Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 32 (2022)

Incompressiblity and Next-Block Pseudoentropy

Iftach Haitner*† Noam Mazor* Jad Silbak*

December 12, 2025

Abstract

A distribution is *k-incompressible*, Yao [FOCS '82], if no efficient compression scheme compresses it to less than k bits. While being a natural measure, its relation to other computational analogs of entropy such as *pseudoentropy*, Hastad, Impagliazzo, Levin, and Luby [SICOMP '99], and to other cryptographic hardness assumptions, was unclear.

We advance towards a better understating of this notion, showing that a k-incompressible distribution has $(k-2)$ bits of *next-block pseudoentropy*, a refinement of pseudoentropy introduced by Haitner, Reingold, and Vadhan [SICOMP '13]. We deduce that a samplable distribution X that is $(H(X)+2)$-incompressible, implies the existence of one-way functions.

Keywords: incompressibility; next-block pseudoentropy; sparse languages

*The Blavatnik School of Computer Science at Tel-Aviv University. E-mail:{iftachh@tauex.tau.ac.il, noammaz@gmail.com, jadsilbak@gmail.com. Research supported by Israel Science Foundation grant 666/19.

†Member of the Check Point Institute for Information Security.

ISSN 1433-8092

Contents

1 Introduction

Computational analogs of information-theoretic notions have given rise to some of the most interesting phenomena in the theory of computation. For example, *computational indistinguishability*, a computational analogue of statistical indistinguishability introduced by Goldwasser and Micali [GM84], enabled the bypassing of Shannon's impossibility results on perfectly secure encryption [Sha49], and provided the basis for the computational theory of pseudorandomness [BM82; Yao82]. *Pseudoentropy*, a computational analogue of entropy introduced by Hastad, Impagliazzo, Levin, and Luby [Has+99], was the key to their fundamental result that established the equivalence of pseudorandom generators and one-way functions and has become a basic concept in complexity theory and cryptography. *Next-block pseudoentropy*, a refinement of pseudoentropy introduced by Haitner, Reingold, and Vadhan [Hai+13] and Vadhan and Zheng [VZ12], has led to simpler and more efficient constructions of pseudorandom generators based on one-way functions. An analogue of entropy for the realm of *unforgeability*, named *inaccessible entropy*, introduced by Haitner, Reingold, Vadhan, and Wee [Hai+19; Hai+20], has led to simpler and more efficient constructions of *statistically hiding commitment* and *universal one-way hash functions* from one-way functions.

In contrast to the above, *incompressibility*, a computational analogue of entropy introduced by Yao [Yao82], was much less explored. Roughly, a random variable X is *k-incompressible*, if there exists no efficient (i.e., poly-time) compression scheme that compresses X to less than k bits. That is, there exists no efficient encoding scheme $(\mathsf{Enc}, \mathsf{Dec})$, i.e., $\mathsf{Dec}(\mathsf{Enc}(x)) = x$ for every $x \in \mathrm{Supp}(X)$, with

$$\mathrm{E}[|\mathsf{Enc}(X)|] < k \ .$$

(Both X and k are functions of a "security parameter" n, which we omit throughout the introduction). It is immediate that a pseudorandom distribution of k bits is $(k - O(1))$-incompressible.[1] More generally, Wee [Wee04] proved that a distribution with k-bits of pseudoentropy, i.e., computationally indistinguishable from a distribution Y of k-bits of (real) Shannon entropy, is $(k-O(\log n))$-incompressible. In contrast, the converse direction is less clear. Barak, Shaltiel, and Wigderson [Bar+03] showed how to extract $k-\omega(\log n)$ (close to uniform) bits from a *k-strongly-incompressible* source X, i.e., $\Pr[|\mathsf{Enc}(X)| < k - t] \leq 2^{-t}$ for every t.[2] Other works showed that incompressibility is unlikely to imply pseudoentropy. Wee [Wee04] showed that proving that incompressibility implies (similar amount of) pseudoentropy cannot be done using *black-box reductions*. Hsiao, Lu, and Reyzin [Hsi+07] proved that under a certain cryptographic assumption, there exists a distribution whose *conditional* incompressibility is much larger than its conditional pseudoentropy, where conditional means that the compression and pseudoentropy are measures with respect to a randomly generated common reference string. Still, the most basic questions remained open:

> Does k-incompressibility imply having a different, natural, type of "pseudoentropy"?
>
> Does *non-trivial* incompressibility, i.e., sufficiently larger than the real entropy, imply the existence of one-way functions?

We give affirmative answers to the above questions, proving that a k-incompressible source has $(k-2)$ bits of next-block pseudoentropy, and thus, a samplable source X that is $(H(X)+2)$-

[1] The $O(1)$ loss is since even the (true) uniform distribution can be compressed by $\Theta(1)$ bits (using non prefix-free schemes) Szpankowski and Verdú [SV11].

[2] I.e., not only that one cannot efficiently compress X to less than k bits, but it cannot compress, non-trivially, even *parts* of X.

incompressible implies the existence of one-way functions. Before stating our results in more details, we recall the notion of next-block pseudoentropy Haitner et al. [Hai+13], focusing on its single-bit block variant, called *next-bit pseudoentropy*.

Next-bit pseudoentropy. Next-bit pseudoentropy measures the *bit-wise unpredictability* of X: how hard is it to predict X_i from $X_{<i} = (X_1, \ldots, X_{i-1})$, for a uniform i. More formally, a random variable $X = (X_1, \ldots, X_m)$ over $\{0,1\}^m$ has next-bit pseudoentropy k, if there exists a set of random variables $\{Y_i\}_{i\in[m]}$, jointly distributed with X, such that

1. $\sum_{i\in[m]} H(Y_i \mid X_{<i}) \geq k$, for H being the Shannon entropy function, and
2. $(X_{<i}, X_i)$ is computationally indistinguishable from $(X_{<i}, Y_i)$, for every i.

That is, X has next-bit pseudoentropy k if predicting X_i from $X_{<i}$ is not easier than predicting Y_i from $X_{<i}$, where the bits of Y have k bits of (real) entropy given the past. It follows from [VZ12] that k-bits of (standard) pseudoentropy implies k-bits of next-bit pseudoentropy,[3] but the converse does not always hold.[4] Yet, Haitner et al. [Hai+13] showed that an efficiently samplable source with non-trivial next-bit pseudoentropy, can be used to construct pseudorandom generators (and thus one-way functions).

1.1 Our Results

In this paper, following [Bar+03; Wee04], we focus on the *non-uniform settings*—the efficient algorithms get non-uniform polynomial-size advice per input length—and defer the uniform version of our results to the future version (see more details in Section 1.2). We state our results with respect to a weaker notion of incompressibility, where the source is only assumed to be incompressible by *prefix-free* schemes: no codeword is a prefix of another.

Lemma 1.1 (Incompressiblity $\to$ next-bit pseudoentropy)**.** *A random variable that is k-incompressible by efficient prefix-free schemes, has next-bit pseudoentropy (at least) $k(n) - 2$.*

That is, incompressibility is a stronger measure of "pseudoentropy" than next-bit pseudoentropy. Since, incompressibility is weaker than pseudoentropy, we now have a rather good understanding about the computational hardness incompressibility induces. By Haitner et al. [Hai+13], Lemma 1.1 yields the following characterization.

Theorem 1.2 (Non-trivial incompressibility implies one-way functions)**.** *Assume there exist an (efficiently) samplable random variable X that is $(H(X) + 2 + 1/p(n))$-incompressible by efficient prefix-free schemes, for some $p \in$ poly, then one-way functions exist.*

That is, if one-way functions do not exist, e.g., we live in "Pessiland" [Imp95], then any samplable distribution can be compressed to its entropy plus two bits.

Theorem 1.2 improves upon previous results that require additional structure from the incompressible distribution. Wee [Wee04] proved that if an m-bit *flat* X (i.e., uniformly distributed over

[3] Indeed, if X has k-bits of pseudoentropy, X is $k - H(X)$ KL-hard to estimate, and thus X_I is $(k - H(X))/m$ KL-hard to estimate given $X_{<I}$. The later implies that X_I has pseudoentropy k/m given $X_{<I}$.

[4] Let g be a pseudorandom generator from n bits to $2n$ bits. Then $Z = (g(U_n), U_n)$ does not have pseudoentropy larger than n (Z is determined by its last n bits), but has $2n$ bits of next-bit pseudoentropy: let $Y_1, \ldots, Y_{2n}$ be uniform and independent bits, and $(Y_{2n+1}, \ldots, Y_{3n}) = (Z_{2n+1}, \ldots, Z_{3n})$.

its support) is $(H(X) + \Omega(\log m))$-incompressible, then one-way functions exist. Where a simple application of Barak et al. [Bar+03], yields the same for $(H(X) + \omega(\log n))$-*strongly*-incompressible X.

Remark 1.3 (Incompressibility and one-way functions)**.** *The common method for proving that a certain primitive implies (the existence of) one-way functions, is to show that the primitive induces a function that if invertible would contradict the security of the primitive (Impagliazzo and Luby [IL89]). Interestingly, such an approach does not seem to work for samplable incompressible distributions; the efficient sampler of an incompressible X might use $\mathrm{poly}(|X|) > H(X)$ random bits. Thus, even if we are able to invert it (assuming that one-way functions don't exist) and output the (long) random string used for sampling, this alone does not contradict the incompressibility of X. Moreover, while the probability of every element in the support of X can be efficiently estimated (assuming that one-way functions do not exist, [IL89]), this also does not suffice for efficient compression.*

So while our main result is establishing a connection between incompressibility and next-bit pseudoentropy, the fact that incompressibility implies one-way functions is an interesting corollary that to the best of our understanding does not immediately follow any standard technique.

Applications to sparse languages. A language $\mathcal{L}$ is *s-sparse* if $|\mathcal{L} \cap \{0,1\}^n| \leq 2^{s(n)}$. Theorem 1.2 yields the following characterization of sparse languages.

Theorem 1.4 (Informal)**.** *Let $\mathcal{L}$ be an s-sparse language and let $\mathcal{D}$ be a samplable distribution over $\mathcal{L}$ (i.e., $\mathrm{Supp}(\mathcal{D}_n) \cap \{0,1\}^n \subseteq \mathcal{L} \cap \{0,1\}^n$). If $\mathcal{D}$ is $(s+3)$-incompressible, then one-way functions exist.*

Remark 1.5 (The two bits gap)**.** *One might wonder whether the annoying two bits gap in Lemma 1.1 and theorems 1.2 and 1.4 is unavoidable. We pay these two bits since in the proof of Lemma 1.1 we use the arithmetic encoding prefix-free compression scheme, which compresses a random variable X to $H(X)+2$ bits. We use arithmetic encoding since it can be implemented efficiently given oracle access to the accumulated distribution function of X (with respect to the lexicographic order of elements). Indeed, the 2 bits gap can be reduced given a better encoding scheme that is efficient in these settings. While there are prefix-free compression schemes that get closer to $H(x)$, cf., Shannon [Sha48], Fano [Fan49], and Huffman [Huf52], these schemes might not be efficient for large alphabets (in our case, the alphabet is $\{0,1\}^m$). So as far as we know, there might be a random variable X that is not compressible to less than $H(x)+2$ by an efficient prefix-free scheme. Since the existence of such a variable is unlikely to imply the existence of one-way functions, the 2 bits gap in our results might be unavoidable.*

One might do better by asking for incompressibility by arbitrary (no prefix-free) efficient schemes (which might even compress to strictly less than $H(X)$ bits). But bearing in mind that compression is used for communicating many samples from the distribution, asking for prefix-freeness seems like the natural definition for incompressibility.

1.2 Our Technique

We explain here the high-level approach for proving Lemma 1.1 (incompressibility $\rightarrow$ next-bit pseudoentropy). Let $X = (X_1, \ldots, X_m)$ be a random variable over $\{0,1\}^m$, and assume X does not have next-bit pseudoentropy $k-2$. We prove that such X can be (efficiently) compressed into

less than k bits, proving the lemma. Recall that next-bit pseudoentropy measures the bit-wise unpredictability of X: how hard is it to predict X_i from $X_{<i}$, for $i \leftarrow [m]$. So X not having $k-2$ bits of next-bit pseudoentropy implies that X is "rather predictable" in an online fashion: predict X_1, then use X_1 to predict X_2, and so on. We use this characterization to design a prefix-free bits encoding of X, of average length less than k, more details below.

Let $I \leftarrow [m]$. Since X does not have $k-2$ bits of next-bit pseudoentropy, the random variable $(X_{<I}, X_I)$ is distinguishable from $(X_{<I}, Y_I)$ for *every* set of random variable $\{Y_i\}_{i\in[m]}$ with $\sum_{i\in[m]} H(Y_i|X_{<i}) \geq (k-2)$. That is, X_I has low entropy given $X_{<I}$, in the eyes of poly-time distinguisher. As discussed above, this implies that X_I is somewhat predictable given $X_{<I}$. Vadhan and Zheng [VZ12] formalized this intuition and proved that for such an X, there exists a poly-time predictor P that predicts X_I from $X_{<I}$ within small KL-divergence. Specifically,

$$\mathrm{KL}(X_{<I}, X_I || X_{<I}, \mathsf{P}(X_{<I})) < (k-2)/m - H(X_I \mid X_{<I}) \tag{1}$$

Let $Y = (Y_1, \ldots, Y_m)$ be the random variable defined inductively by P as follows: $Y_1 = \mathsf{P}(\epsilon)$, and $Y_i = \mathsf{P}(Y_{<i})$. By Equation (1) and the chain-rules of KL-divergence, we deduce that

$$\mathrm{KL}(X||Y) < k - 2 - H(X) \tag{2}$$

The above suggest the following method for compressing X. Use P for designing a good prefix-free encoding scheme for Y, and then apply Equation (2) to deduce that the scheme also compresses X well. The scheme we design for Y is the *arithmetic encoding* scheme. This scheme is useful for compressing any distribution $\mathcal{D}$ for which we know how to compute the *accumulated probability function* $F^{\mathcal{D}}(x) := \sum_{x' \leq x} \Pr_{\mathcal{D}}[x']$. Since Y is defined according to P, it is not hard to see that the accumulated function of Y is efficiently computable, implying that the arithmetic encoding $(\mathsf{Enc}, \cdot)$ of Y can be computed efficiently. Furthermore, since $(\mathsf{Enc}, \cdot)$ is the arithmetic encoding of Y, it holds that

$$|\mathsf{Enc}(y)| \leq -\log(\Pr[Y = y]) + 2 \tag{3}$$

for every $y \in \mathrm{Supp}(Y)$. Using a well-known fact about using "wrong compression", we deduce that

$$\mathrm{E}[|\mathsf{Enc}(X)|] \leq H(X) + \mathrm{KL}(X||Y) + 2 \tag{4}$$

Applying Equation (2), we conclude that $\mathrm{E}[|\mathsf{Enc}(X)|] < k$.

Remark 1.6 (Uniform incompressibility)**.** *In the above proof we assumed that the predictor* P *is deterministic (which can be assumed without loss of generality in the non-uniform setting). This assumption was crucial for the arithmetic encoding to work. Otherwise, the encoder and decoder will not agree on the same accumulated probability function. In the uniform setting, this obstacle can be overcome by letting the encoder and decoder have access to* shared randomness *(independent of the distribution to compress) which they can use as the random coins of* P*. We measure the compression of such shared randomness scheme by the expected encoding length over the shared randomness and the underlying distribution. All the results stated in this paper extend, with essentially the same parameters, to this uniform setting.*

We note that, when compressing more than one sample from the distribution, the shared randomness needs to be sampled only once. Thus, the amortized cost of our compressing scheme is still equal to the entropy of the distribution, even if the decoder sends its randomness (instead of using shared randomness).

1.3 Related Work

Yao [Yao82] used the term *effective entropy* to measure by how much a distribution can be compressed efficiently. So k-incompressiblity is equivalent to having effective entropy at least k. Yao [Yao82] did not require the compression scheme to be prefix-free, but only required that for every $t \in \text{poly}$, the sequence $\mathsf{Enc}(x_1), \ldots, \mathsf{Enc}(x_t)$, where $x_1, \ldots, x_t$ are independent samples from the distribution, are decoded with high probability.

An interesting line of work considered efficient compression of of samplable distributions with (efficient) membership queries. Goldberg and Sipser [GS85] showed that any such distribution can be compressed by $\log n$ bits, and Trevisan, Vadhan, and Zuckerman [Tre+05] gave better schemes for flat distributions, and for distribution generated by log-space machines.

Paper Organization

Basic definitions and notations are given in Section 2. The formal definition of incompressibility and our results relating it to next-bit pseudoentropy, including some results that are not mentioned above, are given in Section 3.

2 Preliminaries

2.1 Notations

All logarithms are taken in base 2. We use calligraphic letters to denote sets and distributions, uppercase for random variables, and lowercase for values and functions. Let poly stand for the set of all polynomials. Let PPT stand for probabilistic poly-time, and n.u.-poly-time stand for non-uniform poly-time. An n.u.-poly-time algorithm A is equipped with a (fixed) poly-size advice string set $\{z_n\}_{n\in\mathbb{N}}$ (that we typically omit from the notation), and we let A_n stand for A equipped with the advice z_n (used for inputs of length n). Let neg stand for a negligible function. For a set $\mathcal{L} \subseteq \{0,1\}^*$, let $\mathcal{L}_n := \mathcal{L} \cap \{0,1\}^n$. A set $\mathcal{S}$ is *prefix free*, if for no $x_1 \neq x_2 \in \mathcal{S}$ it holds that x_1 is a prefix of x_2. Given a vector $v \in \Sigma^n$, let v_i denote its i^{th} entry, let $v_{<i} = (v_1, \ldots, v_{i-1})$ and $v_{\leq i} = (v_1, \ldots, v_i)$. Similarly, for a set $\mathcal{I} \subseteq [n]$, let $v_{\mathcal{I}}$ be the ordered sequence $(v_i)_{i\in\mathcal{I}}$.

2.2 Distributions and Random Variables

When unambiguous, we will naturally view a random variable as its marginal distribution. The support of a finite distribution $\mathcal{P}$ is defined by $\text{Supp}(\mathcal{P}) := \{x \colon \Pr_{\mathcal{P}}[x] > 0\}$. For a (discrete) distribution $\mathcal{P}$, let $x \leftarrow \mathcal{P}$ denote that x was sampled according to $\mathcal{P}$. Similarly, for a set $\mathcal{S}$, let $x \leftarrow \mathcal{S}$ denote that x is drawn uniformly from $\mathcal{S}$. For random variable ensemble $B = \{B_n\}_{n\in\mathbb{N}}$ and $t \colon \mathbb{N} \mapsto \mathbb{N}$, let $B^t = \left\{B_n^{t(n)}\right\}_{n\in\mathbb{N}}$ for $B_n^{t(n)}$ being $t(n)$ independent copies of B_n. For $m \in \mathbb{N}$, we use U_m to denote a uniform random variable over $\{0,1\}^m$ (that is independent from other random variables in consideration). We use the following standard definitions:

Definition 2.1 (Indistinguishability)**.** *Distribution ensembles $\mathcal{P} = \{\mathcal{P}_n\}_{n\in\mathbb{N}}$ and $\mathcal{Q} = \{\mathcal{Q}_n\}_{n\in\mathbb{N}}$ are n.u.-poly-time-indistinguishable, if*

$$\Pr_{x\leftarrow\mathcal{P}_n}[\mathsf{D}(x) = 1] - \Pr_{x\leftarrow\mathcal{Q}_n}[\mathsf{D}(x) = 1] \leq \text{neg}(n)$$

for any n.u.-poly-time algorithm D.

Definition 2.2 (Samplablity)**.** *A distribution ensemble $\mathcal{P} = \{\mathcal{P}_n\}$ is* samplable*, if there exists poly-time algorithm (sampler)* S *and poly-time computable function $m \in \text{poly}$, such that for every $n \in \mathbb{N}$, $\mathsf{S}(1^n; U_{m(n)})$ is distributed according to $\mathcal{P}_n$.*

2.2.1 Entropy and Distance Measures

The *Shannon entropy* of a distribution $\mathcal{P}$ is defined by $H(\mathcal{P}) = \sum_{p \in \text{Supp}(\mathcal{P})} \Pr_{\mathcal{P}}[p] \cdot \log \frac{1}{\Pr_{\mathcal{P}}[p]}$. The conditional entropy of a random variable A given B, is defined as $H(A|B) = \mathrm{E}_{b \leftarrow B}[H(A|_{B=b})]$. We will use the following known facts:

Fact 2.3 (Chain rule for Shannon entropy)**.** *For a random variable $A = (A_1, \ldots, A_n)$, it holds that $H(A_1, \ldots, A_n) = \sum_{i=1}^{n} H(A_i|A_{<i})$.*

The *KL-divergence* (also known as, *Kullback-Leibler* divergence, and *relative entropy*) between distributions $\mathcal{P}$ and $\mathcal{Q}$ is defined by

$$\mathrm{KL}(\mathcal{P}||\mathcal{Q}) = \sum_{a \in \text{Supp}(\mathcal{P})} \Pr_{\mathcal{P}}[a] \log(\frac{\Pr_{\mathcal{P}}[a]}{\Pr_{\mathcal{Q}}[a]}).$$

The KL-divergence also admits a chain rule.

Fact 2.4 (Chain rule for KL-divergence)**.** *For random variables $A = (A_1, \ldots, A_n)$ and $B = (B_1, \ldots, B_n)$, it holds that*

$$\mathrm{KL}(A||B) = \sum_{i \in [n]} \mathrm{E}_{a_1, \ldots, a_n \leftarrow A} \big[\mathrm{KL}(A_i|_{A_{<i}=a_{<i}} || B_i|_{B_{<i}=a_{<i}})\big].$$

Both Entropy and KL-divergence admit data processing inequalities.

Fact 2.5 (Data processing inequality)**.** *For every random variables A, B and function f, it holds that $\mathrm{KL}(f(A)||f(B)) \leq \mathrm{KL}(A||B)$ and $H(f(A)) \leq H(A)$.*

We will use the following observation that bounds the KL-divergence between two distributions using their maximal ratio over a set.

Proposition 2.6 (Bounding KL-divergence)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be finite distributions, and assume $\log \frac{\Pr_{\mathcal{X}}[s]}{\Pr_{\mathcal{Y}}[s]} \geq \alpha$ for every element s of a set $\mathcal{S}$. Then $\mathrm{KL}(\mathcal{X}||\mathcal{Y}) > (\alpha - \log e) \cdot \Pr_{\mathcal{X}}[\mathcal{S}]$.*

Proposition 2.6 is proved in Appendix A.

We will also use the following simple observation to bound the probability of an event in terms of Entropy.

Fact 2.7 (Bounding probability via entropy)**.** *Let C be a Boolean random variable. If $H(C) \geq d$ then $\Pr[C = 1] \geq (d/40)^2$.*

Fact 2.7 is proved in Appendix A.

2.3 Encoding and Compression

We start with the definition of encoding schemes.

Definition 2.8 (Encoding schemes)**.** *A pair of algorithms* $(\mathsf{Enc}, \mathsf{Dec})$ *is an* encoding *for a distribution* $\mathcal{D}$*, if for every* $x \in \mathrm{Supp}(\mathcal{D})$ *it holds that* $\mathsf{Dec}(\mathsf{Enc}(x)) = x$.[5] *The pair is an* encoding *for a distribution ensemble* $\mathcal{D} = \{\mathcal{D}_n\}_{n\in\mathbb{N}}$*, if* $(\mathsf{Enc}(1^n, \cdot), \mathsf{Dec}(1^n, \cdot))$ *is an encoding scheme for* $\mathcal{D}_n$ *for every* n*. The encoding is* fixed-length*, if* $|\mathsf{Enc}(1^n, x)| = \ell(n)$ *for every* n *and* $x \in \mathrm{Supp}(\mathcal{D}_n)$*, for some function* ℓ*, and is* prefix-free *if the set* $\{\mathsf{Enc}(1^n, x) \colon x \in \mathrm{Supp}(\mathcal{D}_n)\}$ *is prefix-free for every* n*.*

Definition 2.9 (Compressing a distribution)**.** *An encoding scheme* $(\mathsf{Enc}, \cdot)$ ℓ-compereses *a distribution ensemble* $\mathcal{D}$*, if* $\mathrm{E}_{x\leftarrow\mathcal{D}_n}[|\mathsf{Enc}(1^n, x)|] \leq \ell(n)$ *for every* $n \in \mathbb{N}$*.*

We will refer to an encoding scheme that compresses the distribution as a *compressing scheme*. When clear from the context, we omit the parameter 1^n given to the encoder and the decoder.

Changing distributions. The following well-known observation bounds the price you pay by using the compressing scheme for the "wrong" distribution.

Proposition 2.10 (Changing distributions)**.** *Let* $\mathcal{P}$ *and* $\mathcal{Q}$ *be finite distributions with* $\mathrm{KL}(\mathcal{P}||\mathcal{Q}) < \infty$*. Let* $(\mathsf{Enc}, \mathsf{Dec})$ *be a compression scheme for* $\mathcal{Q}$*, such that* $|\mathsf{Enc}(q)| \leq -\log(\Pr_{\mathcal{Q}}[q]) + c$ *for every* $q \in \mathrm{Supp}(\mathcal{Q})$ *for some* $c > 0$*. Then* $(\mathsf{Enc}, \mathsf{Dec})$ *is a compression scheme for* $\mathcal{P}$ *with* $\mathrm{E}_{p\leftarrow\mathcal{P}}[|\mathsf{Enc}(p)|] \leq H(\mathcal{P}) + \mathrm{KL}(\mathcal{P}||\mathcal{Q}) + c$*.*

Proposition 2.10 is proved in Appendix A

Arithmetic encoding We use *Arithmetic encoding*, a well-known prefix-free encoding scheme.

Definition 2.11 (Arithmetic encoding)**.** *Let* X *be a finite random variable over* $\mathcal{U}$ *and let* $\prec$ *be a total order over* $\mathcal{U}$*. Let* $F : \mathcal{U} \to [0, 1]$ *by* $F(x) = (\sum_{a\prec x} \Pr[X = a]) + 1/2 \cdot \Pr[X = x]$*. Define* $\mathsf{Enc}(x)$ *as the first* $(\lceil -\log \Pr[X = x] \rceil + 1)$ *bits of* $F(x)$*.*

Arithmetic encoding enjoys the following properties:

Fact 2.12 (Properties of arithmetic encoding, lemma 2.8 in [Tre+05])**.** *For every random variable* X *and order* $\prec$*, the function* Enc *defined in Definition 2.11 is one-to-one and monotone, and the scheme* $(\mathsf{Enc}, \mathsf{Enc}^{-1})$ *is prefix-free and compresses* X *to* $H(X) + 2$ *bits.*

2.4 One-way Functions

Definition 2.13 (One-way functions)**.** *A poly-time computable function* $f \colon \{0, 1\}^n \mapsto \{0, 1\}^n$ *is* n.u.-one-way*, if*

$$\Pr_{x\leftarrow\{0,1\}^n}\left[\mathsf{A}_n(f(x)) \in f^{-1}(f(x))\right] \leq \mathrm{neg}(n)$$

for any n.u.-poly-time A*.*

[5] Our results readily extend to encoding schemes with negligible decoding errors.

2.5 Pseudoentropy and Next-bit Pseudoentropy

In this section we define pseudoentropy and next-bit pseudoentropy, a special case of next-block pseudoentropy defined at Haitner et al. [Hai+13].

Pseudoentropy. We start with recalling the standard notion of pseudoentropy [Has+99].

Definition 2.14 (Pseudoentropy)**.** *A random variable ensemble B has* n.u.-pseudoentropy (at least) *k, if for every $p \in$ poly there exists an ensemble $C = \{C_n\}_{n\in\mathbb{N}}$, such that:*

1. *$H(C_n) \geq k(n)$, and*
2. *B and C are n.u.-poly-time-indistinguishable.*

We also use a conditional version of the above definition.

Definition 2.15 (Conditional pseudoentropy)**.** *Let $B = \{B_n\}_{n\in\mathbb{N}}$ be a random variable ensemble over $\{0,1\}$ jointly distributed with $X = \{X_n\}_{n\in\mathbb{N}}$. We say that B has* n.u.-conditional-pseudoentropy (at least) *k* given *X, if for every $p \in$ poly there exists an ensemble $C = \{C_n\}_{n\in\mathbb{N}}$ over $\{0,1\}$, jointly distributed with (X, B), such that:*

1. *$H(C_n|X_n) \geq k(n) - 1/p(n)$, [6] and*
2. *(X, B) and (X, C) are n.u.-poly-time-indistinguishable.*

Remark 2.16 (Order of quantifiers)**.** *The order of quantifiers in ours definition of conditional pseudoentropy, Definition 2.15, is different from the ones appearing in Vadhan and Zheng [VZ12]. Their definition requires that for every $p \in$ poly there exists an ensemble of random variable $\{C_n\}_{n\in\mathbb{N}}$ such that:*

1. *$H(C_n|X_n) \geq k(n) - 1/p(n)$, and*
2. *(X, B) and (X, C) are $1/p(n)$-indistinguishable for circuits of size $p(n)$.*

Clearly Definition 2.15 implies that of [VZ12], but it turns out that the converse also holds (making the definitions equivalent). Indeed, assuming that the definition of [VZ12] holds, we show that there exists an ensemble $C = \{C_n\}_{n\in\mathbb{N}}$ such that (X, C) is n.u.-poly-time-indistinguishable from (X, B) (satisfying our definition of conditional pseudoentropy). For every n, let C_n be a random variable that fulfills Items 1 and 2 in the definition of [VZ12] with respect to $p(n) = n^c$ for the largest possible value of $c \leq n$. It is not hard to see that $\{C_n\}_{n\in\mathbb{N}}$ satisfies Item 2 in Definition 2.15, and thus the definitions are indeed equivalent.

A similar difference exist between [VZ12]'s and our definitions of (non conditional) pseudoentropy. Also in this case it can be shown that the two definitions are equivalent.

[6] [Hai+13] do not have the $1/p(n)$ term in their definition of conditional pseudoentropy (which is implicit in their definition of next-bit pseudoentropy). Following [VZ12], we add this term to slightly simplify the text.

Next-bit pseudoentropy. We are now ready to define next-bit pseudoentropy. Intuitively, a random variable B over $\{0,1\}^m$ has next-bit pseudoentropy k, if for a uniformly chosen $i \in [m]$, it holds that B_i has pseudoentropy k/m given $B_{<i}$. Formally, this is put using the above notation of conditional pseudoentropy.

Definition 2.17 (Next-bit pseudoentropy)**.** *The random variables ensemble $B = \{B_n\}_{n\in\mathbb{N}}$ over $\{0,1\}^{m(n)}$ has* n.u.-next-bit-pseudoentropy (at least) *k if the following holds: let $I = \{I_n\}_{n\in\mathbb{N}}$ be an ensemble of uniformly distributed random variables over $[m(n)]$, then $\{(B_n)_{I_n}\}_{n\in\mathbb{N}}$ has n.u.-conditional-pseudoentropy k/m given $\{(B_n)_{<I_n}\}_{n\in\mathbb{N}}$.*

In their construction of pseudorandom generator, [Hai+13] has proved that a generator whose next-bit pseudoentropy is larger than its input length, can be used to construct pseudorandom generators and thus one-way functions. The following is the non-uniform variant of their result.

Theorem 2.18 (Extending next-bit pseudoentropy implies one-way functions, [Hai+13])**.** *Assume there is a poly-time computable function $f\colon \{0,1\}^n \to \{0,1\}^{m(n)}$ such that $\{f(U_n)\}_{n\in\mathbb{N}}$ has n.u.-next-bit-pseudoentropy $n + 1/p(n)$ for some $p \in$ poly. Then there exists a n.u.-one-way functions.*

For our needs, we use the following corollary of the above.

Corollary 2.19 (Non-trivial next-bit pseudoentropy implies one-way functions)**.** *Assume there is a poly-time computable function $f\colon \{0,1\}^n \to \{0,1\}^{m(n)}$ such that $\{f(U_n)\}_{n\in\mathbb{N}}$ has n.u.-next-bit-pseudoentropy $H(f(U_n)) + 1/p(n)$ for some $p \in$ poly. Then there exists a n.u.-one-way functions.*

That is, it is enough to show that f has next-bit pseudoentropy larger than its image real entropy, rather than its input size. A proof sketch for Corollary 2.19 is given in Appendix A.

It is easy to see that next-bit pseudoentropy behaves nicely under direct product. It turns out that the converse is also true: if the direct product has kt bits of next-bit pseudoentropy, then a single copy has next-bit pseudoentropy (at least) k.

Proposition 2.20 (Direct product of next-bit pseudoentropy)**.** *For any random variable ensemble B and $t \in$ poly, if B^t has n.u.-next-bit pseudoentropy $(t \cdot k)$, then B has n.u.-next-bit pseudoentropy k.*

Proposition 2.20 is proved in Appendix A.

2.5.1 KL-hardness

In their "hashing free" construction of pseudorandom generators from one-way functions, Vadhan and Zheng [VZ12] introduced the notion of *KL-hardness* of a distribution. Informally, it states that it is hard to approximate the distribution within a small KL divergence. This notion is formally defined using *KL-predictors*.

Definition 2.21 (KL-predictors)**.** *Let (X, B) be a distribution over $\{0,1\}^m \times \{0,1\}$, and let $\mathsf{P}\colon \{0,1\}^m \times \{0,1\} \mapsto (0, +\infty)$ be a deterministic function. We say that P is a δ-*KL-predictor *of B given X, if*

$$\mathrm{KL}(X, B || X, C_{\mathsf{P}}) \leq \delta,$$

for C_{P} being a random variable (jointly distributed with X) with $\Pr[C_{\mathsf{P}} = b \mid X = x] = \frac{\mathsf{P}(x,b)}{\mathsf{P}(x,0)+\mathsf{P}(x,1)}$.

A distribution is KL-hard if it possesses no efficient KL-predictor.

Definition 2.22 (KL-hardness)**.** *Let (X, B) be a distribution ensemble over $\{0,1\}^{m(n)} \times \{0,1\}$. We say that B is δ-n.u.-*KL-hard given *X, if there exists no n.u.-poly-time P and $q \in$ poly such that P_n is a $(\delta - 1/q(n))$-KL-predictor of $\mathcal{B}$ given $\mathcal{X}$, for infinitely many n's.*

We use the following result from [VZ12].

Theorem 2.23 (KL-hardness imply pseudoentropy, [VZ12] Corollary 3.9)**.** *Let $(X, B) = \{(X_n, B_n)\}_{n\in\mathbb{N}}$ be a random variable ensemble over $\{0,1\}^{m(n)} \times \{0,1\}$. If B is δ-n.u.-KL-hard given X, then it has n.u.-conditional-pseudoentropy $H(B_n|X_n) + \delta(n)$ given X.*

That is, KL-hard distribution has non-trivial next-bit pseudoentropy.[7]

3 Incompressibility and Next-bit Pseudoentropy

In this section, we define several notions of incompressibility and relate them to next-bit pseudoentropy (defined in Section 2.5). As said in the introduction, we focus on the non-uniform settings.

Incompressibility. We start with the standard notion of incompressibility that we define with respect to prefix-free compression schemes (it is immediate that a distribution that is incompressible with respect to arbitrary scheme is incompressible according to our definition).

Definition 3.1 (Incompressibility)**.** *A distribution ensemble $\mathcal{B}$ is k-*incompressible*, if for every n.u.-poly-time prefix-free compression scheme $(\mathsf{Enc}, \cdot)$ for $\mathcal{B}$, it holds that*

$$\mathrm{E}_{x\leftarrow\mathcal{B}_n}[|\mathsf{Enc}(x)|] \geq k(n),$$

for all but finitely many n's.

We will also address the following more fine-grain version of incompressibility.

Definition 3.2 (Local incompressibility)**.** *A distribution ensemble $\mathcal{B}$ is (α, β)-*locally-incompressible*, if for every n.u.-poly-time prefix-free compression scheme $(\mathsf{Enc}, \cdot)$ for $\mathcal{B}$, it holds that*

$$\Pr_{x\leftarrow\mathcal{B}_n}\left[|\mathsf{Enc}(x)| \geq \log\frac{1}{\Pr_{\mathcal{B}_n}[x]} + \alpha(n)\right] \geq \beta(n),$$

for all but finitely many n's.

Local incompressibility gets handy when the gap between next-bit pseudoentropy and the real entropy is smaller than 2, settings in which our result for (non-local) incompressibility is not applicable.

We observe the following connections between incompressiblity and local incompressiblity, both proved in Section 3.6.

Proposition 3.3 (Incompressiblity $\to$ local incompressiblity)**.** *d A k-incompressible distribution ensemble $\mathcal{B}$ over $\{0,1\}^{m(n)}$, with $m \in$ poly, is $(k(n) - H(\mathcal{B}_n) - 2, \frac{1}{3m(n)})$-locally-incompressible.*

Proposition 3.4 (local incompressiblity $\to$ incompressiblity)**.** *An (α, β)-locally incompressible distribution ensemble $\mathcal{B}$, is $H(\mathcal{B}_n) + \beta(n)(\alpha(n) - \log e)$-incompressible.*

[7] [VZ12] also proved that the converse direction holds.

Relation to pseudoentropy. We recall the following two facts. The first states that incompressibility is not stronger than pseudoentropy.

Theorem 3.5 (Pseudoentropy $\rightarrow$ incompressibility, [Wee04])**.** *Let $\mathcal{B}$ be a distributions ensemble over $\{0,1\}^{m(n)}$ with n.u.-pseudoentropy k, then $\mathcal{B}$ is $(k - 2\log m)$-incompressible.*[8]

While it is unknown if incompressibility is a weaker notion than pseudoentropy, there is an oracle separation between them, as stated in the next theorem.

Theorem 3.6 (Incompressibility $\not\rightarrow$ pseudoentropy, oracle separation, [Wee04])**.** *There is an oracle $\mathcal{O}$ and a distribution ensemble $\mathcal{B}$ that relative to $\mathcal{O}$, $\mathcal{B}$ is $(n - \omega(\log n))$-incompressible but does not have pseudoentropy larger than $n/2$.*

3.1 Our Results

3.1.1 Incompressibility $\rightarrow$ Next-Bit Pseudoentropy

Our main result states that incompressibility implies next-bit pseudoentropy.

Lemma 3.7 (Incompressibility $\rightarrow$ next-bit pseudoentropy)**.** *The following holds for every distribution ensemble $\mathcal{B}$ over $\{0,1\}^{m(n)}$ with $m \in$ poly.*

1. *$\mathcal{B}$ is k-incompressible $\implies$ $\mathcal{B}$ has n.u.-next-bit pseudoentropy $k(n) - 2$.*
2. *$\mathcal{B}$ is (α, β)-locally-incompressible $\implies$ $\mathcal{B}$ has n.u.-next-bit pseudoentropy $H(\mathcal{B}_n) + \beta(n)(\alpha(n) - 2 - \log e)$.*

Lemma 3.7 is proved in Section 3.3. Combining its first part with Corollary 2.19, yields the following informative theorem.

Theorem 3.8 (Incompressiblity $\rightarrow$ one-way functions)**.** *Assume there exists a samplable distribution ensemble $\mathcal{B}$ over $\{0,1\}^{m(n)}$ that is $(H(\mathcal{B}_n) + 2 + 1/p(n))$-incompressible for some $p \in$ poly, then n.u.-one-way functions exit.*

Amortization. When amortizing over several instances, Lemma 3.7 yields the following tighter characterization.

Lemma 3.9 (Incompressiblity $\rightarrow$ next-bit pseudoentropy, multiples copies)**.** *Let $\mathcal{B}$ be a distribution ensemble over $\{0,1\}^{m(n)}$ with $m \in$ poly, such that $\mathcal{B}^t$ for some $t \in$ poly, is $(t \cdot k)$-incompressible. Then $\mathcal{B}$ has n.u.-next-bit pseudoentropy $k(n) - 2/t(n)$.*

Proof. Since $\mathcal{B}^t$ has $t \cdot k$-incompressibility, by Lemma 3.7 it has n.u.-next-bit pseudoentropy $t \cdot k - 2$. Hence, Proposition 2.20 yields that $\mathcal{B}$ has n.u.-next-bit pseudoentropy $k - 2/t$. $\square$

[8] This result holds also for non-prefix free compressing schemes.

3.1.2 Strong-Next-Bit Pseudoentropy

It is easy to see that next-bit pseudoentropy does not imply incompressibility.

Proposition 3.10 (Next-bit pseudoentropy $\not\Rightarrow$ incompressibility)**.** *Assuming n.u.-one-way function exists, then there exists a samplable distribution ensemble with n.u.-next-bit pseudoentropy $2n$, that is not $(n+1)$-incompressible.*

Proof sketch. Let $g\colon \{0,1\}^n \mapsto \{0,1\}^{2n}$ be a pseudorandom generator. The distribution ensemble $\mathcal{B} = \{(g(U_n), U_n)\}_{n\in\mathbb{N}}$ has $2n$ n.u.-next-bit pseudoentropy. But $\mathcal{B}$ can be trivially compressed to n bits by $\mathsf{Enc}(f(x), x) = x$. □

In contrast, the following variant of next-bit pseudoentropy does imply (and is equivalent to) incompressibility.

Definition 3.11 (Strong-next-bit pseudoentropy)**.** *A random variable ensemble B has n.u.-strong-next-bit pseudoentropy k, if for every n.u.-poly-time, fixed-length encoding $(\mathsf{Enc}, \cdot)$, the ensemble $\{\mathsf{Enc}(B)\}_{n\in\mathbb{N}}$ has n.u.-next-bit pseudoentropy k.*[9]

That is, B has strong-next-bit pseudoentropy if every encoding of B has next-bit pseudoentropy. Lemma 3.7 easily extends to strong-next-bit pseudoentropy.

Lemma 3.12 (Incompressibility → strong-next-bit pseudoentropy)**.** *A k-incompressible distribution ensemble has n.u.-strong-next-bit pseudoentropy $k-2$.*

Proof. Let $\mathcal{B}$ be a k-incompressible distribution ensemble, and let B be a random variable ensemble distributed according to $\mathcal{B}$. It follows that for every n.u.-poly-time fixed-length encoding scheme $(\mathsf{Enc}, \cdot)$, it holds that $\mathsf{Enc}(B_n)$ is k-incompressible. Otherwise, one can efficiently compress $\mathcal{B}_n$ by first encode it according to Enc, and then compress the output. Thus, by Lemma 3.7, $\mathsf{Enc}(B_n)$ has *n.u.*-next-bit pseudoentropy at least $k-2$. □

More interestingly, strong-next-bit pseudoentropy does imply incompressibility.

Lemma 3.13 (Strong-next-bit pseudoentropy → incompressibility)**.** *Let $\mathcal{B}$ be a distribution ensemble with $\mathrm{Supp}(\mathcal{B}_n) = \{0,1\}^{m(n)}$. If $\mathcal{B}$ has n.u.-strong-next-bit pseudoentropy $k + 1/p$ for some $p \in \mathrm{poly}$, then $\mathcal{B}$ is k-incompressible.*

Without requiring that $\mathrm{Supp}(\mathcal{B}_n) = \{0,1\}^{m(n)}$, we would only get that $\mathcal{B}$ is incompressible by an encoding schemes that is prefix-free over $\{0,1\}^{m(n)}$. Interestingly, the proof of Lemma 3.12 readily yields that this type of incompressibility is sufficient for next-bit pseudoentropy. Lemma 3.13 is proved in Section 3.5.

3.2 Applications to Sparse Languages

We use the following definition of sparse language. (Recall that for a set $\mathcal{L} \subseteq \{0,1\}^*$, $\mathcal{L}_n := \mathcal{L} \cap \{0,1\}^n$.)

Definition 3.14. *A language $\mathcal{L} \in \{0,1\}^*$ is s-*sparse *if $|\mathcal{L}_n| \le 2^{s(n)}$ for every $n \in \mathbb{N}$.*

[9] Note that in this definition, Enc is not necessarily a compressing encoding.

The results of Section 3 immediately yield that unless one-way functions exist, any samplable distribution over s-sparse language can be compressed to $s+2$ bits.

Theorem 3.15. *For every samplable distribution ensemble $\mathcal{B} = \{\mathcal{B}_n\}_{n\in\mathbb{N}}$ and s-sparse language $\mathcal{L}$, such that $\mathrm{Supp}(\mathcal{B}_n) \subseteq \mathcal{L}_n$ for every $n \in \mathbb{N}$, if $\mathcal{B}$ is $(s+2+1/p)$-incompressible for some $p \in \mathrm{poly}$, then n.u.-one-way functions exist.*

Proof. Since $\mathcal{B}$ is over $\mathcal{L}$, $s(n) \geq H(\mathcal{B}_n)$, and thus $\mathcal{B}$ is $(H(\mathcal{B}_n)+2+1/p)$-incompressible. Thus by Theorem 3.8, n.u.-one-way functions exist. □

3.3 Proving Lemma 3.7—Incompressibility implies Next-Bit Pseudoentropy

In this part we prove Lemma 3.7. We use the following lemma, proved in Section 3.4.

Lemma 3.16 (Next-bit predictor to compression)**.** *There exists a pair of oracle-aided algorithms $(\mathsf{Enc}, \mathsf{Dec})$ such that the following holds: let $\mathsf{P}\colon \{0,1\}^* \to (0,+\infty)$ be deterministic algorithm, and for $m \in \mathbb{N}$ let $\mathcal{D}^{\mathsf{P}}_m$ be the distribution over $\{0,1\}^m$ defined by:*

$$\Pr_{\mathcal{D}^{\mathsf{P}}_m}[x] = \prod_{i\in[m]} \frac{\mathsf{P}(x_{<i}, x_i)}{\mathsf{P}(x_{<i},0) + \mathsf{P}(x_{<i},1)}$$

Then $(\mathsf{Enc}^{\mathsf{P}}(1^m,\cdot), \mathsf{Dec}^{\mathsf{P}}(1^m,\cdot))$ is a prefix-free compressing scheme for $\mathcal{D}^{\mathsf{P}}_m$ with $|\mathsf{Enc}(x)| \leq \lceil -\log \Pr_{\mathcal{D}^{\mathsf{P}}}[x]\rceil + 1$ for every $x \in \{0,1\}^m$. The running-time of $\mathsf{Enc}^{\mathsf{P}}(1^m,)$ and $\mathsf{Dec}^{\mathsf{P}}(1^m,\cdot)$ is polynomial in m and in the output length of P on inputs of length at most m.

Given Lemma 3.16, we are ready to prove Lemma 3.7.

Proof of Lemma 3.7. Let $\mathcal{B}$ be as in Lemma 3.7, and assume it does not have next-bit pseudoentropy q (we will chose q later). We start by proving that there exists n.u.-poly-time algorithm P such that the distribution ensemble $\left\{\mathcal{D}^{\mathsf{P}}_{m(n)}\right\}_{n\in\mathbb{N}}$ is close in KL-divergence to $\mathcal{B}$, for $\mathcal{D}^{\mathsf{P}}_m$ being according to Lemma 3.16.

Let I_n be a uniform random variable over $[m(n)]$, and let B be an ensemble of random variables distributed according to $\mathcal{B}$. By assumption and Definition 2.17, $\{(B_n)_{I_n}\}_{n\in\mathbb{N}}$ has no q/m conditional pseudoentropy given $\{(B_n)_{<I_n}\}_{n\in\mathbb{N}}$. Thus, Theorem 2.23 implies that $\{(B_n)_{I_n}\}_{n\in\mathbb{N}}$ is not $\delta(n) := \big(q/m - H\big((B_n)_{I_n} \mid (B_n)_{<I_n}\big)\big)$ KL-hard given $\{(B_n)_{<I_n}\}_{n\in\mathbb{N}}$. Namely, (see, Definition 2.22) there exists an infinite set $\mathcal{I} \subseteq \mathbb{N}$, $c>0$ and a n.u.-poly-time algorithm $\mathsf{P}\colon \{0,1\}^* \to (0,+\infty)$, such that P_n is a $(\delta - 1/n^c)$-KL-predictor of $\{(B_n)_{I_n}\}_{n\in\mathbb{N}}$ given $\{(B_n)_{<I_n}\}_{n\in\mathbb{N}}$, for every $n \in \mathcal{I}$. Fix $n \in \mathcal{I}$, and omit n from the notation, and let D^{P}_m be a random variable distributed according to $\mathcal{D}^{\mathsf{P}}_m$. By definition of KL-predictor (Definition 2.21), it holds that:

$$\mathrm{KL}(B_{<I}, B_I || B_{<I}, \mathsf{P}_n(B_{<I})) \leq \delta - 1/n^c < \delta = q/m - H(B_I \mid B_{<I}) \tag{5}$$

It follows that

$$\begin{aligned}
q/m &> \mathrm{E}_{i\leftarrow I}[\mathrm{KL}(B_{<i}, B_i||B_{<i}, \mathsf{P}(B_{<i})) + H(B_i \mid B_{<i})] \\
&= 1/m \cdot \sum_{i\in[m]} (\mathrm{KL}(B_{<i}, B_i||B_{<i}, \mathsf{P}(B_{<i})) + H(B_i \mid B_{<i})) \\
&= 1/m \cdot \sum_{i\in[m]} (\mathrm{KL}(B_{<i}||B_{<i}) + \mathrm{E}_{b\leftarrow B}\left[\mathrm{KL}(B_i|_{B_{<i}=b_{<i}}||\mathsf{P}(B_{<i})|_{B_{<i}=b_{<i}})\right] + H(B_i \mid B_{<i})) \\
&= 1/m \cdot \sum_{i\in[m]} (\mathrm{E}_{b\leftarrow B}\left[\mathrm{KL}(B_i|_{B_{<i}=b_{<i}}||(D_m^{\mathsf{P}})_i|_{(D_m^{\mathsf{P}})_{<i}=b_{<i}})\right] + H(B_i \mid B_{<i})) \\
&= 1/m \cdot (\mathrm{KL}(B||D_m^{\mathsf{P}}) + H(B)).
\end{aligned}$$

The second equality is due to chain-rule of KL-divergence. The third equality by the definition of $\mathcal{D}_m^{\mathsf{P}}$ and since $\mathrm{KL}(X||X) = 0$ for every random variable X. The last equality holds by chain-rule of KL-divergence and Shannon entropy. We deduce that

$$\mathrm{KL}(B||D_m^{\mathsf{P}}) < q - H(B) \tag{6}$$

Let $(\mathsf{Enc}, \mathsf{Dec})$ be the compressing scheme guaranteed by Lemma 3.16. Lemma 3.16 implies that

$$\left|\mathsf{Enc}^{\mathsf{P}}(x)\right| \leq \left\lceil -\log \Pr_{\mathcal{D}_m^{\mathsf{P}}}[x] \right\rceil + 1 \leq -\log \Pr_{\mathcal{D}_m^{\mathsf{P}}}[x] + 2 \tag{7}$$

for every $x \in \{0,1\}^m$. Given the above, we separately prove each part of the lemma.

$\mathcal{B}$ is k-incompressible. Let $q(n) = k(n) - 2$. By Proposition 2.10 and Equations (6) and (7),

$$\mathrm{E}_{x\leftarrow \mathcal{B}_n}\left[\left|\mathsf{Enc}^{\mathsf{P}}(x)\right|\right] < q(n) + 2 = k(n)$$

We conclude that $\mathcal{B}$ is k-compressible, yielding a contradiction.

$\mathcal{B}$ is (α, β)-locally-incompressible . By Equation (7),

$$\left|\mathsf{Enc}^{\mathsf{P}}(x)\right| \leq -\log \Pr_{\mathcal{D}_m^{\mathsf{P}}}[x] + 2 \tag{8}$$

Let $\mathcal{S} = \left\{x \in \{0,1\}^m \colon \left|\mathsf{Enc}^{\mathsf{P}}(x)\right| \geq -\log \Pr_{\mathcal{B}}[x] + \alpha\right\}$ and let $\eta = \Pr_{\mathcal{B}}[\mathcal{S}]$. Equation (8) yields that $-\log \Pr_{\mathcal{B}}[x] + \alpha \leq -\log \Pr_{\mathcal{D}_m^{\mathsf{P}}}[x] + 2$ for every $x \in \mathcal{S}$, implying that $\alpha - 2 \leq \log \frac{\Pr_{\mathcal{B}}[x]}{\Pr_{\mathcal{D}_m^{\mathsf{P}}}[x]}$ for every $x \in \mathcal{S}$. Applying Proposition 2.6 with respect to $\mathcal{S}$, yields that

$$\mathrm{KL}(B||D_m^{\mathsf{P}}) > \eta \cdot (\alpha - 2 - \log e)$$

Applying Equation (6) for $q = H(B) + \beta \cdot (\alpha - 2 - \log e)$, yields that

$$\mathrm{KL}(B||D_m^{\mathsf{P}}) < \beta \cdot (\alpha - 2 - \log e)$$

We deduce that that $\beta > \eta = \Pr_{\mathcal{B}}[\mathcal{S}]$, yielding that $\mathcal{B}$ is not (α, β)-locally incompressible.

□

3.4 Proving Lemma 3.16—Next-bit Predictor to Compression

Proof. Let P, m and $\mathcal{D} = \mathcal{D}_m^{\mathsf{P}}$ be according to the lemma statement. Our encoder defined below, encodes $\mathcal{D}$ according to the arithmetic encoding, see Definition 2.11, with respect to the lexicographic order. Recall that on input x, the arithmetic encoding should output $e(x)$: the first $(\lceil \log 1/\Pr_{\mathcal{D}}[x] \rceil + 1)$ bits of $F(x) := (\sum_{y<x} \Pr_{\mathcal{D}}[y]) + 1/2 \cdot \Pr_{\mathcal{D}}[x]$.

Algorithm 3.17 (Enc)**.**

Oracle: Predictor P.

Input: $x \in \{0,1\}^m$.

Operation:

1. *Let* $p_{\mathsf{eq}} = 1$ *and* $p_{\mathsf{less}} = 0$.
2. *For every* $i \in [m]$*:*
   (a) *If* $x_i = 1$: $p_{\mathsf{less}} = p_{\mathsf{less}} + p_{\mathsf{eq}} \cdot \frac{\mathsf{P}(x_{<i},0)}{\mathsf{P}(x_{<i},0)+\mathsf{P}(x_{<i},1)}$.
   (b) $p_{\mathsf{eq}} = p_{\mathsf{eq}} \cdot \frac{\mathsf{P}(x_{<i},x_i)}{\mathsf{P}(x_{<i},0)+\mathsf{P}(x_{<i},1)}$.
3. *Output the first* $(\lceil -\log p_{\mathsf{eq}} \rceil + 1)$ *bits of* $p_{\mathsf{less}} + p_{\mathsf{eq}}/2$.

By induction, at the end of the i^{th} iteration of Enc it holds that $p_{\mathsf{less}} = \Pr_{y\leftarrow\mathcal{D}}[y_{\le i} < x_{\le i}]$, and $p_{\mathsf{eq}} = \Pr_{y\leftarrow\mathcal{D}}[y_{\le i} = x_{\le i}]$. Hence, when Enc reaches Step 3., it holds that $p_{\mathsf{less}} = \Pr_{y\leftarrow\mathcal{D}}[y < x]$, and $p_{\mathsf{eq}} = \Pr_{\mathcal{D}}[x]$, stipulating that $\mathsf{Enc}^{\mathsf{P}}(x) = e(x)$. Thus by Fact 2.12, we deduce that $\mathsf{Enc}^{\mathsf{P}}$ is a prefix-free compressing scheme for $\mathcal{D}$ with $|\mathsf{Enc}(x)| = \lceil \log 1/\Pr_{\mathcal{D}}[x] \rceil + 1$, for every $x \in \{0,1\}^m$.

Regarding efficiency, since P only outputs positive numbers, the running time of $\mathsf{Enc}^{\mathsf{P}}$ is polynomial time in m and output size of P on inputs of length at most m. In addition, a decoding procedure $\mathsf{Dec}^{\mathsf{P}}(1^m, \cdot)$ for $\mathsf{Enc}^{\mathsf{P}}(1^m, \cdot)$ can be implemented with the same efficiently using a straightforward binary search over $\{0,1\}^m$. □

3.5 Proving Lemma 3.13 — Strong-Next-Bit Pseudoentropy implies Incompressibility

Proof. Assume that $\mathcal{B}$ is not k-incompressible and let $(\mathsf{Enc}, \mathsf{Dec})$ be the n.u.-poly-time compressing scheme that k'-compresses $\mathcal{B}$, with $k'(n) < k(n)$ for infinite many n's. Let $q(n)$ for $q \in$ poly be a bound on the output length of Enc on inputs of length $m(n)$, and let $\mathsf{Enc}'(1^n, x) = (\mathsf{Enc}(x), 0^{q(n)-|\mathsf{Enc}(x)|})$.

Let B be a random variable ensemble distributed according to $\mathcal{B}$. We claim that $\mathsf{Enc}'(1^n, B_n)$ does not have next-bit pseudoentropy $k(n) + 1/p(n)$. Indeed, consider the following distinguisher $\mathsf{D} = \{\mathsf{D}_n\}_{n\in\mathbb{N}}$:

Algorithm 3.18 (D_n)**.**

Input: $y \in \{0,1\}^*$, $b \in \{0,1\}$.

Operation:

1. *If there is no* $i \le |y|$ *such that* $\mathsf{Dec}(y_{\le i}) \in \{0,1\}^{m(n)}$ *and* $\mathsf{Enc}(\mathsf{Dec}(y_{\le i})) = y_{\le i}$, *output* 0.

2. Otherwise, output b.

We now show that D contradicts the $k(n)+1/p(n)$-next-bit pseudoentropy of $\{\mathsf{Enc}'(B_n)\}_{n\in\mathbb{N}}$ (Definition 2.17). It is easy to see that if (y,b) is a prefix of $\mathsf{Enc}'(x)$ for some $x\in\{0,1\}^{m(n)}$, then D_n outputs 0. Hence, we conclude the proof by showing that D_n outputs 1 with noticeable probability over $(\mathsf{Enc}'(B_n)_{<I_n}, C_n)$, for $I_n \leftarrow [q(n)]$ and *any* random variable C_n with

$$H(C_n \mid \mathsf{Enc}'(B_n)_{<I_n}) \geq \frac{k(n)+1/p(n)}{q(n)} - \frac{1}{2p(n)q(n)} = \frac{k(n)}{q(n)} + \frac{1}{2p(n)q(n)}.$$

Indeed, fix $n\in\mathbb{N}$ with $k'(n)<k(n)$, and let C_n be such random variable and let $\delta(n) := \frac{1}{2p(n)q(n)} \geq 1/\mathrm{poly}(n)$. Let W be the indicator for the event $I_n > |\mathsf{Enc}(B_n)|$ (that is, $W=1$ if $I_n > |\mathsf{Enc}(B_n)|$ and $W=0$ otherwise). Compute,

$$\begin{aligned}
&k(n)/q(n)+\delta(n)\\
&\leq H(C_n \mid \mathsf{Enc}'(B_n)_{<I_n})\\
&= H(C_n \mid \mathsf{Enc}'(B_n)_{<I_n}, W)\\
&= \Pr[W=1]\cdot H\big(C_n \mid \mathsf{Enc}'(B_n)_{<I_n}, W=1\big) + \Pr[W=0]\cdot H\big(C_n \mid \mathsf{Enc}'(B_n)_{<I_n}, W=0\big)\\
&\leq \Pr[W=1]\cdot H\big(C_n \mid \mathsf{Enc}'(B_n)_{<I_n}, W=1\big) + \Pr[W=0]\\
&= \Pr[I_n > |\mathsf{Enc}(B_n)|]\cdot H\big(C_n \mid \mathsf{Enc}'(B_n)_{<I_n}, I > |\mathsf{Enc}(B_n)|\big) + \Pr[I_n \leq |\mathsf{Enc}(B_n)|]\\
&\leq \Pr[I_n > |\mathsf{Enc}(B_n)|]\cdot H\big(C_n \mid \mathsf{Enc}'(B_n)_{<I_n}, I > |\mathsf{Enc}(B_n)|\big) + k(n)/q(n).
\end{aligned}$$

The first equality holds since $\mathsf{Enc}'(B_n)_{<I_n}$ determines the value of W. The second inequality holds since $H(C_n)\leq 1$. It follows that

$$\Pr[I_n > |\mathsf{Enc}(B_n)|]\cdot H\big(C_n \mid \mathsf{Enc}'(B_n)_{<I_n}, I > |\mathsf{Enc}(B_n)|\big) \geq \delta(n) \tag{9}$$

In particular,

$$\Pr[I_n > |\mathsf{Enc}(B_n)|] \geq \delta(n) \tag{10}$$

and

$$H(C_n \mid I_n > |\mathsf{Enc}(B_n)|) \geq H\big(C_n \mid \mathsf{Enc}'(B_n)_{<I_n}, I_n > |\mathsf{Enc}(B_n)|\big) \geq \delta(n) \tag{11}$$

Hence, Fact 2.7 yields that

$$\Pr[C_n = 1 \mid I_n > |\mathsf{Enc}(B_n)|] \geq (\delta(n)/40)^2 \tag{12}$$

Since D_n outputs 1 when $C_n=1$ and $I_n > |\mathsf{Enc}(B_n)|$, we deduce that

$$\Pr\big[\mathsf{D}_n(\mathsf{Enc}'(B_n)_{<I_n}, C_n) = 1 \mid I_n > |\mathsf{Enc}(B_n)|\big] \geq (\delta(n)/40)^2 \tag{13}$$

Combining the above with Equation (10), yields that,

$$\begin{aligned}
&\Pr\big[\mathsf{D}_n(\mathsf{Enc}'(B_n)_{<I_n}, C_n) = 1\big]\\
&= \Pr\big[\mathsf{D}_n(\mathsf{Enc}'(B_n)_{<I_n}, C_n) = 1 \mid I_n > |\mathsf{Enc}(B_n)|\big]\cdot \Pr[I_n > |\mathsf{Enc}(B_n)|]\\
&\geq (\delta(n)/40)^3.
\end{aligned}$$

It follows that D_n distinguishes C_n from $\mathsf{Enc}'(B_n)_{I_n}$ given $\mathsf{Enc}'(B_n)_{<I_n}$ with probability $(\delta(n)/40)^3 \geq 1/\mathrm{poly}(n)$. □

3.6 Additional Missing Proofs

Proposition 3.3.

Proof of Proposition 3.3. Assume toward contradiction that $\mathcal{B}$ is not (α, β)-locally-incompressible, for $\alpha(n) = k(n) - H(\mathcal{B}_n) - 2$ and $\beta = 1/3m(n)$.

Let $(\mathsf{Enc}, \cdot)$ be a compression scheme that violates the local-incompressibility of $\mathcal{B}$, and consider the following encoder Enc':

$$\mathsf{Enc}'(1^n, x) = \begin{cases} 0, \mathsf{Enc}(x) & |\mathsf{Enc}(x)| \leq m(n) \\ 1, x & \text{o/w.} \end{cases}$$

It follows that $\left|\mathsf{Enc}'(1^n, x)\right| \leq m(n) + 1$ for every $x \in \{0,1\}^{m(n)}$, and by (local) compressibility

$$\Pr_{x \leftarrow \mathcal{B}_n}\left[\left|\mathsf{Enc}'(1^n, x)\right| \geq \log \frac{1}{\Pr_{\mathcal{B}_n}[x]} + \alpha(n) + 1\right] < \beta(n) \tag{14}$$

It follows that,

$$\begin{aligned} \mathrm{E}_{x \leftarrow \mathcal{B}_n}\left[\left|\mathsf{Enc}'(1^n, x)\right|\right] &\leq \mathrm{E}_{x \leftarrow \mathcal{B}_n}[-\log \Pr_{\mathcal{B}_n}[x] + \alpha(n) + 1] + \beta(n)(m(n)+1) \\ &\leq H(\mathcal{B}_n) + \alpha(n) + 1 + \beta(n)(m(n)+1) \\ &= k(n) - 1 + \beta(n)(m(n)+1). \\ &< k(n) \end{aligned} \tag{15}$$

The first equation holds by Equation (14), the equality holds by our choice of α and the last inequality follows by our choice of β. This conclude that proof since by Equation (15), $\mathcal{B}_n$ is not k-incompressible. □

Proposition 3.4.

Proof of Proposition 3.4. Let $(\mathsf{Enc}, \cdot)$ be a prefix-free compression scheme for $\mathcal{B}$, and let $\mathcal{D} = \{\mathcal{D}_n\}_{n \in \mathbb{N}}$ be the distribution ensemble over $\mathrm{Supp}(\mathcal{B}_n) \cup \{\perp\}$, defined by $\Pr_{\mathcal{D}_n}[x] = 2^{-|\mathsf{Enc}(x)|}$ for $x \in \mathrm{Supp}(\mathcal{B}_n)$, $\Pr_{\mathcal{D}_n}[\perp] = 1 - \sum_{x \in \mathrm{Supp}(\mathcal{B}_n)} 2^{-|\mathsf{Enc}(x)|}$.

Since $\mathcal{B}$ is locally-incompressible, it holds that $\Pr_{x \leftarrow \mathcal{B}_n}\left[\log \frac{\Pr_{\mathcal{B}_n}[x]}{\Pr_{\mathcal{D}_n}[x]} \geq \alpha(n)\right] \geq \beta(n)$. Thus, Proposition 2.6 yields that $\mathrm{KL}(\mathcal{B}_n || \mathcal{D}_n) \geq \beta(n)(\alpha(n) - \log e)$, and therefore

$$\begin{aligned} \mathrm{E}_{x \leftarrow \mathcal{B}_n}[|\mathsf{Enc}(x)|] &= \mathrm{E}_{x \leftarrow \mathcal{B}_n}[-\log \Pr_{\mathcal{D}_n}[x]] \\ &= \mathrm{E}_{x \leftarrow \mathcal{B}_n}\left[-\log \Pr_{\mathcal{B}_n}[x] + \log \frac{\Pr_{\mathcal{B}_n}[x]}{\Pr_{\mathcal{D}_n}[x]}\right] \\ &= H(\mathcal{B}_n) + \mathrm{KL}(\mathcal{B}_n || \mathcal{D}_n) \\ &\geq H(\mathcal{B}_n) + \beta(n)(\alpha(n) - \log e). \end{aligned}$$

□

Acknowledgments

We thank Geoffroy Couteau, Ronen Shaltiel and Ofer Shayevitz for many useful discussions.

References

[Bar+03] Boaz Barak, Ronen Shaltiel, and Avi Wigderson. "Computational analogues of entropy". In: *Approximation, Randomization, and Combinatorial Optimization.. Algorithms and Techniques (APPROX)*. 2003, pp. 200–215 (cit. on pp. 1–3).

[BM82] Manuel Blum and Silvio Micali. "How to Generate Cryptographically Strong Sequences of Pseudo Random Bits". In: *Annual Symposium on Foundations of Computer Science (FOCS)*. 1982, pp. 112–117 (cit. on p. 1).

[Fan49] Robert M Fano. *The transmission of information*. Massachusetts Institute of Technology, Research Laboratory of Electronics, 1949 (cit. on p. 3).

[GM84] Shafi Goldwasser and Silvio Micali. "Probabilistic Encryption". In: *Journal of Computer and System Sciences* (1984), pp. 270–299 (cit. on p. 1).

[GS85] Andrew Goldberg and Michael Sipser. "Compression and ranking". In: *Annual ACM Symposium on Theory of Computing (STOC)*. 1985, pp. 440–448 (cit. on p. 5).

[Hai+13] Iftach Haitner, Omer Reingold, and Salil Vadhan. "Efficiency Improvements in Constructing Pseudorandom Generators from One-Way Functions". In: *SIAM Journal on Computing* 42.3 (2013), pp. 1405–1430 (cit. on pp. 1, 2, 8, 9).

[Hai+19] Iftach Haitner, Omer Reingold, Salil Vadhan, and Hoeteck Wee. *Inaccessible Entropy I: Inaccessible Entropy Generators and Statistically Hiding Commitments from One-Way Functions*. Tech. rep. 2010.05586. Preliminary version in STOC '09. arXiv, 2019 (cit. on p. 1).

[Hai+20] Iftach Haitner, Thomas Holenstein, Omer Reingold, Salil P. Vadhan, and Hoeteck Wee. "Inaccessible Entropy II: IE Functions and Universal One-Way Hashing". In: *Theory of Computing* (2020). Preliminary version in Eurocrypt '10 (cit. on p. 1).

[Has+99] Johan Hastad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. "A pseudorandom generator from any one-way function". In: *SIAM Journal on Computing* (1999), pp. 1364–1396 (cit. on pp. 1, 8).

[Hsi+07] Chun-Yuan Hsiao, Chi-Jen Lu, and Leonid Reyzin. "Conditional computational entropy, or toward separating pseudoentropy from compressibility". In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques (EUROCRYPT)*. 2007, pp. 169–186 (cit. on p. 1).

[Huf52] David A Huffman. "A method for the construction of minimum-redundancy codes". In: *Proceedings of the IRE* 40.9 (1952), pp. 1098–1101 (cit. on p. 3).

[IL89] Russell Impagliazzo and Michael Luby. "One-way Functions are Essential for Complexity Based Cryptography". In: *Annual Symposium on Foundations of Computer Science (FOCS)*. 1989, pp. 230–235 (cit. on p. 3).

[Imp95] Russell Impagliazzo. "A Personal View of Average-Case Complexity." In: *Proceedings of the Tenth Annual Structure in Complexity Theory Conference*. IEEE Computer Society, 1995, pp. 134–147 (cit. on p. 2).

[Sha48] Claude Elwood Shannon. "A mathematical theory of communication". In: *The Bell system technical journal* 27.3 (1948), pp. 379–423 (cit. on p. 3).

[Sha49] Claude Shannon. "Communication Theory of Secrecy Systems". In: *Bell System Technical Journal* (1949), pp. 656–715 (cit. on p. 1).

[SV11] Wojciech Szpankowski and Sergio Verdú. "Minimum expected length of fixed-to-variable lossless compression without prefix constraints". In: *IEEE Transactions on Information Theory* 57.7 (2011), pp. 4017–4025 (cit. on p. 1).

[Tre+05] Luca Trevisan, Salil Vadhan, and David Zuckerman. "Compression of samplable sources". In: *Computational Complexity* 14.3 (2005), pp. 186–227 (cit. on pp. 5, 7).

[VZ12] Salil Vadhan and Colin Jia Zheng. "Characterizing pseudoentropy and simplifying pseudorandom generator constructions". In: *Annual ACM Symposium on Theory of Computing (STOC)*. 2012, pp. 817–836 (cit. on pp. 1, 2, 4, 8–10).

[Wee04] Hoeteck Wee. "On pseudoentropy versus compressibility". In: *Annual IEEE Conference on Computational Complexity (COMPLEXITY)*. 2004, pp. 29–41 (cit. on pp. 1, 2, 11).

[Yao82] Andrew C. Yao. "Theory and Applications of Trapdoor Functions". In: *Annual Symposium on Foundations of Computer Science (FOCS)*. 1982, pp. 80–91 (cit. on pp. 1, 5).

A Missing Proofs

We will make use of the following Fact:

Fact A.1. $(1-p)\log(1-p) \geq -p\log e$ *for* $p \in [0,1)$.

Proof. Let $f(p) := (1-p)\log(1-p) + p\log e$, and note that $f(0) = 0$ and $\lim_{p\to 1} f(p) = \log e$. Moreover,

$$f'(p) = -\log(1-p) - (1-p)\cdot\frac{1}{(1-p)\ln 2} + \log e = -\log(1-p).$$

Thus, the only extreme point in this interval is 0, implying that the inequality holds. □

A.1 Fact 2.7

Fact A.2 (Bound on the probability via entropy, restatement of Fact 2.7)**.** *Let C be a Boolean random variable. If $H(C) \geq d$ then* $\Pr[C=1] \geq (d/40)^2$.

Proof of Fact 2.7. Let $\Pr[C=1] = p$. For every $p \leq 1/100$, by definition,

$$\begin{aligned} H(C) &= p\log(1/p) + (1-p)\log(1/(1-p)) \\ &\leq p\log(1/p) + p\log e \\ &\leq 4\sqrt{p} \end{aligned}$$

Where the penultimate inequality follows by Fact A.1. Thus it follows that,

$$\Pr[C=1] \geq \max\{1/100, (H(C)/4)^2\} \geq d^2/(4^2\cdot 100) \geq (d/40)^2.$$

□

A.2 Proposition 2.6

Proposition A.3 (Bounding KL-divergence, restatement of Proposition 2.6)**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be finite distributions, and assume $\log \frac{\Pr_{\mathcal{X}}[s]}{\Pr_{\mathcal{Y}}[s]} \geq \alpha$ for every element s of a set $\mathcal{S}$. Then $\mathrm{KL}(\mathcal{X}||\mathcal{Y}) > (\alpha - \log e) \cdot \Pr_{\mathcal{X}}[\mathcal{S}]$.*

Proof. Let $\widehat{X}$ be the indicator for $\mathcal{X} \in \mathcal{S}$ and define $\widehat{Y}$ similarly with respect to $\mathcal{Y}$. Let $p = \Pr[\mathcal{X} \in \mathcal{S}] = \Pr\left[\widehat{X} = 1\right]$ and let $q = \Pr[\mathcal{Y} \in \mathcal{S}] = \Pr\left[\widehat{Y} = 1\right]$. Notice that by definition of $\mathcal{S}$, it holds that $p \geq 2^{\alpha} q$. Therefore,

$$\begin{aligned}\mathrm{KL}(\mathcal{X}||\mathcal{Y}) \geq \mathrm{KL}(\widehat{X}||\widehat{Y}) &= p \log \frac{p}{q} + (1-p) \log \frac{1-p}{1-q} \\ &> p\alpha + (1-p)\log(1-p) \geq p(\alpha - \log e)\end{aligned}$$

The first inequality is by data-processing of KL-divergence. The last inequality holds by Fact A.1 since $(1-p)\log(1-p) \geq -p \log e$ for every $p \in [0,1)$. □

A.3 Proposition 2.10

Proposition A.4 (Changing distributions, restatement of Proposition 2.10)**.** *Let $\mathcal{P}$ and $\mathcal{Q}$ be finite distributions with $\mathrm{KL}(\mathcal{P}||\mathcal{Q}) < \infty$. Let $(\mathsf{Enc}, \mathsf{Dec})$ be a compression scheme for $\mathcal{Q}$, such that $|\mathsf{Enc}(q)| \leq -\log(\Pr_{\mathcal{Q}}[q]) + c$ for every $q \in \mathrm{Supp}(\mathcal{Q})$ for some $c > 0$. Then $(\mathsf{Enc}, \mathsf{Dec})$ is a compression scheme for $\mathcal{P}$ with $\mathrm{E}_{p \leftarrow \mathcal{P}}[|\mathsf{Enc}(p)|] \leq H(\mathcal{P}) + \mathrm{KL}(\mathcal{P}||\mathcal{Q}) + c$.*

Proof. Let $\mathcal{P}, \mathcal{Q}$ and $(\mathsf{Enc}, \mathsf{Dec})$ be as above. Since $\mathrm{KL}(\mathcal{P}||\mathcal{Q}) < \infty$, it holds that $\mathrm{Supp}(\mathcal{P}) \subseteq \mathrm{Supp}(\mathcal{Q})$, and thus it is clear that $(\mathsf{Enc}, \mathsf{Dec})$ is a compression scheme for $\mathcal{P}$.

$$\begin{aligned}\mathrm{E}_{p \leftarrow \mathsf{P}}[|\mathsf{Enc}(p)|] &\leq \sum_{p \in \mathrm{Supp}(\mathcal{P})} \Pr_{\mathcal{P}}[p] \cdot (-\log(\Pr_{\mathcal{Q}}[p]) + c) \\ &= c + \sum_{p \in \mathrm{Supp}(\mathcal{P})} \Pr_{\mathcal{P}}[p] \cdot (\log \frac{\Pr_{\mathcal{P}}[p]}{\Pr_{\mathcal{Q}}[p]} - \log \Pr_{\mathcal{P}}[p]) \\ &= c + \mathrm{KL}(\mathcal{P}||\mathcal{Q}) + H(\mathcal{P}).\end{aligned}$$

□

A.4 Corollary 2.19

Corollary A.5 (Non-trivial next-bit pseudoentropy implies one-way functions, restatement of Corollary 2.19)**.** *Assume there is a poly-time computable function $f\colon \{0,1\}^n \to \{0,1\}^{m(n)}$ such that $\{f(U_n)\}_{n \in \mathbb{N}}$ has n.u.-next-bit-pseudoentropy $H(f(U_n)) + 1/p(n)$ for some $p \in \mathrm{poly}$. Then there exists a n.u.-one-way functions.*

Proof sketch. Let f and p be as in Corollary 2.19, and define $g\colon \{0,1\}^n \mapsto \{0,1\}^{m(n)+n}$ by $g(x) = (f(x), x)$. We will show that $\{g(U_n)\}_{n \in \mathbb{N}}$ has n.u. next-bit pseudoentropy $n + 1/p(n)$, and the proof follows by Theorem 2.18.

For every n, let $X_n \leftarrow U_n$ and $I_n \leftarrow [m(n)]$ be a random variable. Fix $q \in \text{poly}$. Since $f(U_n)$ has non-trivial next-bit pseudoentropy, there is a random variable ensemble C (jointly distributed with X_n and I_n), such that

$$H(C_n \mid f(X_n)_{<I}) \geq \frac{H(f(U_n)) + 1/p(n)}{m(n)} - 1/q(n),$$

and $\{(f(X_n)_{<I}, C_n)\}_{n\in\mathbb{N}}$ is indistinguishable from $\{(f(X_n)_{<I_n}, f(X_n)_{I_n})\}_{n\in\mathbb{N}}$.

Let $I'_n \leftarrow [m(n)+n]$, and define C'_n (jointly distributed with X_n, I'_n) as following:

$$C'_n|_{X_n=x, I'_n=i} = \begin{cases} C_n|_{X_n=x, I_n=i} & i \leq m(n) \\ x_{(i-m(n))} & \text{o/w.} \end{cases}$$

Since $(C'_n, I'_n, X_n)|_{I'_n \leq m(n)} \equiv (C_n, I_n, X_n)$, and $(g(X_n), C')|_{I'_n > m(n)} \equiv (g(X_n), g(X_n)_{I'_n})|_{I'_n > m(n)}$, it is clear that $(g(X_n)_{<I'_n}, C'_n)$ and $(g(X_n)_{<I'_n}, g(X_n)_{I'_n})$ are indistinguishable.

We conclude the proof by bounding the entropy of C'_n. Fix $n \in \mathbb{N}$ and omit it from the notation. Let W be the indicator for the event $I' \leq m$ (that is, $W = 1$ if $I' \leq m$ and $W = 0$ otherwise). It holds that,

$$\begin{aligned}
H(C'|g(X)_{<I'}) &= H(C'|g(X)_{<I'}, W) \\
&= \Pr[W=1] \cdot H\big(C'|g(X)_{<I'}, W=1\big) + \Pr[W=0] \cdot H\big(C'|g(X)_{<I'}, W=0\big) \\
&= \Pr[W=1] \cdot H(C|f(X)_{<I}) + \Pr[W=0] \cdot H(X_{I'-m}|f(X), X_{<I'-m}, W=0) \\
&\geq \Pr[W=1] \cdot \left(\frac{H(f(X)) + 1/p}{m} - 1/q\right) + \Pr[W=0] \cdot H(X|f(X))/n \\
&= \frac{1}{m+n} \cdot (H(f(X)) + 1/p - m/q + H(X \mid f(X))) \\
&= (H(X, f(X)) + 1/p)/(m+n) - m/(m+n)q \\
&\geq (n + 1/p)/(m+n) - 1/q,
\end{aligned}$$

where the first equality holds since the value of I' (and W) is determined by $g(X)_{<I'}$, the first inequality holds by the bound on the entropy of C, and chain-rule of entropy. The last equality holds by chain-rule of entropy, and the last since $H(X, f(X)) = n$. The above concludes the proof, since it holds for every $q \in \text{poly}$. □

A.5 Proposition 2.20

Fact A.6 (Direct product of next-bit pseudoentropy, restatement of Proposition 2.20)**.** *For any random variable ensemble B and $t \in \text{poly}$, if B^t has n.u.-next-bit pseudoentropy $(t \cdot k)$, then B has n.u.-next-bit pseudoentropy k.*

Proof. Let $I = \{I_n\}_{n\in\mathbb{N}}$ be ensemble of uniform random variables over $[t(n) \cdot m(n)]$. Fix $p \in \text{poly}$, and let $C = \{C_n\}_{n\in\mathbb{N}}$ be the ensemble of random variables over $\{0,1\}^{t\cdot m}$ promised by the definition of next-bit pseudoentropy. That is:

1. $H(C_n \mid (B^t_n)_{<I_n}) \geq k(n)/m(n) - 1/p(n)$, and

2. $\left\{(B_n^{t(n)})_{\leq I_n}\right\}_{n\in\mathbb{N}}$ and $\left\{(B_n^{t(n)})_{<I_n}, C_n\right\}_{n\in\mathbb{N}}$ are n.u.-poly-time indistinguishable.

Fix n. For $j \in [t(n)]$, let $\mathcal{Z}_j$ be j^{th} $m(n)$-size block of $[t(n)m(n)]$, i.e., $\mathcal{Z}_j = \{(j-1)m(n)+1, \ldots, j \cdot m(n)\}$. The guarantee of C yields that

$$H(C_n \mid (B_n^t)_{<I_n}) = H(C_n \mid (B_n^t)_{<I_n}, I_n) = \mathrm{E}_{j\leftarrow[t(n)]}\big[H(C_n \mid (B_n^t)_{<I_n}, I_n \in \mathcal{Z}_j)\big] \tag{16}$$

In particular, there exists $j = j(n) \in [t(n)]$ such that for $\mathcal{Z}_n := \mathcal{Z}_{j(n)}$ it holds that

$$H(C_n \mid (B_n^t)_{<I_n}, I_n \in \mathcal{Z}_n) \geq H(C_n \mid (B_n^t)_{<I_n}) \geq k(n)/m(n) - 1/p(n) \tag{17}$$

In addition, since $H(A \mid B) \geq H(A \mid B, C)$ for every random variables A, B and C,

$$H(C_n \mid (B_n^t)_{[I_n-1]\cap\mathcal{Z}_n}, I_n \in \mathcal{Z}_n) \geq H(C_n \mid (B_n^t)_{<I_n}, I_n \in \mathcal{Z}_n) \tag{18}$$

We prove the claimed next-bit pseudoentropy of B by considering the following ensemble of random variables $C' = \{C'_n\}_{n\in\mathbb{N}}$ defined by

$$(C'_n, B'_n, I'_n) := \Big(C_n, (B_n^{t(n)})_{\mathcal{Z}_n}, (I_n - (j_n - 1)m(n))\Big)|_{I_n\in\mathcal{Z}_n}.$$

Notice that, (B'_n, I'_n) has the same distribution as (B_n, I''_n) for $I''_n \leftarrow [m(n)]$. By Equations (17) and (18),

$$H(C'_n \mid (B'_n)_{<I'_n}) \geq k(n)/m(n) - 1/p(n) \tag{19}$$

In addition, since the event $I_n \in \mathcal{Z}_n$ is noticeable, the ensembles $\left\{(B_n^{t(n)})_{\leq I_n} \mid I_n \in \mathcal{Z}_n\right\}_{n\in\mathbb{N}}$ and $\left\{(B_n^{t(n)})_{<I_n}, C_n \mid I_n \in \mathcal{Z}_n\right\}_{n\in\mathbb{N}}$ are n.u.-poly-time indistinguishable. And by a simple date-processing argument, so do the ensembles $\left\{(B_n^{t(n)})_{[I_n]\cap\mathcal{Z}_n} \mid I_n \in \mathcal{Z}_n\right\}_{n\in\mathbb{N}}$ and $\left\{(B_n^{t(n)})_{[I_n-1]\cap\mathcal{Z}_n}, C_n \mid I_n \in \mathcal{Z}_n\right\}_{n\in\mathbb{N}}$. We conclude that C' realizes the $k - 1/p$ next-bit pseudoentropy of B. □

ECCC ISSN 1433-8092
https://eccc.weizmann.ac.il