The Role of Mrs. Gerber's Lemma for Evaluating the Information Leakage of Secret Sharing Schemes

Olivier Rioul[0000−0002−8681−8916] and
Julien Béguinot[0009−0000−0729−6965]

Abstract In the context of secret sharing computation in some finite Abelian group, given noisy observations of each share, how can one measure the information leakage of the secret? We review various instances of this problem, where it boils down to establishing some variation of a "Mrs. Gerber's lemma" (MGL): Find a lower bound on some randomness measure of a sum of discrete random variables in the Abelian group in terms of the product of individual randomnesses of each discrete random variable. We focus on Sibson's α-information for positive and negative orders α. The MGL is obtained for all orders, except in the interval $[\frac{3}{2}, 2]$, solving a conjecture by Hirche. We also compare the resulting various security bounds from the literature.

1 Introduction

Consider a random variable K representing a secret, such as a cryptographic key or a password. In many instances of privacy and secrecy problems, an adversary inevitably observes some leakage random variable Y as the output of a side channel $X \to Y$ where X is some computed sensitive variable that depends on the secret K. This is prone to side channel attacks where the secret K may be recovered from many side channel uses (leakage measurements) [8].

During the execution of a typical cryptographic algorithm, each sensitive variable X is known to depend on the bitwise XOR (modulo 2 addition) of the secret K and a (plain or cipher) text T which is uniformly distributed and independent of K, and can be publicly known. The same secret K is combined with the every component in the text sequence $T^m = (T_1, T_2, \ldots, T_m)$, where m is the number of measurements. We let $X^m = (X_1, X_2, \ldots, X_m)$ and $Y^m = (Y_1, Y_2, \ldots, Y_m)$ be the corresponding vectors of m sensitive variables and measurements, respectively. Therefore, the adversary aims at recovering K from the knowledge of the text sequence T^m and the observed leakage sequence Y^m.

Télécom Paris, Institut Polytechnique de Paris e-mail: firstname.name@telecom-paris.fr

To mitigate such a threat, a quite common countermeasure is *secret sharing*, also known as *masking* [10]. We assume that both K and the sensitive variable X take values in an Abelian group $\mathcal{G}$ of finite order M, and that they are uniformly distributed over $\mathcal{G}$: $K, X \sim \mathcal{U}(\mathcal{G})$. The Abelian group usually depends on the cryptographic implementation. During the sensitive computations, X is split into multiple *shares* $X_0, X_1, \ldots, X_d$ such that X is independent of any subset of at most d shares, yet it can be fully recovered by combining *all* shares: $X = X_0 + X_1 + \ldots + X_d$ in additive notation in the Abelian group $\mathcal{G}$. For example, $X_1, X_2, \ldots, X_d$ are chosen i.i.d. $\sim \mathcal{U}(\mathcal{G})$ independent of X and X_0 is set to $X_0 = X - X_1 - \cdots - X_d$. Here d is the *masking order*: $d = 0$ means no protection at all and first-order masking is the usual secret sharing scheme with two shares.

In this situation, the adversary has to rely on all possible leakage observations $Y_0, Y_1, \ldots, Y_d$ and on the publicly known texts T in order to retrieve the secret. A practical attack would exploit many (say, m) measurements for each of the $d + 1$ independent side channels $X_i \to Y_i$, $i = 0, 1, \ldots, d$. A fairly common assumption is that all side channels are stationary and memoryless. Thus, the adversary performs m measurements (side channel uses) of the vector channel

$$X \to \mathbf{X} = (X_0, \ldots, X_d) \to \boxed{\prod_{i=0}^{d} P_{Y_i|X_i}} \to \mathbf{Y} = (Y_1, \ldots, Y_d). \tag{1}$$

where the corresponding m vector inputs (secret shares) and outputs (leakages) are denoted by $\mathbf{X}^m = (\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m)$ and $\mathbf{Y}^m = (\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_m)$, respectively.

Because shares are refreshed each time, the plaintext are independent and uniformly distributed, and the channel is memoryless, all m components of $(X^m, \mathbf{X}^m, \mathbf{Y}^m)$ are assumed i.i.d. Upon observation $\mathbf{Y}^m$, the adversary performs a guess $\hat{K}$ to recover the secret K. The overall attack can be seen as a Markov chain

$$K \to (X^m, T^m) \to (\mathbf{X}^m, T^m) \to (\mathbf{Y}^m, T^m) \to \hat{K} \tag{2}$$

and the attack is considered successful if $\hat{K} = K$.

The primary goal of the secret sharing scheme of order $d > 0$ is to ensure that the number of measurements that the adversary has to make to recover the secret successfully is prohibitively large, for any type of attack, given the available information leakage. In order to quantify the robustness of this countermeasure, the defender can choose any specific criterion (figure of merit) for the attack success, as well as any specific information measure to evaluate the amount of leakage.

1.1 Figure of Merit: Probability of Success

The most common criterion to quantify the successfulness of an attack is the probability of success [4]

$$\mathbb{P}_s = \mathbb{P}_s(K|T^m, \mathbf{Y}^m) \tag{3}$$

given by the *maximum a posteriori* (MAP) rule [19, 9], which maximizes $\mathbb{P}(\hat{K} = K)$. Under this criterion, the optimal attack given observation $\mathbf{y}^m$ takes the form

$$\hat{k} = \arg\max_k p_{K|T^m,\mathbf{Y}^m}(k|t^m, \mathbf{y}^m). \tag{4}$$

Since the secret is uniformly distributed, this amounts to maximizing likelihood $\hat{k} = \arg\max_k p_{T^m,\mathbf{Y}^m|K}(t^m, \mathbf{y}^m|k)$. It should be noted that the MAP rule corresponds to the most unfavorable situation for the defender, which is in line with the desire to establish the best possible level of security.

Another quite common criterion in the side channel analysis literature is the *guesswork* [32, 33] or *guessing entropy* $G(K|T^m, \mathbf{Y}^m)$ [29], which can be generalized to ρth-order guessing moments $G_\rho(K|T^m, \mathbf{Y}^m)$ [2]. To simplify the presentation we restrict ourselves to the success probability $\mathbb{P}_s$. Most of the derivations presented below, however, can be also carried out for the guessing entropies G_ρ instead.

1.2 Information Leakage Measure: α-Information

The defender may choose any measure of information leakage $\mathcal{L}(K; \mathbf{Y}^m, T^m)$ which is theoretically defined from the joint distribution of the secret and the leakages. Sometimes, it is convenient to consider an equivalent measure of the form $\mathcal{L}' = \varphi(\mathcal{L})$ instead of $\mathcal{L}$, where φ is a differentiable increasing function satisfying $\varphi(0) = 0$. This essentially does not change the methodology presented below.

In the sequel, we put emphasis of Sibson's α-information [44] $I_\alpha(K; \mathbf{Y}^m, T^m)$ (α-information in short) because it satisfies all the ideal properties needed to achieve the defender's objective, as seen below. Here α is a real-valued parameter that can be positive or negative. Typical values encountered in the side channel literature are $\alpha = 1$ (Shannon's mutual information), $\alpha = 2$ (quadratic leakage [23, 26]), $\alpha = +\infty$ (maximal leakage [22]) and $\alpha = -\infty$ (maximal cost leakage [22, 15], essentially equivalent to the complementary Doeblin coefficient [5]). Section 2 lists useful properties of α-information.

Other information measures, different from $\mathcal{L} = I_\alpha$ or $\varphi_\alpha(I_\alpha)$, were also proposed in the literature [37, 12, 36]. Since they do not have all the properties of I_α, specific inequalities (such as Pinsker and reverse Pinsker inequalities) must be used to establish a link with α-information [13, 30], so that the methodology presented below can be applied. These other information measures are reviewed in Section 5.

1.3 General Methodology

The defender's objective is to apply information theoretic tools to evaluate the *minimum number of measurements* m that can achieve a given *figure of merit* $\mathbb{P}_s$.

That evaluation depends on the choice of the information leakage measure $\mathcal{L}$. Ideally, one can relay on the following information-theoretic ingredients:

1.3.1 (Generalized) Fano Inequality

This inequality allows one to lower bound the information leakage $\mathcal{L}(K;T^m,\mathbf{Y}^m)$ by some "distance" between the probability of success $\mathbb{P}_s(K|T^m,\mathbf{Y}^m)$ given observation $T^m,\mathbf{Y}^m$, and the probability of success $\mathbb{P}_s(K)=\frac{1}{M}$ in the case of blind estimation of the secret (without any observation). Fano's inequality has the following general form:

$$d(\mathbb{P}_s,\tfrac{1}{M}) \le \mathcal{L}(K\,;\,T^m,\mathbf{Y}^m) \tag{5}$$

The classical Fano inequality [16] corresponds to the case $\mathcal{L}=I$ (mutual information) and d = Kullback-Leibler divergence. In general, Fano's inequality implies that if leakage $\mathcal{L}$ is small enough, then the probability of success $\mathbb{P}_s > \frac{1}{M}$ cannot be too large compared to its minimum value $\frac{1}{M}$. This is essentially the reason why the countermeasure is implemented. Similar "Fano inequalities" may also be derived for the guessing entropy G_ρ [7].

1.3.2 Text Switching Inequality

In order to handle the presence of the (publicly known) text sequence T^m which is independent of the secret K, the leakage measure should satisfy the following "text switching" property

$$\mathcal{L}(K\,;\,T^m,\mathbf{Y}^m) \le \mathcal{L}(K,T^m\,;\,\mathbf{Y}^m). \tag{6}$$

This ensures that a small information upon the observation of the leakage will enforce a small information on the secret. This is easily checked in the case of mutual information $\mathcal{L}=I$ (where $\alpha=1$) using the chain rule property and the fact that T^m is independent of K: $I(K;T^m,Y^m)=I(K;T^m)+I(K;Y^m|T^m)=I(K;Y^m|T^m)=I(K,T^m;Y^m)-I(T^m;Y)\le I(K,T^m;Y^m)$. We show in the sequel that the results also holds for Sibson's α-information even though it does not verify the chain rule property.

1.3.3 (Pre) Data Processing Inequality (DPI)

The pre-DPI applied to the Markov chain $(K,T^m)\to X^m\to\mathbf{Y}^m$ yields

$$\mathcal{L}(K,T^m\,;\,\mathbf{Y}^m) \le \mathcal{L}(X^m;\mathbf{Y}^m) \tag{7}$$

Thus, overall information leakage will be small if all shares do not leak too much information.

1.3.4 Tensorization Property

A nice property of the information measure is when it "tensorizes", e.g., the information measure of a product joint distribution equals the sum of individual information measures [35, Chapter 6]. Then for the i.i.d. vector $(X^m, \mathbf{Y}^m)$ we simply have

$$\mathcal{L}(X^m; \mathbf{Y}^m) = m\mathcal{L}(X; \mathbf{Y}). \tag{8}$$

This allows one to work with a single-letter expression $\mathcal{L}(X; \mathbf{Y})$. Notice that only the inequality $\mathcal{L}(X^m; \mathbf{Y}^m) \leq m\mathcal{L}(X; \mathbf{Y})$ would be sufficient to obtain a lower bound on the number of measurements.

1.3.5 Mrs. Gerber's Lemma (a.k.a. Discrete Entropy Power Inequality)

The crucial step is now to upper bound the single letter expression of the information leakage measure $\mathcal{L}(X; \mathbf{Y})$, where $X = X_0 + \ldots + X_d$ and $\mathbf{Y} = (Y_0, \ldots, Y_d)$, in terms of the information leakage measures *for each share* $\mathcal{L}(X_i; Y_i)$, $i = 0, 1, \ldots, d$. Typically—possibly with an appropriate $\varphi(\mathcal{L})$ in place of $\mathcal{L}$—the (generalized) Mrs. Gerber's lemma (MGL) writes

$$\mathcal{L}(X; \mathbf{Y}) \leq \prod_{i=0}^{d} \mathcal{L}(X_i; Y_i). \tag{9}$$

As seen below, this reduces to the classical Mrs. Gerber's Lemma [47] (a.k.a. the binary analog [43] of the entropy power inequality [28]) when $\mathcal{G} = \mathbb{Z}_2$ is the binary group.

In general, the MGL step is crucial because it allows one to efficiently evaluate $\mathcal{L}(X_i; Y_i)$ for a single side-channel (such as a AWGN channel), avoiding a cumbersome evaluation $\mathcal{L}(X; \mathbf{Y})$ due to the $(d+1)$-dimensional noise present in $\mathbf{Y}$ ("curse of dimensionality"). Thus, a security bound can be derived without having to mount the complete attack.

Furthermore, for small values of $\mathcal{L}(X_i; Y_i)$, it can be easily checked that the share by share evaluation with MGL has much greater precision than a direct evaluation. In fact, if (say) $\mathcal{L}(X_i; Y_i) = \delta$ and each term is evaluated up to an additive error term of the same order $\approx \delta$, the error on $\mathcal{L}(X; \mathbf{Y})$ is $\approx \delta$ with a direct evaluation, but only $\approx \delta^{d+1}$ using MGL.

1.3.6 Putting All Ingredients Together

Combining all previous steps:

$$\begin{aligned} d(\mathbb{P}_s, \tfrac{1}{M}) &\le \mathcal{L}(K\ ;\ T^m, \mathbf{Y}^m) && \text{(Fano)} \\ &\le \mathcal{L}(K, T^m\ ;\ \mathbf{Y}^m) && \text{(text switch)} \\ &\le \mathcal{L}(X^m; \mathbf{Y}^m) && \text{(pre-DPI)} \\ &= m\mathcal{L}(X; \mathbf{Y}) && \text{(tensorization)} \\ &\le m \textstyle\prod_{i=0}^{d} \mathcal{L}(X_i; Y_i). && \text{(Mrs. Gerber's Lemma)} \end{aligned} \tag{10}$$

we obtain an inequality of the form

$$m \ge \frac{d(\mathbb{P}_s, \frac{1}{M})}{\prod_{i=0}^{d} \mathcal{L}(X_i; Y_i)}. \tag{11}$$

This gives a lower bound on the number of side channel uses (measurements) required to achieve a given success probability $\mathbb{P}_s$ (in the numerator). The secret sharing scheme will be all the more efficient as a countermeasure as the denominator is small, i.e., the information leakage per share is sufficiently small.

Remark 1 It is also possible to proceed in an *hybrid* way. If $\mathcal{L}_1$ and $\mathcal{L}_2$ are two leakage measures such that $\mathcal{L}_1 \le \mathcal{L}_2$, $\mathcal{L}_1$ verifies the first properties and $\mathcal{L}_2$ the remaining ones then a similar result holds, where the divergence on the left hand-side is associated to $\mathcal{L}_1$ and the $\mathcal{L}$ measure in the right-hand side is $\mathcal{L}_2$. This is simply done by inserting the extra step $\mathcal{L}_1 \le \mathcal{L}_2$ in the derivation above. Section 5 contains multiple examples.

2 Preliminaries on α-Information

Throughout this chapter, we assume that all considered probability distributions P are dominated by a common dominating measure μ and write the corresponding densities as the Radon-Nykodym derivatives $p = \frac{\mathrm{d}P}{\mathrm{d}\mu}$. All considered integrals with respect to μ are integrals of positive measurable functions and as such are well defined (possibly infinite). We also use natural logarithms $\log = \log_e$ (expressed in *nats*).

2.1 α-Divergence

For any $\alpha \in \mathbb{R}$ we use the special notation [27]

$$\langle p \| q \rangle_\alpha \triangleq \left(\int p^\alpha q^{1-\alpha} d\mu \right)^{\frac{1}{\alpha}}. \tag{12}$$

which is easily seen to be independent of the choice of the dominating measure μ. Since $\int p^\alpha q^{1-\alpha} d\mu = \int q^{1-\alpha} p^{1-(1-\alpha)} d\mu$, one has the following *reflection formula*:

$$\langle p\|q\rangle_\alpha^\alpha = \langle q\|p\rangle_{1-\alpha}^{1-\alpha}. \tag{13}$$

Because the usual Rényi α-divergence $D_\alpha(p\|q) = \frac{\alpha}{\alpha-1}\log\langle p\|q\rangle_\alpha$ is typically negative for negative α [45], we use the following definition.

Definition 1 (Signed Rényi Divergence)

$$\widetilde{D}_\alpha(p\|q) \triangleq \frac{|\alpha|}{\alpha-1}\log\langle p\|q\rangle_\alpha = \frac{\operatorname{sgn}(\alpha)}{\alpha-1}\log\langle p\|q\rangle_\alpha^\alpha \tag{14}$$

where sgn is the sign function. For binary (Bernoulli) distributions,

$$\tilde{d}_\alpha(p\|q) \triangleq \widetilde{D}_\alpha(\mathcal{B}(p)\|\mathcal{B}(q)) = \frac{\operatorname{sgn}(\alpha)}{\alpha-1}\log(p^\alpha q^{1-\alpha} + (1-p)^\alpha(1-q)^{1-\alpha}) \tag{15}$$

In particular $\tilde{d}_\infty(p\|q) = \log\max\{\frac{p}{q}, \frac{1-p}{1-q}\}$ and $\tilde{d}_{-\infty}(p\|q) = -\log\min\{\frac{p}{q}, \frac{1-p}{1-q}\}$.

The limiting case $\alpha = 1$ gives the usual Kullback-Leibler divergence $D(p\|q)$. In particular $d(p\|q) \triangleq D_{\mathrm{KL}}(\mathcal{B}(p)\|\mathcal{B}(q)) = p\log\frac{p}{q} + (1-p)\log\frac{1-p}{1-q}$.

In terms for the signed Rényi divergence, the reflection formula (13) reads

$$\frac{\widetilde{D}_\alpha(p\|q)}{|\alpha|} = \frac{\widetilde{D}_{1-\alpha}(q\|p)}{|1-\alpha|}. \tag{16}$$

This is equivalent to the skew symmetry formula of [45, Lemma 10].

It is well known [45, Theorem 2] by Jensen's inequality that the usual Rényi α-divergence is nonnegative for $\alpha > 0$. As a result of the reflection formula (16), since either α or $1-\alpha$ is positive, the signed Rényi divergence is always nonnegative for any $\alpha \in \mathbb{R}$:

$$\widetilde{D}_\alpha(p\|q) \geq 0 \tag{17}$$

with equality if and only if $p = q$ μ-a.e. (see Proposition 1 of [15] for negative α).

2.2 α-Information

Let p_X denote the probability density (w.r.t. μ) of a random variable X. For two random variables X, Y one has:

$$\widetilde{D}_\alpha(p_{X|Y=y}\|p_X) = \frac{|\alpha|}{\alpha-1}\log\langle p_{X|Y=y}\|p_X\rangle_\alpha$$

for fixed $Y = y$. Taking the expectation over Y *inside* the logarithm we obtain Sibson's α-mutual information [44, 46]:

Definition 2 (Sibson's α-Information)

$$I_\alpha(X;Y) \triangleq \frac{|\alpha|}{\alpha-1}\log\mathbb{E}_y\langle p_{X|Y=y}\|p_X\rangle_\alpha. \tag{18}$$

In particular, for negative α this definition is equivalent to the one of [15]. The limiting case $\alpha = 1$ gives the usual mutual information $I(X;Y)$. Considering a discrete random variable X for simplicity, the limiting case $\alpha = +\infty$ is known as the *maximal leakage* [22, Def. 1]:

$$I_\infty(X;Y) = \log \int_y \max_{x:p_X(x)>0} p_{Y|X} d\mu_Y \tag{19}$$

while the limiting case $\alpha = -\infty$ is the *maximal cost leakage* [22, Def. 11]:

$$I_{-\infty}(X;Y) = -\log \int_y \min_{x:p_X(x)>0} p_{Y|X} d\mu_Y. \tag{20}$$

The latter quantity turns out to be equivalent to the *complementary Doeblin coefficient* [5, Definition 9]

$$\overline{\mathcal{E}}(X \to Y) = 1 - \exp(-I_{-\infty}(X;Y)) \tag{21}$$

that was used in [5] to derive side channel security proofs.

2.3 α-Entropy and Conditional α-Entropy

We use the "α-norm" notation [27]

$$\|p\|_\alpha = \left(\int p^\alpha d\mu \right)^{1/\alpha} \tag{22}$$

to define the (conditional) Rényi α-entropy [38, 1] as follows.

Definition 3 (Signed Arimoto-Rényi α-Entropy) Similarly as for α-divergence, we use a slightly modified definition:

$$\widetilde{H}_\alpha(X) = \widetilde{H}_\alpha(p_X) \triangleq \frac{|\alpha|}{1-\alpha} \log \|p_X\|_\alpha = \frac{\operatorname{sgn}(\alpha)}{1-\alpha} \log \|p_X\|_\alpha^\alpha. \tag{23}$$

$$\widetilde{H}_\alpha(X|Y) \triangleq \frac{|\alpha|}{1-\alpha} \log \mathbb{E}_Y \|p_{X|Y}\|_\alpha. \tag{24}$$

The limiting case $\alpha = 1$ gives the usual entropy $H(X)$ and equivocation $H(X|Y)$. Note that for a binary (Bernoulli) random variable,

$$\tilde{h}_\alpha(p) \triangleq \widetilde{H}_\alpha(\mathcal{B}(p)) = \frac{|\alpha|}{\alpha-1} \log((p^\alpha + (1-p)^\alpha)^{\frac{1}{\alpha}}). \tag{25}$$

In the limiting case $\alpha = 1$, $h(p) \triangleq H(\mathcal{B}(p)) = -p \log p - (1-p)\log(1-p)$.

We may also consider Hayashi's conditional α-entropy [18], generalized to include negative values of α:

Definition 4 (Signed Hayashi's α-Entropy)

$$\widetilde{H}^H_\alpha(X|Y) \triangleq \frac{\operatorname{sgn}(\alpha)}{1-\alpha} \log \mathbb{E}_Y \|p_{X|Y}\|_\alpha^\alpha. \tag{26}$$

2.4 Properties of α-Information

Sibson's α-information may also be defined via the so called Sibson's identity:

Lemma 1 (Sibson's Identity) *Let $q^*_Y(y) = p_Y(y)\frac{\langle p_{X|Y=y}\|p_X\rangle_\alpha}{\mathbb{E}_y\langle p_{X|Y=y}\|p_X\rangle_\alpha}$. Then for any q_Y we have*

$$\widetilde{D}_\alpha(p_{XY}\|p_Xq_Y) = \widetilde{D}_\alpha(q^*_Y\|q_Y) + I_\alpha(X;Y) \tag{27}$$

As a consequence, since $\widetilde{D}_\alpha(q^*_Y\|q_Y) \geq 0$ with equality if and only if $q^*_Y = q_Y$ a.e.,

$$I_\alpha(X;Y) = \min_{q_Y} \widetilde{D}_\alpha(p_{XY}\|p_Xq_Y) \tag{28}$$

where the minimum is achieved when $q_Y = q^*_Y$.

Proof. Easy calculation. See [46, Def. 4] for positive α and [15, Def. 3] for negative α. □

Lemma 2 (Uniform Expansion Property [27]) *If X is uniformly distributed then*

$$I_\alpha(X;Y) = \widetilde{H}_\alpha(X) - \widetilde{H}_\alpha(X|Y). \tag{29}$$

Proof. Easy calculation. □

Lemma 3 (Data Processing Inequality) *For any Markov chain $X \to Y \to Z \to T$,*

$$I_\alpha(X;T) \leq I_\alpha(Y;Z). \tag{30}$$

Let P, Q be two distributions and W a channel. Let P^W and Q^W be the respective output distributions of random variable with distribution P (resp. Q) passed through the channel W. Then

$$\widetilde{D}_\alpha(P^W\|Q^W) \leq \widetilde{D}_\alpha(P\|Q). \tag{31}$$

Proof. Equation (30) is proved in [34] for $\alpha > 0$, which applies verbatim to $\alpha < 0$. Equation (31) for positive α is well known and proved for instance in the reference article [45, Theorem 1]. Equation (31) is proved for negative α in [15, Proposition 1.4] using the results for positive α together with the the reflection formula (13). □

Lemma 4 (Text Switching Inequality) *Let T and K be two independent random variables and Y be some side information about the pair (K, T). For any $\alpha \neq 0$,*

$$I_\alpha(K\,;\,T,Y) \leq I_\alpha(K,T\,;\,Y). \tag{32}$$

This was proved directly with involved calculations for $\alpha > 0$ in [26, Lemma 2]. The proof carries over verbatim for $\alpha < 0$. We provide a simpler proof:

Proof. Since K and T are independent, using the representation (28),

$$\begin{aligned} I_\alpha(K\,;\,T,Y) &= \min_{Q_{TY}} \widetilde{D}_\alpha(P_{KTY}\|P_K Q_{TY}) && (33)\\ &= \min_{Q_{TY}} \widetilde{D}_\alpha(P_{KTY}\|P_{K|T} Q_{TY}) && (34)\\ &\le \min_{Q_Y} \widetilde{D}_\alpha(P_{KTY}\|P_{K|T} P_T Q_Y) && (35)\\ &= \min_{Q_Y} \widetilde{D}_\alpha(P_{KTY}\|P_{KT} Q_Y) && (36)\\ &= I_\alpha(K,T\,;\,Y). && (37) \end{aligned}$$

□

Remark 2 Notice that the quantity in equation (34) corresponds to a definition of the *conditional Sibson's α-information* which is the suitable definition for leakage evaluation as explained in [27].

Lemma 5 (Tensorization) *If $X_1, \dots, X_m$ is and i.i.d sequence and the channel $X_i \to Y_i$ is stationary and memoryless then*

$$I_\alpha(X_1, \dots, X_m; Y_1, \dots, Y_m) = m I_\alpha(X_1; Y_1). \tag{38}$$

Proof. Easy calculation. □

3 Generalized Fano Inequalities for α-Information

One important ingredient in our framework is Fano's inequality. The classical Fano inequality can be written as a lower bound on mutual information, and can be generalized to α-information for positive or negative α.

3.1 The Classical Fano Inequality

Let X be a M-ary random variable and $X \to Y \to \hat{X}$ be a Markov chain. Fano's inequality [16] is classically seen as upper bound on the equivocation $H(X|Y)$ in terms of the probability of error $\mathbb{P}_e(X|Y) = \mathbb{P}(X \neq \hat{X}) = 1 - \mathbb{P}_s(X|Y)$:

$$H(X|Y) \le h(\mathbb{P}_e(X|Y)) + \mathbb{P}_e(X|Y) \log(M-1). \tag{39}$$

For uniformly distributed X, it can be rewritten as a lower bound on mutual information [17]:

$$I(X;Y) \ge d(\mathbb{P}_s(X|Y)\|\mathbb{P}_s(X)) = d(\mathbb{P}_s(X|Y)\|\tfrac{1}{M}). \tag{40}$$

This second reformulation in terms of KL divergence is perhaps more intuitive. The probability of success $\mathbb{P}_s(X|Y)$ cannot be too different from a blind guess $\mathbb{P}_s(X)$ if mutual information $I(X;Y)$ is small. This is illustrated in Figure 1.

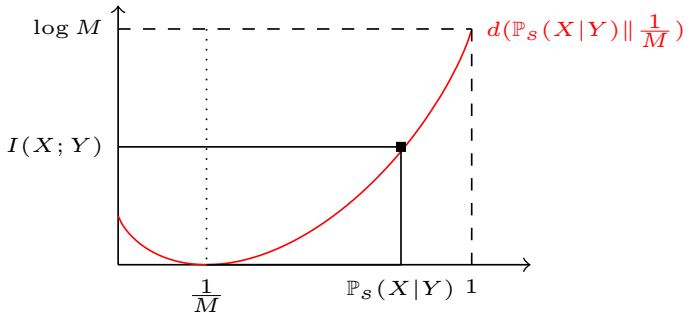

Fig. 1: Illustration of Fano's inequality seen as a lower bound on mutual information

3.2 Extension to α-Information

Fano's inequality can be generalized to Sibson's α-Information as shown for $\alpha > 0$ in [42, Theorem 3] for uniform X and in [39, Theorem 1] for arbitrary X. The proof easily carries over to negative values of α as well:

Lemma 6 (α-Fano Inequality) *For any $\alpha \in \mathbb{R}$,*

$$I_\alpha(X;Y) \geq \tilde{d}_\alpha(\mathbb{P}_s(X|Y)\|\mathbb{P}_s(X)) \tag{41}$$

Proof. The proof from [39] derived for $\alpha > 0$ using the data processing inequality for α-information and for α-divergence, can be applied verbatim to negative α. □

Lemma 6 is illustrated in Figure 2 for $\alpha = 2$ and $\alpha = -1$, respectively. While the binary divergence is always bounded by $\log M$ for $\alpha > 0$, it tends to $+\infty$ for $\alpha < 0$ as success increases. As a result, when $\alpha < 0$, the Fano bound on the probability of success will never be vacuous.

In particular, for a uniformly distributed M-ary random variable X,

$$I_\infty(X;Y) \geq \tilde{d}_\infty(\mathbb{P}_s(X|Y)\|\mathbb{P}_s(X)) = \log(M\mathbb{P}_s(X|Y)) \tag{42}$$

and

$$I_{-\infty}(X;Y) \geq \tilde{d}_{-\infty}(\mathbb{P}_s(X|Y)\|\mathbb{P}_s(X)) = -\log\left(\frac{1-\frac{1}{M}}{1-\mathbb{P}_s(K|Y)}\right) \tag{43}$$

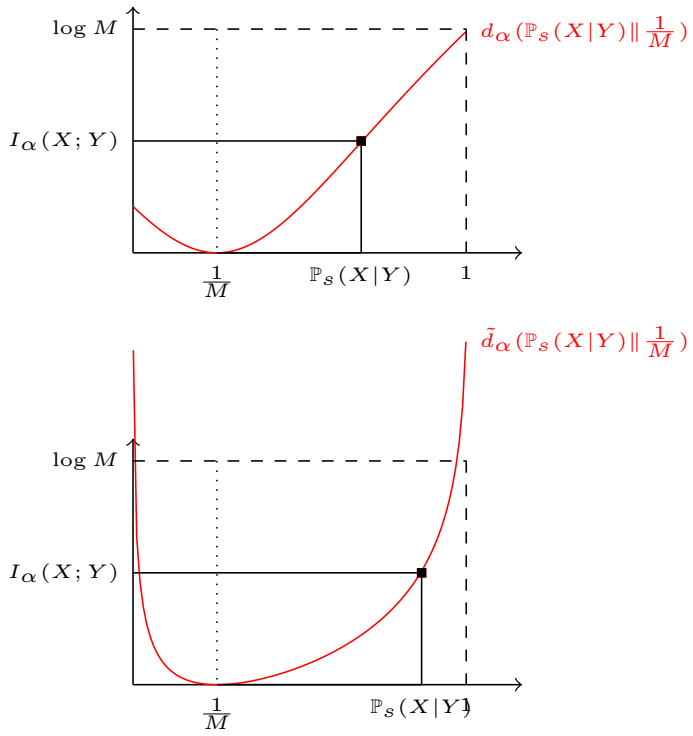

Fig. 2: Illustration of Fano's inequality seen as a lower bound on α-information. Top: $\alpha = 2$. Bottom: $\alpha = -1$.

One recovers the bound on the success rate derived in [5, Proposition 1] in terms of complementary Doeblin coefficient.

4 Mrs. Gerber's Lemma or Discrete Entropy Power Inequalities

In the context of secret sharing, the crucial ingredient in our framework is Mrs. Gerber's Lemma. The classical Mrs. Gerber's Lemma in the binary group can be written as a lower bound on equivocation (or an upper bound of mutual information) in term of the individual equivocations for each share. Thanks to a pivotal lemma derived in this section, MGL can be generalized to conditional α-entropies for positive or negative α.

4.1 Original Mrs. Gerber's Lemma

First consider the case of two binary shares. Let X_0, X_1 be two independent Bernoulli random variables (taking values in $\mathbb{Z}_2 = \{0, 1\}$) with respective parameters p_0, p_1. Then $X = X_0 + X_1$ (addition modulo 2) is a Bernoulli random variable with parameter $p = p_0(1 - p_1) + (1 - p_0)p_1 = p_0 * p_1$ where $*$ denotes *binary convolution* : $x * y = x(1 - y) + (1 - x)y$. In particular,

$$H(X) = h(p) = h(h^{-1}(H(X_0) * h^{-1}(H(X_1))). \tag{44}$$

If each share leaks through a side-channel $X_i \to Y_i$ $(i = 0, 1)$, letting $Y = (Y_0, Y_1)$, we obtain $H(X|Y = y) = h(h^{-1}(H(X_0|Y_0 = y_0) * h^{-1}(H(X_1|Y_1 = y_1)))$ for every $y = (y_0, y_1)$, hence

$$H(X|Y) = \mathbb{E}_{y_0,y_1}[h(h^{-1}(H(X_0|Y_0 = y_0)) * h^{-1}(H(X_1|Y_1 = y_1)))]. \tag{45}$$

The expectation can be moved inside this expression to obtain an expression in terms of the equivocations $H(X_i|Y_i)$, thanks to the original Mrs. Gerber's Lemma of Wyner and Ziv [47]:

Lemma 7 (Original Mrs. Gerber's Lemma [47]) *The function $h(h^{-1}(x) * h^{-1}(y))$ is convex in one variable when the other is fixed (convex in x for fixed y and vice versa).*

A simple proof will be given below as a particular case.

As an immediate consequence one obtains the following lower bound:

$$H(X|Y) \geq h(h^{-1}(H(X_0|Y_0)) * h^{-1}(H(X_1|Y_1))). \tag{46}$$

Shamai and Ziv [43] reformulated this as a binary analog of the *entropy-power inequality*:

$$\sigma(X|Y) \geq \sigma(X_0|Y_0) * \sigma(X_1|Y_1) \tag{47}$$

where X_0, X_1 are binary random variable and $\sigma(X|Y) \triangleq h^{-1}(H(X|Y))$. By induction, we obtain that for a dth-order secret sharing scheme,

$$\sigma(X|Y) \geq \sigma(X_0|Y_0) * \ldots * \sigma(X_d|Y_d). \tag{48}$$

We give yet another attractive reformulation in terms of mutual information:

Theorem 1 (Reformulated Mrs. Gerber's Lemma)

$$H(X|Y) \geq \phi\left(\prod_{i=0}^{d} \phi^{-1}(H(X_i|Y_i))\right) \tag{49}$$

where $\phi(x) = h(\frac{1-x}{2}) \in [0, \log 2]$. Equivalently, for any uniformly distributed X,

$$I(X;Y) \leq \varphi\left(\prod_{i=0}^{d} \varphi^{-1}(I(X_i;Y_i))\right) \tag{50}$$

where $\varphi(x) = \log 2 - \phi(x)$.

Proof. The binary convolution $*$ is transformed into a multiplicative product under the discrete Fourier transform (DFT) over $\mathbb{Z}_2$. Since $\mathrm{DFT}(1-p, p) = (1, 1-2p)$, it follows that $\mathrm{DFT}(1-p*q, p*q) = (1, (1-2p)(1-2q))$ and similarly for several factors. Taking the inverse DFT yields

$$\begin{pmatrix} 1-\sigma(X_0|Y_0) * \ldots * \sigma(X_d|Y_d) \\ \sigma(X_0|Y_0) * \ldots * \sigma(X_d|Y_d) \end{pmatrix} = \frac{1}{2} \begin{pmatrix} 1+\prod_{i=0}^{d}(1-2\sigma(X_i|Y_i)) \\ 1-\prod_{i=0}^{d}(1-2\sigma(X_i|Y_i)) \end{pmatrix}. \tag{51}$$

Therefore,

$$H(X|Y) \geq h\left(\frac{1-\prod_{i=0}^{d}(1-2h^{-1}(H(X_i|Y_i)))}{2}\right) = \phi\left(\prod_{i=0}^{d}\phi^{-1}(H(X_i|Y_i))\right) \tag{52}$$

Finally $I(X;Y) = H(X) - H(X|Y) = \log 2 - H(X|Y)$ for uniformly distributed X. □

Remark 3 Equation (49) may also be reformulated as

$$\sum_{i=0}^{d} N(H(X_i|Y_i)) \leq N(H(X|Y)) \tag{53}$$

where $N(x) = -\log \phi^{-1}(x) = -\log(1-2h^{-1}(x))$. In a sense, N corresponds better than σ to the "*entropy power*" for binary random variables.

Remark 4 Hirche proved a "Mr. Gerber's Lemma" in the opposite direction [20, Eqn. 4] which reformulates for uniform X as

$$I(X;Y) \geq (\log 2) \prod_{i=0}^{d} \tfrac{I(X_i;Y_i)}{\log 2} = \psi\left(\prod_{i=0}^{d}\psi^{-1}(I(X_i;Y_i))\right) \tag{54}$$

where $\psi(x) = x \cdot \log 2$. A simple proof will also be given below as a particular case.

4.2 A Pivotal Lemma

The original MGL was noted to hinge on the convexity of the function $h(p * h^{-1}(x))$ for any fixed $p \in [0, \frac{1}{2}]$ by Cheng [11]. It is easily seen to be equivalent to the convexity of the mapping $w_a(x) = \phi(a\phi^{-1}(x))$ for any fixed $a \in [0,1]$.

More generally, for any twice differentiable function ϕ defined over $[0,1]$, consider the two auxiliary functions

$$w_a(x) = \phi(a\phi^{-1}(x)) \tag{55}$$

for any fixed $a \in [0,1]$, and

$$u(x) = \frac{\phi'(x)}{x\phi''(x)}. \tag{56}$$

In order to prove Mrs. Gerber's lemma (or Mr. Gerber's lemma [21] in the opposite direction), the following technical result is pivotal since it characterizes the convexity of an infinite family of functions by the behavior of a single function:

Lemma 8 *Let $\phi(x) \geq 0$ be twice differentiable.*

Assume that $\phi(x)$ is strictly increasing and convex. Then $w_a(x)$ is convex for all $a \in [0,1]$ if and only if $u(x)$ is increasing; and $w_a(x)$ is concave for all $a \in [0,1]$ if and only if $u(x)$ is decreasing.

If instead $\phi(x)$ is strictly decreasing and concave, then $w_a(x)$ is convex for all $a \in [0,1]$ if and only if $u(x)$ is decreasing; and $w_a(x)$ is concave for all $a \in [0,1]$ if and only if $u(x)$ is increasing.

Proof. Assume that $\phi(x) \geq 0$ is strictly increasing and convex. Then $w_a(x)$ is convex for all $a \in [0,1]$ if and only if its derivative

$$\frac{\mathrm{d}w_a}{\mathrm{d}x} = a\frac{\phi'(a\phi^{-1}(x))}{\phi'(\phi^{-1}(x))} \tag{57}$$

is increasing, i.e., since ϕ^{-1} is increasing, $(\log)\frac{\phi'(ax)}{\phi(x)}$ is increasing, that is

$$\frac{\mathrm{d}}{\mathrm{d}x}\log\frac{\phi'(ax)}{\phi(x)} \geq 0, \tag{58}$$

which computing the derivative is equivalent to

$$\frac{a\phi''(ax)}{\phi'(ax)} \geq \frac{\phi''(x)}{\phi'(x)} (\geq 0) \tag{59}$$

or taking the reciprocals and dividing by x,

$$\frac{\phi'(x)}{\phi''(x)x} \geq \frac{\phi'(ax)}{\phi''(ax)ax} \tag{60}$$

for any $a \in [0,1]$, hence for any $x' = ax$ not greater than x. In other words, $u(x)$ is decreasing. The proof for concave $w_a(x)$ is similar with all inequalities reversed.

The case where $\phi(x)$ is strictly decreasing and concave follows similarly with again all inequalities reversed because ϕ^{-1} is decreasing. □

As an illustration, we obtain a simple proof of the original MGL:

Proof of Lemma 7. With the reformulation of Theorem 1, Lemma 7 is equivalent to the convexity of $w_a(x) = \phi(a\phi^{-1}(x))$ for all $a \in [0,1]$, where $\phi(x) = h(\frac{1-x}{2})$.

Now from Lemma 8, one checks that $\phi(x)$ is strictly decreasing and concave, with $\phi'(x) = -\frac{1}{2}\log\frac{1+x}{1-x} < 0$ and $\phi''(x) = \frac{-1}{1-x^2} \leq 0$. Thus $u(x) = \frac{1-x^2}{2x}\log\frac{1+x}{1-x}$, which is easily checked to be decreasing for $x \in [0,1]$. □

4.3 Extension to Sibson's α-Information

As shown by Hirche [20], Mrs. Gerber's lemma can be extended to Rényi entropies of various orders. We present the corresponding derivations based on Lemma 8 and extend them to negative orders $\alpha < 0$. This will allow one to refine and generalize Hirche's results.

If each share leaks through a side-channel $X_i \to Y_i$, $i = 0, 1, \ldots, d$, we obtain, with the same notations as above,

$$\widetilde{H}_\alpha(X|Y=y) = \tilde{h}_\alpha(\tilde{h}_\alpha^{-1}(\widetilde{H}_\alpha(X_0|Y_0=y_0)) * \ldots * \tilde{h}_\alpha^{-1}(\widetilde{H}_\alpha(X_d|Y_d=y_d))). \quad (61)$$

This is equivalent to

$$K_\alpha(X|Y=y) = k_\alpha(k_\alpha^{-1}(K_\alpha(X_0|Y_0=y_0)) * \ldots * k_\alpha^{-1}(K_\alpha(X_d|Y_d=y_d))) \quad (62)$$

where

$$K_\alpha(X|Y) = \exp(\frac{1-\alpha}{|\alpha|}\widetilde{H}_\alpha(X|Y)) \quad (63)$$

and

$$k_\alpha(p) \triangleq K_\alpha(\mathcal{B}(p)) = \|(p, 1-p)\|_\alpha \in [\delta_\alpha \triangleq 2^{\frac{1-\alpha}{\alpha}}, \frac{1+\operatorname{sgn}(\alpha)}{2}]. \quad (64)$$

As a consequence,

$$K_\alpha(X|Y) = \mathbb{E}_Y[k_\alpha(k_\alpha^{-1}(K_\alpha(X_0|Y_0=y_0)) * \ldots * k_\alpha^{-1}(K_\alpha(X_d|Y_d=y_d)))]. \quad (65)$$

Again using the discrete Fourier transform over $\mathbb{Z}_2$ we can diagonalize the convolution to obtain the product

$$K_\alpha(X|Y) = \mathbb{E}_Y[\phi_\alpha(\textstyle\prod_{i=0}^d \phi_\alpha^{-1}(K_\alpha(X_i|Y_i=y_i)))] \quad (66)$$

where

$$\phi_\alpha(x) = k_\alpha(\tfrac{1-x}{2}) \in [\delta_\alpha, \frac{1+\operatorname{sgn}(\alpha)}{2}]. \quad (67)$$

We obtain again an inequality depending on the concavity or convexity of the mapping $w_a(x) = \phi_\alpha(a\phi_\alpha^{-1}(x))$ for any $a \in [0,1]$.

Note that $\phi_\alpha(x) \geq 0$ is defined over $[0,1]$, with

$$\phi'_\alpha(x) = \frac{1}{2}\big((1+x)^{\alpha-1} - (1-x)^{\alpha-1}\big)\big((1+x)^\alpha + (1-x)^\alpha\big)^{\frac{1}{\alpha}-1}. \quad (68)$$

In particular, $\phi'_\alpha(0) = 0$ and ϕ'_α has the same sign as $\alpha - 1$ so that $\phi_\alpha(x)$ is increasing for $\alpha > 1$ and decreasing for $\alpha < 1$. Next we have

$$\phi''_\alpha(x) = 2(\alpha-1)(1-x^2)^{\alpha-2}((1-x)^\alpha + (1+x)^\alpha)^{\frac{1}{\alpha}-2} \quad (69)$$

In particular, $\phi''_\alpha(0) = (\alpha-1)\delta_\alpha$ and ϕ''_α has the same sign than $\alpha - 1$ so that ϕ_α is convex for $\alpha > 1$ and concave for $\alpha < 1$.

Furthermore, since ϕ_α and ϕ_α^{-1} have the same monotonicity, it is obvious that for any $\alpha \neq 1$ and $a \in [0,1]$, $w_{a,\alpha}(x) = \phi_\alpha(a\phi_\alpha^{-1}(x))$ is increasing.

There are some interesting special cases $\alpha = -1$ and infinite α.

4.3.1 $\alpha = -1$

$$\phi_{-1}(x) = \frac{1-x}{2}\frac{1+x}{2} = \frac{1-x^2}{4} \tag{70}$$

so that for $x \in [\delta_{-1} = \frac{1}{4}, 0]$,

$$\phi_{-1}^{-1}(x) = \sqrt{1-4x}. \tag{71}$$

In particular, for any $a \in [0,1]$ we obtain

$$\phi_{-1}(a\phi_{-1}^{-1}(x)) = \frac{1-(a\sqrt{1-4x})^2}{4} = \frac{1-a^2(1-4x)}{4}. \tag{72}$$

Since this is linear in x we obtain a MGL which holds with equality:

$$K_{-1}(X|Y) = \phi_{-1}(\prod_{i=0}^{d}\phi_{-1}^{-1}(K_{-1}(X_i|Y_i))). \tag{73}$$

4.3.2 Limiting Cases $\alpha = \pm\infty$

Christoph Hirche showed a MGL for binary random variables with $\alpha = +\infty$ and noticed it holds with equality [20, Theorem IV] while Béguinot et al. [5] proved it for M-ary random variables with $\alpha = -\infty$. When $\alpha = \pm\infty$ we obtain two MGLs in the limiting case that holds with equality:

Lemma 9 ((Reformulation) MGL for Maximal Leakage Cost Leakage)

$$I_\infty(X;Y) = \log\Big(1 + \textstyle\prod_{i=0}^{d}(e^{I_\infty(X_i;Y_i)} - 1)\Big). \tag{74}$$

$$I_{-\infty}(X;Y) = -\log\Big(1 - \textstyle\prod_{i=0}^{d}(1 - e^{-I_{-\infty}(X_i;Y_i)})\Big). \tag{75}$$

In fact,

$$e^{-I_{-\infty}(X;Y)} + e^{I_\infty(X;Y)} = \int_y \left(\max_{x:p_X(x)>0} p_{Y|X} + \min_{x:p_X(x)>0} p_{Y|X}\right) d\mu_Y \tag{76}$$

If X is a non constant binary random variable then for all y, $\max_{x:p_X(x)>0} p_{Y|X}(y|x) + \min_{x:p_X(x)>0} p_{Y|X}(y|x) = p_{Y|X}(y|1) + p_{Y|X}(y|0)$ hence

$$\int_y \left(\max_{x:p_X(x)>0} p_{Y|X} + \min_{x:p_X(x)>0} p_{Y|X} \right) d\mu_Y = 2 \tag{77}$$

and we obtain the identity (see [42, Eqn. 24])

$$e^{-I_{-\infty}(X;Y)} + e^{I_\infty(X;Y)} = 2. \tag{78}$$

Or equivalently

$$1 - e^{-I_{-\infty}(X;Y)} = e^{I_\infty(X;Y)} - 1. \tag{79}$$

This shows that both MGLs for maximal leakage and maximal cost leakage are equivalent.

4.3.3 Mrs. Gerber's Lemma in the General Case

Christoph Hirche showed for $\alpha > 0$ a MGL for binary random variables [20, Theorem IV]. We extend his results to all real values of α. Our derivation hinges on the convexity or concavity of

$$w_{a,\alpha}(x) = \phi_\alpha(a\phi_\alpha^{-1}(x)) \tag{80}$$

for any $a \in [0,1]$.

Theorem 2 *If $w_{a,\alpha}$ is convex and $\alpha > 1$ or $w_{a,\alpha}$ is concave and $\alpha < 1$ then*

$$\frac{|\alpha|}{1-\alpha} \log\Big(\phi_\alpha(0) + (\phi_\alpha(1) - \phi_\alpha(0)) \prod_{i=0}^d \frac{\exp(\frac{1-\alpha}{|\alpha|} \widetilde{H}_\alpha(X_i|Y_i)) - \phi_\alpha(0)}{\phi_\alpha(1) - \phi_\alpha(0)} \Big) \tag{81}$$

$$\le \widetilde{H}_\alpha(X|Y) \le \theta_\alpha\big(\textstyle\prod_{i=0}^d \theta_\alpha^{-1}(\widetilde{H}_\alpha(X_i|Y_i)) \big) \tag{82}$$

where $\theta_\alpha(x) = \tilde{h}_\alpha(\frac{1-x}{2})$. If $w_{a,\alpha}$ is convex for all $a \in [0,1]$ and $\alpha < 1$ or $w_{a,\alpha}$ is concave for all $a \in [0,1]$ and $\alpha > 1$, the inequalities are reversed.

Proof. Recall that

$$K_\alpha(X|Y) = \mathbb{E}_{y_0,\dots,y_d} \phi_\alpha\big(\textstyle\prod_{i=0}^d \phi_\alpha^{-1}(K_\alpha(X_i|Y_i = y_i)) \big). \tag{83}$$

Assume that $w_{a,\alpha}$ is convex then by Jensen's inequality

$$K_\alpha(X|Y) \ge \phi_\alpha\big(\textstyle\prod_{i=0}^d \phi_\alpha^{-1}(K_\alpha(X_i|Y_i)) \big).$$

Also since $w_{a,\alpha}$ is convex

$$w_{a,\alpha}(y) \le w_{a,\alpha}(\phi_\alpha(0)) + \frac{w_{a,\alpha}(\phi_\alpha(1)) - w_{a,\alpha}(\phi_\alpha(0))}{\phi_\alpha(1) - \phi_\alpha(0)} (y - w_{a,\alpha}(\phi_\alpha(0))) \tag{84}$$

$$= \phi_\alpha(0) + \frac{\phi_\alpha(a) - \phi_\alpha(0)}{\phi_\alpha(1) - \phi_\alpha(0)} (y - \phi_\alpha(0)). \tag{85}$$

As a consequence

$$K_\alpha(X|Y) \leq \phi_\alpha(0) + (\phi_\alpha(1) - \phi_\alpha(0)) \prod_{i=0}^{d} \frac{K_\alpha(X_i|Y_i) - \phi_\alpha(0)}{\phi_\alpha(1) - \phi_\alpha(0)}. \tag{86}$$

Now if $\alpha > 1$ then

$$x \mapsto \frac{|\alpha|}{1-\alpha} \log(x)$$

is decreasing and we obtain

$$\widetilde{H}_\alpha(X|Y) \leq \tfrac{|\alpha|}{1-\alpha} \log(\phi_\alpha(\textstyle\prod_{i=0}^{d} \phi_\alpha^{-1}(K_\alpha(X_i|Y_i)))) = \theta_\alpha\big(\textstyle\prod_{i=0}^{d} \theta_\alpha^{-1}(\widetilde{H}_\alpha(X_i|Y_i))\big) \tag{87}$$

and

$$\widetilde{H}_\alpha(X|Y) \geq \tfrac{|\alpha|}{1-\alpha} \log\Big(\phi_\alpha(0) + (\phi_\alpha(1) - \phi_\alpha(0)) \textstyle\prod_{i=0}^{d} \frac{\exp(\frac{1-\alpha}{|\alpha|}\widetilde{H}_\alpha(X_i|Y_i)) - \phi_\alpha(0)}{\phi_\alpha(1) - \phi_\alpha(0)}\Big). \tag{88}$$

1. We obtained sandwiching bounds when $w_{a,\alpha}$ is convex and $\alpha > 1$. Clearly from the derivation the same bounds holds if $w_{a,\alpha}$ is concave and $\alpha < 1$.
2. If $w_{a,\alpha}$ is convex and $\alpha < 1$ or $w_{a,\alpha}$ is concave and $\alpha > 1$ the inequalities are flipped. □

Now let

$$u_\alpha : x \in [0,1] \mapsto \frac{\phi'_\alpha(x)}{x\phi''_\alpha(x)} \in [0, \infty]. \tag{89}$$

Plugging (68) and (69) into (89), we obtain

$$u_\alpha(x) = \frac{((1+x)^{\alpha-1} - (1-x)^{\alpha-1})((1+x)^\alpha + (1-x)^\alpha)}{4(\alpha-1)(1-x^2)^{\alpha-2}x}. \tag{90}$$

Observe that the assertion (($w_{a,\alpha}$ is convex for all $a \in [0,1]$ and $\alpha > 1$) or ($w_{a,\alpha}$ is concave for all $a \in [0,1]$ and $\alpha < 1$)) is equivalent to the fact that u_α is decreasing. Similarly, the assertion (($w_{a,\alpha}$ is concave for all $a \in [0,1]$ and $\alpha > 1$) or ($w_{a,\alpha}$ is convex for all $a \in [0,1]$ and $\alpha < 1$)) is equivalent to the fact that u_α is increasing. Hence Theorem 2 admits the following simpler reformulation when X is uniformly distributed:

Theorem 3 (MGL for Binary Variables, General Case) *Let $\alpha \neq 1$ and*

$$\Lambda_\alpha = \frac{\phi_\alpha(0)}{\phi_\alpha(1) - \phi_\alpha(0)} = \frac{\delta_\alpha}{\frac{1+\mathrm{sgn}(\alpha)}{2} - \delta_\alpha} = \begin{cases} -1 & \text{if } \alpha < 0 \\ \frac{2^{\frac{1-\alpha}{\alpha}}}{1-2^{\frac{1-\alpha}{\alpha}}} & \text{if } \alpha > 0 \end{cases}. \tag{91}$$

If u_α is decreasing then

$$\tfrac{|\alpha|}{\alpha-1} \log\Big(1+\Lambda_\alpha^d \textstyle\prod_{i=0}^{d} \Big(e^{\frac{\alpha-1}{|\alpha|} I_\alpha(X_i;Y_i)} - 1\Big)\Big) \leq I_\alpha(X;Y) \leq \varphi_\alpha(\textstyle\prod_{i=0}^{d} \varphi_\alpha^{-1}(I_\alpha(X_i;Y_i))). \tag{92}$$

where $\varphi_\alpha(x) = \tilde{h}_\alpha(\frac{1}{2}) - \theta_\alpha(x)$. Let $\psi_\alpha(x) = \frac{|\alpha|}{\alpha-1} \log(1 + \Lambda_\alpha^{-1} x)$ this rewrites as

$$\psi_\alpha\left(\prod_{i=0}^d \psi_\alpha^{-1}(I_\alpha(X_i;Y_i))\right) \le I_\alpha(X;Y) \le \varphi_\alpha\left(\prod_{i=0}^d \varphi_\alpha^{-1}(I_\alpha(X_i;Y_i))\right). \tag{93}$$

If u_α is increasing then the inequalities holds in the reverse order.

Remark 5 We expect both lower and upper bounds to coincide for $\alpha = -1$. We check that $\psi_{-1}(x) = -\frac{1}{2}\log(1-x)$ while $\varphi_{-1}(x) = \psi_{-1}(x^2)$. While φ_{-1} and ψ_{-1} does not coincide we do obtain matching bounds since $x \mapsto x^2$ is monomial.

Example 1 We can simplify further the expression for some values of α, for instance:

- $u_3(x) = \frac{1+3x^2}{1-x^2}$ is increasing,
- $u_2(x) = (1+x^2)$ is increasing,
- $u_{\frac{7}{4}}(x) = -\frac{1}{3x}(2x(1-x^2)^{\frac{3}{4}} - (x+1)^{\frac{9}{4}} + (1-x)^{\frac{9}{4}})(1-x^2)^{\frac{1}{4}}$ is not monotonic;
- $u_{\frac{3}{2}}(x) = \sqrt{1-x^2}(2-\sqrt{1-x^2})$ is decreasing,
- $u_0(x) = u_{\frac{1}{2}}(x) = (1-x^2)$ is decreasing,
- $u_{-\frac{1}{2}}(x) = \frac{1}{3x}\sqrt{1-x^2}(\sqrt{1-x^2}+2)x$ is decreasing,
- $u_{-1}(x) = 1$ is constant,
- $u_{-2}(x) = \frac{1}{3}\frac{3+4x^2+x^4}{1-x^2}$ is increasing.

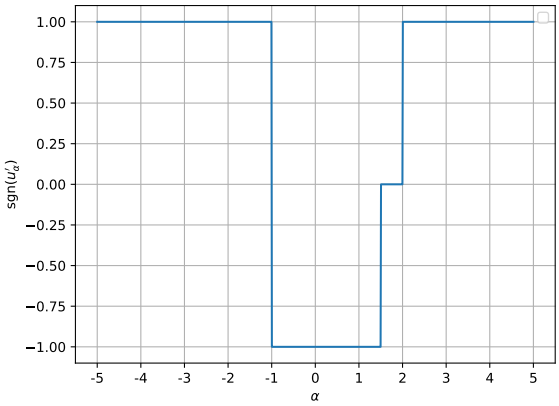

Fig. 3: Sign of u'_α verified numerically with SageMath.

Proposition 1 *It is easily checked numerically (see Figure 3) that:*

- *If $\alpha \le -1$ or $\alpha \ge 2$ then u_α is increasing;*
- *if $\alpha \in [-1, 1.5]$ then u_α is decreasing;*
- *if $\alpha \in (1,5,2)$ then u_α is not monotonic.*

Remark 6 This Proposition answers the open question left as a conjecture by Hirche [20, Conjecture IV.6.], where we established that $\hat{\alpha} = \frac{3}{2}$ in his conjecture. This also strengthens [20, Lemma IV] to the larger interval $(1.5, 2)$.

4.4 Extension to Hayashi's α-Entropy

We can proceed similarly to the previous section to recover the results from [20, Thm. 4.9] in terms of Hayashi's α-entropy and extend them to negative values of α. In this setting it is natural to introduce

$$K_\alpha^H(X|Y) = \exp(\frac{1-\alpha}{\mathrm{sgn}(\alpha)}\widetilde{H}_\alpha^H(X|Y)). \tag{94}$$

Let

$$k_\alpha^H : \begin{cases} x & \mapsto x^\alpha + (1-x)^\alpha \\ [0,1] & \to [2^{1-\alpha}, t_\alpha] \end{cases} \tag{95}$$

and

$$\phi_\alpha^H : \begin{cases} x & \mapsto k_\alpha^H(\frac{1-x}{2}) = (\frac{1+x}{2})^\alpha + (\frac{1-x}{2})^\alpha \\ [0,1] & \to [2^{1-\alpha}, t_\alpha] \end{cases} \tag{96}$$

where $t_\alpha = 1$ if $\alpha > 0$, $t_0 = 2$ and $t_\alpha = +\infty$ if $\alpha < 0$.

- $$\phi_\alpha^{H'}(x) = \frac{\alpha}{2}\left((\frac{x+1}{2})^{\alpha-1} - (\frac{1-x}{2})^{\alpha-1}\right) \tag{97}$$

  In particular,

  - If $0 < \alpha < 1$ then ϕ_α^H is decreasing.
  - If $\alpha = 0$ then ϕ_α^H is constant.
  - If $\alpha > 1$ or $\alpha < 0$ then ϕ_α^H is increasing.

- $$\phi_\alpha^{H''}(x) = \frac{\alpha(\alpha-1)}{4}\left((\frac{1+x}{2})^{\alpha-2} + (\frac{1-x}{2})^{\alpha-2}\right) \tag{98}$$

  - If $0 \le \alpha < 1$ then ϕ_α^H is concave.
  - If $\alpha > 1$ or $\alpha \le 0$ then ϕ_α^H is convex.

As in the previous section, the MGL for Hayashi's entropy hinges on the concavity/convextity of the mapping

$$w_{a,\alpha}^H : x \in [2^{1-\alpha}, t_\alpha] \mapsto \phi_\alpha^H(a\phi_\alpha^{H-1}(x)) \in [2^{1-\alpha}, t_\alpha]$$

for all a. By the pivotal lemma (Lemma 8), this characterized by the monotonicity of

$$u_\alpha^H : x \in [0,1] \mapsto \frac{\phi_\alpha^{H'}(x)}{x\phi_\alpha^{H''}(x)} = \frac{1}{\alpha-1}\frac{1}{x}\frac{(1+x)^{\alpha-1} - (1-x)^{\alpha-1}}{(1+x)^{\alpha-2} + (1-x)^{\alpha-2}} \in [0, +\infty].$$

Proposition 2 *It can be checked numerically (see Figure 4) that:*

- *If $\alpha \in (2,3)$ then u_α^H is decreasing;*

- *If* $\alpha \in (-\infty, 2) \cup (3, +\infty)$ *then* u_α^H *is increasing;*
- *If* $\alpha \in \{2, 3\}$ *then* u_α^H *is constant equal to* 1.

This confirms [20, Lemma IV.10 and IV.11] derived for positive values of α and extends it to negative values of α.

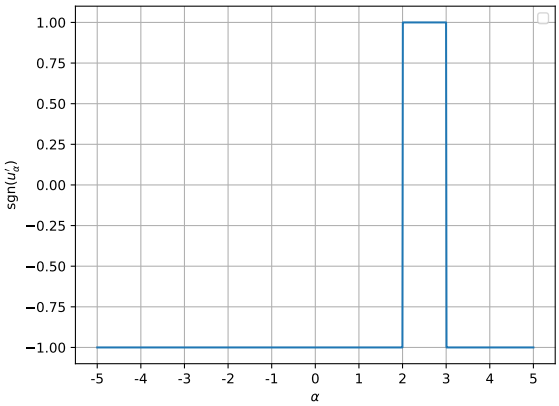

Fig. 4: Sign of u'_α verified numerically with SageMath.

In particular, we can generalize [20, Thm. IV.9] to negative values of α and reformulate it as

Theorem 4 *Let* $\alpha \neq 1$. *If* $\alpha \in (2, 3)$ *then*

$$\frac{\operatorname{sgn}(\alpha)}{1-\alpha} \log\Big(\phi_\alpha^H(0) + (\phi_\alpha^H(1) - \phi_\alpha^H(0)) \prod_{i=0}^d \frac{\exp(\frac{1-\alpha}{\operatorname{sgn}(\alpha)} \widetilde{H}_\alpha^H(X_i|Y_i)) - \phi_\alpha^H(0)}{\phi_\alpha^H(1) - \phi_\alpha^H(0)}\Big) \tag{99}$$

$$\leq \widetilde{H}_\alpha^H(X;Y) \leq \theta_\alpha\big(\textstyle\prod_{i=0}^d \theta_\alpha^{-1}(\widetilde{H}_\alpha(X_i;Y_i))\big). \tag{100}$$

If $\alpha \in (0, 2) \cup (3, +\infty)$ *the inequalities holds in the reverse order. If* $\alpha \in (-\infty, 0)$ *the upper bounds becomes a lower bound but we have no upper bound. If* $\alpha \in \{2, 3\}$ *the upper and lower bounds matches so that we have equality.*

If further X is uniformly distributed we may rewrite $\widetilde{H}_\alpha^H(X|Y)$ as $\tilde{h}_\alpha(\frac{1}{2}) - I_\alpha^R(X;Y)$ where I_α^R is termed Rényi α-mutual information [26].

Theorem 5 *Let* $\alpha \neq 1$, $\psi_\alpha^H(x) = \frac{1}{\alpha-1}\log(1 + (2^{\alpha-1} - 1)x)$ *if* $\alpha > 0$ *and* $\psi_\alpha^H(x) = 0$ *otherwise. Then if* $\alpha \in (-\infty, 2) \cup (3, +\infty)$,

$$\psi_\alpha^H\big(\textstyle\prod_{i=0}^d \psi_\alpha^{H-1}(I_\alpha^R(X_i;Y_i))\big) \leq I_\alpha^R(X;Y) \leq \varphi_\alpha\big(\textstyle\prod_{i=0}^d \varphi_\alpha^{-1}(I_\alpha^R(X_i;Y_i))\big). \tag{101}$$

If $\alpha \in (2, 3)$ *the inequalities holds in the reverse order. If* $\alpha \in \{2, 3\}$ *both upper and lower bounds matches and we have equality.*

4.5 Extension Beyond Binary Random Variables

Jog and Anantharam [24, 25] extended Mrs. Gerber's lemma to any Abelian group of order 2^n. Their MGL can be formulated as follows:

Theorem 6 *Let $X_0, \ldots, X_d$ be $d+1$ shares of X uniformly distributed in a an Abelian group of order $M = 2^n$. Let Y_i be the side information associated to the share X_i. Without loss of generality we can assume that $H(X_0|Y_0) \geq H(X_1|Y_1) \geq \ldots \geq H(X_d|Y_d)$. Let $k \triangleq \left\lfloor \frac{H(X_0|Y_0)}{\log 2} \right\rfloor$ and $t = \max\{i \in \{0, \ldots, d\} | H(X_i|Y_i) \geq k \log 2\}$. Then*

$$H(X|Y) \geq k \log 2 + h(h^{-1}((H(X_0|Y_0) - k \log 2) * \ldots * (H(X_t|Y_t) - k \log 2)) \quad (102)$$

where $h(h^{-1}((H(X_0|Y_0) - k \log 2) * \ldots * (H(X_t|Y_t) - k \log 2)) + k \log \leq (k+1) \log 2$.

In this lemma t can be seen as the effective masking order. If share does not contain enough entropy (or it is revealed) then it does not contribute to the security of the masked encoding.

Béguinot et al. [3] leveraged this result to improve the side channel security bounds of masked encodings. Masure and Standaert [31] then extended this result to the more general setting of masked computations. We propose to reformulate the MGL as a product using the discrete Fourier transform as follows:

Lemma 10 (Mrs. Gerber's Lemma (Revisited)) *If the group is Abelian of order 2^n then*

$$I(X; \mathbf{Y}) \leq k \log 2 + \varphi\Big(\prod_{i=0}^{t} \varphi^{-1}(I(X_i; Y_i) - k \log 2)\Big) \quad (103)$$

where $\varphi(x) = \log 2 - h(\frac{1-x}{2})$ and $I(X_i; Y_i) < (k+1) \log 2$ for $i = 0, \ldots, t$.

A weakened version of Mrs.Gerber's Lemma can also be obtained without restriction on the M using Pinkser and reverse Pinsker inequality together with the XOR Lemma as shown by Masure et al. [30, Theorem 3]:

Lemma 11 ((Weak) Mrs. Gerber's Lemma) *For any additive group $\mathcal{G}$,*

$$I(X; \mathbf{Y}) \leq \log\Big(1 + \frac{M}{2} \prod_{0}^{d} \frac{2}{\log e} I(X_i; Y_i)\Big) \leq M\big(\frac{2}{\log e}\big)^d \prod_{i=0}^{d} I(X_i; Y_i) \quad (104)$$

Proof. This is an easy corollary of Lemma 14 below. □

Béguinot et al. [5, Lemma 10] derived a MGL for the *complementary Doeblin coefficient* in any finite Abelian group by leveraging its relation to the stochastic degradation into an erasure channel. It turns out that this is equivalent to a MGL in terms of maximal cost leakage:

Lemma 12 (MGL for Maximal Cost Leakage (Reformulated)) *Let X be uniformly distributed M-ary random variable shared into $X_0, \ldots, X_d$ with corresponding leakages $Y_0, \ldots, Y_d$. Then*

$$I_{-\infty}(X;Y) \le -\log\Big(1 - \prod_{i=0}^{d}(1 - e^{-I_{-\infty}(X_i;Y_i)})\Big). \tag{105}$$

Proof. This is a reformulation of the MGL derived for the complementary Doeblin coefficient [5, Lemma 10] observing that $\overline{\mathcal{E}}(X \to Y) = 1 - \exp(-I_{-\infty}(X;Y))$. □

Under some restrictive assumption, Béguinot and Rioul also derived a MGL for maximal leakage in any finite Abelian group [6] using majorization theory:

Lemma 13 (MGL for Min-Entropy) *For any additive group $\mathcal{G}$, if for all $i \in \{0,\dots,d\}$ we have $I_\infty(X_i;Y_i) \le \log(\frac{M}{M-1})$ then*

$$I_\infty(X;\mathbf{Y}) \le \log\Big(1 + (M-1)^d \prod_{i=0}^{d}\Big(e^{I_\infty(X_i;Y_i)} - 1\Big)\Big). \tag{106}$$

Under some conditions given in [6] the results can be strengthened by removing the $(M-1)^d$ constant but these conditions are hard to verify in practice. Let $\psi_{\infty,M}(x) = \log(1 + (M-1)^{-1}x)$ then this rewrites as

$$I_\infty(X;\mathbf{Y}) \le \psi_{\infty,M}\Big(\prod_{i=0}^{d}\psi_{\infty,M}^{-1}(I_\infty(X_i;Y_i))\Big). \tag{107}$$

In this general case, we do not have an identity to express maximal leakage in terms of maximal cost leakage. Hence, both MGLs are not equivalent anymore contrary to the binary setting.

5 Other Information Leakage Measures

A number of alternative leakage measures have also been introduced in the side channel analysis literature to evaluate the security of masked random variables: Rényi α-information based on α-divergence, total variation information (TVI) based on total variation distance, (squared) Euclidean normal bias, relative error (RE) and average relative error (ARE). Fano inequalities and MGLs are established for all these, while tensorization can be seen only in some cases, albeit making use of Pinsker/reverse Pinsker inequalities.

To simplify the presentation in this section, we consider any side channel $X \to Y$ where X is generally uniformly distributed over the Abelian group $\mathcal{G}$ (e.g., the secret key K) with corresponding leakage Y observed by the attacker.

5.1 Information Leakage Metrics

- Rényi's α-Mutual-Information [26] is another generalization of mutual information. For any $\alpha \neq 1$, it is defined by

$$I_\alpha^R(X;Y) = \widetilde{D}_\alpha(P_{XY} \| P_X \otimes P_Y). \tag{108}$$

  Contrarily to Sibson's α-information, it is symmetric in X and Y which is why it can be said to be *mutual*. Explicitly

$$I_\alpha^R(X;Y) = \frac{\operatorname{sgn}(\alpha)}{\alpha - 1} \log \int p_Y(y) \int p_{X|Y}(x|y)^\alpha p_X(x)^{1-\alpha} d\mu_X(x) d\mu_Y(y). \tag{109}$$

- Total Variation Information [12] is pivotal in cryptography and is commonly used with the so-called simulation arguments:

$$\Delta_1(X;Y) = \frac{1}{2} \int_{x,y} |p_{XY}(xy) - p_X(x)p_Y(y)| d\mu_{XY}(x,y). \tag{110}$$

  This can be seen as $\Delta_1(X;Y) = \Delta_1(p_{XY}; p_X p_Y) = \mathbb{E}_Y \Delta_1(p_{X|Y}; p_X)$ where $\Delta_1(p;q) = \frac{1}{2}\|p-q\|_1$ is half the L^1-norm.
- Squared-Euclidean Norm Bias [26]:

$$\Delta_2(X;Y) = \int_{x,y} p_Y(y)(p_{X|Y}(x|y) - p_X(x))^2 d\mu_{XY}(x,y) \tag{111}$$

  This can be seen as $\Delta_2(X;Y) = \mathbb{E}_Y \Delta_2(p_{X|Y}, p_X)$ where $\Delta_2(p,q) = \|p-q\|_2^2$ is the squared L^2-norm.
- Euclidean Norm Bias [37] is more marginal and has been introduced in the first security for side channel analysis considering computations:

$$\beta(X;Y) = \not{\sum}_y p_Y(y) \sqrt{\sum_x (p_{X|Y}(x|y) - p_X(x))^2} \tag{112}$$

  This can be seen as $\beta(X;Y) = \mathbb{E}_Y \beta(p_{X|Y}, p_X)$ where $\beta(p,q) = \|p-q\|_2$ is the L^2-norm.
- Relative Error [36]:

$$\mathrm{RE}(X;Y) = \sup_{x,y} \left| 1 - \frac{p_{XY}(xy)}{p_X(x)p_Y(y)} \right| \tag{113}$$

- Average Relative Error [36], which is again symmetric in X and Y:

$$\mathrm{ARE}(X;Y) = \int p_Y(y) \sup_x \left| 1 - \frac{p_{XY}(xy)}{p_X(x)p_Y(y)} \right| d\mu_Y(y) \tag{114}$$

It turns out that for all these metrics, the analysis falls down to an *hybrid* argument as explained above in Remark 1.

5.2 Fano-Type Inequalities

5.2.1 Fano's Inequality for Rényi α-Mutual Information

Since

$$I_\alpha^R(X;Y) = \widetilde{D}_\alpha(p_{XY}\|p_X \otimes p_Y) \geq \min_{q_Y} \widetilde{D}_\alpha(p_{XY}\|p_X \otimes p_Y) = I_\alpha(X;Y), \quad (115)$$

we can plug this inequality into Lemma 6 to obtain the following Fano inequality for Rényi's α-information:

$$\tilde{d}_\alpha(\mathbb{P}_s(X|Y)\|\mathbb{P}_s(X)) \leq I_\alpha(X;Y) \leq I_\alpha^R(X;Y) \quad (116)$$

which holds for any positive or negative α.

5.2.2 Fano's Inequality for (Squared) Euclidean Norm Bias

For uniformly distributed X,

$$I_2^R(X;Y) = \log(1 + M\Delta_2(X;Y)) \leq \log(1 + M\beta(X;Y)). \quad (117)$$

where the first equality was by observed by Liu et al. [26, Lemma 3] and the inequality $\Delta_2(X;Y) \leq \beta(X;Y)$ is due to the fact that $\sqrt{x} \geq x$ when $x \in [0,1]$. Plugging this into Fano's inequality for Rényi's 2-mutual information above yields Fano's inequality for (squared) Euclidean norm Bias:

$$\tilde{d}_2(\mathbb{P}_s(X|Y)\|\mathbb{P}_s(X)) \leq \log(1 + M\Delta_2(X;Y)) \leq \log(1 + M\beta(X;Y)). \quad (118)$$

5.2.3 Fano's Inequality for (Average) Relative Error

As observed in [5], one has

$$1 - \exp(-I_{-\infty}(X;Y)) = \overline{\mathcal{E}}(X \to Y) \leq \mathrm{ARE}(X;Y) \leq \mathrm{RE}(X;Y) \quad (119)$$

which also writes

$$I_{-\infty}(X;Y) \leq -\log(1 - \mathrm{ARE}(X;Y)) \leq -\log(1 - \mathrm{RE}(X;Y)). \quad (120)$$

Plugging these inequalities into Lemma 6 with $\alpha = -\infty$ we obtain the following Fano inequality for both ARE and RE:

$$\tilde{d}_{-\infty}(\mathbb{P}_s(X|Y)\|\mathbb{P}_s(X)) \leq -\log(1-\mathrm{ARE}(X;Y)) \leq -\log(1-\mathrm{RE}(X;Y)). \quad (121)$$

5.3 Fano's Inequality for Total Variation Information

For total variation information one can obtain the following inequality from [40, Example 21, Eqn. 106] for a uniformly distributed X,

$$\mathbb{P}_s(X|Y) \leq \frac{1}{M} + \Delta_1(X;Y). \quad (122)$$

Remark 7 The intermediate reduction $1-\exp(-I_{-\infty}(X;Y)) = \overline{\mathcal{E}}(X \to Y) \leq M\Delta_1(X;Y)$ yields a looser bound $\mathbb{P}_s(X|Y) \leq \frac{1}{M} + (M-1)\Delta_1(X;Y)$.

5.4 Tensorizations

5.4.1 Tensorization for Total Variation Information

While Fano's inequality is easily obtained for Δ_1, tensorization is harder to establish. However, by Pinsker's and reverse Pinsker's inequality [41, Thm .28], one has

$$2(\log e)\Delta_1^2(X^m;Y^m) \leq I(X^m;Y^m) = mI(X;Y) \leq m\log(1+2M\Delta_1^2(X;Y)). \quad (123)$$

Hence, we convert total variation distance into mutual information which tensorizes for an i.i.d sequence (X^m, Y^m). However, this back and forth reduction necessarily implies a degraded bound with respect to a direct bound in terms of mutual information.

We can also observe that maximal cost leakage can be tensorized and then bounded in terms of total variation information by an analog of the reverse Pinsker inequality (stated in terms of complementary Doeblin coefficient [5, Lemma 9] and implicitly in [12]):

$$I_{-\infty}(X^m;Y^m) = mI_{-\infty}(X;Y) \leq -m\log(1-M\Delta_1(X;Y)). \quad (124)$$

Again a direct bound in terms of maximal cost leakage is tighter.

5.4.2 Tensorization for (Average) Relative Error

It is not clear how to tensorize (A)RE. (A)RE can be used with only one trace ($m=1$) to upper bound $I_{-\infty}$ which itself tensorizes.

5.5 Mrs. Gerber's Lemmas

5.5.1 MGL for Total Variation Information

Dziembowski et al. [14] leveraged the following XOR Lemma (analog of the MGL) in terms of total variation information:

Lemma 14 (XOR Lemma for Δ_1) *For any additive group $\mathcal{G}$ with uniformly distributed $X \sim \mathcal{U}(\mathcal{G})$,*

$$\Delta_1(p_{X|\mathbf{Y}}, p_X) \le 2^d \prod_{i=0}^{d} \Delta_1(p_{X_i|Y_i}, p_{X_i}) \qquad \Delta_1(X;\mathbf{Y}) \le 2^d \prod_{i=0}^{d} \Delta_1(X_i;Y_i)$$

We provide a simple proof.

Proof. Let $p_X = u$ be the uniform distribution. By Young's convolution inequality for two distributions, $\|p * q - u\|_1 = \|(p-u) * (q-u)\|_1 \le \|p-u\|_1 \cdot \|q-u\|_1$, i.e., $\Delta_1(p*q,u) \le 2\Delta_1(p,u)\Delta_1(q,u)$. The first inequality then follows by induction. The inequality for $\Delta_1(X;\mathbf{Y})$ follows by taking the expectation over $\mathbf{Y}$. □

As a corollary, one obtains the weak MGL of Lemma 11 above, following the proof of Masure et al [30]:

Proof of Lemma 11. We combine the XOR Lemma (Lemma 14) with Pinsker/reverse Pinsker inequalities: $2(\log e)\Delta_1^2 \le D \le \log(1+2M\Delta_1^2)$. This gives $D(p_{X|\mathbf{Y}}\|p_X) \le \log(1+2M(2^d \prod_{i=0}^{d} \Delta_1(p_{X_i|Y_i}, p_{X_i}))^2) \le \log(1+\frac{M}{2}\prod_{i=0}^{d} \frac{2}{\log e} D(p_{X_i|Y_i}\|p_{X_i}))$. Taking the expectation over Y and applying Jensen's inequality (concavity of the logarithm) gives $I(X;\mathbf{Y}) \le \log(1+\frac{M}{2}\prod_{i=0}^{d}\frac{2}{\log e} I(X_i;Y_i))$. □

5.5.2 MGL for (Average) Relative Error

Prest et al. [36, Theorem 3] proved a MGL for RE:

Lemma 15 (MGL for Relative Error) *For any additive group $\mathcal{G}$ with uniformly distributed $X \sim \mathcal{U}(\mathcal{G})$,*

$$RE(X,\mathbf{Y}) \le \prod_{i=0}^{d} \mathrm{RE}(X_i,Y_i) \tag{125}$$

We provide a very simple proof of Lemma 15 compared to the original:

Proof. Let u be the uniform distribution and $M = |\mathcal{G}|$. Letting $RE(p,q) = \max|\frac{p}{q}-1|$, one has by definition $RE(X,Y) = RE(p_{X|Y}, p_X) = RE(p_{X|Y}, u)$ where $RE(p,u) = \max|Mp-1| = \|Mp-1\|_\infty$. Now $\|Mp*q-1\|_\infty = M\|(p*q-u)\|_\infty = M\|(p-u)*(q-u)\|_\infty \le M\|(p-u)\|_1\|(q-u)\|_\infty \le M^2\|(p-u)\|_\infty\|(q-u)\|_\infty =$

$\|Mp-1\|_\infty \|Mq-1\|_\infty$ where we have used Young's convolutional inequality. The inequality follows by induction. □

Although Prest et al. [36] did not derive a MGL for *average* relative error ARE, its derivation turns out to be very similar using our simplified proof:

Lemma 16 (MGL for Average Relative Error) *For any additive group $\mathcal{G}$ with uniformly distributed $X \sim \mathcal{U}(\mathcal{G})$,*

$$ARE(X, \mathbf{Y}) \leq \prod_{i=0}^{d} ARE(X_i, Y_i) \tag{126}$$

Proof. Same as above for RE, replacing the maximum over Y by the average (expectation over Y) instead. □

5.5.3 MGL for (Squared) Euclidean Norm Bias

Prouff & Rivain implicitly derived a MGL for β [37]. Liu et al. [26, Lemma 4] also proved a MGL for Δ_2. Both results can in fact be proved similarly using the Cauchy-Schwarz inequality.

Lemma 17 (MGL/XOR Lemma for (Squared) Euclidean Norm Bias) *For any additive group $\mathcal{G}$ with uniformly distributed $X \sim \mathcal{U}(\mathcal{G})$,*

$$\Delta_2(p_{X|\mathbf{Y}}, p_X) \leq M^d \prod_{i=0}^{d} \Delta_2(p_{X_i|Y_i}, p_{X_i}) \qquad \Delta_2(X; \mathbf{Y}) \leq M^d \prod_{i=0}^{d} \Delta_2(X_i; Y_i) \tag{127}$$

and

$$\beta(p_{X|\mathbf{Y}}, p_X) \leq M^{d/2} \prod_{i=0}^{d} \beta(p_{X_i|Y_i}, p_{X_i}) \qquad \beta(X; \mathbf{Y}) \leq M^{d/2} \prod_{i=0}^{d} \beta(X_i; Y_i). \tag{128}$$

Proof. Let u be the pmf of the uniform distribution. For two distributions p, q, one has $\|p*q-u\|_2^2 = \|(p-u)*(q-u)\|_2^2$, where by Cauchy-Schwarz, $|(p-u)*(q-u)| \leq \|p-u\|_2 \|q-u\|_2$, hence $\|p*q-u\|_2^2 \leq M\|p-u\|_2^2\|q-u\|_2^2$, that is, $\|p*q-u\|_2 \leq \sqrt{M}\|p-u\|_2\|q-u\|_2$. The result then follows as in the above proofs. □

Liu et al. [26, Eqn. 30] derived a MGL for Rényi's 2-mutual information which is equivalent to Lemma 17 and matches Theorem 5 for binary random variables:

Lemma 18 (Mrs. Gerber's Lemma for I_2^R) *For uniformly distributed X,*

$$I_2^R(X; \mathbf{Y}) \leq \log\Big(1 + \prod_{i=0}^{d} (\exp I_2^R(X_i; Y_i) - 1)\Big) \tag{129}$$

Let $\psi_{2,R}(x) = \log(1+x)$ this rewrites as

$$I_2^R(X;\mathbf{Y}) \le \psi_{2,R}\Big(\prod_{i=0}^{d} \psi_{2,R}^{-1}(I_2^R(X_i;Y_i))\Big) \tag{130}$$

Proof. $I_2^R(X;\mathbf{Y}) = \log(1 + M\Delta_2(X;\mathbf{Y}))$ and similarly for $I_2^R(X_i;Y_i)$. □

6 Resulting Security Bounds and Numerical Simulations

By combining all steps as explained in the Introduction, for binary secrets we obtain an inequality of the type

$$m \ge \frac{\tilde{d}_\alpha(\mathbb{P}_s \| \frac{1}{M})}{f_\alpha(\prod_{i=0}^{d} f_\alpha^{-1}(I_\alpha(X_i;Y_i)))} \tag{131}$$

provided that α is not in $[\frac{3}{2}, 1]$ and where f_α is either ψ_α or φ_α depending on the value of α. For groups of order $M = 2^n$ one obtains the following that unifies all results from the literature:

Theorem 7 *For small enough leakages[1], we obtain the following series of lower bounds on m for targeted probability of success $\mathbb{P}_s$:*

$$m \ge \frac{d(\mathbb{P}_s \| \frac{1}{M})}{\varphi(\prod_{i=0}^{d} \varphi^{-1}(I(X_i;Y_i)))} \tag{132}$$

$$m \ge \frac{d_\infty(\mathbb{P}_s \| \frac{1}{M})}{\psi_{\infty,M}(\prod_{i=0}^{d} \psi_{\infty,M}^{-1}(I_\infty(X_i;Y_i)))} \tag{133}$$

$$m \ge \frac{d_2(\mathbb{P}_s \| \frac{1}{M})}{\psi_{2,R}(\prod_{i=0}^{d} \psi_{2,R}^{-1}(I_2^R(X_i;Y_i)))} \tag{134}$$

$$m \ge \frac{d_{-\infty}(\mathbb{P}_s \| \frac{1}{M})}{\psi_{-\infty}(\prod_{i=0}^{d} \psi_{-\infty}^{-1}(I_{-\infty}(X_i;Y_i)))} \tag{135}$$

$$m \ge \frac{d_{-\infty}(\mathbb{P}_s \| \frac{1}{M})}{-\log(1 - M^{d+1} \prod_{i=0}^{d} \Delta_1(X_i;Y_i))} \tag{136}$$

$$m \ge \frac{d_{-\infty}(\mathbb{P}_s \| \frac{1}{M})}{-\log(1 - (\frac{M}{\sqrt{2\log e}})^{d+1} \prod_{i=0}^{d} \sqrt{I(X_i;Y_i)})} \tag{137}$$

$$m \ge \frac{d_{-\infty}(\mathbb{P}_s \| \frac{1}{M})}{-\log(1 - \prod_{i=0}^{d} \mathrm{RE}(X_i;Y_i))} \tag{138}$$

[1] So that the expression makes sense e.g. the terms in the logarithm should be positive, I_∞ should be less than $\log \frac{M}{M-1}$ and I should be less than $\log 2$.

$$m \geq \frac{d_{-\infty}(\mathbb{P}_s \| \frac{1}{M})}{-\log(1 - \prod_{i=0}^{d} \mathrm{ARE}(X_i; Y_i))} \tag{139}$$

$$m \geq \frac{d_{-\infty}(\mathbb{P}_s \| \frac{1}{M})}{-\log(1 - M^{d+1} \prod_{i=0}^{d} \beta(X_i; Y_i))} \tag{140}$$

Theorem 7 can be used by an evaluator to evaluate the level of (in)security of a device protected by masking against side channel attack. The lower bound on m may be used to assess the required time to mount an attack which useful for certification such as the common criteria. It can also be combined with re-keying techniques that changes the secret key regularly based on the value of m to avoid the attack. An interesting extension for Theorem 7 is to also consider multiplications of two protected sensitive values as in [31, 5].

To conclude, we compare the results of Theorem 7 in Figure 5 assuming $M = 2^8$ (one-byte texts and key) and using the commonly adopted *Hamming weight model*: the Hamming weight of each share leaks under additive white Gaussian noise of variance σ^2:

$$X_i \to Y_i = w_H(X_i) + \sigma \mathcal{N}(0, 1) \tag{141}$$

It appears that for $m = 1$ the best bound is for $\alpha = \infty$. Then on most of the range of values of m, cases $\alpha = 1, 2$ outperform the other bounds.

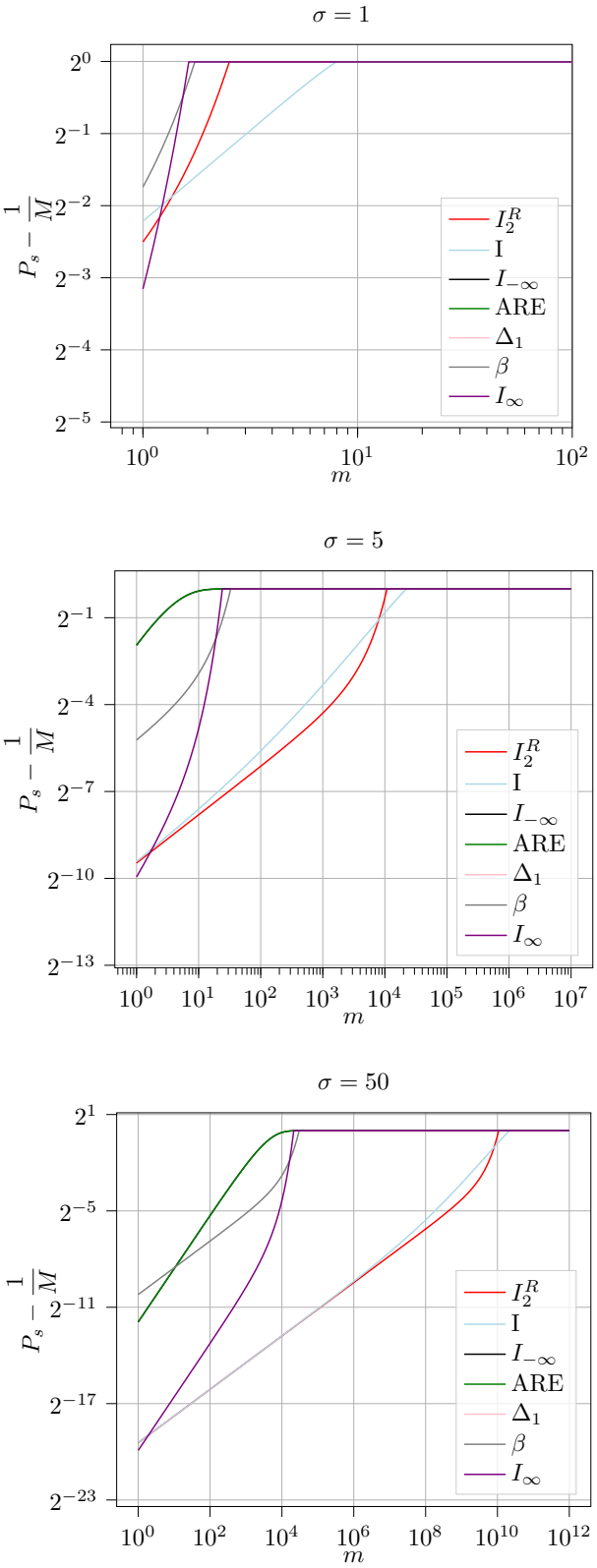

Fig. 5: Upper bounds on the success rate advantage $\mathbb{P}_s - \frac{1}{M}$ vs. number of measurements for several noise levels $\sigma = 1$, 5 and 50.

References

1. Suguru Arimoto. Information measures and capacity of order α for discrete memoryless channels. *Topics in information theory*, 1977.
2. Erdal Arıkan. An inequality on guessing and its application to sequential decoding. *Proceedings of 1995 IEEE International Symposium on Information Theory*, pages 322–, 1995.
3. Julien Béguinot, Wei Cheng, Sylvain Guilley, Yi Liu, Loïc Masure, Olivier Rioul, and François-Xavier Standaert. Removing the field size loss from Duc et al.'s conjectured bound for masked encodings. In *Proc. 14th International Workshop on Constructive Side-Channel Analysis and Secure Design (COSADE 2023), Munich, Germany, April 3-4, 2023*, volume 13979 of *Lecture Notes in Computer Science*, pages 86–104. Springer, 2023.
4. Julien Béguinot, Wei Cheng, Sylvain Guilley, and Olivier Rioul. Be my guesses: The interplay between side-channel leakage metrics. *Microprocessors and Microsystems*, 107:105045, 2024.
5. Julien Béguinot, Wei Cheng, Sylvain Guilley, and Olivier Rioul. Formal security proofs via Doeblin coefficients: Optimal side-channel factorization from noisy leakage to random probing. In *44th Annual International Cryptology Conference (CRYPTO 2024), Santa Barbara, USA, Aug. 18-22, 2024*, 2024.
6. Julien Béguinot, Yi Liu, Olivier Rioul, Wei Cheng, and Sylvain Guilley. Maximal leakage of masked implementations using Mrs. Gerber's lemma for min-entropy. In *IEEE International Symposium on Information Theory (ISIT 2023), Taipei, Taiwan, June 25-30, 2023*, volume abs/2305.06276, 2023.
7. Julien Béguinot and Olivier Rioul. What can information guess? Guessing advantage vs. Rényi entropy for small leakages. In *IEEE International Symposium on Information Theory (ISIT 2024), Athens, Greece, July 7-12, 2024*, 2024.
8. Matthieu R. Bloch, Onur Günlü, Aylin Yener, Frédérique E. Oggier, H. Vincent Poor, L. Sankar, and Rafael F. Schaefer. An overview of information-theoretic security and privacy: Metrics, limits and applications. *IEEE Journal on Selected Areas in Information Theory*, 2:5–22, 2021.
9. Nicolas Bruneau, Sylvain Guilley, Annelie Heuser, and Olivier Rioul. Masks will fall off - Higher-order optimal distinguishers. In Palash Sarkar and Tetsu Iwata, editors, *Advances in Cryptology - ASIACRYPT 2014 - 20th International Conference on the Theory and Application of Cryptology and Information Security, Kaoshiung, Taiwan, R.O.C., December 7-11, 2014, Proceedings, Part II*, volume 8874 of *Lecture Notes in Computer Science*, pages 344–365. Springer, 2014.
10. Suresh Chari, Charanjit S. Jutla, Josyula R. Rao, and Pankaj Rohatgi. Towards sound approaches to counteract power-analysis attacks. In *Annual International Cryptology Conference*, 1999.
11. Fan Cheng. Generalization of Mrs. Gerber's lemma. *Commun. Inf. Syst.*, 14(2):79–86, 2014.
12. Alexandre Duc, Stefan Dziembowski, and Sebastian Faust. Unifying Leakage Models: From Probing Attacks to Noisy Leakage. In Phong Q. Nguyen and Elisabeth Oswald, editors, *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*, volume 8441 of *Lecture Notes in Computer Science*, pages 423–440. Springer, 2014.
13. Alexandre Duc, Sebastian Faust, and François-Xavier Standaert. Making masking security proofs concrete (or how to evaluate the security of any leaking device), extended version. *Journal of Cryptology*, 32:1263 – 1297, 2015.
14. Stefan Dziembowski, Sebastian Faust, and Maciej Skorski. Noisy leakage revisited. In Elisabeth Oswald and Marc Fischlin, editors, *Advances in Cryptology - EUROCRYPT 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part II*, volume 9057 of *Lecture Notes in Computer Science*, pages 159–188. Springer, 2015.
15. Amedeo Roberto Esposito, Adrien Vandenbroucque, and Michael Gastpar. On Sibson's α-mutual information. *2022 IEEE International Symposium on Information Theory (ISIT)*, pages 2904–2909, 2022.

16. Robert M. Fano. *Class notes for course 6.574: Transmission of information*. MIT, 1952.
17. Te Sun Han and Sergio Verdú. Generalizing the Fano inequality. *IEEE Trans. Inf. Theory*, 40(4):1247–1251, 1994.
18. Masahito Hayashi. Exponential decreasing rate of leaked information in universal random privacy amplification. *IEEE Transactions on Information Theory*, 57(6):3989–4001, June 2011.
19. Annelie Heuser, Olivier Rioul, and Sylvain Guilley. Good is not good enough - Deriving optimal distinguishers from communication theory. In Lejla Batina and Matthew Robshaw, editors, *Cryptographic Hardware and Embedded Systems - CHES 2014 - 16th International Workshop, Busan, South Korea, September 23-26, 2014. Proceedings*, volume 8731 of *Lecture Notes in Computer Science*, pages 55–74. Springer, 2014.
20. Christoph Hirche. Rényi bounds on information combining. In *IEEE International Symposium on Information Theory, ISIT 2020, Los Angeles, CA, USA, June 21-26, 2020*, pages 2297–2302. IEEE, 2020.
21. Hsiang Hsu, Shahab Asoodeh, Salman Salamatian, and Flávio P. Calmon. Generalizing bottleneck problems. In *2018 IEEE International Symposium on Information Theory, ISIT 2018, Vail, CO, USA, June 17-22, 2018*, pages 531–535. IEEE, 2018.
22. Ibrahim Issa, Aaron B. Wagner, and Sudeep Kamath. An operational approach to information leakage. *IEEE Trans. Inf. Theory*, 66(3):1625–1657, 2020.
23. Akira Ito, Rei Ueno, and Naofumi Homma. On the success rate of side-channel attacks on masked implementations: Information-theoretical bounds and their practical usage. In Heng Yin, Angelos Stavrou, Cas Cremers, and Elaine Shi, editors, *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security, CCS 2022, Los Angeles, CA, USA, November 7-11, 2022*, pages 1521–1535. ACM, 2022.
24. Varun Jog and Venkat Anantharam. The entropy power inequality and Mrs. Gerber's lemma for groups of order 2^n. *IEEE Transactions on Information Theory*, 60(7):3773–3786, 2014.
25. Varun S. Jog and Venkat Anantharam. The entropy power inequality and Mrs. Gerber's lemma for groups of order 2^n. In *Proceedings of the 2013 IEEE International Symposium on Information Theory, Istanbul, Turkey, July 7-12, 2013*, pages 594–598. IEEE, 2013.
26. Yi Liu, Julien Béguinot, Wei Cheng, Sylvain Guilley, Loïc Masure, Olivier Rioul, and François-Xavier Standaert. Improved alpha-information bounds for higher-order masked cryptographic implementations. In *IEEE Information Theory Workshop, ITW 2023, Saint-Malo, France, April 23-28, 2023*, pages 81–86. IEEE, 2023.
27. Yi Liu, Wei Cheng, Sylvain Guilley, and Olivier Rioul. On conditional alpha-information and its application to side-channel analysis. In *IEEE Information Theory Workshop, ITW 2021, Kanazawa, Japan, October 17-21, 2021*, pages 1–6. IEEE, 2021.
28. Mokshay M. Madiman and Andrew R. Barron. Generalized entropy power inequalities and monotonicity properties of information. *IEEE Transactions on Information Theory*, 53:2317–2329, 2006.
29. James L. Massey. Guessing and entropy. *Proceedings of 1994 IEEE International Symposium on Information Theory*, pages 204–, 1994.
30. Loïc Masure, Olivier Rioul, and François-Xavier Standaert. A nearly tight proof of Duc et al.'s conjectured security bound for masked implementations. In Ileana Buhan and Tobias Schneider, editors, *Smart Card Research and Advanced Applications - 21st International Conference, CARDIS 2022, Birmingham, UK, November 7-9, 2022, Revised Selected Papers*, volume 13820 of *Lecture Notes in Computer Science*, pages 69–81. Springer, 2022.
31. Loïc Masure and François-Xavier Standaert. Prouff and Rivain's formal security proof of masking, revisited - Tight bounds in the noisy leakage model. In Helena Handschuh and Anna Lysyanskaya, editors, *Advances in Cryptology - CRYPTO 2023 - 43rd Annual International Cryptology Conference, CRYPTO 2023, Santa Barbara, CA, USA, August 20-24, 2023, Proceedings, Part III*, volume 14083 of *Lecture Notes in Computer Science*, pages 343–376. Springer, 2023.
32. John O. Pliam. Guesswork and variation distance as measures of cipher security. In Howard M. Heys and Carlisle M. Adams, editors, *Selected Areas in Cryptography, 6th Annual International*

Workshop, SAC'99, Kingston, Ontario, Canada, August 9-10, 1999, Proceedings, volume 1758 of *Lecture Notes in Computer Science*, pages 62–77. Springer, 1999.
33. John O. Pliam. On the incomparability of entropy and marginal guesswork in brute-force attacks. In Bimal K. Roy and Eiji Okamoto, editors, *Progress in Cryptology - INDOCRYPT 2000, First International Conference in Cryptology in India, Calcutta, India, December 10-13, 2000, Proceedings*, volume 1977 of *Lecture Notes in Computer Science*, pages 67–79. Springer, 2000.
34. Yury Polyanskiy and Sergio Verdú. Arimoto channel coding converse and Rényi divergence. *2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pages 1327–1333, 2010.
35. Yury Polyanskiy and Yihong Wu. *Information theory: From coding to learning*. Cambridge university press, 2024.
36. Thomas Prest, Dahmun Goudarzi, Ange Martinelli, and Alain Passelègue. Unifying leakage models on a Rényi day. In *Advances in Cryptology - CRYPTO 2019 - 39th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2019, Proceedings, Part I*, pages 683–712, 2019.
37. Emmanuel Prouff and Matthieu Rivain. Masking against side-channel attacks: A formal security proof. In *Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings*, pages 142–159, 2013.
38. Alfréd Rényi. Az információelmélet néhány alapveto vkérdése (some basic questions in information theory). *Magyar Tud. Akad. Mat. Fiz. Oszt. Közl*, 10(3):251–282, 1960.
39. Olivier Rioul. A primer on alpha-information theory with application to leakage in secrecy systems. In Frank Nielsen and Frédéric Barbaresco, editors, *Geometric Science of Information - 5th International Conference, GSI 2021, Paris, France, July 21-23, 2021, Proceedings*, volume 12829 of *Lecture Notes in Computer Science*, pages 459–467. Springer, 2021.
40. Olivier Rioul. The interplay between error, total variation, alpha-entropy and guessing: Fano and Pinsker direct and reverse inequalities. *Entropy*, 25(7):978, 2023.
41. Igal Sason and Sergio Verdú. f-Divergence inequalities. *IEEE Trans. Inf. Theory*, 62(11):5973–6006, 2016.
42. Igal Sason and Sergio Verdú. Arimoto-Rényi conditional entropy and Bayesian M-ary hypothesis testing. *IEEE Trans. Inf. Theory*, 64(1):4–25, 2018.
43. Shlomo Shamai and Aaron D. Wyner. A binary analog to the entropy-power inequality. *IEEE Trans. Inf. Theory*, 36(6):1428–1430, 1990.
44. Robin Sibson. Information radius. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14:149–160, 1969.
45. Tim van Erven and Peter Harremos. Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory*, 60(7):3797–3820, 2014.
46. Sergio Verdú. α-Mutual information. *2015 Information Theory and Applications Workshop (ITA)*, pages 1–6, 2015.
47. Aaron D. Wyner and Jacob Ziv. A theorem on the entropy of certain binary sequences and applications -I. *IEEE Trans. Inf. Theory*, 19(6):769–772, 1973.