y_n$ are conditionally i.i.d. given X. Composition theorems in differential privacy consider the worst-case distributions given leakage levels for each of $Y_1, \ldots, Y_n$ individually, assuming only conditional independence.

Although our motivation is averaging attacks in side channels, the results may have some use in capacity studies of channels with multiple conditionally i.i.d. outputs given the input [16, Prob. 7.20].

2 Sibson, Arimoto, Rényi, and Chernoff

This study relies on both Sibson's and Arimoto's tunable mutual information metrics as well as α-maximal leakage. All random variables in the paper are assumed discrete.

Definition 1 ([13, 14]). *The* Sibson mutual information of order α *between random variables X and Y is defined by*

$$I_\alpha^S(X;Y) = \frac{\alpha}{\alpha - 1} \log \sum_{y \in \mathcal{Y}} \Big(\sum_{x \in \mathcal{X}} P(x) P(y|x)^\alpha \Big)^{1/\alpha}, \tag{3}$$

for $\alpha \in (0,1) \cap (1, \infty)$ and for $\alpha = 1$ and $\alpha = \infty$ by its continuous extensions. These are

$$I_1^S(X;Y) = I(X;Y)$$
$$I_\infty^S(X;Y) = \mathcal{L}(X \to Y),$$

defined in (1)-(2) above.

Definition 2 ([15]). *The* Arimoto mutual information of order α *between random variables X and Y is defined by*

$$I_\alpha^A(X;Y) = \frac{\alpha}{\alpha - 1} \log \sum_{y \in \mathcal{Y}} \Big(\frac{\sum_{x \in \mathcal{X}} P(x)^\alpha P(y|x)^\alpha}{\sum_{x \in \mathcal{X}} P(x)^\alpha} \Big)^{1/\alpha} \tag{4}$$

for $\alpha \in (0,1) \cap (1, \infty)$ and for $\alpha = 1$ and $\alpha = \infty$ by its continuous extensions. Note that [15]

$$I_1^A(X;Y) = I(X;Y)$$

but

$$I_\infty^A(X;Y) \ne \mathcal{L}(X \to Y).$$

2

Definition 3 ([11]). *The α-maximal leakage for $\alpha \in (1, \infty]$ is equivalently defined using either Sibson or Arimoto mutual information as:*[1]

$$\mathcal{L}_\alpha^{max}(X \to Y) = \max_{Q(X)} I_\alpha^S(X;Y) = \max_{Q(X)} I_\alpha^A(X;Y), \tag{5}$$

where the maxima are over all distributions of X that have full support. For $\alpha = 1$, we have

$$\mathcal{L}_\alpha^{max}(X \to Y) = I(X;Y). \tag{6}$$

as opposed to the (Shannon) capacity

$$\mathcal{C}(X;Y) = \max_{Q(X)} I(X;Y). \tag{7}$$

Liao *et al.* [11] define α-maximal leakage operationally. The identities in (5)-(6) are a theorem in that work, which we shall take as a definition. Likewise, Issa *et al.* [7] define maximal leakage operationally, and (2) is a theorem that we take as a definition.

We are interested in how $I_\alpha^S(X;Y^n)$, $I_\alpha^A(X;Y^n)$, and $\mathcal{L}_\alpha^{max}(X \to Y^n)$ grow with n when $Y_1, \ldots, Y_n$ are conditionally i.i.d. given X for $\alpha \geq 1$. The question for $\alpha < 1$ is meaningful in all cases but is not considered here because we are interested in the behavior of operational leakage measures, and the $\alpha < 1$ regime is not known to be relevant to measuring leakage. We do not consider the mutual information meaures put forward by Csiszár [17] and Lapidoth and Pfister [18, 19] for the same reason. For the quantities under study, we shall see that the limits are given by *Rényi entropy*. As they will be needed for proof later, we also define *Arimoto-Rényi conditional entropy* and *Rényi divergence*.

Definition 4. *The* Rényi entropy *of order α of a random variable X is given by:*

$$H_\alpha(X) = \frac{1}{1-\alpha} \log \sum_{x \in \mathcal{X}} P(x)^\alpha \tag{8}$$

for $\alpha \in (0,1) \cap (1, \infty)$ and for $\alpha = 0$, $\alpha = 1$, and $\alpha = \infty$ by its continuous extensions. These are

$$H_0(X) = \log |\{x : P(x) > 0\}| \tag{9}$$
$$H_1(X) = H(X) \tag{10}$$

$$H_\infty(X) = \log \frac{1}{\max_x P(x)}. \tag{11}$$

where $H(X)$ is the regular Shannon entropy.

Definition 5. *The* Arimoto-Rényi conditional entropy *of order α of a random variable X given Y is defined as:*

$$H_\alpha(X|Y) = \frac{\alpha}{1-\alpha} \log \sum_{y \in \mathcal{Y}} \left(\sum_{x \in \mathcal{X}} P(x)^\alpha P(y|x)^\alpha \right)^{\frac{1}{\alpha}}. \tag{12}$$

Remark. *One can verify that it holds*

$$I_\alpha^A(X;Y) = H_\alpha(X) - H_\alpha(X|Y). \tag{13}$$

Definition 6. *The* Rényi divergence *of order α between probability distributions P and Q is defined for $\alpha \in [0, \infty)$, $\alpha \neq 1$ as:*

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \sum_{x \in \mathcal{X}} P(x)^\alpha Q(x)^{1-\alpha}, \tag{14}$$

where the continuous extension at $\alpha = 1$ is given by the standard Kullback-Leibler divergence

$$D(P||Q) = \sum_{x \in \mathcal{X}} P(x) \log \frac{P(x)}{Q(x)}. \tag{15}$$

The speed of convergence of $I_\alpha^S(X;Y^n)$, $I_\alpha^A(X;Y^n)$, $\mathcal{L}_\alpha^{max}(X \to Y^n)$, and $\mathcal{C}(X;Y^n)$ and to their respective limits turns out to be governed by *Chernoff information*.

---

[1]The second equality for $1 < \alpha < \infty$ in (5) is apparent from (3) and (4) since the tilting of $P(x)$ in the latter can be absorbed into the maximization.

Definition 7 ([16]). *The* Chernoff information *between two probability mass functions, P_1 and P_2, over the same alphabet $\mathcal{X}$ is given as follows. First, for all $x \in \mathcal{X}$ and $\lambda \in [0,1]$, let:*

$$P_\lambda(x) = P_\lambda(P_1, P_2, x) = \frac{P_1(x)^\lambda P_2(x)^{1-\lambda}}{\sum_{x' \in \mathcal{X}} P_1(x')^\lambda P_2(x')^{1-\lambda}}. \tag{16}$$

Then the Chernoff information is given by

$$\mathscr{C}(P_1 || P_2) = D(P_{\lambda^*} || P_1) = D(P_{\lambda^*} || P_2), \tag{17}$$

where λ^* is any value of λ such that the above two relative entropies are equal. Equivalently, the Chernoff information is also given by:

$$\mathscr{C}(P_1 || P_2) = - \min_{0 \leq \lambda < 1} \log \left(\sum_x P_1(x)^\lambda P_2(x)^{1-\lambda} \right) \tag{18}$$

Since we consider finite alphabets, the Chernoff information is infinite if and only if P_1 and P_2 have disjoint support.

Other Notation: We use $\mathcal{P}_n$ to denote the set of all possible empirical distributions of Y^n. We let $\mathcal{P}$ denote the set of all possible probability distributions over $\mathcal{Y}$ For any $P \in \mathcal{P}$, let

$$T(P) = \{y^n \in \mathcal{Y}^n | P_{y^n} = P\},$$

where P_{y^n} is the empirical distribution of y^n. Note that $T(P)$ is empty if $P \notin \mathcal{P}_n$. We use $Q(\cdot)$ to denote the true distributions of X and Y^n. We let Q_x denote the distribution of Y given x for a given $x \in \mathcal{X}$. For any $P \in \mathcal{P}$, let $x_k(P)$ denote $x \in \mathcal{X}$ such that $D(P || Q_x)$ is the kth smallest relative entropy across all elements of $\mathcal{X}$. Ties can be broken by the ordering of $\mathcal{X}$.

We also define x-domains for fixed n in two slightly different ways. Let

$$D_x = \{P \in \mathcal{P} | D(P || Q_x) < D(P || Q_{x'}) \; \forall x' \neq x\} \tag{19}$$
$$\bar{D}_x = \{P \in \mathcal{P} | D(P || Q_x) \leq D(P || Q_{x'}) \; \forall x' \in \mathcal{X}\} \tag{20}$$

Note that for any $P \in \bar{D}_x$, $D(P || Q_x) = \min_{x' \in \mathcal{X}} D(P || Q_{x'})$.

3 The Result

Let X be a random variable with alphabet $\mathcal{X} = \{x_1, x_2, ... x_{|\mathcal{X}|}\}$. Let $Y^n = (Y_1, Y_2, ... Y_n)$ be a vector of discrete random variables with a shared alphabet $\mathcal{Y} = \{y_1, y_2, ... y_{|\mathcal{Y}|}\}$. We assume that $Y_1, Y_2, \ldots, Y_n$ are conditionally i.i.d. given X. We may assume, without loss of generality, that X and Y have full support. We will also assume that the distributions $P_{Y|X}(\cdot | x)$ are unique over x, which we call the *unique row assumption*. For Sibson mutual information and α-max leakage, this is without loss of generality, since we can divide $\mathcal{X}$ into equivalence classes based on their respective $P_{Y|X}(\cdot | x)$ distributions and define $\tilde{X}$ to be the equivalence class of X. Then both Markov chains $X \leftrightarrow \tilde{X} \leftrightarrow Y^n$ and $\tilde{X} \leftrightarrow X \leftrightarrow Y^n$ hold and so

$$I_\alpha^S(X; Y^n) = I_\alpha^S(\tilde{X}; Y^n) \tag{21}$$
$$\mathcal{L}_\alpha^{max}(X \to Y^n) = \mathcal{L}_\alpha^{max}(\tilde{X} \to Y^n), \tag{22}$$

by the data processing inequality for Sibson mutual information [20] and α-maximal leakage [11, Thm. 3]. We may then replace X with $\tilde{X}$ in the case of these measures. For Arimoto mutual information, the chain rule does not hold, and in fact an arbitrarily large discrepancy can exist between $I_\alpha^A(X; Y)$ and $I_\alpha^A(\tilde{X}; Y)$, as shown in Appendix B, where it is also shown that the unique row assumption is nonetheless still without loss of generality.

Our measures of interest satisfy the following upper bounds:

$$I(X; Y^n) \leq H(X) \tag{23}$$
$$\mathcal{C}(X; Y^n) \leq \log |\mathcal{X}| \tag{24}$$
$$I_\alpha^S(X; Y^n) \leq H_{1/\alpha}(X) \quad \text{[14, Ex. 2 and Thm. 3]} \tag{25}$$
$$I_\alpha^A(X; Y^n) \leq H_\alpha(X) \quad \text{[21, Prop. 3]} \tag{26}$$
$$\mathcal{L}_\alpha^{max}(X \to Y^n) \leq \begin{cases} H(X) & \text{if } \alpha = 1 \\ \log |\mathcal{X}| & \text{if } \alpha > 1 \end{cases} \quad \text{[11, Thm. 3]} \tag{27}$$
$$=: \mathcal{L}_\alpha(X),$$

where each inequality holds for all n and all $\alpha \in [1, \infty]$. Comparing (25) and (26) suggests that perhaps the Arimoto mutual information of order α should be associated with the Sibson mutual information of order $1/\alpha$; the identity in (5) suggests otherwise.

Our main result describes how fast these upper bounds are approached as $n \to \infty$.

Theorem 1. *Under the unique row assumption, for all $\alpha \in [1, \infty]$,*

$$\min_{x \neq x'} \mathscr{C}(Q_x \| Q_{x'}) = \lim_{n \to \infty} -\frac{1}{n} \log \Big(H(X) - I(X; Y^n) \Big) \tag{28}$$

$$= \lim_{n \to \infty} -\frac{1}{n} \log \Big(\log |\mathcal{X}| - \mathcal{C}(X; Y^n) \Big) \tag{29}$$

$$= \lim_{n \to \infty} -\frac{1}{n} \log \Big(H_{1/\alpha}(X) - I_\alpha^S(X; Y^n) \Big) \tag{30}$$

$$= \lim_{n \to \infty} -\frac{1}{n} \log \Big(H_\alpha(X) - I_\alpha^A(X; Y^n) \Big) \tag{31}$$

$$= \lim_{n \to \infty} -\frac{1}{n} \log \Big(\mathcal{L}_\alpha(X) - \mathcal{L}_\alpha^{max}(X \to Y^n) \Big). \tag{32}$$

Thus the Chernoff information governs the exponential rate-of-approach for all measures and for all values of α. This Chernoff information is infinite if Q_x and $Q_{x'}$ have disjoint support for all $x \neq x'$; in this case, the bounds in (23)-(27) are met with equality already for $n = 1$. Channels with this property arise naturally in certain applications [22].

Observe that (30)-(32) coincide with (28) when $\alpha = 1$. Also, (30) and (32) coincide for $\alpha = \infty$; otherwise the assertions are independent.

For continuous random variables, it is meaningful and interesting to study how $I_\alpha^S(X; Y^n)$, $\mathcal{C}(X; Y^n)$, and $\mathcal{L}_\alpha^{max}(X \to Y^n)$ grow with n. The behavior would be fundamentally different from the discrete case, however. See Aishwarya and Madiman [23] for a discussion of Arimoto mutual information in the continuous case.

The remainder of the paper is devoted to proving the various assertions contained within Theorem 1. The assertions are evidently asymptotic in nature, and our proofs are not opimtized to provide the best finite-n bounds. Numerical experiments show that in many cases our lower and upper bounds are quite far apart for moderate values of n.

4 Proof for Mutual Information and Capacity

We begin by proving (28) and (29), starting with the former. For this, we derive separate upper and lower bounds on $-H(X|Y^n)$. For the lower bound,

$$-H(X|Y^n) \equiv \sum_{y^n \in \mathcal{Y}^n} Q(y^n) \sum_{x \in \mathcal{X}} Q(x|y^n) \log Q(x|y^n) \tag{33}$$

$$= \sum_{P \in \mathcal{P}_n} \sum_{y^n \in T(P)} Q(y^n) \sum_{x \in \mathcal{X}} \frac{Q(y^n|x)Q(x)}{Q(y^n)} \log \frac{Q(y^n|x)Q(x)}{Q(y^n)} \tag{34}$$

$$= \sum_{P \in \mathcal{P}_n} \sum_{y^n \in T(P)} \sum_{x \in \mathcal{X}} \frac{1}{|T(P)|} Q(T(P)|x)Q(x)$$
$$\cdot \log \frac{\frac{1}{|T(P)|} Q(T(P)|x)Q(x)}{\sum_{x' \in \mathcal{X}} \frac{1}{|T(P)|} Q(T(P)|x')Q(x')} \tag{35}$$

$$= \sum_{P \in \mathcal{P}_n} \sum_{x \in \mathcal{X}} Q(T(P)|x)Q(x) \log \frac{Q(T(P)|x)Q(x)}{\sum_{x' \in \mathcal{X}} Q(T(P)|x')Q(x')} \tag{36}$$

$$= - \sum_{\substack{P \in \mathcal{P}_n: \\ Q(T(P)) > 0}} \Big[Q(T(P)|x_1(P))Q(x_1(P))$$
$$\cdot \log \frac{\sum_{x' \in \mathcal{X}} Q(T(P)|x')Q(x')}{Q(T(P)|x_1(P))Q(x_1(P))}$$

5

$$+ \sum_{\substack{x \neq x_1(P): \\ Q(T(P)|x)>0}} Q(T(P)|x)Q(x)$$

$$\cdot \log \frac{\sum_{x' \in \mathcal{X}} Q(T(P)|x')Q(x')}{Q(T(P)|x)Q(x)} \Bigg], \tag{37}$$

due to the convention that $0 \log 0 = 0$. Then, replacing weighted sums over x with their largest summand gives

$$\geq - \sum_{\substack{P \in \mathcal{P}_n: \\ Q(T(P))>0}} \Bigg[Q(T(P)|x_1(P))Q(x_1(P))$$

$$\cdot \log \left(1 + \frac{\sum_{x' \neq x_1(P)} Q(T(P)|x')Q(x')}{Q(T(P)|x_1(P))Q(x_1(P))} \right)$$

$$+ \max_{\substack{x \neq x_1(P): \\ Q(T(P)|x)>0}} \left\{ Q(T(P)|x) \log \frac{\max_{x' \in \mathcal{X}} Q(T(P)|x')}{Q(T(P)|x)Q(x)} \right\} \Bigg]. \tag{38}$$

Note that the entire expression inside the summation over P is 0 if $Q(T(P)|x_2(P)) = 0$. Letting $Q_{\min}(X) = \min_{x \in \mathcal{X}} Q(x)$ and using $\ln(1+x) \leq x$ for the $x = x_1(P)$ term,

$$\geq - \sum_{\substack{P \in \mathcal{P}_n: \\ Q(T(P))>0}} \Bigg[\frac{1}{\ln 2} \sum_{x' \neq x_1(P)} Q(T(P)|x')Q(x')$$

$$+ \max_{\substack{x \neq x_1(P): \\ Q(T(P)|x)>0}} \left\{ Q(T(P)|x) \right\}$$

$$\cdot \log \frac{1}{\min_{\substack{x \neq x_1(P): \\ Q(T(P)|x)>0}} Q(T(P)|x) \cdot Q_{\min}(X)} \Bigg] \tag{39}$$

$$\geq - \sum_{\substack{P \in \mathcal{P}_n: \\ Q(T(P))>0}} \Bigg[\frac{1}{\ln 2} 2^{-nD(P||Q_{x_2(P)})} + 2^{-nD(P||Q_{x_2(P)})}$$

$$\cdot \left[nD_{sup} + \log \frac{(n+1)^{|\mathcal{Y}|}}{Q_{\min}(X)} \right] \Bigg] \tag{40}$$

where

$$D_{sup} \equiv \sup_{\substack{x, P' \in \mathcal{P} \\ D(P'||Q_x)<\infty}} D(P'||Q_x) \tag{41}$$

$$= \sup_{\substack{x, P' \in \mathcal{P}: \\ D(P'||Q_x)<\infty}} \sum_{y \in \mathcal{Y}} P'(y) \log \frac{P'(y)}{Q(y|x)} \tag{42}$$

$$= \sup_{\substack{x, P' \in \mathcal{P}: \\ D(P'||Q_x)<\infty}} \sum_{y \in \mathcal{Y}} P'(y) \log \frac{1}{Q(y|x)} - H(P') \tag{43}$$

$$\leq \sup_x \log \frac{1}{\min_{Q(y|x)>0} Q(y|x)} < \infty. \tag{44}$$

Hence,

$$- H(X|Y^n)$$

$$\geq -(n+1)^{|\mathcal{Y}|} 2^{-nD_n^*} \Big[\frac{1}{\ln 2} + \log \frac{(n+1)^{|\mathcal{Y}|}}{Q_{\min}(X)} + nD_{sup} \Big] \tag{45}$$

where

$$D_n^* = \min_{P \in \mathcal{P}_n} D(P||Q_{x_2(P)}) \tag{46}$$

6

and P_n^* is its minimizer.

For the upper bound,

$$-H(X|Y^n)$$

$$= \sum_{P \in \mathcal{P}_n} \sum_{x \in \mathcal{X}} Q(T(P)|x)Q(x) \log \frac{Q(T(P)|x)Q(x)}{\sum_{x' \in \mathcal{X}} Q(T(P)|x')Q(x')} \tag{47}$$

$$\leq \sum_{x \in \mathcal{X}} Q(T(P_n^*)|x)Q(x) \log \frac{Q(T(P_n^*)|x)Q(x)}{\sum_{x' \in \mathcal{X}} Q(T(P_n^*)|x')Q(x')} \tag{48}$$

$$\leq Q(T(P_n^*)|x_1(P_n^*))Q(x_1(P_n^*))$$
$$\cdot \log \frac{Q(T(P_n^*)|x_1(P_n^*))Q(x_1(P_n^*))}{\sum_{x' \in \mathcal{X}} Q(T(P_n^*)|x')Q(x')} \tag{49}$$

$$= Q(T(P_n^*)|x_1(P_n^*))Q(x_1(P_n^*))$$
$$\cdot \log \left[1 - \frac{\sum_{x' \neq x_1(P_n^*)} Q(T(P_n^*)|x')Q(x')}{\sum_{x' \in \mathcal{X}} Q(T(P_n^*)|x')Q(x')} \right] \tag{50}$$

recalling that $-\ln(1-x) \geq x$,

$$\leq -Q(T(P_n^*)|x_1(P_n^*))Q(x_1(P_n^*))$$
$$\cdot \frac{\sum_{x' \neq x_1(P_n^*)} Q(T(P_n^*)|x')Q(x')}{\sum_{x' \in \mathcal{X}} Q(T(P_n^*)|x')Q(x')} \cdot \frac{1}{\ln 2} \tag{51}$$

$$\leq -Q(T(P_n^*)|x_1(P_n^*))Q(x_1(P_n^*))$$
$$\cdot \frac{Q(T(P_n^*)|x_2(P_n^*))Q(x_2(P_n^*))}{\max_{x' \in \mathcal{X}} Q(T(P_n^*)|x')} \cdot \frac{1}{\ln 2} \tag{52}$$

$$\leq -\frac{1}{(n+1)^{|\mathcal{Y}|}} 2^{-nD(P_n^*||Q_{x_1(P_n^*)})} Q(x_1(P_n^*))$$
$$\cdot \frac{2^{-nD_n^*} Q(x_2(P_n^*))}{(n+1)^{|\mathcal{Y}|} 2^{-nD(P_n^*||Q_{x_1(P_n^*)})}} \cdot \frac{1}{\ln 2} \tag{53}$$

$$= -\frac{Q(x_1(P_n^*))Q(x_2(P_n^*))}{(n+1)^{2|\mathcal{Y}|} \ln 2} 2^{-nD_n^*}. \tag{54}$$

As we have now shown that mutual information is upper and lower bounded by expressions of the form $H(X) - K_n \cdot 2^{-nD_n^*}$ for some subexponential sequence K_n, it remains to be shown that this exponent approaches the minimum Chernoff information as $n \to \infty$.

First, it can be shown using standard continuity arguments that

$$\lim_{n \to \infty} \inf_{P \in \mathcal{P}_n} D(P||Q_{x_2(P)}) = \inf_{P \in \mathcal{P}} D(P||Q_{x_2(P)}) \tag{55}$$

since $D(P||Q_{x_2(P)})$ is a continuous function of P. Finally, we arrive at the desired result using Lemma 4 in Appendix A.

Turning to the result for capacity, let Q_u denote the uniform distribution over $\mathcal{X}$. Then by (28) we have

$$\liminf_{n \to \infty} -\frac{1}{n} \log \left(\log |\mathcal{X}| - C(X; Y^n) \right) \tag{56}$$

$$\geq \liminf_{n \to \infty} -\frac{1}{n} \log \left(\log |\mathcal{X}| - I(X; Y^n) \Big|_{Q_u} \right) \tag{57}$$

$$= \min_{x \neq x'} \mathscr{C}(Q_x||Q_{x'}). \tag{58}$$

For the reverse inequality, for each n, let Q_n be a maximizer of $I(X; Y^n)$. Then from the previous observation, eventually we have

$$H(X) \Big|_{Q_n} - H(X|Y^n) \Big|_{Q_n} \geq I(X; Y^n) \Big|_{Q_u} \tag{59}$$

$$\geq \log |\mathcal{X}| - e^{-\frac{n}{2} \min_{x \neq x'} \mathscr{C}(Q_x||Q_{x'})}. \tag{60}$$

Dropping the second term from the left-hand side and using the fact that

$$D(Q_n||Q_u) = \log |\mathcal{X}| - H(X)\Big|_{Q_n} \tag{61}$$

this implies that, eventually,

$$e^{-\frac{n}{2}\min_{x \neq x'} \mathscr{C}(Q_x||Q_{x'})} \geq D(Q_n||Q_u). \tag{62}$$

Thus Q_n tends to Q_u as $n \to \infty$. Combining this fact with the bound in (54), we have that, eventually,

$$C(X;Y^n) = I(X;Y^n)\Big|_{Q_n} \tag{63}$$

$$= H(X)\Big|_{Q_n} - H(X|Y)\Big|_{Q_n} \tag{64}$$

$$\leq H(X)\Big|_{Q_n} - \frac{Q_n(x_1(P_n^*))Q(x_2(P_n^*))}{(n+1)^{2|\mathcal{Y}|}\ln 2}2^{-nD_n^*}, \tag{65}$$

$$\leq H(X)\Big|_{Q_n} - \frac{1}{4|\mathcal{X}|^2(n+1)^{2|\mathcal{Y}|}\ln 2}2^{-nD_n^*} \tag{66}$$

$$\leq \log|\mathcal{X}| - \frac{1}{4|\mathcal{X}|^2(n+1)^{2|\mathcal{Y}|}\ln 2}2^{-nD_n^*}, \tag{67}$$

which establishes the result since D_n^* converges to the Chernoff information as shown above.

5   Proof for Sibson ($\alpha \in (1,\infty)$)

We turn to (30), focusing on the regime $\alpha \in (1,\infty)$, since the $\alpha = 1$ case is established in (28) and the $\alpha = \infty$ case will be proven subsequently. First, we derive a lower bound of $I_\alpha^S(X;Y^n)$ for $\alpha > 1$ that will be useful in this and subsequent proofs.

Lemma 2.

$$I_\alpha^S(X;Y^n) \geq H_{1/\alpha}(X) - \frac{\alpha}{(\alpha-1)\ln 2}\left(\Gamma_n + \frac{\Gamma_n^2}{2(1-\Gamma_n)}\right) \tag{68}$$

for $\alpha > 1$, where

$$\Gamma_n = \min(1, (n+1)^{|\mathcal{Y}|} \cdot 2^{-n \cdot \min_{x \neq x'} \mathscr{C}(Q_x||Q_{x'})}). \tag{69}$$

Remark. *If Q_x and $Q_{x'}$ have disjoint support for every $x \neq x'$, then $\Gamma_n = 0$ and this lemma establishes that $I_\alpha^S(X;Y^n) = H_{1/\alpha}(X)$ for any $n \geq 1$.*

Proof. We use the D_x sets defined in (19) and (20):

$$I_\alpha^S(X;Y^n) \equiv \frac{\alpha}{\alpha-1}\log\sum_{y^n \in \mathcal{Y}^n}\left(\sum_{x \in \mathcal{X}}Q(x)Q(y^n|x)^\alpha\right)^{1/\alpha} \tag{70}$$

$$= \frac{\alpha}{\alpha-1}\log\sum_{P \in \mathcal{P}_n}\left(\sum_{x \in \mathcal{X}}Q(x)Q(T(P)|x)^\alpha\right)^{1/\alpha} \tag{71}$$

$$\geq \frac{\alpha}{\alpha-1}\log\sum_{x \in \mathcal{X}}\sum_{P \in D_x \cap \mathcal{P}_n}\left(\sum_{x' \in \mathcal{X}}Q(x')Q(T(P)|x')^\alpha\right)^{1/\alpha} \tag{72}$$

$$\geq \frac{\alpha}{\alpha-1}\log\sum_{x \in \mathcal{X}}Q(x)^{1/\alpha}\sum_{P \in D_x \cap \mathcal{P}_n}Q(T(P)|x) \tag{73}$$

$$= \frac{\alpha}{\alpha-1}\log\sum_{x \in \mathcal{X}}Q(x)^{1/\alpha}\left(1 - \sum_{P \in \mathcal{P}_n \setminus D_x}Q(T(P)|x)\right) \tag{74}$$

$$= \frac{\alpha}{\alpha-1}\log\Big(\sum_{x \in \mathcal{X}}Q(x)^{1/\alpha} \tag{75}$$

$$- \sum_{x \in \mathcal{X}}\sum_{P \in \mathcal{P}_n \setminus D_x}Q(x)^{1/\alpha}Q(T(P)|x)\Big).$$

Define

$$\gamma_n = \frac{\sum_{x \in \mathcal{X}} \sum_{P \in \mathcal{P}_n \setminus D_x} Q(x)^{1/\alpha} Q(T(P)|x)}{\sum_{x \in \mathcal{X}} Q(x)^{1/\alpha}} \le 1. \tag{76}$$

Then we can write

$$I_\alpha^S(X; Y^n) \ge \frac{\alpha}{\alpha - 1} \log \Big\{ \Big(\sum_{x \in \mathcal{X}} Q(x)^{1/\alpha} \Big)(1 - \gamma_n) \Big\} \tag{77}$$

$$= H_{1/\alpha}(X) + \frac{\alpha}{\alpha - 1} \log(1 - \gamma_n). \tag{78}$$

Note that

$$\ln(1 - \epsilon) = -\sum_{i=1}^{\infty} \frac{\epsilon^i}{i} \tag{79}$$

$$\ge -\epsilon - \frac{\epsilon}{2} \Big(\sum_{i=1}^{\infty} \epsilon^i \Big) = -\epsilon - \frac{\epsilon^2}{2(1 - \epsilon)} \tag{80}$$

for $0 < \epsilon < 1$. Hence,

$$I_\alpha^S(X; Y^n) \ge H_{1/\alpha}(X) + \frac{\alpha}{(\alpha - 1)\ln 2}\Big(-\gamma_n - \frac{\gamma_n^2}{2(1 - \gamma_n)}\Big). \tag{81}$$

The right-hand side in decreasing in γ_n over $[0, 1]$. We also have

$$\gamma_n \le \frac{\sum_{x \in \mathcal{X}} Q(x)^{1/\alpha}(n+1)^{|\mathcal{Y}|} \cdot \max_{P \in \mathcal{P}_n \setminus D_x} Q(T(P)|x)}{\sum_{x \in \mathcal{X}} Q(x)^{1/\alpha}} \tag{82}$$

$$\le \frac{\sum_{x \in \mathcal{X}} Q(x)^{1/\alpha}(n+1)^{|\mathcal{Y}|} \cdot \max_{x' \in \mathcal{X}} \max_{P \in \mathcal{P}_n \setminus D_{x'}} Q(T(P)|x')}{\sum_{x \in \mathcal{X}} Q(x)^{1/\alpha}} \tag{83}$$

$$= (n+1)^{|\mathcal{Y}|} \cdot \max_{x \in \mathcal{X}} \max_{P \in \mathcal{P}_n \setminus D_x} Q(T(P)|x) \tag{84}$$

$$\le (n+1)^{|\mathcal{Y}|} \cdot 2^{-n(\min_{x \in \mathcal{X}} \min_{P \in \mathcal{P}_n \setminus D_x} D(P||Q_x))} \tag{85}$$

$$\le (n+1)^{|\mathcal{Y}|} \cdot 2^{-n(\min_{x \ne x'} \inf_{P \in \bar{D}_{x'}} D(P||Q_x))} \tag{86}$$

$$= (n+1)^{|\mathcal{Y}|} \cdot 2^{-n \cdot \min_{x \ne x'} \mathscr{C}(Q_x||Q_{x'})}, \tag{87}$$

where we have used Lemma 4 in Appendix A. $\qquad\qquad\qquad\qquad\qquad \square$

We next prove an analogous upper bound.

Lemma 3. *For $\alpha > 1$, define*

$$F(x, P) = Q(x)Q(T(P)|x)^\alpha. \tag{88}$$

For each n, let $\{E_{x_i}^{(n)}\}_{i=1}^{|\mathcal{X}|}$ be a partition of $\mathcal{P}_n$ such that $P \in E_x^{(n)}$ implies $F(x, P) = \max_{x' \in \mathcal{X}} F(x', P)$. Then

$$I_\alpha^S(X; Y^n) \le H_{1/\alpha}(X) + \frac{\alpha}{(\alpha - 1)\ln 2} \frac{1}{\sum_{x \in \mathcal{X}} Q(x)^{1/\alpha}} \sum_{x \in \mathcal{X}} \sum_{P \notin E_x^{(n)}}$$
$$\cdot (F(x_1(P), P)^{1/\alpha - 1} - F(x, P)^{1/\alpha - 1})F(x, P)), \tag{89}$$

where for the remainder of this section we redefine $x_k(P)$ so that they are ordered by $F(x, P)$ instead of relative entropy. Note that this ordering now depends on n.

Proof. We have

$$I_\alpha^S(X; Y^n) = \frac{\alpha}{\alpha - 1} \log \sum_{x \in \mathcal{X}} \sum_{P \in E_x^{(n)}} \Big(\sum_{x' \in \mathcal{X}} F(x', P) \Big)^{1/\alpha} \tag{90}$$

$$= \frac{\alpha}{\alpha - 1} \log \sum_{x \in \mathcal{X}} \sum_{P \in E_x^{(n)}} F(x, P)^{1/\alpha} \Big(1 + \sum_{x' \ne x} \frac{F(x', P)}{F(x, P)} \Big)^{1/\alpha} \tag{91}$$

9

Using the Taylor series expansion of $(1 + x)^{1/\alpha}$ and discarding x^2 and higher-order terms (since $\frac{1}{\alpha} < 1$), we have

$$\leq \frac{\alpha}{\alpha - 1} \log \sum_{x \in \mathcal{X}} \sum_{P \in E_x^{(n)}} F(x, P)^{1/\alpha} \left(1 + \frac{1}{\alpha} \sum_{x' \neq x} \frac{F(x', P)}{F(x, P)}\right) \tag{92}$$

$$\leq \frac{\alpha}{\alpha - 1} \log \sum_{x \in \mathcal{X}} \sum_{P \in E_x^{(n)}} \Big(F(x, P)^{1/\alpha}$$
$$+ F(x, P)^{1/\alpha - 1} \sum_{x' \neq x} F(x', P)\Big), \tag{93}$$

where we have used the fact that $\alpha > 1$. Continuing,

$$= \frac{\alpha}{\alpha - 1} \log \sum_{x \in \mathcal{X}} \Big(\sum_{P \in E_x^{(n)}} F(x, P)^{1/\alpha}$$
$$+ \sum_{P \notin E_x^{(n)}} F(x_1(P), P)^{1/\alpha - 1} F(x, P) \Big) \tag{94}$$

$$= \frac{\alpha}{\alpha - 1} \log \sum_{x \in \mathcal{X}} \Big(\sum_{P \in E_x^{(n)}} F(x, P)^{1/\alpha}$$
$$+ \sum_{P \notin E_x^{(n)}} F(x_1(P), P)^{1/\alpha - 1} F(x, P)$$
$$+ \sum_{P \notin E_x^{(n)}} F(x, P)^{1/\alpha} - \sum_{P \notin E_x^{(n)}} F(x, P)^{1/\alpha} \Big) \tag{95}$$

$$= \frac{\alpha}{\alpha - 1} \log \sum_{x \in \mathcal{X}} \Big(\sum_{P \in \mathcal{P}_n} F(x, P)^{1/\alpha}$$
$$+ \sum_{P \notin E_x^{(n)}} \big(F(x_1(P), P)^{1/\alpha - 1} - F(x, P)^{1/\alpha - 1}\big) F(x, P) \Big) \tag{96}$$

Using $\ln(1 + x) \leq x$ then gives the result. $\qquad \square$

The lower bound in (30) for $\alpha \in (1, \infty)$ follows directly from Lemma 2. For the upper bound, pick $x_a \neq x_b$ and $P^* \in D_{x_b}$. Let $\{P_n\}_{n=1}^{\infty}$ be a sequence of types converging to P^*. From Lemma 3 we have

$$I_\alpha^S(X; Y^n) \leq H_{1/\alpha}(X) + \frac{\alpha}{(\alpha - 1) \ln 2} \frac{1}{\sum_{x \in \mathcal{X}} Q(x)^{1/\alpha}} \sum_{x \in \mathcal{X}} \sum_{P \notin E_x^{(n)}}$$
$$\cdot (F(x_1(P), P)^{1/\alpha - 1} - F(x, P)^{1/\alpha - 1}) F(x, P). \tag{97}$$

Note that eventually $P_n \in E_{x_b}^{(n)}$, $x_1(P_n) = x_b$ and $F(x_b, P_n)^{1/\alpha - 1} < \frac{1}{2} F(x_a, P_n)^{1/\alpha - 1}$. Thus, eventually,

$$\leq H_{1/\alpha}(X) + \frac{\alpha}{(\alpha - 1) \ln 2} \frac{1}{\sum_{x \in \mathcal{X}} Q(x)^{1/\alpha}}$$
$$\cdot (F(x_1(P_n), P_n)^{1/\alpha - 1} - F(x_a, P_n)^{1/\alpha - 1}) F(x_a, P_n). \tag{98}$$

$$\leq H_{1/\alpha}(X) - \frac{\alpha}{2(\alpha - 1) \ln 2} \frac{1}{\sum_{x \in \mathcal{X}} Q(x)^{1/\alpha}} F(x_a, P_n)^{1/\alpha} \tag{99}$$

$$\leq H_{1/\alpha}(X) - \frac{\alpha}{2(\alpha - 1) \ln 2} \frac{1}{\sum_{x \in \mathcal{X}} Q(x)^{1/\alpha}} Q_{\min}(X)^{1/\alpha}$$
$$\cdot \frac{1}{(n + 1)^{|\mathcal{Y}|}} 2^{-n D(P_n || Q_{x_a})} \tag{100}$$

where $Q_{\min}(X) = \min_{x \in \mathcal{X}} Q(x)$. This implies:

$$\limsup_{n \to \infty} -\frac{1}{n} \log \left(H_{1/\alpha}(X) - I_\alpha^S(X; Y^n) \right)$$
$$\leq \lim_{n \to \infty} D(P_n \| Q_{x_a}) = D(P^* \| Q_{x_a}). \tag{101}$$

Since $x_a \neq x_b$ and $P \in D_{x_b}$ were arbitrarily chosen, this implies:

$$\limsup_{n \to \infty} -\frac{1}{n} \log \left(H_{1/\alpha}(X) - I_\alpha^S(X; Y^n) \right)$$
$$\leq \min_{x \neq x'} \inf_{P \in \bar{D}_x} D(P \| Q_{x'}) = \min_{x \neq x'} \mathscr{C}(Q_x \| Q_{x'}), \tag{102}$$

where the last step used Lemma 4 in Appendix A.

6 Proof for Maximal Leakage

We turn to proving (30) for the case $\alpha = \infty$. While the lower bound on $I_\infty^S(X; Y^n)$ can be proven directly, we will instead note that it can be obtained from Lemma 2 by letting $\alpha \to \infty$ and then $n \to \infty$.

For the upper bound, recalling the x-domains defined in (19) and (20), fix $x_a \neq x_b \in \mathcal{X}$ and a $P \in D_{x_b}$ and let $\{P_n\}_{n=1}^\infty$ be a sequence such that $P_n \in \mathcal{P}_n$ for each n and $P_n \to P$. Then $P_n \in D_{x_b}$ eventually and

$$I_\infty^S(X; Y^n) \leq \log \sum_{x \in \mathcal{X}} \sum_{P \in \bar{D}_x \cap \mathcal{P}_n} Q(T(P)|x) \tag{103}$$

$$= \log \left[|\mathcal{X}| - \sum_{x \in \mathcal{X}} \sum_{P \in \mathcal{P}_n \setminus \bar{D}_x} Q(T(P)|x) \right] \tag{104}$$

$$\leq \log \left[|\mathcal{X}| - \sum_{P \in \mathcal{P}_n \setminus \bar{D}_{x_a}} Q(T(P)|x_a) \right] \tag{105}$$

$$\leq \log \left[|\mathcal{X}| - Q(T(P_n)|x_a) \right], \tag{106}$$

eventually. Thus for sufficiently large n,

$$I_\infty^S(X; Y^n)$$
$$\leq \log \left[|\mathcal{X}| - \frac{1}{(n+1)^{|\mathcal{Y}|}} 2^{-nD(P_n \| Q_{x_a})} \right] \tag{107}$$

$$\leq \log \left[|\mathcal{X}| \right] - \frac{1}{(\ln 2)|\mathcal{X}|(n+1)^{|\mathcal{Y}|}} 2^{-nD(P_n \| Q_{x_a})} \tag{108}$$

and

$$\limsup_{n \to \infty} -\frac{1}{n} \log \left(|\mathcal{X}| - I_\infty^S(X; Y^n) \right)$$
$$\leq \lim_{n \to \infty} D(P_n \| Q_{x_a}) = D(P \| Q_{x_a}). \tag{109}$$

Since $x_a \neq x_b$ and P were arbitrary, the result follows by Lemma 4 in Appendix A.

7 Proof for Arimoto

Note that (31) for the case $\alpha = 1$ has already been proven. We prove the lower and upper bounds for the $\alpha > 1$ case as follows.

7.1 Proof of Lower Bound

Proof. Let $|\mathcal{X}| = M$ and

$$\epsilon_{X|Y^n} = \min_{f:\mathcal{Y}^n \to X} P(X \neq f(Y^n)) \tag{110}$$

$$= 1 - E_{Y^n}[\max_x Q(x|Y^n)] \tag{111}$$

$$\leq 1 - p_{max} \text{ where } p_{max} = \max_X Q(X) \tag{112}$$

$$\leq 1 - \frac{1}{M}. \tag{113}$$

For $1 < \alpha < \infty$,

$$H_\alpha(X|Y^n) \leq \log M - d_\alpha(\epsilon_{X|Y^n}||1 - \frac{1}{M}) \tag{114}$$

where $d_\alpha(p||q)$ is the binary Renyi divergence ([24], Thm. 3):

$$d_\alpha(p||q) = \frac{1}{\alpha - 1} \log(p^\alpha q^{1-\alpha} + (1-p)^\alpha (1-q)^{1-\alpha}). \tag{115}$$

So,

$$I_\alpha^A(X;Y^n) = H_\alpha(X) - H_\alpha(X|Y^n) \tag{116}$$

$$\geq H_\alpha(X) - \log M + d_\alpha(\epsilon_{X|Y^n}||1 - \frac{1}{M}) \tag{117}$$

$$= H_\alpha(X) - \log M$$
$$+ \frac{1}{\alpha - 1} \log\left(\epsilon_{X|Y^n}^\alpha (1 - \frac{1}{M})^{1-\alpha}\right.$$
$$\left. + (1 - \epsilon_{X|Y^n})^\alpha (\frac{1}{M})^{1-\alpha}\right) \tag{118}$$

$$\geq H_\alpha(X) - \log M$$
$$+ \frac{1}{\alpha - 1} \log\left((1 - \epsilon_{X|Y^n})^\alpha (\frac{1}{M})^{1-\alpha}\right) \tag{119}$$

$$= H_\alpha(X) + \frac{\alpha}{\alpha - 1} \log(1 - \epsilon_{X|Y^n}) \tag{120}$$

Hence,

$$I_\alpha^A(X;Y^n) - H_\alpha(X) \geq \frac{\alpha}{\alpha - 1} \log(1 - \epsilon_{X|Y^n}) \tag{121}$$

which gives

$$\frac{\alpha - 1}{\alpha}[H_\alpha(X) - I_\alpha^A(X;Y^n)] \leq \log \frac{1}{1 - \epsilon_{X|Y^n}}. \tag{122}$$

For $0 < \epsilon \leq 1/2$,

$$\log \frac{1}{1 - \epsilon} = \log(1 + \frac{\epsilon}{1 - \epsilon}) \tag{123}$$

$$\leq \frac{\epsilon}{1 - \epsilon} \frac{1}{\ln 2} \tag{124}$$

$$\leq \frac{2\epsilon}{\ln 2}. \tag{125}$$

For all sufficiently large n, we have $\epsilon_{X|Y^n} \leq 1/2$ by the unique row assumption. Thus, combining (122) and (125), for all $1 < \alpha < \infty$,

$$\frac{2\epsilon_{X|Y^n}}{\ln 2} \geq \frac{\alpha - 1}{\alpha}[H_\alpha(X) - I_\alpha^A(X;Y^n)] \tag{126}$$

$$-\frac{1}{n} \log(\frac{2\epsilon_{X|Y^n}}{\ln 2}) \leq -\frac{1}{n} \log(\frac{\alpha - 1}{\alpha}[H_\alpha(X) - I_\alpha^A(X;Y^n)]), \tag{127}$$

and, taking $\alpha \to \infty$ in (127),

$$-\frac{1}{n}\log(\frac{2\epsilon_{X|Y^n}}{\ln 2}) \leq -\frac{1}{n}\log(H_\infty(X) - I_\infty^A(X;Y^n)). \tag{128}$$

Note that $\epsilon_{X|Y^n}$ is bounded as [24, Thm. 15]

$$\epsilon_{X|Y^n} \leq (M-1)\exp\left(-\min_{x \neq x'}\mathscr{C}(Q_x^n||Q_{x'}^n)\right). \tag{129}$$

Then, using Lemma 5 in Appendix A, for any $\alpha \in (1, \infty]$,

$$\min_{x \neq x'}\mathscr{C}(P_x||P_{x'})$$

$$\leq \liminf_{n \to \infty} -\frac{1}{n}\log\left[H_\alpha(X) - I_\alpha^A(X;Y^n)\right] \tag{130}$$

$\square$

7.2 Proof of Upper Bound

Proof. For $\alpha \in [0, \infty]$ [24, (165)],

$$H_\alpha(X|Y^n) \geq \log\frac{1}{1 - \epsilon_{X|Y^n}}. \tag{131}$$

Thus,

$$I_\alpha^A(X;Y^n) \leq H_\alpha(X) - \log\frac{1}{1 - \epsilon_{X|Y^n}} \tag{132}$$

$$\log\frac{1}{1 - \epsilon_{X|Y^n}} \leq H_\alpha(X) - I_\alpha^A(X;Y^n) \tag{133}$$

$$\frac{\epsilon_{X|Y^n}}{\ln 2} \leq H_\alpha(X) - I_\alpha^A(X;Y^n) \tag{134}$$

and so

$$\limsup_{n \to \infty} -\frac{1}{n}\log\epsilon_{X|Y^n} \geq \limsup_{n \to \infty} -\frac{1}{n}\log[H_\alpha(X) - I_\alpha^A(X;Y^n)]. \tag{135}$$

It remains to show that

$$\limsup_{n \to \infty} -\frac{1}{n}\log\epsilon_{X|Y^n} \leq \min_{x \neq x'}\mathscr{C}(Q_x||Q_{x'}). \tag{136}$$

To this end, for any $i \neq j$, we have

$$\epsilon_{X|Y^n} = E_{Y^n}[1 - \max_x Q(x|Y^n)] \tag{137}$$

$$\geq \sum_{y^n} Q(y^n)\min(Q(x_i|y^n), Q(x_j|y^n)) \tag{138}$$

$$\geq \min(Q(x_i), Q(x_j))\sum_{y^n} Q(y^n)\min\left(\frac{Q(x_i|y^n)}{Q(x_i)}, \frac{Q(x_j|y^n)}{Q(x_j)}\right) \tag{139}$$

$$= 2\min(Q(x_i), Q(x_j))\epsilon_{n,i,j}, \tag{140}$$

where

$$\epsilon_{n,i,j} = \frac{1}{2}\sum_{y^n}\min(Q(y^n|x_i), Q(y^n|x_j)) \tag{141}$$

is the error probability for the alternative problem in which X assumes only two values, x_i and x_j, which are equally likely, and we seek to guess X from Y^n. By [16, Thm. 11.9.1], we have

$$\lim_{n \to \infty} -\frac{1}{n}\log\epsilon_{n,i,j} = \mathscr{C}(Q_{x_i}||Q_{x_j}). \tag{142}$$

But i and j were arbitrary.

$\square$

13

8 Proof for α-Maximal Leakage

Note that for $\alpha = 1$, α-maximal leakage is given by regular mutual information, so that case is already proven.

8.1 Proof of Lower Bound

Proof. We obtain the lower bound by choosing $X \sim Q_u$, where $Q_u(X)$ denotes the uniform distribution over $\mathcal{X}$. Then

$$\mathcal{L}_\alpha^{max}(X \to Y) = \max_{Q(X)} I_\alpha^S(X; Y^n) \geq I_\alpha^S(X; Y^n)|_{Q_u(X)}. \tag{143}$$

Then by (30),

$$\liminf_{n \to \infty} -\frac{1}{n} \log(\log|\mathcal{X}| - \mathcal{L}_\alpha^{max}(X \to Y)) \geq \min_{x \neq x'} \mathscr{C}(Q_x \| Q_{x'}) \tag{144}$$

$\square$

8.2 Proof of Upper Bound

Proof. As with the proof for Shannon capacity, the idea is to show that the maximizing $Q(X)$ must eventually be contained in a neighborhood of the uniform distribution. Over this neighborhood, we can use Lemma 3 to uniformly bound the difference

$$\log|\mathcal{X}| - \max_{Q(X)} I_\alpha^S(X : Y^n). \tag{145}$$

First, for each n, let

$$Q_n(X) \in \arg\max_{Q(X)} I_\alpha^S(X; Y^n). \tag{146}$$

We have [14, Ex. 2 and Thm. 3]

$$H_{1/\alpha}(X)|_{Q_n(X)} \geq I_\alpha^S(X; Y^n)|_{Q_n(X)}, \tag{147}$$

and thus, by Lemma 2,

$$H_{1/\alpha}(X)|_{Q_n(X)} \geq I_\alpha^S(X; Y^n)|_{Q_u(X)} \tag{148}$$

$$\geq H_{1/\alpha}(X)|_{Q_u(X)}$$

$$- \frac{\alpha}{(\alpha - 1)\ln 2}(\Gamma_n + \frac{\Gamma_n^2}{2(1 - \Gamma_n)}). \tag{149}$$

Then,

$$H_{1/\alpha}(X)|_{Q_n(X)} \geq H_{1/\alpha}(X)|_{Q_u(X)}$$

$$- \frac{\alpha}{(\alpha - 1)\ln 2}(\Gamma_n + \frac{\Gamma_n^2}{2(1 - \Gamma_n)}) \tag{150}$$

$$H_{1/\alpha}(X)|_{Q_u(X)} - H_{1/\alpha}(X)|_{Q_n(X)}$$

$$\leq \frac{\alpha}{(\alpha - 1)\ln 2}(\Gamma_n + \frac{\Gamma_n^2}{2(1 - \Gamma_n)}) \tag{151}$$

$$D_{1/\alpha}(Q_n(X)\|Q_u(X))$$

$$\leq \frac{\alpha}{(\alpha - 1)\ln 2}(\Gamma_n + \frac{\Gamma_n^2}{2(1 - \Gamma_n)}) \equiv \epsilon_n, \tag{152}$$

where we have used the fact that $H_{1/\alpha}(X)|_{Q_u(X)} - H_{1/\alpha}(X)|_{Q_n(X)} = D_{1/\alpha}(Q_n(X)\|Q_u(X))$. Note that $\lim_{n \to \infty} \epsilon_n = 0$. Then, using the Rényi version of Pinsker's Inequality ([25, Thm. 31]),

$$D_{1/\alpha}(Q_u(X)\|Q_n(X)) \geq \frac{2}{\alpha} \sup_A |Q_n(A) - Q_u(A)|^2 \tag{153}$$

$$\geq \frac{2}{\alpha} \sup_x |Q_n(x) - Q_u(x)|^2 \tag{154}$$

and so

$$\epsilon_n \geq \frac{2}{\alpha} \sup_x |Q_n(x) - Q_u(x)|^2. \tag{155}$$

14

It also follows that, under this constraint,

$$\epsilon_n \geq \frac{2}{\alpha}(Q_u(x) - \min_{x'} Q_n(x'))^2 \tag{156}$$

$$\sqrt{\frac{\alpha\epsilon_n}{2}} \geq Q_u(x) - \min_{x'} Q_n(x') \tag{157}$$

$$\min_{x'} Q_n(x') \equiv Q_{\min,n}(X) \geq \frac{1}{|\mathcal{X}|} - \sqrt{\frac{\alpha\epsilon_n}{2}} \tag{158}$$

and similarly,

$$\max_{x'} Q_n(x') \equiv Q_{\max,n}(X) \leq \frac{1}{|\mathcal{X}|} + \sqrt{\frac{\alpha\epsilon_n}{2}} \tag{159}$$

Let A_n be the set of distributions over X that satisfy both (158) and (159) and note that $Q_n \in A_n$ eventually. Recalling (88), define

$$F(x, P, \tilde{Q}) = \tilde{Q}(x)Q(T(P)|x)^\alpha, \tag{160}$$

where we now indicate the dependence on the input distribution $\tilde{Q}(x)$. Similarly, we let $\{E^{(n)}_{x_i,\tilde{Q}}\}$ be a partition of $\mathcal{P}_n$ such that $P \in E^{(n)}_{x,\tilde{Q}}$ implies $F(x, P, \tilde{Q}) = \max_{x'} F(x', P, \tilde{Q})$ and we let $x_1(P, \tilde{Q})$, $x_2(P, \tilde{Q})$, ..., denote the letters of $\mathcal{X}$ in decreasing order of (160). By Lemma 3, we have, eventually

$$\max_{\tilde{Q}} I^S_\alpha(X; Y^n)$$

$$= \max_{\tilde{Q} \in A_n} I^S_\alpha(X; Y^n)$$

$$\leq \max_{\tilde{Q} \in A_n} H_{1/\alpha}(X) + \frac{\alpha}{(\alpha-1)\ln 2} \frac{1}{\sum_{x \in \mathcal{X}} \tilde{Q}(x)^{1/\alpha}} \sum_{x \in \mathcal{X}} \sum_{P \notin E^{(n)}_{x,\tilde{Q}}}$$

$$\cdot (F(x_1(P, \tilde{Q}), P, \tilde{Q})^{1/\alpha-1} - F(x, P, \tilde{Q})^{1/\alpha-1})F(x, P, \tilde{Q})). \tag{161}$$

Fix $x_a \neq x_b$ and $P^* \in D_{x_b}$ and let P_n be a sequence of types converging to P^*. Then for all sufficiently large n, we have that $P_n \in E^{(n)}_{x_b,\tilde{Q}}$ for all $\tilde{Q} \in A_n$. Then because the summands in (161) are nonpositive, we have

$$\max_{\tilde{Q} \in A_n} I^S_\alpha(X; Y^n)$$

$$\leq \max_{\tilde{Q} \in A_n} H_{1/\alpha}(X) + \frac{\alpha}{(\alpha-1)\ln 2} \frac{1}{\sum_{x \in \mathcal{X}} \tilde{Q}(x)^{1/\alpha}}$$

$$\cdot (F(x_1(P_n, \tilde{Q}), P_n, \tilde{Q})^{1/\alpha-1} - F(x_a, P_n, \tilde{Q})^{1/\alpha-1})F(x_a, P_n, \tilde{Q})). \tag{162}$$

Note that, eventually, $x_1(P_n, \tilde{Q}) = x_b$ for all $\tilde{Q} \in A_n$ and $F(x_b, P_n, \tilde{Q})^{1/\alpha-1} < \frac{1}{2}F(x_a, P_n, \tilde{Q})^{1/\alpha-1}$ for all $\tilde{Q} \in A_n$. The remainder of the argument proceeds analogously to the Sibson proof. Eventually, we have

$$\max_{\tilde{Q} \in A_n} I^S_\alpha(X; Y^n)$$

$$\leq \max_{\tilde{Q} \in A_n} H_{1/\alpha}(X) - \frac{1}{2}\frac{\alpha}{(\alpha-1)\ln 2} \cdot \frac{1}{\sum_{x \in \mathcal{X}} \tilde{Q}(x)^{1/\alpha}} \tag{163}$$

$$\cdot F(x_a, P_n, \tilde{Q})^{1/\alpha} \tag{164}$$

$$\leq \max_{\tilde{Q} \in A_n} H_{1/\alpha}(X) - \frac{1}{2}\frac{\alpha}{(\alpha-1)\ln 2} \cdot \frac{1}{|\mathcal{X}|\left(\frac{1}{|\mathcal{X}|} + \sqrt{\frac{\alpha\epsilon_n}{2}}\right)^{1/\alpha}} \tag{165}$$

$$\cdot \left(\frac{1}{|\mathcal{X}|} - \sqrt{\frac{\alpha\epsilon_n}{2}}\right)^{1/\alpha} \frac{1}{(n+1)^{|\mathcal{Y}|}} 2^{-nD(P_n||Q_{x_a})} \tag{166}$$

$$\leq \log|\mathcal{X}| - \frac{1}{2}\frac{\alpha}{(\alpha-1)\ln 2} \frac{1}{|\mathcal{X}|\left(\frac{1}{|\mathcal{X}|} + \sqrt{\frac{\alpha\epsilon_n}{2}}\right)^{1/\alpha}} \tag{167}$$

$$\cdot \left(\frac{1}{|\mathcal{X}|} - \sqrt{\frac{\alpha\epsilon_n}{2}}\right)^{1/\alpha} \frac{1}{(n+1)^{|\mathcal{Y}|}} 2^{-nD(P_n||Q_{x_a})}. \tag{168}$$

15

This implies that

$$\lim_{n \to \infty} -\frac{1}{n} \log \left(\log |\mathcal{X}| - \max_{\tilde{Q}(X)} I_\alpha^S(X; Y^n) \right) \leq \min_{x \neq x'} \mathscr{C}(Q_x || Q_{x'})$$ (169)

by Lemma 4 in Appendix A, which implies the result for $1 < \alpha < \infty$. The $\alpha = \infty$ case follows from (30) since $I_\infty^S(X; Y^n)$ does not depend on $Q(X)$, and $H_{1/\alpha}(X) = \log |\mathcal{X}|$ in that case. $\quad\square$

Acknowledgment

A An Ancillary Lemma

Recall that Q_x denotes the distribution of Y given x, and for any $P \in \mathcal{P}$, $x_k(P)$ denotes $x \in \mathcal{X}$ such that $D(P||Q_x)$ is the kth smallest relative entropy across all elements of $\mathcal{X}$.

Lemma 4.

$$\inf_{P \in \mathcal{P}} D(P||Q_{x_2(P)}) = \min_{x \neq x'} \mathscr{C}(Q_x || Q_{x'}),$$ (170)

where both quantities may be infinite.

Proof. We will separately prove that

$$\inf_{P \in \mathcal{P}} D(P||Q_{x_2(P)}) \leq \min_{x \neq x'} \mathscr{C}(Q_x || Q_{x'})$$ (171)

and

$$\inf_{P \in \mathcal{P}} D(P||Q_{x_2(P)}) \geq \min_{x \neq x'} \mathscr{C}(Q_x || Q_{x'}).$$ (172)

To prove the upper bound, fix $x \neq x'$ and consider $P_\lambda(y) = P_\lambda(Q_x, Q_{x'}, y)$ as defined in (16). Choose λ^* such that $D(P_{\lambda^*}||Q_x) = D(P_{\lambda^*}||Q_{x'})$. Then, certainly

$$D(P_{\lambda^*}||Q_{x_2(P_{\lambda^*})}) \leq \mathscr{C}(Q_x || Q_{x'})$$ (173)

since we know of two X-values whose corresponding $Q(Y|X)$ distributions are equidistant to P_{λ^*}, from which (171) follows.

For the lower bound, we first define subsets of $\mathcal{P}$:

$$E_x = \{P \in \mathcal{P} \mid D(P||Q_x) \leq \mathscr{C}(Q_x || Q_{x'})\}$$ (174)
$$E_{x'} = \{P \in \mathcal{P} \mid D(P||Q_{x'}) \leq \mathscr{C}(Q_x || Q_{x'})\}$$ (175)

Note that E_x and $E_{x'}$ are convex sets since $D(\cdot||\cdot)$ is convex and that P_{λ^*} achieves the minimum distance to $Q_{x'}$ in E_x and the minimum distance to Q_x in $E_{x'}$ [16, Sec. 11.9].

Choose any $P \in \mathcal{P}$. There are three cases to consider, depending on the location of P in $\mathcal{P}$-space.

Case 1: $P \notin E_x$ and $P \notin E_{x'}$. By construction, $D(P||Q_x) \geq \mathscr{C}(Q_x || Q_{x'})$ and $D(P||Q_{x'}) \geq \mathscr{C}(Q_x || Q_{x'})$.

Case 2: $P \in E_x$. Using the Pythagorean theorem for relative entropy [16, Thm. 11.6.1],

$$D(P||Q_{x'}) \geq D(P||P_{\lambda^*}) + D(P_{\lambda^*}||Q_{x'})$$ (176)

Case 3: $P \in E_{x'}$. By the same argument,

$$D(P||Q_x) \geq D(P||P_{\lambda^*}) + D(P_{\lambda^*}||Q_x)$$ (177)

Hence, for any $P \in \mathcal{P}$,

$$\max\{D(P||Q_x), D(P||Q_{x'})\} \geq \mathscr{C}(Q_x || Q_{x'})$$ (178)

Since $D(P||Q_{x_2(P)}) = \min_{x \neq x'} \max\{D(P||Q_x), D(P||Q_{x'})\}$,

$$\inf_{P \in \mathcal{P}} D(P||Q_{x_2(P)}) \geq \min_{x \neq x'} \mathscr{C}(Q_x || Q_{x'}).$$ (179)

$\square$

The following result is standard; we provide a proof for completeness.

Lemma 5. *For any discrete distributions P_1 and P_2 on a common alphabet $\mathcal{X}$,*

$$\mathscr{C}(P_1^n \| P_2^n) = n\mathscr{C}(P_1 \| P_2) \tag{180}$$

Proof. From (18),

$$\mathscr{C}(P_1 \| P_2) = -\min_{0 \le \lambda < 1} \log \left(\sum_x P_1(x)^\lambda P_2(x)^{1-\lambda} \right). \tag{181}$$

Furthermore,

$$\log \left(\sum_{x^n} P_1(x^n)^\lambda P_2(x^n)^{1-\lambda} \right) \tag{182}$$

$$= \log \left(\sum_{x_1} \sum_{x_2} \cdots \sum_{x_n} \prod_i^n P_1(x_i)^\lambda P_2(x_i)^{1-\lambda} \right) \tag{183}$$

$$= \log \left(\prod_i^n \sum_{x_i} P_1(x_i)^\lambda P_2(x_i)^{1-\lambda} \right) \tag{184}$$

$$= \log \left(\sum_{x \in \mathcal{X}} P_1(x)^\lambda P_2(x)^{1-\lambda} \right)^n. \tag{185}$$

Hence,

$$\mathscr{C}(P_1^n \| P_2^n) = -\min_{0 \le \lambda < 1} \log \left(\sum_{x^n} P_1(x^n)^\lambda P_2(x^n)^{1-\lambda} \right) \tag{186}$$

$$= -\min_{0 \le \lambda < 1} n \log \left(\sum_{x \in \mathcal{X}} P_1(x)^\lambda P_2(x)^{1-\lambda} \right) \tag{187}$$

$$= n\mathscr{C}(P_1 \| P_2). \tag{188}$$

$\square$

B   Data Processing for Arimoto Mutual Information

As a generalization of Shannon conditional entropy, Arimoto-Rényi conditional entropy satisfies a number of desirable properties. In particular, the rule that conditioning cannot increase entropy carries over to the Arimoto-Rényi version [15], [21, Thm. 2], [23, Corr. 1], [26, Prop. 2]:

$$H_\alpha(X|Y,Z) \le H_\alpha(X|Y). \tag{189}$$

It follows from the definition of Arimoto mutual information that a "right-hand" data processing inequality therefore holds: if $X \leftrightarrow Y \leftrightarrow Z$ form a Markov chain, then

$$I_\alpha^A(X;Z) \le I_\alpha^A(X;Y). \tag{190}$$

To reduce our problem to an instance satisfying the distinct row assumption using the technique in Section 3, we require a "left-hand" version of the inequality, i.e.,

$$I_\alpha^A(X;Z) \le I_\alpha^A(Y;Z)? \tag{191}$$

In fact, this inequality can fail dramatically.

Proposition 1. *For any $1 < \alpha < \infty$, there exist random variables X, Y, and Z such that $X \leftrightarrow Y \leftrightarrow Z$ and $Y \leftrightarrow X \leftrightarrow Z$ with $I_\alpha^A(X;Z)$ being arbitrarily small and $I_\alpha^A(Y;Z)$ being arbitrarily large.*

Proof. Fix positive integers K and L and $0 < \epsilon < 1/L$. Let Y and Z be jointly distributed as

$$P(Y = i) = \begin{cases} \epsilon & \text{if } i \in \{1, \dots, L\} \\ \frac{1-L\epsilon}{K} & \text{if } i \in \{L+1, \dots, L+K\} \end{cases} \tag{192}$$

$$P(Z = j | Y = i) = \begin{cases} 1 & \text{if } j = i \text{ and } i \in \{1, \dots, L\} \\ \frac{1}{L} & \text{if } i \in \{L+1, \dots, L+K\} \\ 0 & \text{otherwise.} \end{cases} \tag{193}$$

We then couple X to Y and Z via

$$X = \min(Y, L+1). \tag{194}$$

From (4), as $\epsilon \to 0$, we have that $I_\alpha^A(X;Z) \to 0$. Fix ϵ so that $I_\alpha^A(X;Z)$ is as small as desired. If we then let $K \to \infty$, we have

$$I_\alpha^A(Y;Z) \to \frac{\alpha}{\alpha-1} \log L. \tag{195}$$

But L was arbitrary. $\qquad\square$

For Sibson mutual information and α-maximal leakage, we could reduce our problem to one satisfying the unique row assumption by dividing $\mathcal{X}$ into equivalence classes based on $P_{Y|X}(\cdot|x)$ and assigning to a "leader" realization in each equivalence class the probability of all of the x realizations in that class. This approach fails for Arimoto mutual information, due to the above result, but the reduction is still possible if one accounts for the exponential tilting of $P(x)$ in (4).

Proposition 2. *Fix $\alpha > 0$. If (X, Y) does not satisfy the unique row assumption then there exists $\tilde{X}$ such that*

(i) *The support of $\tilde{X}$ is strictly contained within the support of X;*

(ii) *$P_{Y|X}(y|x) = P_{Y|\tilde{X}}(y|x)$ for all x and y;*

(iii) *$(\tilde{X}, Y)$ satisfies the unique row assumption; and*

(iv) *$I_\alpha^A(X;Y) = I_\alpha^A(\tilde{X};Y)$.*

Proof. For $\alpha = 1$, this follows directly from the chain rule for mutual information. For $\alpha \neq 1$, without loss of generality, we may assume that there exists a $k < |\mathcal{X}|$ such that

$$P_{Y|X}(\cdot|x_j) \neq P_{Y|X}(\cdot|x_i) \tag{196}$$

for all $1 \leq i < j \leq k$, and for all $k < j \leq |\mathcal{X}|$ there exists $1 \leq i \leq k$ such that

$$P_{Y|X}(y|x_j) = P_{Y|X}(y|x_i) \text{ for all } y. \tag{197}$$

That is, the first k rows of $P_{Y|X}$, viewed as a stochastic matrix, are unique, and every other row is a copy of one of those k rows. For each $1 \leq i \leq k$, define the set of X realizations

$$C_i = \left\{ x \in \mathcal{X} : P_{Y|X}(y|x) = P_{Y|X}(y|x_i) \text{ for all } y \right\}, \tag{198}$$

and note that $C_1, \ldots, C_k$ are nonempty and form a partition of $\mathcal{X}$. Define $\tilde{X}$ to have support $\{x_1, \ldots, x_k\}$ with marginal distribution

$$P(\tilde{X} = x_i) = \frac{1}{\Gamma} \left(\sum_{x \in C_i} P(X = x)^\alpha \right)^{1/\alpha}, \tag{199}$$

where

$$\Gamma = \sum_{i=1}^k \left(\sum_{x \in C_i} P(X = x)^\alpha \right)^{1/\alpha}. \tag{200}$$

Define the joint distribution between $\tilde{X}$ and Y through (*ii*). Then (*i*)-(*iii*) clearly hold and we have

$$I_\alpha^A(X;Y) \tag{201}$$

$$= \frac{\alpha}{\alpha-1} \log \sum_y \left(\frac{\sum_{i=1}^k \sum_{x \in C_i} P(x)^\alpha P(y|x)^\alpha}{\sum_{i=1}^k \sum_{x \in C_i} P(x)^\alpha} \right)^{1/\alpha} \tag{202}$$

$$= \frac{\alpha}{\alpha-1} \log \sum_y \left(\frac{\sum_{i=1}^k \sum_{x \in C_i} (P(x)^\alpha/\Gamma^\alpha) P(y|x)^\alpha}{\sum_{i=1}^k \sum_{x \in C_i} (P(x)^\alpha/\Gamma^\alpha)} \right)^{1/\alpha} \tag{203}$$

$$= \frac{\alpha}{\alpha-1} \log \sum_y \left(\frac{\sum_{i=1}^k P(\tilde{X} = x_i)^\alpha P(y|x)^\alpha}{\sum_{i=1}^k P(\tilde{X} = x_i)^\alpha} \right)^{1/\alpha} \tag{204}$$

$$= I_\alpha^A(\tilde{X};Y). \tag{205}$$

$\qquad\square$

References

[1] P. C. Kocher, "Timing attacks on implementations of Diffie-Hellman, RSA, DSS, and other systems," in *Lecture Notes in Computer Science*, vol. 1109, 1996, pp. 104–113.

[2] C. Wampler, S. Uluagac, and R. Beyah, "Information leakage in encrypted IP video traffic," in *2015 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2015, pp. 1–7.

[3] Y. Zhu, Y. Lu, and A. Vikram, "On privacy of encrypted speech communications," *IEEE Transactions on Dependable and Secure Computing*, vol. 9, no. 4, pp. 470–481, Jul. 2012.

[4] P. Kairouz, S. Oh, and P. Viswanath, "The composition theorem for differential privacy," *IEEE Trans. Inf. Theory*, vol. 63, no. 6, pp. 4037–4049, Jun. 2017.

[5] C. Dwork and G. N. Rothblum, "Concentrated differential privacy." [Online]. Available: arXiv:1603.01887

[6] I. Mironov, "Rényi differential privacy," in *Proc. IEEE Comp. Sec. Found. Symp.*, 2017, pp. 263–275.

[7] I. Issa, A. B. Wagner, and S. Kamath, "An operational measure of information leakage," *IEEE Trans. Inf. Theory*, vol. 66, no. 3, pp. 1625–1657, Mar. 2020.

[8] I. Issa, S. Kamath, and A. B. Wagner, "Maximal leakage minimization for the Shannon cipher system," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2016, pp. 520–524.

[9] I. Issa and A. B. Wagner, "Operational definitions for some common information leakage metrics," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2017, pp. 769–773.

[10] I. Issa, S. Kamath, and A. B. Wagner, "An operational measure of information leakage," in *Proc. Conf. Inf. Sci. and Sys. (CISS)*, 2016, pp. 234–239.

[11] J. Liao, O. Kosut, L. Sankar, and F. P. Calmon, "A tunable measure for information leakage," *arXiv:1806.03332 [cs, math]*, Jun. 2018, arXiv: 1806.03332. [Online]. Available: http://arxiv.org/abs/1806.03332

[12] D. M. Smith and G. Smith, "Tight bounds on information leakage from repeated independent runs," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*, Aug. 2017, pp. 318–327, iSSN: 2374-8303.

[13] R. Sibson, "Information radius," *Zeitschrift for Wahrscheinlichkeitstheorie und Verwandte Gebiete*, vol. 14, no. 2, pp. 149–160, 1969. [Online]. Available: http://link.springer.com/10.1007/BF00537520

[14] S. Verdú, "α-mutual information," in *Proc. Inf. Theory and Appl. (ITA) Workshop*, 2015.

[15] S. Arimoto, "Information measures and capacity of order α for discrete memoryless channels," *Topics in Information Theory Proc. Coll. Math Soc. Janos Bolyai*, pp. 41–52, 1975.

[16] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., 2006.

[17] I. Csiszár, "Generalized cutoff rates and Rényi's information measures," *IEEE Transactions on Information Theory*, vol. 41, no. 1, pp. 26–34, 1995.

[18] A. Lapidoth and C. Pfister, "Two measures of dependence," in *2016 IEEE International Conference on the Science of Electrical Engineering (ICSEE)*, 2016, pp. 1–5.

[19] ——, "Testing against independence and a Rényi information measure," in *2018 IEEE Information Theory Workshop (ITW)*, 2018, pp. 1–5.

[20] Y. Polyanskiy and S. Verdú, "Arimoto channel coding converse and Rényi divergence," in *Proc. Ann. Allerton Conf. on Comm., Control, and Computing*, 2010, pp. 1327–1333.

[21] S. Fehr and S. Berens, "On the conditional Rényi entropy," *IEEE Transactions on Information Theory*, vol. 60, no. 11, pp. 6801–6810, 2014.

[22] B. Wu, A. B. Wagner, and G. E. Suh, "Optimal mechanisms under maximal leakage," in *2020 IEEE Conference on Communications and Network Security (CNS)*, 2020, pp. 1–6.

[23] G. Aishwarya and M. Madiman, "Remarks on Rényi versions of conditional entropy and mutual information," in *2019 IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1117–1121.

[24] I. Sason and S. Verdú, "Arimoto–Rényi conditional entropy and Bayesian M-ary hypothesis testing," *IEEE Transactions on Information Theory*, vol. 64, no. 1, pp. 4–25, Jan. 2018.

[25] T. van Erven and P. Harremoës, "Rényi divergence and Kullback-Leibler divergence," *IEEE Transactions on Information Theory*, vol. 60, no. 7, pp. 3797–3820, Jul. 2014, arXiv: 1206.2459. [Online]. Available: http://arxiv.org/abs/1206.2459

[26] E. Arikan, "An inequality on guessing and its application to sequential decoding," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 99–105, 1996.