Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 49 (2022)

Non-Adaptive Universal One-Way Hash Functions from Arbitrary One-Way Functions

Xinyu Mao* Noam Mazor† Jiapeng Zhang‡

April 4, 2022

Abstract

Two of the most useful cryptographic primitives that can be constructed from one-way functions are *pseudorandom generators* (PRGs) and *universal one-way hash functions* (UOWHFs). The three major efficiency measures of these primitives are: *seed length*, *number of calls* to the one-way function, and *adaptivity* of these calls. Although a long and successful line of research studied these primitives, their optimal efficiency is not yet fully understood: there are gaps between the known upper bounds and the known lower bounds for black-box constructions.

Interestingly, the first construction of PRGs by Håstad, Impagliazzo, Levin, and Luby [SICOMP '99], and the UOWHFs construction by Rompel [STOC '90] shared a similar structure. Since then, there was an improvement in the efficiency of both constructions: The state of the art construction of PRGs by Haitner, Reingold, and Vadhan [STOC '10] uses $O(n^4)$ bits of random seed and $O(n^3)$ non-adaptive calls to the one-way function, or alternatively, seed of size $O(n^3)$ with $O(n^3)$ adaptive calls (Vadhan and Zheng [STOC '12]). Constructing a UOWHF with similar parameters is still an open question. Currently, the best UOWHF construction by Haitner, Holenstein, Reingold, Vadhan, and Wee [Eurocrypt '10] uses $O(n^{13})$ adaptive calls and a key of size $O(n^5)$.

In this work we give the first *non-adaptive* construction of UOWHFs from arbitrary one-way functions. Our construction uses $O(n^9)$ calls to the one-way function, and a key of length $O(n^{10})$. By the result of Applebaum, Ishai, and Kushilevitz [FOCS '04], the above implies the existence of UOWHFs in NC0, given the existence of one-way functions in NC1. We also show that the PRG construction of Haitner et al., with small modifications, yields a relaxed notion of UOWHFs. In order to analyze this construction, we introduce the notion of *next-bit unreachable entropy*, which replaces the next-bit pseudoentropy notion, used in the PRG construction above.

Keywords: universal one-way hash function; one-way function; non-adaptive.

*Shanghai Jiao Tong University. E-mail: `xinyumao.tcs@gmail.com`.

†The Blavatnik School of Computer Science at Tel-Aviv University. E-mail: `noammaz@gmail.com`. Research supported by Israel Science Foundation grant 666/19.

‡Department of Computer Science, University of Southern California. E-mail: `jiapengz@usc.edu`. Research supported by NSF CAREER award 2141536.

ISSN 1433-8092

Contents

1 Introduction

A wide class of cryptographic primitives can be constructed from *one-way functions*, which is the minimal assumption for cryptography. Two important such primitives are *pseudorandom generators* (PRGs) [Yao82; BM84] and *universal one-way hash functions* (UOWHFs) also known as, target-collision resistant (TCR) [NY89]. These two primitives are useful for constructing even more powerful primitives such as encryption, digital signatures and commitments. Yet, the optimal efficiency of these two basic primitives is not fully understood.

There are several important efficiency measures to account for when considering UOWHFs and PRGs. For PRG constructions, one aims to minimize the seed length and the number of calls to the one-way function f. For UOWHF constructions, there is a need to minimize the key length and the number of calls to f. Besides these two measurements, another important parameter is the *adaptivity* of the calls. That is, whether the invocations of the one-way function are independent of the output of previous calls. A non-adaptive construction naturally gives rise to a, more efficient, parallel algorithm. By contrast, if the calls are adaptive, one must make them sequentially.

In this paper, we focus on constructions of UOWHF, a relaxation of *collision-resistant hash function* (CRHF) introduced by Naor and Yung [NY89]. Informally, a keyed function family $\mathcal{F} = \{f_k \colon \{0,1\}^n \to \{0,1\}^m\}_k$ is an UOWHF if $m < n$, and, for every poly-time algorithm A, and for every input $x \in \{0,1\}^n$, the following holds: with high probability over the choice of the key k, $\mathsf{A}(k, x)$ cannot find a collision $x' \neq x$ with $f_k(x) = f_k(x')$.

The first UOWHF construction from arbitrary one-way functions is due to Rompel [Rom90]. The efficiency was then improved by Haitner, Holenstein, Reingold, Vadhan, and Wee [Hai+10], who gave a construction of UOWHF using $O(n^{13})$ *adaptive* calls with a key of size $O(n^5)$.[1,2] Notably, prior to the work presented here, there was no non-adaptive UOWHF construction.

The above construction of [Hai+10] uses ideas similar to the ones used in the constructions of PRGs. Still, the best PRGs constructions from arbitrary one-way functions are more efficient. Currently, the state-of-the-art construction of PRGs uses $O(n^4)$ bits of random seed and $O(n^3)$ *non-adaptive* calls to the one-way function, or alternatively seed of size $O(n^3)$ with $O(n^3)$ adaptive calls [Hai+13; VZ12]. Constructing a UOWHF using $O(n^3)$ calls to the one-way function is still an interesting open question.

These efficiency gaps between UOWHFs and PRGs constructions are even more surprising in the light of the similarities between the constructions. Specially, for more structured one-way functions such as permutations or regular functions, there is essentially no efficiency gap between PRG and UOWHF constructions.[3] Moreover, the constructions are very similar to each other and use similar techniques. For example, the method of *randomized iterate* is used for the constructions of both primitives from unknown-regular one-way functions [Hai+06; Yu+15; Ame+12]. Recently, Mazor and Zhang [MZ21] introduced non-adaptive constructions for both UOWHF and PRG from

[1] We ignore low order terms for this introduction.

[2] [Hai+10] showed an UOWHF construction with key length $O(n^7)$. However, we observe that this can be reduced easily to $O(n^5 \log n)$. In their construction, [Hai+10] first construct $\kappa = \Omega(n^2/\log n)$ function families, that at least one of them is UOWHF. In the last stage of the construction (removing non-uniformity), the final UOWHF is obtained by concatenating the output of all functions on the same input x. The key of the final construction is the concatenation of κ keys of all families. We observe that it is possible to use the same key for all families.

[3] f is called *regular* if for every n and x, x' with $|x| = |x'| = n$ it holds that $\left|f^{-1}(f(x))\right| = \left|f^{-1}(f(x'))\right|$. We say that the function is *unknown-regular* if the *regularity parameter*, $\left|f^{-1}(f(x))\right|$, may not be a computable function of n.

an unknown-regular one-way function. Their constructions for both primitives have in common a similar formula shape and are composed of the same building-block operations.

Furthermore, the first constructions from (unstructured) arbitrary one-way functions of PRGs, by Håstad, Impagliazzo, Levin, and Luby [Hås+99], and the constructions of UOWHFs by Rompel [Rom90] and Haitner et al. [Hai+10], shared a similar framework. This framework includes first constructing a non-uniform version of the desired primitive, and then eliminating the non-uniform (short) advice by enumerating over all possible advices, and combining the constructions together. This enumeration and combining step has a significant efficiency cost for both primitives.

By contrast, in their beautiful work, Haitner et al. [Hai+13] introduced a simpler and more efficient framework to construct PRGs from arbitrary one-way functions. By introducing a notion called *next-bit pseudoentropy*, they gave a very efficient and simple non-adaptive construction of PRGs from one-way functions. As stated above, this construction has $O(n^4)$ random seeds with $O(n^3)$ calls, which is a significant improvement over [Hås+99]. One main reason for this efficiency improvement is that this framework no longer requires the non-uniformity elimination step. Unfortunately, there is no analog to this construction for UOWHFs. Adapting the framework of Haitner et al. [Hai+13] to improve the efficiency of UOWHF constructions is still an interesting open question.

1.1 Our Contribution

In this paper, we partially answer the last question above. Our first result is (the first) non-adaptive construction of UOWHF from arbitrary one-way functions. We achieve this by introducing a construction that does not have the non-uniformity elimination step. By the result of Applebaum, Ishai, and Kushilevitz [App+06], the above implies the existence of UOWHFs in NC0, assuming the existence of one-way functions in NC1. In addition, our construction reduces the call complexity over Haitner et al. [Hai+10], and uses $O(n^9)$ calls to the one-way function instead of $O(n^{13})$. On the negative side, the key length of our construction is $O(n^{10})$, instead of $O(n^5)$.

Next, aiming to close the still remaining gap between PRG and UOWHF constructions, we show that small modifications to the PRG construction of [Hai+13] yield a relaxed notion of UOWHF, which we call "almost-UOWHF". Informally, a function family is almost-UOWHF if by changing the functions on a negligible fraction of the inputs, we can convert it into a (perfect) UOWHF. To analyze the almost-UOWHF construction, we introduce the notion of *next-bit unreachable entropy*, an analogue of *next-bit pseudoentropy* used in [Hai+13]. Similarly to the PRG construction, our almost-UOWHF construction uses $\tilde{O}(n^3)$ non-adaptive calls to the one-way function and has a key of size $\tilde{O}(n^4)$. More details below.

1.1.1 Non-Adaptive UOWHF from One-Way Functions

In their construction of UOWHFs from one-way functions, Haitner, Holenstein, Reingold, Vadhan, and Wee [Hai+10] defined the notion of *accessible entropy*.[4] Informally, for a function g, the accessible entropy of g^{-1} is a bound on the entropy of the output of every *collision finder for* g (i.e., of every poly-time algorithm that, given an input x, always outputs a pre-image of $g(x)$).

[4]We use the term accessible entropy to denote accessible entropy of functions. Somewhat different notions of accessible entropy used in other contexts, for example to construct statistically-hiding commitments from one-way functions ([Hai+09]).

Haitner et al. [Hai+10] showed how a one-way function $f\colon \{0,1\}^n \to \{0,1\}^n$ can be used to construct a function $\rho\colon \{0,1\}^{n^5} \to \{0,1\}^{n^5}$ such that, for a uniformly chosen input $X \leftarrow \{0,1\}^{n^5}$, there is a gap between the *entropy* of X given $\rho(X)$, and the *accessible entropy* of ρ^{-1}. Namely, there exists some $\ell \in \mathbb{N}$, such that for every collision finder A for ρ, the following holds with all but a negligible probability: the size of $\rho^{-1}(\rho(X))$ is at least $2^{\ell+\omega(\log n)}$, while the max-entropy of the output of $A(X)$, given the input X, is at most ℓ.[5] When ℓ is known, such a function can be converted easily to UOWHF, but here the parameter ℓ depends on f and may be unknown. To overcome this obstacle, Haitner et al. [Hai+10] constructed from ρ a UOWHF candidate C_ℓ for n^2 possible values of ℓ, and then combined them together. This part, used also by Rompel [Rom90], introduces the adaptivity of the construction and blows up the number of calls.

By viewing ℓ as an unknown regularity parameter of ρ, we replace the last step by applying the recent construction of [MZ21] of non-adaptive UOWHF from (unknown) regular one-way functions. The above gives rise to the following result.

Theorem 1.1 (Non-Adaptive UOWHF from OWF, informal)**.** *Assuming one-way functions exist, there exists a non-adaptive UOWHF. The construction uses $O(n^9)$ calls to the one-way function, and has a key length and output length of $O(n^{10})$.*

We note that, since ρ is not a regular function (indeed, there is a negligible fraction of inputs for which ρ may have fewer collisions), the use of [MZ21] is not straightforward, and the security proof requires a new analysis.

1.1.2 Efficient Almost-UOWHF from One-Way Functions

Our second construction is inspired by the work of Haitner, Reingold, and Vadhan [Hai+13] on PRGs. We show that small modifications to the PRG of [Hai+13] yield an *almost-UOWHF*. Informally, a function family is an almost-UOWHF, if there exists a negligible-sized set of inputs $\mathcal{B}$, such that no poly-time adversary can find a collision x' to an input x, unless the collision x' is a member of $\mathcal{B}$.[6] That is, a function family is an almost-UOWHF, if by changing the functions on a negligible fraction of the inputs, we can convert it into a UOWHF.

We note that, similarly to the above definition of almost-UOWHF, we can also define an "almost-PRG". However, unlike UOWHF, it is easy to see that an almost-PRG is a (perfect) PRG. Hence, viewing [Hai+13]'s construction as an "almost-PRG", we believe that the UOWHF analog of [Hai+13]'s constructions is essentially our almost-UOWHF. While we do not know if an almost-UOWHF can be converted efficiently into a UOWHF, we believe this construction sheds a light on the reasons to the efficiency gap between PRGs and UOWHFs constructions. We get the following theorem.

Theorem 1.2. *[Almost-UOWHF from OWF, informal] Assuming one-way functions exist, there exists an almost-UOWHF with key length $\tilde{O}(n^4)$. The almost-UOWHF makes $\tilde{O}(n^3)$ non-adaptive calls to the underlying one-way function.*

We also note that our (perfect) non-adaptive UOWHF construction can be seen, to some extent, as an instantiation of the almost-UOWHF construction (see Remark 2.2 for more details).

[5] The max-entropy of a random variable X is defined to be the log of the support size of X. The actual definition of inaccessible entropy ignores some events that have a negligible probability.

[6] The set $\mathcal{B}$ is independent of the adversary, but can be dependent on the key. For the formal definition, see Definition 5.2.

This suggests that with an improved analysis, the parameter of the UOWHF construction can be improved, and that our (perfect) UOWHF construction is a step towards the adjustment of the framework of [Hai+13] for UOWHFs.

Next-bit unreachable entropy. In their work, Haitner et al. [Hai+13] defined the notion of *next-bit pseudoentropy*. Roughly, a function $g\colon \{0,1\}^m \to \{0,1\}^\ell$ has next-bit pseudoentropy k, if for random $X \leftarrow \{0,1\}^m$ and $I \leftarrow [\ell]$ the bit $g(X)_I$ has pseudoentropy k/ℓ given $g(X)_{<I}$. Haitner et al. [Hai+13] used a one-way function to construct a function $g\colon \{0,1\}^m \to \{0,1\}^\ell$ with non-trivial (i.e., larger than m) next-bit pseudoentropy. This function g is then used to construct an efficient and simple PRG.

To replace the notion of next-bit pseudoentropy in our construction, we define the notion of *next-bit unreachable entropy*, a variant of inaccessible entropy, defined by Haitner et al. [Hai+10], that allows us to achieve almost-UOWHF using similar construction to the above PRG. We say that a function $g\colon \{0,1\}^m \to \{0,1\}^\ell$ has next-bit unreachable entropy v, if for every index $i \in \ell$, there is a sets $\mathcal{U}_i \subseteq \{0,1\}^m$ of *unreachable inputs*, such that for random x and i, the probability that $x \notin \mathcal{U}_i$ is smaller than $(m-v)/\ell$, and the following holds.

1. For every $x \notin \mathcal{U}_i$, and every PPT A, it holds that

$$\Pr_{x' \leftarrow \mathsf{A}(x)}\left[x' \in \mathcal{U}_i \wedge g(x)_{<i} = g(x')_{<i}\right] = \operatorname{neg}(n).$$

   That is, it is hard to find a collision for $g(x)_{<i}$ inside $\mathcal{U}_i$.

2. Moreover, even if $x \in \mathcal{U}_i$, it is hard to flip the i-th bit of g while staying inside $\mathcal{U}_i$. Formally, we require that

$$\Pr_{x' \leftarrow \mathsf{A}(x)}\left[x' \in \mathcal{U}_i \wedge g(x)_{<i} = g(x')_{<i} \wedge g(x)_i \neq g(x')_i\right] = \operatorname{neg}(n).$$

For example, for every permutation $p\colon \{0,1\}^m \to \{0,1\}^m$, the function $g\colon \{0,1\}^m \to \{0,1\}^\ell$ defined by $g(x) = p(x)0^{\ell-m}$ has next-bit unreachable entropy 0, as can been seen by setting $\mathcal{U}_i = \{0,1\}^m$ for every $i > m$, or the empty set for $i \le m$. More generally, for every *injective* function $g\colon \{0,1\}^m \to \{0,1\}^\ell$, we can define $\mathcal{U}_i$ to be the set of all inputs $x \in \{0,1\}^m$, such that there is no $x' \in \{0,1\}^m$ with $g(x)_{<i} = g(x')_{<i}$ while $g(x)_i \neq g(x')_i$.[7] In this case, it is not hard to see that the probability that a random x is outside of $\mathcal{U}_i$ (for any fixed i) is at least the entropy of $g(X)_i$ given $g(X)_{<i}$ (i.e., $H(g(X)_i \mid g(X)_{<i})$).[8] Using the chain rule of entropy, we get that for a random index I, the probability that X is outside of $\mathcal{U}_I$ is at least

$$1/\ell \cdot \sum_{i \in \ell} H(g(X)_i \mid g(X)_{<i}) = 1/\ell \cdot H(g(X)) = m/\ell.$$

Note also that, without assuming computational hardness, the above sets $\mathcal{U}_i$ are the maximal that respect the definition of unreachable entropy. By the above observations, it follows that a function g has $v > 0$ next-bit unreachable entropy if the "reachable entropy" of $g(X)_I$ given $g(X)_{<I}$ is smaller

[7] If the function g is not injective, it is natural to consider $g'(x) = (g(x), x)$. We use similar construction in Section 5.

[8] Indeed, observe that $H(g(X)_i \mid g(X)_{<i} = g(x)_{<i})$ is zero iff $x \in \mathcal{U}_i$. Additionally, $H(g(X)_i \mid g(X)_{<i} = g(x)_{<i}) \le 1$ for every x. It follows that $H(g(X)_i \mid g(X)_{<i}) = \mathbf{E}_{x \leftarrow \{0,1\}^m}[H(g(X)_i \mid g(X)_{<i} = g(x)_{<i})] \le \mathbf{E}_{x \leftarrow \{0,1\}^m}[1_{x \notin \mathcal{U}_i}] = \Pr[X \notin \mathcal{U}_i]$.

than its real entropy. In this sense, our definition is a dual version of the next-bit pseudoentropy definition. We show that a very similar function to the function g used by [Hai+13] has non-trivial next-bit unreachable entropy. More details on the constructions and the security proof are given in Section 2.

1.2 Additional Related Work

Next-block pseudoentropy and inaccessible entropy. A different variant of inaccessible entropy, for online generator, was defined and used by Haitner, Reingold, Vadhan, and Wee [Hai+09] to construct statistically hiding commitments. Agrawal, Chen, Horel, and Vadhan [Agr+19] pointed out that the above notions of accessible entropy and next-block pseudoentropy are deeply related to each other.

UOWHFs from regular one-way functions. The constructions of UOWHF from regular one-way functions are more efficient. Besides the works mentioned above, Naor and Yung [NY89] showed an efficient construction from 1-to-1 one way functions.

Additionally, a few refinements of regularity were considered in past works. [BM12] showed a construction for UOWHF that uses $O(ns^6(n))$ key-length under the assumption that $\left|f^{-1}(f(x))\right|$ is concentrated in an interval of size $2^{s(n)}$. [Yu+15] showed adaptive constructions for a number of regularity refinements.

UOWHFs from arbitrary one-way functions. As mentioned above, [Rom90] gave the first constraction of UOWHF from arbitrary one-way functions. Katz and Koo [KK05] gave the first full proof for Rompel [Rom90]'s construction.

Lower bounds. The lower bounds for black-box constructions are relatively far from the upper bounds. In this line of work, there are two incomparable types of results. The first type, due to [Gen+05] is stated with terms of the stretching and compression of the PRG and UOWHF, respectively. Specifically, [Gen+05] showed that any black-box PRG construction $G\colon \{0,1\}^m \to \{0,1\}^{m+s}$ from f must use $\Omega(s/\log n)$ calls to f. Similarly, any black box UWOHF construction with input size m and output size $m-s$ must use $\Omega(s/\log n)$ calls. In the second type of results [HS12] showed that any black-box PRG construction from f must use $\Omega(n/\log n)$ calls to f, even for 1-bit stretching. [BH13] showed similar results for 1-bit compressing UWOHF. These lower bounds hold even when the one-way function f is unknown-regular. In this case, these bounds are known to be tight.

Collision resistant hash functions (CRHFs). UOWHF is a relaxation of CRHF. In the latter, we require that for a random function from the family, no adversary can find a collision (x, x'). Constructing a CRHF is a more challenging task, and its complexity is still not clear. Asharov and Segev [AS16] showed that there is no black-box construction of CRHFs even from indistinguishable obfuscation (iO) additionally to a one-way permutation. Naor and Yung [NY89] showed that UOWHFs are sufficient for the task of constructing digital signatures.

1.3 Paper Organisation

In Section 2, we give a high-level description of our constructions and proof technique. Formal definitions are given in Section 3. The UOWHF construction and the proof of Theorem 1.1 are in Section 4. The formal definition of almost-UOWHF and next-bit unreachable entropy, and the almost-UOWHF construction and the proof of Theorem 1.2 are in Section 5.

2 Our Technique

In this section, we provide a rather elaborate description of our constructions and proof technique.

2.1 Non-Adaptive UOWHF

We start with a high-level description of the constructions of [Hai+10] and [MZ21].

2.1.1 UOWHF from Unknown-Regular One-Way Functions ([MZ21])

Mazor and Zhang [MZ21] showed how to construct a non-adaptive UOWHF from an unknown-regular one-way function $f : \{0,1\}^n \to \{0,1\}^n$. For hash functions $h_1, \ldots, h_{n-1} \colon \{0,1\}^n \to \{0,1\}^{n-\log n}$ from a hash family $\mathcal{H}$, and n inputs $x_1, \ldots, x_n \in \{0,1\}^n$, [MZ21] showed that, for the right choice of $\mathcal{H}$, the function

$$C(h_1, \ldots, h_n, x_1, \ldots, x_n) = h_1, \ldots, h_n, f(x_1), h_1(x_1, f(x_2)), \ldots, h_{n-1}(x_{n-1}, f(x_n)), x_n$$

is collision resistant on random inputs. Since this function is also shrinking, it can be converted into an UOWHF easily by a standard construction.[9]

Intuitively, for a regular function f and i.i.d uniform random variables X_1, X_2 over $\{0,1\}^n$, given any fixing of $f(X_1)$, the entropy of the pair $X_1, f(X_2)$ is exactly n. To see the above, recall that for a regular f with an (unknown) regularity parameter Δ, it holds that there are exactly Δ possible values for X_1 given $f(X_1)$, and exactly $2^n/\Delta$ possible values for $f(X_2)$. Thus, the regularity parameter Δ "cancels out" when considering the number of possible values (given $f(X_1)$) of the pair $X_1, f(X_2)$, as this number is $\Delta \cdot 2^n/\Delta = 2^n$. It follows that the compression of the pair $X_1, f(X_2)$ does not create too many collisions. This fact can be used in order to reduce the problem of inverting f, into finding a collision for C.

2.1.2 Inaccessble Entropy from One-Way Functions ([Hai+10])

In order to construct an UOWHF from an arbitrary one-way function, given a one-way function $f \colon \{0,1\}^n \to \{0,1\}^n$, Haitner et al. [Hai+10] started with constructing a function g, that takes as input an index $i \in [n]$, string $x \in \{0,1\}^n$ and a description of a random hash function h from a 3-wise independent hash family $\mathcal{H}$, and outputs h, together with the i first bits of $h(f(x))$. That is, $g(i, x, h) = (h, h(x)_{\leq i})$. Haitner et al. [Hai+10] showed that for every collision finder algorithm A, there are sets $\{\mathcal{L}_w\}_{w \in ([n]\times\{0,1\}^n\times\mathcal{H})}$, such that, for a random input $W \leftarrow ([n] \times \{0,1\}^n \times \mathcal{H})$,

1. $\mathbf{Pr}[\mathsf{A}(W) \notin \mathcal{L}_W] = neg(n)$, and,

[9] [MZ21] actually showed it is enough to use a single hash function. The number of repetitions n is in order to make the function shrinking.

2. $H(W \mid g(W)) - \mathbf{E}[\log(|\mathcal{L}_W|)] \geq \log n/n$,

where H is the entropy function. The above $\log n/n$ is a gap between the *entropy* of W given $g(W)$, to its *accessible average max entropy.*

Haitner et al. [Hai+10] then showed that for $\rho = g^{n^4}$ (i.e., $\rho(w_1, \ldots, w_{n^4}) = g(w_1), \ldots, g(w_{n^4})$, the concatenation of the outputs of n^4 independent invocations of g), both the entropy and the accessible entropy are highly concentrated around their means. That is, there exist some $\ell \in \mathbb{N}$, $s = \omega(\log n)$ and sets $\{\mathcal{L}_z\}_{z \in \mathrm{Domain}(\rho)}$ that satisfy the following: (1) first, $\left|\rho^{-1}(\rho(z))\right| \geq 2^{\ell+s}$, (2) $|\{\mathcal{L}_z\}| \leq 2^{\ell}$ for all but negligible fraction of z's, and (3), $\mathbf{Pr}[\mathsf{A}(z) \notin \mathcal{L}_z] = neg(n)$ for every collision finder A for ρ.

We now proceed to describing our construction. In the following we view ρ as a function from $\{0,1\}^m$ to $\{0,1\}^k$, for $m, k = O(n^5)$ (using a proper encoding of the input).

2.1.3 Our Construction

Thinking of ℓ as the regularity parameter of the function ρ, we use the construction of [MZ21] in order to get a non-adaptive UOWHF. That is, for hash functions $h_1, \ldots, h_{m-1} \colon \{0,1\}^m \times \{0,1\}^k \to \{0,1\}^{m - \log n}$ from a universal family $\mathcal{H}$, and inputs $z_1, \ldots, z_m$, let

$$C(h_1, \ldots, h_{m-1}, z_1, \ldots, z_m) = h_1, \ldots, h_{m-1}, \rho(z_1), h_1(z_1, \rho(z_2)), \ldots, h_{m-1}(z_{m-1}, \rho(z_m)), z_m.$$

We show that C is collision resistant on random inputs. Indeed, assume that $\left|\rho^{-1}(\rho(z))\right| \geq 2^{\ell+\omega(\log n)}$ for every $z \in \{0,1\}^m$. We get that the image size of ρ is at most $2^m \cdot 2^{-\ell-\omega(\log n)}$. Thus, for $Z_1, Z_2 \leftarrow \{0,1\}^m$ and $H_1 \leftarrow \mathcal{H}$, any poly-time algorithm cannot find a collision for $\rho(Z_1), H_1(Z_1, \rho(Z_2))$, since it only has

$$|\mathcal{L}_{Z_1}| \cdot |\mathrm{Image}(\rho)| \leq 2^{\ell} \cdot (2^m \cdot 2^{-\ell-\omega(\log n)}) = 2^{m-\omega(\log n)}$$

possible values to choose from, and the probability for each such value to collide with $Z_1, \rho(Z_2)$ on H_1 is $2^{-m+\log n}$. Thus, by the union bound, the probability that there is a collision for Z_1, Z_2 inside the set $\mathcal{L}_{Z_1} \times \mathrm{Image}(\rho)$ is negligible. By a similar argument, the analysis shows that it is impossible to find a collision for the entire function C.

However, there is an issue with the above idea. Note that the condition concerning the pre-image size of an image of ρ holds only with overwhelming probability, which may pose a problem. Indeed, let $\mathcal{B} = \left\{z \in \{0,1\}^m \colon \left|\rho^{-1}(\rho(z))\right| < 2^{\ell+\omega(\log n)}\right\}$ be the set of all untypical inputs. The size of $\rho(\mathcal{B})$ can be much larger than $2^m \cdot 2^{-\ell-\omega(\log n)}$, the number of "typical" images. Thus, by choosing X'_2 from this set, the adversary might be able to find a collision (Z'_1, Z'_2) for $\rho(Z_1), H_1(Z_1, \rho(Z_2))$. Fortunately, it turns out that this issue can be resolved by a more careful analysis, which yields the following key insight: for every collision $(z'_1, \ldots, z'_m)$ for C found by an efficient algorithm, it holds that if $z'_i \in \mathcal{B}$ for some i, it must holds that $z'_{i+1} \in \mathcal{B}$ as well. It follows from the above that in this case, z'_m is also in $\mathcal{B}$. Since C outputs its last input z_m, and with all but a negligible probability $z_m \notin \mathcal{B}$, we have that $(z'_1, \ldots, z'_m)$ is not a valid collision.

Remark 2.1 (Using a more shrinking hashing)**.** *The actual gap s between the accessible and real entropy of ρ^{-1} is $s \approx n^3$. Thus, the first part of the argument above will work even if the hash functions will output only $m - n^3 + \omega(\log n)$ bits. In this case, however, we will not be able to show* *that it is infeasible to find a collision inside $\mathcal{B}$. The above suggests the following construction of* almost-UOWHF*: let $t \approx n^2$, and for $h_1, \ldots, h_{t-1}, z_1, \ldots, z_t$, consider*

$$C(h_1, \ldots, h_{t-1}, z_1, \ldots, z_t) = h_1, \ldots, h_{t-1}, \rho(z_1), h_1(z_1, \rho(z_2), \ldots, h_{t-1}(z_{t-1}, \rho(z_t)), z_t,$$

for $h_i \colon \{0,1\}^{m+k} \to \{0,1\}^{m-n^3/2}$.

For large enough t, the above function is shrinking, and, for a random input $(h_1, \ldots, h_{t-1}, z_1, \ldots, z_t)$ it is hard to find a collision $(z'_1, \ldots, z'_t)$, such that $z'_i \notin \mathcal{B}$ for every i. The latter implies that all the collisions that can be found by an efficient algorithm come from a negligible-sized set. Such a function can easily be converted into an almost-UOWHF, which yields a construction with $O(n^6)$ non-adaptive calls, and key length of $O(n^7)$ bits. It turns out, see next section, that there are better approaches for constructing almost-UOWHFs.

2.2 Almost-UOWHF

We start with a high-level description of the one-way function based pseudorandom generator of [Hai+13]. The main building block of the construction is a function $g \colon \{0,1\}^m \to \{0,1\}^\ell$, with $k > m$ next-bit pseudoentropy. On a given input, their PRG starts by using g to construct the following matrix-like structure (see Figure 1): the structure is composed of $q \approx m^2$ rows, where each row contains $t \approx m$ independent copies of $g(X)$, for $X \leftarrow \{0,1\}^m$, shifted by a random offset between 0 to ℓ. Every fully populated column is then hashed by a hash function $h \colon \{0,1\}^q \to \{0,1\}^a$, for $a \approx q \cdot k/\ell > q \cdot m/\ell$. Finally, the output of the PRG is the concatenation of the outputs of the hash function applied to every fully populated column.

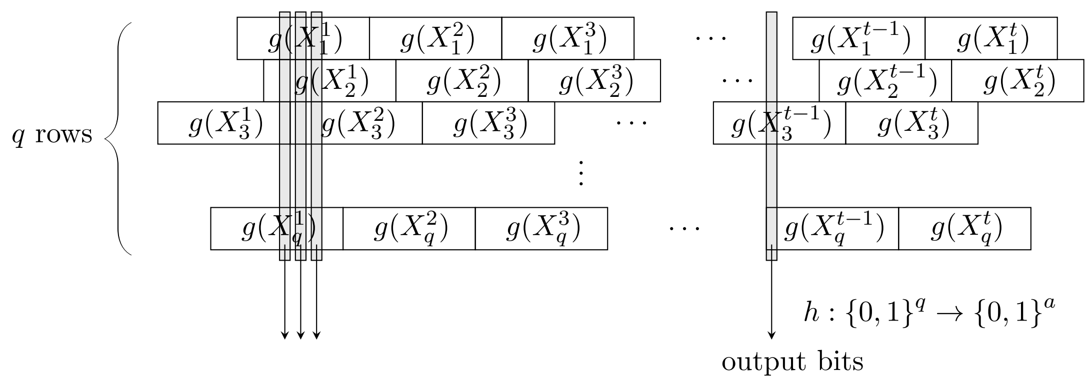

Figure 1: The PRG construction of [Hai+13], $G \colon \mathcal{H} \times (\{0,1\}^m)^{t \cdot q} \to \mathcal{H} \times (\{0,1\}^a)^{(t-1)\ell}$. there are $q \approx m^2$ rows, each row has $t \approx m$ i.i.d copies of $g(X)$, shifted by a random offset. Every fully populated column, marked in grey, is hashed by $h \in \mathcal{H}$. The almost-UOWHF construction also outputs the columns that are not fully populated.

We prove that slightly tweaking the above construction, and using a different function g, yields a function that is almost collision-resistant on random inputs. Specifically, the output of our construction contains not only the hashed fully populated columns, but also all the columns that are not fully populated (without hashing). Additionally, we choose the parameter a to be smaller than $q \cdot n/m$, in order to make the function length-decreasing. The function g we are using in our construction, is defined by

$$g(h_1, h_2, x) = (h_1, h_2, h_1(f(x)), h_2(x)),$$

for hash functions $h_1, h_2 \colon \{0,1\}^n \to \{0,1\}^n$ from a 3-wise independent family. We prove in Section 5.2 that if f is a one-way function, the above function g has next-bit unreachable entropy $\log n$.[10]

Remark 2.2 (Similarities between our constructions)**.** *We note that the function ρ, defined in Section 2.1, is composed of n^4 independent repetitions of a simpler function with random shifts. Thus, our first construction of non-adaptive UOWHF can be seen as an instantiation of the second (almost-UOWHF) construction, described above, where we apply the hash function on blocks of m columns, instead of hashing every single column (and by taking the number of rows to be larger).*

In the rest of this section we give some details on the security proof. Consider the function σ induced by taking the first hashed column in our almost-UOWHF construction (Figure 1) together with the columns to the left of it. That is,

$$\sigma(h, i_1, \ldots, i_q, x_1, \ldots, x_q) = h, (g(x_1)_{<i_1}, \ldots, g(x_q)_{<i_q}), h(g(x_1)_{i_1}, \ldots, g(x_q)_{i_q}),$$

for a hash function $h \colon \{0,1\}^q \to \{0,1\}^a$ from a universal family $\mathcal{H}$.

Additionally, consider the function $\hat{\sigma}$, defined similarly to σ, but without applying the hash h on the column. That is,

$$\hat{\sigma}(h, i_1, \ldots, i_q, x_1, \ldots, x_q) = h, (g(x_1)_{<i_1}, \ldots, g(x_q)_{<i_q}), (g(x_1)_{i_1}, \ldots, g(x_q)_{i_q}).$$

It turns out, see detail below, that the following holds for a right choice of the parameter a and for some negligible-sized set of inputs $\mathcal{B}$: for a random input, every collision found by a collision finder to the function σ is either a collision for $\hat{\sigma}$, or it is inside the set $\mathcal{B}$. That is, the function h does not make the task of finding a collision (outside of $\mathcal{B}$) easier.

To see that the above is enough to prove the security of the construction, let C be the almost-UOWHF construction described above, and let $\widehat{C}$ be the function defined by the raw matrix-like structure (without applying the hash function on every fully-populated column). Observe that since the function g is (close to) injective, the function $\widehat{C}$ is (not shrinking) collision-resistant on random inputs. A simple hybrid argument yields that every collision finder that, given an input w for C, is able to find a collision $w' \neq w$ for C that is not a collision for $\widehat{C}$ (namely, $C(w) = C(w')$ but $\widehat{C}(w) \neq \widehat{C}(w')$), can be used to find a collision for σ which is not a collision for $\widehat{\sigma}$.[11] Since the latter is hard to find, and since $\widehat{C}$ is collision-resistant, the above concludes the proof.

σ is (almost) as hard as $\hat{\sigma}$. It thus left to prove that it is hard to find a collision for σ which is not a collision for $\hat{\sigma}$, outside of the negligible sized set $\mathcal{B}$. let A be a collision finder for σ, and let $w' = (h, i_1, \ldots, i_q, x'_1 \ldots, x'_q)$ be a collision found by $\mathsf{A}(w)$, for some $w = (h, i_1, \ldots, i_q, x_1 \ldots, x_q)$. We show that either $\hat{\sigma}(w) = \hat{\sigma}(w')$, or w' is a member of a small set $\mathcal{B}$. To do so, we use the next-bit unreachable entropy property of g.

Let $\{\mathcal{U}_i\}_{i \in [\ell]}$ be the sets guaranteed by the next-bit unreachable entropy of g (these sets are independent from the choice of A). By the definition of next-bit unreachable entropy it holds that:

1. for every j such that $x_j \notin \mathcal{U}_{i_j}$, no collision finder can find $x'_j \in \mathcal{U}_{i_j}$ such that $g(x_j)_{<i_j} = g(x'_j)_{<i_j}$, and thus it must hold that $x'_j \notin \mathcal{U}_{i_j}$.

[10] For the PRG construction, [Hai+13] used $g(h, x) = (h, f(x), h(x))$ and [VZ12] used $g(x) = (f(x), x)$. Observe that, since $h_1(f(x))$ is also a one-way function, our g can be used in the PRG construction.

[11] Furthermore, the hybrid argument yields that w' must be from a small set if the collision for $\widehat{\sigma}$ is.

2. Similarly, for every j with $x_j \in \mathcal{U}_{i_j}$, it holds that $g(x_j)_{i_j} = g(x'_j)_{i_j}$, unless $x'_j \notin \mathcal{U}_{i_j}$.

Let $\mathcal{J}_w$ be the set of indices for which x_j is inside the set $\mathcal{U}_{i_j}$. Formally,

$$\mathcal{J}_w = \left\{ j \in [q] \colon x_j \in \mathcal{U}_{i_j} \right\}.$$

By Item 1 above, it holds that $\mathcal{J}_{w'} \subseteq \mathcal{J}_w$. Moreover, Item 2 implies that $g(x_j)_{i_j} = g(x'_j)_{i_j}$ for every $j \in \mathcal{J}_{w'} \cap \mathcal{J}_w$. The above yields the key observation of the proof:

Claim 2.3. *For any collision $w' = (h, i_1, \ldots, i_q, x'_1 \ldots, x'_q)$ found by a collision finder A, unless $|\mathcal{J}_{w'}|$ is smaller than $|\mathcal{J}_w|$, there are $|\mathcal{J}_w|$ bits in $g(x'_1)_{i_1}, \ldots, g(x'_q)_{i_q}$ that get the exact same value as in $g(x_1)_{i_1}, \ldots, g(x_q)_{i_q}$ (namely, $g(x_j)_{i_j} = g(x'_j)_{i_j}$ for every $j \in \mathcal{J}_w$).*

Observe that for large enough q, the size of $\mathcal{J}_w$ (for a random w) is concentrated around its mean. Since g has $\log n$ next-bit unreachable entropy, its mean is at least $q \cdot (1 - \frac{m - \log n}{\ell})$. In the following, assume for simplicity that the size of $\mathcal{J}_w$ is equal to its mean, and that this mean is exactly $q \cdot (1 - \frac{m - \log n}{\ell})$. To conclude the proof, let $\mathcal{B}$ be the negligible-sized set of all inputs $w' = (h, i_1, \ldots, i_q, x'_1 \ldots, x'_q)$ for which $|\mathcal{J}_{w'}|$ is (much) smaller than $q \cdot (1 - \frac{m - \log n}{\ell})$, and set the length of the output of the hash function h to be $a \approx q \cdot \frac{m - \log n}{\ell} < q \cdot m/\ell$. It follows that the output of every collision finder for σ is either in $\mathcal{B}$, or agrees with $g(x_1)_{i_1}, \ldots, g(x_q)_{i_q}$ on (almost) all the indices in $\mathcal{J}_w$. However, with all but a negligible probability, there is no string y' that agrees with $y = (g(x_1)_{i_1}, \ldots, g(x_q)_{i_q})$ on $q \cdot (1 - \frac{m - \log n}{\ell})$ bits, for which $h(y) = h(y')$, unless $y = y'$. In other words, any such collision for σ is also a collision for $\hat{\sigma}$.

3 Preliminaries

3.1 Notations

We use calligraphic letters to denote sets, uppercase for random variables, and lowercase for values and functions. For $n \in \mathbb{N}$, let $[n] := \{1, \ldots, n\}$. Given a vector $s \in \{0,1\}^n$, let s_i denote its i-th entry, and $s_{\leq i}$ denote its first i entries. Define $s_{<i}$, $s_{>i}$ and $s_{\geq i}$ similarly.

The support of a distribution P over a finite set $\mathcal{S}$ is defined by $\mathrm{Supp}(P) := \{x \in \mathcal{S} : P(x) > 0\}$. For a (discrete) distribution D let $d \leftarrow D$ denote that d was sampled according to D. Similarly, for a set $\mathcal{S}$, let $s \leftarrow \mathcal{S}$ denote that s is drawn uniformly from $\mathcal{S}$. For an event W, we use $\overline{W}$ to denote the compliment event. For a function $f \colon \{0,1\}^n \to \{0,1\}^n$, let $\mathsf{Im}(f) := \{f(x) \colon x \in \{0,1\}^n\}$ be the image of f.

Let poly denote the set of all polynomials, and let PPT stand for probabilistic polynomial time. A function $\mu \colon \mathbb{N} \to [0,1]$ is negligible, denoted $\mu(n) = neg(n)$, if $\mu(n) < 1/p(n)$ for every $p \in \mathrm{poly}$ and large enough n. For a security parameter n, a function $f \colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}$ is efficiently computable if it is computable in polynomial time in n.

3.2 One-Way Functions

We now formally define basic cryptographic primitives. We start with the definition of one-way function.

Definition 3.1 (One-way function)**.** *A polynomial-time computable function $f : \{0,1\}^* \to \{0,1\}^*$ is called a* one-way function *if for every probabilistic polynomial time algorithm* A*, there is a negligible function $\mu : \mathbb{N} \to [0,1]$ such that for every $n \in \mathbb{N}$*

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

For simplicity we assume that the one-way function f is length-preserving. That is, $|f(x)| = |x|$ for every $x \in \{0,1\}^*$. This can be assumed without loss of generality, and is not crucial for our constructions.

Immediately from the definition of a one-way function, we get the following simple observation.

Claim 3.2. *For every one-way function $f \colon \{0,1\}^n \to \{0,1\}^n$ there exists a negligible function $\mu(n)$ such that for every input $x \in \{0,1\}^n$ it holds that $\left|f^{-1}(f(x))\right| \leq 2^n \cdot \mu(n)$.*

3.3 Universal One Way Hash Functions

We now formally define UOWHF.

Definition 3.3 (Universal one-way hash function)**.** *Let n be a security parameter. A family of functions $\mathcal{F} = \left\{f_z \colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}\right\}_{z \in \{0,1\}^{k(n)}}$ is a family of* universal one-way hash functions *(UOWHFs) if it satisfies:*

1. *Efficiency: Given $z \in \{0,1\}^{k(n)}$ and $x \in \{0,1\}^{m(k)}$, $f_z(x)$ can be evaluated in time* $\mathrm{poly}(n)$*.*
2. *Shrinking: $\ell(n) < m(n)$.*
3. *Target Collision Resistance: For every probabilistic polynomial-time adversary* A*, the probability that* A *succeeds in the following game is negligible in n:*
   (a) *Let $(x, state) \leftarrow \mathsf{A}(1^n) \in \{0,1\}^{m(n)} \times \{0,1\}^*$.*
   (b) *Choose $z \leftarrow \{0,1\}^{k(n)}$.*
   (c) *Let $x' \leftarrow A(state, z) \in \{0,1\}^{m(n)}$.*
   (d) A *succeeds if $x \neq x'$ and $f_z(x) = f_z(x')$.*

A relaxation of the target collision resistance property can be done by requiring the function to be collision resistant only on random inputs.

Definition 3.4 (Collision resistance on random inputs)**.** *Let n be a security parameter. A function $f \colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}$ is* collision resistant on random inputs *if for every probabilistic polynomial-time adversary* A*, the probability that* A *succeeds in the following game is negligible in n:*

1. *Choose $x \leftarrow \{0,1\}^{m(n)}$.*
2. *Let $x' \leftarrow A(1^n, x) \in \{0,1\}^{m(n)}$.*
3. A *succeeds if $x \neq x'$ and $f(x) = f(x')$.*

The following lemma states that it is enough to construct a function that is collision resistant on random inputs, in order to get UOWHF.

Lemma 3.5 (From random inputs to targets, folklore)**.** *Let n be a security parameter. Let $F\colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}$ be an efficiently computable length-decreasing function. Suppose F is collision-resistant on random inputs. Then $\left\{F_y\colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}\right\}_{y\in\{0,1\}^{m(n)}}$, for $F_y(x) := F(y \oplus x)$, is an UOWHF.*

3.4 Hash Families

2-universal and t-wise independent hash families are an important ingredient in our constructions. In this section, we formally define this notion, together with some useful properties of such families.

Definition 3.6 (2-universal and t-wise independent families)**.**
A family of function $\mathcal{F} = \left\{f\colon \{0,1\}^n \to \{0,1\}^\ell\right\}$ is 2-universal *if for every $x \neq x' \in \{0,1\}^n$ it holds that $\mathbf{Pr}_{f\leftarrow\mathcal{F}}[f(x) = f(x')] \leq 2^{-\ell}$. $\mathcal{F}$ is t*-wise independent *if for all $x_1 \neq \ldots \neq x_t \in \{0,1\}^n$, the random variables $F(x_1), \ldots, F(x_t)$ for $F \leftarrow \mathcal{F}$ are independent and uniformly distributed over $\{0,1\}^\ell$.*

A family is explicit *if given a description of a function $f \in \mathcal{F}$ and $x \in \{0,1\}^n$, $f(x)$ can be computed in polynomial time (in n, ℓ). Such family is* constructible *if it is explicit and there is a* PPT *algorithm that given $x \in \{0,1\}^n$ and $y \in \{0,1\}^\ell$ outputs a uniform $f \in \mathcal{F}$, such that $f(x) = y$.*

It is well-known that there are constructible families of t-wise independent functions with description size $O(t \cdot (n + \ell))$. The next lemma, proven in Appendix A, will be useful in the proof.

Lemma 3.7. *Let $f\colon \{0,1\}^n \to \{0,1\}^n$ be a function, and $\mathcal{H} = \{h\colon \{0,1\}^n \to \{0,1\}^n\}$ a two-wise independent family. For every $x \in \{0,1\}^n$ and $c \in \mathbb{N}$ the following holds.*

$$\Pr_{h\leftarrow\mathcal{H}}\left[\left|\left\{x'\colon h(f(x')) = h(f(x))\right\}\right| \geq \left|f^{-1}(f(x))\right| + n^{2c}\right] \leq 2/n^c.$$

3.5 Entropy and Accessible Entropy

The *Shannon entropy* of a random variable X is defined by

$$H(X) = -\sum_{x\in\mathrm{Supp}(X)} \mathbf{Pr}[X = x] \cdot \log(\mathbf{Pr}[X = x]).$$

The conditional entropy of a random variable X given Y, is defined as $H(X \mid Y) = \mathbf{E}_{y\leftarrow Y}[H(X|_{Y=y})]$. For a number $p \in [0,1]$, we will use $H(p)$ to denote the entropy of a random variable distributed according to *Bernoulli*(p). That is $H(p) = -p\log p - (1-p)\log(1-p)$.

The *min entropy* of a random variable X is defined by $\mathrm{H}_\infty(X) = \min_{x\in\mathrm{Supp}(X)} \log\frac{1}{\mathbf{Pr}[X=x]}$, and the *max entropy* of X is defined by $H_0(X) = \log|\mathrm{Supp}(X)|$.

Lastly, for random variables X and Y, the sample entropy of $x \in \mathrm{Supp}(X)$ (with respect to X) is defined by $H_X(x) = -\log\mathbf{Pr}[X = x]$, and the sample entropy of x given $y \in \mathrm{Supp}(Y)$ is defined by $H_{X|Y}(x \mid y) = -\log\mathbf{Pr}[X = x \mid Y = y]$. The following equality is immediate from the definitions above.

$$H(X \mid Y) = \mathop{\mathbf{E}}_{x\leftarrow X, y\leftarrow Y}\left[H_{X|Y}(x \mid y)\right] \tag{1}$$

For a function g, we also use the following notation, defined in [Hai+10], for the entropy of g^{-1}.

Definition 3.8 (Real entropy). *Let n be a security parameter and $g\colon \{0,1\}^n \to \{0,1\}^m$ be a function.*

We say that g^{-1} has real Shannon entropy *k if $H(X|g(X)) = k$, where X is uniformly distributed on $\{0,1\}^n$.*

We say that g^{-1} has real min-entropy *at least k if there is a negligible function $\varepsilon = \varepsilon(n)$ such that $\mathbf{Pr}_{x\leftarrow X}\left[H_{X|g(X)}(x|g(x)) \geq k\right] \geq 1 - \varepsilon(n)$.*

We say that g^{-1} has real max-entropy *at most k if there is a negligible function $\varepsilon = \varepsilon(n)$ such that $\mathbf{Pr}_{x\leftarrow X}\left[H_{X|g(X)}(x|g(x)) \leq k\right] \geq 1 - \varepsilon(n)$.*

[Hai+10] also introduced the notion of accessible max-entropy. A collision finder for a function g is an algorithm that, given input x, always output x' such that $g(x) = g(x')$. g^{-1} has small accessible entropy, if the output of every collision finder for g comes from a small set.

Definition 3.9 (Collision finder). *For a function $g\colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}$, an algorithm A is a* g-collision finder *if for every $x \in \{0,1\}^{m(n)}$ it holds that $\mathbf{Pr}[g(\mathsf{A}(1^n, x)) = g(x)] = 1$.*

Definition 3.10 (accessible max-entropy). *Let n be a security parameter and $g\colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}$ be a function. We say that g^{-1} has accessible max-entropy at most k if for every* PPT *g-collision finder A and for every $n \in \mathbb{N}$, there exists a family of sets $\{\mathcal{L}(x)\}_{x\in\{0,1\}^{m(n)}}$ each of size at most $2^{k(n)}$ such that $x \in \mathcal{L}(x)$ for all x, and $\mathbf{Pr}_{x\leftarrow\{0,1\}^{m(n)}}[\mathsf{A}(1^n, x) \in \mathcal{L}(x)] \geq 1 - neg(n)$.*

The following theorems are implicit in [Hai+10] and will be useful in our constructions.

Theorem 3.11 (Entropy gap, implicit in [Hai+10]). *Let $f\colon \{0,1\}^n \to \{0,1\}^n$ be a one-way function. Then there exists $\ell = \ell(n)$, $s = \omega(\log n)$ and an efficiently computable function $g\colon \{0,1\}^{n^5} \to \{0,1\}^{n^5}$ such that:*

1. *g^{-1} has real min-entropy at least $\ell + s$.*
2. *g^{-1} has accessible max-entropy at most ℓ.*
3. *The evaluation procedure of g makes $O(n^4)$ non-adaptive calls to the one-way function.*

Theorem 3.12 (Implied by Claim 4.9, [Hai+10]). *Let $f\colon \{0,1\}^n \to \{0,1\}^n$ be a one-way function and let $\mathcal{H} = \{h : \{0,1\}^n \to \{0,1\}^n\}$ be a family of constructible, three-wise independent hash functions.[12] Then, for every ppt A, every constant $c > 0$ and every $i \in [n]$, it holds that:*

$$\mathop{\mathbf{Pr}}_{\substack{h\leftarrow\mathcal{H},\\ x\leftarrow\{0,1\}^n,\\ x'\leftarrow\mathsf{A}(1^n,h,x,i)}}\left[(f(x') \neq f(x)) \wedge (h(f(x))_{<i} = h(f(x'))_{<i}) \wedge i > n - (\log\left|f^{-1}(f(x'))\right| - c\log n)\right] = neg(n).$$

[12] Actually, the proof in [Hai+10] only requires two-wise independence.

3.6 Useful Facts

We will use the well known Chernoff bound in our proof.

Fact 3.13 (Chernoff bound)**.** *Let $A_1, ..., A_n$ be independent random variables s.t. $A_i \in \{0,1\}$ and let $\widehat{A} = \Sigma_{i=1}^n A_i$. For every $\epsilon \in [0,1]$ It holds that:*

$$\mathbf{Pr}\left[\left|\widehat{A} - \mathbf{E}\left[\widehat{A}\right]\right| \geq \epsilon \cdot \mathbf{E}\left[\widehat{A}\right]\right] \leq 2 \cdot e^{-\epsilon^2 \cdot \mathbf{E}[\widehat{A}]/3}.$$

Additionally, the following inequalities will be useful.

Fact 3.14 ([Gal14])**.** *For every $\epsilon < 1/2$, $\binom{n}{\leq \epsilon n} \leq 2^{nH(\epsilon)}$, for $\binom{n}{\leq \epsilon n} = \sum_{i \leq \epsilon n} \binom{n}{i}$.*

Fact 3.15. *For every $\epsilon \leq 1/2$, $H(\epsilon) \leq -2\epsilon \log \epsilon$.*

4 Non-adaptive UOWHF From One-Way Functions

In this part we construct and prove the security of our non-adaptive UOWHF. This is done by combining the construction of [Hai+10] with the non-adaptive construction of UOWHF for unknown-regular one-way functions of [MZ21].

We start with the construction. Let $g\colon \{0,1\}^{m(n)} \to \{0,1\}^{k(n)}$ be a function with a sufficient gap between the real min-entropy and the max accessible entropy of g^{-1}. Let $\mathcal{H}_n = \left\{h\colon \{0,1\}^{m(n)+k(n)} \to \{0,1\}^{m(n)-\log n}\right\}$ be 2-universal hash family. For every $t \in \mathbb{N}$, define the function $C_t\colon \mathcal{H}_n^{t-1} \times (\{0,1\}^{m(n)})^t \to \mathcal{H}_n^{t-1} \times \{0,1\}^{k(n)} \times (\{0,1\}^{m(n)-\log n})^{t-1} \times \{0,1\}^{m(n)}$, by

$$C_t(h_1, \ldots, h_{t-1}, x_1, \ldots, x_t) := h_1, \ldots, h_{t-1}, g(x_1), h_1(x_1, g(x_2)), \ldots h_{t-1}(x_{t-1}, g(x_t)), x_t.$$

Note that the above function is length decreasing when $(t-1)\log n > k(n)$. The next theorem states that, for the right choice of parameters, C_t is also collision resistant.

Theorem 4.1. *Let $\ell = \ell(n)$, $s = \omega(\log n)$ and let $g\colon \{0,1\}^{m(n)} \to \{0,1\}^{k(n)}$ be a function. Assume that g^{-1} has real min-entropy at least $\ell + s$ and accessible max entropy at most ℓ. Then the function C_t is collision resistant on random inputs, for every $t \in \mathrm{poly}$.*

Corollary 4.2. *Assuming that one-way functions exist, there exists a non-adaptive UOWHF. Moreover, the construction uses $O(n^9)$ calls to the one-way function, and has key length and output length of $O(n^{10})$.*

Proof. Let $k = m = n^5$ and $t = k/\log n + 2$. The proof is immediate from Theorems 3.11 and 4.1 and Lemma 3.5, together with the fact that there is an explicit 2-universal family $\mathcal{H} = \left\{h\colon \{0,1\}^{m+k} \to \{0,1\}^{m-\log n}\right\}$ with description size $O(m+k)$. □

We now prove Theorem 4.1. Let m, k, ℓ, g and t be as in Theorem 4.1. We will need the following two claims. The first, which is straight-forward from the definition of accessible entropy, states that every collision for C_t comes from a small set.

Claim 4.3. *For every collision-finder algorithm for C_t it holds that there exists a family of sets $\{\mathcal{L}(x)\}_{x\in\{0,1\}^m}$ each of size at most 2^ℓ such that*

$$\Pr_{\substack{h:=(h_1,\ldots,h_{t-1})\leftarrow\mathcal{H}_n^{t-1},\\ x:=(x_1,\ldots,x_t)\leftarrow(\{0,1\}^{m(n)})^{t(n)}\\ (h,(x'_1,\ldots,x'_t))\leftarrow\mathsf{A}(1^n,h,x)}}\left[\exists i\in[t] \text{ s.t. } g(x'_i)=g(x_i)\wedge x'_i\notin\mathcal{L}(x_i)\right]=neg(n).$$

For the second claim we will need the following definition. Let

$$\mathcal{T}_n:=\left\{x\in\{0,1\}^{m(n)}\colon H_{X|g(X)}(x|g(x))\geq\ell+s\right\}=\left\{x\in\{0,1\}^m\colon\left|g^{-1}(g(x))\right|\geq 2^{\ell+s}\right\}.$$

That is, $\mathcal{T}$ is the set of all "typical" inputs x for g, for which $H_{X|g(X)}(x|g(x))$ is large.

The second claim consider the function C_d for every $d\in\text{poly}$. It states that for typical inputs, i.e., $x_1,\ldots,x_d\in\mathcal{T}$, there is no collision $x'_1,\ldots,x'_d$ for C_d such that x'_1 is from a small set $\mathcal{G}$.

Claim 4.4. *For every $d,n\in\mathbb{N}$, set $\mathcal{G}\subseteq\{0,1\}^{m(n)}$ of size at most $2^{\ell(n)}$ and $x=(x_1,\ldots,x_d)\in\mathcal{T}_n^d$ it holds that*

$$\Pr_{h=(h_1,\ldots,h_{d-1})\leftarrow\mathcal{H}_n^{d-1}}\left[\exists x'=(x'_1,\ldots,x'_d)\text{ s.t } x'_1\in\mathcal{G}\wedge(x'_1,g(x'_2))\neq(x_1,g(x_2))\wedge C_d(h,x)=C_d(h,x')\right]$$
$$\leq d\cdot\mu(n),$$

for some negligible function μ.

We prove Claims 4.3 and 4.4 below, but first we use them in order to prove Theorem 4.1.

Proof of Theorem 4.1. Let A be a PPT collision-finder algorithm of C_t such that

$$\Pr_{\substack{h=(h_1,\ldots,h_{t-1})\leftarrow\mathcal{H}_n^{t-1},\\ x=(x_1,\ldots,x_t)\leftarrow(\{0,1\}^{m(n)})^{t(n)}\\ (h,x')\leftarrow\mathsf{A}(1^n,h,x)}}\left[x\neq x'\wedge C_t(h,x)=C_t(h,x')\right]=\alpha(n). \tag{2}$$

We will show that α must be negligible.

For $n\in\mathbb{N}$, let $\{\mathcal{L}(x)\}_{x\in\{0,1\}^{m(n)}}$ be the family promised by Claim 4.3. Let $H=(H_1,\ldots,H_{t-1})\leftarrow\mathcal{H}_n^{t-1}$ and $X=(X_1,\ldots,X_t)\leftarrow(\{0,1\}^{m(n)})^{t(n)}$ be random variables, and let $(\cdot,X')\leftarrow\mathsf{A}(1^n,H,X)$ be the output of A. Let W_1^n be the event that A found a valid collision. By construction, this event can be written as follows: There exists $i\in[t(n)]$, such that,

1. $(X'_i,g(X'_{i+1}))\neq(X_i,g(X_{i+1}))$, and
2. $(g(X_i),H_i(X_i,g(X_{i+1})),\ldots,H_{t-1}(X_{t-1},g(X_t)),X_t)=(g(X'_i),H_i(X'_i,g(X'_{i+1})),\ldots,H_{t-1}(X'_{t-1},g(X'_t)),X'_t)$.

Observe that, by definition of the function C, the last condition is equivalent to $C_{t-i+1}(H_{i,\ldots,t-1},X_{i,\ldots,t-1})=C_{t-i+1}(H_{i,\ldots,t-1},X'_{i,\ldots,t-1})$.

Additionally, we define the following two events. Let W_2^n be the event that exists $i\in[t(n)]$ such that $X_i\notin\mathcal{T}_n$, and let W_3^n be the event that there exists $i\in[t(n)]$ such that $g(X_i)=g(X'_i)$ and $X'_i\notin\mathcal{L}(X_i)$.

It holds that,

$$\alpha\leq\Pr[W_2^n]+\Pr[W_3^n]+\Pr\left[W_1^n\wedge\overline{W_2^n}\wedge\overline{W_3^n}\right].$$

Finally, observe that $\Pr[W_2^n]=\text{neg}(n)$ by the assumption that g^{-1} has min-entropy at least $\ell+s$ and the union bound, and $\Pr[W_3^n]=\text{neg}(n)$ by Claim 4.3. Additionally, $\Pr\left[W_1^n\wedge\overline{W_2^n}\wedge\overline{W_3^n}\right]=\text{neg}(n)$ by Claim 4.4 and the union bound (choosing $\mathcal{G}=\mathcal{L}(X_i)$). □

4.1 Proving Claim 4.3

Proof of Claim 4.3. Let A be a collision-finder algorithm for C. Consider R, a collision finder for g:

Algorithm 1: The g-collision finder R

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

Oracle: A.

Operation:

1. Sample $i \leftarrow [t(n)-1]$, $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{t(n)} \leftarrow \{0,1\}^{m(n)}$ and $h_1, \ldots, h_{t(n)-1} \leftarrow \mathcal{H}_n$, and let $x_i := x$.
2. Apply $\mathsf{A}(1^n, h_1, \ldots, h_{t(n)-1}, x_1, \ldots, x_t)$ to get $x'_1, \ldots, x'_{t(n)}$.
3. If $g(x'_i) = g(x_i)$ output x'_i. Otherwise output x_i.

Note that by definition it always holds that $g(x) = g(\mathsf{R}^{\mathsf{A}}(1^n, x))$. Thus, by Theorem 3.11, there exists a family of sets $\{\mathcal{L}(x)\}_{x \in \{0,1\}^{m(n)}}$ such that $\mathbf{Pr}_{x \leftarrow \{0,1\}^{m(n)}}\left[\mathsf{R}^{\mathsf{A}}(1^n, x) \notin \mathcal{L}(x)\right] \leq neg(n)$ and $|\mathcal{L}(x)| \leq 2^{\ell(n)}$ for every $x \in \{0,1\}^{m(n)}$. Observe that,

$$\begin{aligned}
&\mathop{\mathbf{Pr}}_{x \leftarrow \{0,1\}^{m(n)}}\left[\mathsf{R}^{\mathsf{A}}(1^n, x) \notin \mathcal{L}(x)\right] \\
&\geq 1/t(n) \cdot \mathop{\mathbf{Pr}}_{\substack{h := (h_1, \ldots, h_{t(n)-1}) \leftarrow \mathcal{H}_n^{t(n)-1}, \\ x := (x_1, \ldots, x_{t(n)}) \leftarrow (\{0,1\}^{m(n)})^{t(n)} \\ (h, (x'_1, \ldots, x'_{t(n)})) \leftarrow \mathsf{A}(1^n, h, x)}}\left[\exists i \in [t(n)] \ s.t. \ g(x'_i) = g(x_i) \wedge x'_i \notin \mathcal{L}(x_i)\right]
\end{aligned}$$

and thus the claim holds by combining the two equation above, and since $t \in \text{poly}$. □

4.2 Proving Claim 4.4

Fix n, and omit it from the notation. Let $\mathcal{T} = \mathcal{T}_n$ and $\mathcal{B} := \{0,1\}^m \backslash \mathcal{T}$. Recall that, by Theorem 3.11 and the definition of real min-entropy, it holds that $|\mathcal{B}| = \varepsilon(n) \cdot 2^m$ for some $\varepsilon \in neg(n)$. Let $g(\mathcal{T}) := \{g(x) \colon x \in \mathcal{T}\}$. The next claim is the main part of the proof of Claim 4.4. It states that for every small set $\mathcal{G}$ and strings x_1, x_2, the following holds with overwhelming probability over $h \in \mathcal{H}$. For every x'_1, x'_2 such that $x'_1 \in \mathcal{G}$ and $h(x_1, g(x_2)) = h(x'_1, g(x'_2))$ it holds that x'_2 is non-typical (that is, $x'_2 \in \mathcal{B}$). Moreover, the number of such collision is small.

Claim 4.5. *Let $x_1, x_2 \in \{0,1\}^m$, and let $\mathcal{G} \subseteq \{0,1\}^m$ be a set of size at most 2^ℓ. For $h \in \mathcal{H}$, let*

$$\mathcal{G}_h = \left\{x'_2 \colon \exists x'_1 \in \mathcal{G} \ s.t. \ (x'_1, g(x'_2)) \neq (x_1, g(x_2)) \wedge h(x_1, g(x_2)) = h(x'_1, g(x'_2))\right\}.$$

Then, $\mathbf{Pr}_{h \leftarrow \mathcal{H}}\left[\mathcal{G}_h \subseteq \mathcal{B} \wedge |\mathcal{G}_h| \leq 2^\ell\right] \geq 1 - n(\varepsilon(n) + 2^{-s(n)})$.

Proof of Claim 4.5. We start with showing that $\mathbf{Pr}_{h\leftarrow\mathcal{H}}[\mathcal{G}_h \subseteq \mathcal{B}] \geq 1 - n \cdot 2^{-s(n)}$. Indeed,

$$\begin{aligned}
&\Pr_{h\leftarrow\mathcal{H}}[\mathcal{G}_h \nsubseteq \mathcal{B}] \\
&= \Pr_{h\leftarrow\mathcal{H}}\big[\exists(x_1', x_2') \in \mathcal{G} \times \mathcal{T} \ s.t. \ (x_1', g(x_2')) \neq (x_1, g(x_2)) \wedge h(x_1, g(x_2)) = h(x_1', g(x_2'))\big] \\
&= \Pr_{h\leftarrow\mathcal{H}}\big[\exists(x_1', y') \in \mathcal{G} \times g(\mathcal{T}) \ s.t. \ (x_1', y') \neq (x_1, g(x_2)) \wedge h(x_1, g(x_2)) = h(x_1', y')\big] \\
&\leq n \cdot 2^{-m} \cdot |\mathcal{G}| \cdot |g(\mathcal{T})| \\
&\leq n \cdot 2^{-m} \cdot 2^{\ell} \cdot 2^m / 2^{\ell+s} \\
&= n \cdot 2^{-s(n)}
\end{aligned}$$

where the first inequality holds since $\mathbf{Pr}_{h\leftarrow\mathcal{H}}[h(x_1, g(x_2)) = h(x_1', y')] \leq n \cdot 2^{-m}$ for every $(x_1', y') \neq (x_1, g(x_2))$ together with the union bound. The second inequality holds since by definition of $\mathcal{T}$ it must hold that $|g(\mathcal{T})| \leq 2^m/2^{\ell+s}$.

We next show that $\mathbf{Pr}_{h\leftarrow\mathcal{H}}\big[|\mathcal{G}_h \cap \mathcal{B}| \geq 2^{\ell}\big] \leq n \cdot \varepsilon(n)$, which concludes the proof. We start with computing the expectation of $|\mathcal{G}_h \cap \mathcal{B}|$:

$$\begin{aligned}
\mathop{\mathbf{E}}_{h\leftarrow\mathcal{H}}[|\mathcal{G}_h \cap \mathcal{B}|] &\leq n \cdot 2^{-m} \cdot |\mathcal{G}| \cdot |\mathcal{B}| \\
&\leq n \cdot 2^{-m} \cdot 2^{\ell} \cdot \varepsilon(n) \cdot 2^m \\
&\leq n \cdot \varepsilon(n) \cdot 2^{\ell}.
\end{aligned}$$

The claim now follows by Markov and the Union bound. □

We are now ready to prove Claim 4.4 using Claim 4.5. Intuitively, Claim 4.5 shows that if x_1' is from a small set, x_2' is from a small set too. Thus, we can continue by induction, to prove that also x_d' is from the set $\mathcal{B}$. It follows that, $x_d' \neq x_d$ with overwhelming probability (as $x_d \in \mathcal{T}$), which is enough since the output of C_d includes x_d.

Proof of Claim 4.4. Fix $n \in \mathbb{N}$, $x = (x_1, \ldots, x_d) \in \mathcal{T}_n^d$ and a set $\mathcal{G} \subseteq \{0,1\}^m$. For $h = (h_1, \ldots, h_{d-1}) \in H^{d-1}$, let

$$\mathcal{COL}(h, x) = \Big\{x' = (x_1', \ldots, x_d') \in \mathcal{G} \times (\{0,1\}^m)^{d-1} \colon (x_1, g(x_2)) \neq (x_1', g(x_2')) \wedge C_d(h, x) = C_d(h, x')\Big\}$$

be the set containing all the possible collision of h, x with $x_1' \in \mathcal{G}$ and $(x_1, g(x_2)) \neq (x_1', g(x_2'))$. Similarly, for every $i \in \{0, \ldots, d-1\}$, let

$$\mathcal{COL}_i(h_1, \ldots, h_i, x) = \Big\{x' \in \mathcal{G} \times (\{0,1\}^m)^{d-1} \colon {\scriptstyle \begin{array}{c} (x_1, g(x_2)) \neq (x_1', g(x_2')) \\ \wedge \forall j \in [i] \ \ h_j(x_j, g(x_{j+1})) = h_j(x_j', g(x_{j+1}')) \end{array}}\Big\}$$

That is, all inputs with $x_1' \in \mathcal{G}$ and $(x_1, g(x_2)) \neq (x_1', g(x_2'))$ that collide with i blocks of C_d. It is clear that for every x and h,

$$\mathcal{COL}(h, x) \subseteq \mathcal{COL}_{d-1}(h, x) \subseteq \ldots \subseteq \mathcal{COL}_0(x) \tag{3}$$

We want to show that with high probability over the choice of h, it holds that $\mathcal{COL}(h, x)$ is empty.

For every $i \in [d-1]$, let W_i be the event (over the choice of $h_1, \dots h_{i-1} \leftarrow \mathcal{H}^{i-1}$) that there exists a set $\mathcal{G}_i$ of size at most 2^ℓ, such that for every $x' \in \mathcal{COL}_{i-1}(h_1, \dots, h_{i-1}, x)$, it holds that $(x'_i, g(x'_{i+1})) \neq (x_i, g(x_{i+1}))$ and $x'_i \in \mathcal{G}_i$.

For $i \in [d]$, let $\widehat{W}_i$ be the event that there exists a set $\mathcal{G}_i$ of size at most 2^ℓ such that for every $x' \in \mathcal{COL}_{i-1}(h_1, \dots, h_{i-1}, x)$, it holds that $x'_i \neq x_i$ and $x'_i \in \mathcal{G}_i$.

Observe that $\mathbf{Pr}\left[W_i \mid \widehat{W}_i\right] = 1$. We will show that, for every $1 \leq i < d$, it holds that

$$\mathbf{Pr}\left[\widehat{W}_{i+1} \mid W_{\leq i}\right] \geq 1 - n(\varepsilon(n) + 2^{-s(n)}) \tag{4}$$

Furthermore, $\mathbf{Pr}[W_1] = 1$. Indeed, let $\mathcal{G}_1 = \mathcal{G}$. By assumption $x_1 \in \mathcal{G}$ and $(x'_1, g(x'_2)) \neq (x_1, g(x_2))$ for every $x'_1, x'_2 \in \mathcal{COL}_0(x)$.

To see that Equation (4) holds, fix $1 \leq i < d$. Let $H' \leftarrow \mathcal{H}^{d-1}|_{W_{\leq i}}$, and observe that H'_i is uniformly distributed over $\mathcal{H}$. By the definition of W_i, it holds that for every $x' \in \mathcal{COL}_{i-1}(H'_{<i}, x)$ it holds that $(x'_i, g(x'_{i+1})) \neq (x_i, g(x_{i+1}))$ and $x'_i \in \mathcal{G}_i$ for some set $\mathcal{G}_i$ of size at most 2^ℓ. Define

$$\mathcal{G}_{i+1} := \left\{x'_{i+1} \colon \exists x'_i \in \mathcal{G}_i \ s.t.\ (x'_i, g(x'_{i+1})) \neq (x_i, g(x_{i+1})) \wedge H'_i(x_i, g(x_{i+1})) = H'_i(x'_i, g(x'_{i+1}))\right\}.$$

By definition $x'_{i+1} \in \mathcal{G}_{i+1}$ for every $x' \in \mathcal{COL}_i(H'_{\leq i}, x)$. Applying Claim 4.5 we get that with all but $n(\varepsilon(n) + 2^{-s(n)})$ probability over the choice of H'_i, it holds that $|\mathcal{G}_{i+1}| \leq 2^\ell$. Moreover, with the same probability $\mathcal{G}_{i+1} \subseteq \mathcal{B}$, which implies that $x'_{i+1} \neq x_{i+1}$ (since by assumption, $x_{i+1} \in \mathcal{T}_n$).

To conclude, we get that for every $x' \in \mathcal{COL}(h, x) \subseteq \mathcal{COL}_{d-1}(h, x)$ it holds that $x'_d \neq x_d$ with probability at least

$$\begin{aligned}
\mathbf{Pr}\left[\widehat{W}_d\right] &\geq \mathbf{Pr}\left[\widehat{W}_d \mid W_{<d}\right] \cdot \prod_{1<i\leq d-1} \mathbf{Pr}[W_i \mid W_{<i}] \\
&\geq \mathbf{Pr}\left[\widehat{W}_d \mid W_{<d}\right] \cdot \prod_{1<i\leq d-1} \mathbf{Pr}\left[W_i, \widehat{W}_i \mid W_{<i}\right] \\
&= \mathbf{Pr}\left[\widehat{W}_d \mid W_{<d}\right] \cdot \prod_{1<i\leq d-1} \left(\mathbf{Pr}\left[W_i \mid \widehat{W}_i, W_{<i}\right] \cdot \mathbf{Pr}\left[\widehat{W}_i \mid W_{<i}\right]\right) \\
&\geq (1 - n(\varepsilon(n) + 2^{-s(n)}))^d \\
&\geq 1 - d \cdot n(\varepsilon(n) + 2^{-s(n)}) \\
&= 1 - d \cdot \mathrm{neg}(n).
\end{aligned}$$

Where the penultimate inequality holds by Equation (4) and the fact that $\mathbf{Pr}\left[W_i \mid \widehat{W}_i, W_{<i}\right] = 1$. Recall that C_d outputs x_d. Thus, the above implies that $C_d(h, x) \neq C_d(h, x')$, which implies that $\mathcal{COL}(h, x) = \emptyset$ with the same probability.

□

5 Almost-UOWHF From One-Way Functions

In this section we formally define almost-UOWHF and next-bit unreachable entropy, and show how to construct them from one-way functions.

5.1 Almost-UOWHF

In this part we formally define almost-UOWHF. We start with the definition of almost collision resistance on random input.

Definition 5.1 (Almost collision resistance on random inputs). *Let n be a security parameter. A function $f\colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}$ is* almost collision resistant on random inputs *if there exists a set $\mathcal{B}_n \subseteq \{0,1\}^{m(n)}$, such that $|\mathcal{B}_n|/2^{m(n)} = \mathrm{neg}(n)$, and for every probabilistic polynomial-time adversary A, the probability that A succeeds in the following game is negligible in n:*

1. *Choose $x \leftarrow \{0,1\}^{m(n)}$.*
2. *Let $x' \leftarrow A(1^n, x) \in \{0,1\}^{m(n)}$.*
3. *A succeeds if $x' \notin \mathcal{B}_n$, $x \neq x'$ and $f(x) = f(x')$.*

The definition of almost-UOWHF is similar.

Definition 5.2 (Almost universal one-way hash function). *Let n be a security parameter. A family of functions $\mathcal{F} = \left\{f_z\colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}\right\}_{z\in\{0,1\}^{k(n)}}$ is a family of* almost universal one-way hash functions *(UOWHFs) if it satisfies:*

1. *Efficiency: Given $z \in \{0,1\}^{k(n)}$ and $x \in \{0,1\}^{m(n)}$, $f_z(x)$ can be evaluated in time $\mathrm{poly}(n)$.*
2. *Shrinking: $\ell(k) < m(k)$.*
3. *Almost Target Collision Resistance: There exist sets $\{\mathcal{B}_z\}_{z\in\{0,1\}^{k(n)}}$ such that $|\mathcal{B}_z|/2^{m(n)} = \mathrm{neg}(n)$, and for every probabilistic polynomial-time adversary A, the probability that A succeeds in the following game is negligible in n:*
   - *(a) Let $(x, state) \leftarrow \mathsf{A}(1^n) \in \{0,1\}^{m(n)} \times \{0,1\}^*$.*
   - *(b) Choose $z \leftarrow \{0,1\}^{k(n)}$.*
   - *(c) Let $x' \leftarrow A(state, z) \in \{0,1\}^{m(k)}$.*
   - *(d) A succeeds if $x' \notin \mathcal{B}_z$, $x \neq x'$ and $f_z(x) = f_z(x')$.*

The next lemma follows easily from Lemma 3.5

Lemma 5.3 (From random inputs to targets, almost version). *Let n be a security parameter. Let $F\colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}$ be an efficiently computable length-decreasing function. Suppose F is almost collision-resistant on random inputs. Then $\left\{F_y\colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}\right\}_{y\in\{0,1\}^{m(n)}}$, for $F_y(x) := F(y \oplus x)$, is an almost-UOWHF.*

5.2 Next-Bit Unreachable Entropy

In this section we present the notion of next-bit unreachable entropy, and construct a function with next-bit unreachable entropy from one-way functions. Intuitively, we say that a function $g\colon \{0,1\}^m \to \{0,1\}^\ell$ has next-bit unreachable entropy v if for every $i \in [\ell]$, there is a set of $\mathcal{U}_i \subseteq \{0,1\}^m$, such that every x is a member of $(\ell - m + v)$ such sets, and, given $x \in \{0,1\}^m$, a poly-time algorithm cannot find $x' \in \mathcal{U}_i$ with $g(x)_{<i} = g(x')_{<i}$, but $g(x)_i \neq g(x')_i$.

Definition 5.4. *A function* $g\colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}$ *has next-bit unreachable entropy* v*, if the following holds. For every* $n \in \mathbb{N}$ *and* $i \in [\ell(n)]$ *there exists a set* $\mathcal{U}_{i,n} \subseteq \{0,1\}^{m(n)}$*, such that*

1. $\mathcal{U}_{i,n}$ are large: *For every* $n \in \mathbb{N}$*,*
$$\Pr_{x \leftarrow \{0,1\}^{m(n)}, i \leftarrow [\ell(n)]}[x \notin \mathcal{U}_{i,n}] \leq (m(n) - v(n))/\ell(n).$$

2. Hard to get inside $\mathcal{U}_{i,n}$: *For every* PPT A*,*
$$\Pr_{x \leftarrow \{0,1\}^{m(n)}, i \leftarrow [\ell(n)], x' \leftarrow \mathsf{A}(1^n, x, i)}\left[((x' \in \mathcal{U}_{i,n})) \wedge (g(x)_{<i} = g(x')_{<i}) \wedge (x \notin \mathcal{U}_{i,n})\right] = \operatorname{neg}(n).$$

3. The entropy inside $\mathcal{U}_{i,n}$ is unreachable: *For every* PPT A*,*
$$\Pr_{x \leftarrow \{0,1\}^{m(n)}, i \leftarrow [\ell(n)], x' \leftarrow \mathsf{A}(1^n, x, i)}\left[((x' \in \mathcal{U}_{i,n})) \wedge (g(x)_{<i} = g(x')_{<i}) \wedge (g(x)_i \neq g(x')_i)\right] = \operatorname{neg}(n).$$

The definition above is especially useful when the function g is close to be injective. Formally,

Definition 5.5. *A function* g *is almost-injective if* $\mathbf{Pr}_{x \leftarrow \{0,1\}^m}\left[\left|g^{-1}(g(x))\right| > 1\right] = neg(n)$*.*

We use the above definition for the construction of almost-UOWHF in Section 5.3, but first we prove that a function with non-trivial next-bit unreachable entropy exists.

Theorem 5.6. *Let* f *and* $\mathcal{H} = \{h : \{0,1\}^n \to \{0,1\}^n\}$ *be as in Theorem 3.12. Let* $g\colon \mathcal{H}^2 \times \{0,1\}^n \to \mathcal{H}^2 \times \{0,1\}^{2n}$ *be defined by* $g(h_1, h_2, x) = (h_1, h_2, h_1(f(x)), h_2(x))$*. Then* g *is almost-injective function with next-bit unreachable entropy* $c\log(n)$*, for every constant* $c > 0$*.*

Moreover, the input and output size of g *are of length* $O(n)$*.*

The proof of Theorem 5.6 is by the next two claims. The first, proven on Appendix A, states that the function g defined in Theorem 5.6 is indeed almost injective.

Claim 5.7. *g is almost injective function.*

Next, in order to prove that g has next-bit unreachable entropy, we define the sets $\mathcal{U}_{i,n}$. Let d be the description size of a function in $\mathcal{H}$. In the following we view g as a function from $\{0,1\}^{2d+n}$ to $\{0,1\}^{2d+2n}$. Fix c and n, and define, for every $i \in [n]$,

$$\mathcal{U}_{2d+i,n} = \left\{(h_1, h_2, x) \in \mathcal{H}^2 \times \{0,1\}^{m(n)} \colon i > n - \log\left|f^{-1}(f(x))\right| - (c+4)\log n\right\}$$

and

$$\mathcal{U}_{2d+n+i,n} = \left\{(h_1, h_2, x) \in \mathcal{H}^2 \times \{0,1\}^{m(n)} \colon \forall x' \neq x, (h_1(f(x)), h_2(x)_{<i}) \neq (h_1(f(x')), h_2(x')_{<i})\right\}.$$

Lastly, for every $i \leq 2d$, let $\mathcal{U}_{i,n} = \emptyset$. Note that by definition, the above $\mathcal{U}_{2d+n+i,n}$ fulfil Items 2 and 3 of Definition 5.4 for any (unbounded) adversary. By Theorem 3.12, it holds that also $\mathcal{U}_{2d+i,n}$ fulfil Items 2 and 3 of Definition 5.4 (observe that, for x and x' with $f(x) = f(x')$, it holds that $x \in \mathcal{U}_{2d+i,n}$ iff $x' \in \mathcal{U}_{2d+i,n}$). The next claim states that also Item 1 in Definition 5.4 holds.

Claim 5.8 ($\mathcal{U}_{i,n}$ is large)**.** *Let $\ell = 2d + 2n$. In holds that*

$$\Pr_{(h_1,h_2,x)\leftarrow\mathcal{H}^2\times\{0,1\}^n, i\leftarrow[\ell(n)]}[(h_1, h_2, x) \in \mathcal{U}_{i,n}] \geq (n + c\log n)/\ell.$$

Proof. It holds that,

$$\begin{aligned}
&\Pr_{(h_1,h_2,x)\leftarrow\mathcal{H}^2\times\{0,1\}^n, i\leftarrow[\ell(n)]}[(h_1, h_2, x) \in \mathcal{U}_{i,n}] \\
&= \mathop{\mathbf{E}}_{x\leftarrow\{0,1\}^n, i\leftarrow[\ell(n)]}\left[\Pr_{(h_1,h_2)\leftarrow\mathcal{H}^2}[(h_1, h_2, x) \in \mathcal{U}_{i,n}]\right] \\
&= 1/\ell \cdot \mathop{\mathbf{E}}_{x\leftarrow\{0,1\}^n}\left[\sum_{i\in[\ell(n)]} \Pr_{(h_1,h_2)\leftarrow\mathcal{H}^2}[(h_1, h_2, x) \in \mathcal{U}_{i,n}]\right] \\
&= 1/\ell \cdot \mathop{\mathbf{E}}_{x\leftarrow\{0,1\}^n}\left[\sum_{i\leq 2d+n} \Pr_{(h_1,h_2)\leftarrow\mathcal{H}^2}[(h_1, h_2, x) \in \mathcal{U}_{i,n}] + \sum_{i>2d+n} \Pr_{(h_1,h_2)\leftarrow\mathcal{H}^2}[(h_1, h_2, x) \in \mathcal{U}_{i,n}]\right] \\
&\geq 1/\ell \cdot \mathop{\mathbf{E}}_{x\leftarrow\{0,1\}^n}\left[(c+4)\log n + \log\left|f^{-1}(f(x))\right| - 1 + \sum_{i>2d+n} \Pr_{(h_1,h_2)\leftarrow\mathcal{H}^2}[(h_1, h_2, x) \in \mathcal{U}_{i,n}]\right]
\end{aligned}$$

We finish the proof by showing that for every x,

$$\sum_{i>2d+n} \Pr_{(h_1,h_2)\leftarrow\mathcal{H}^2}[(h_1, h_2, x) \in \mathcal{U}_{i,n}] \geq n - \log\left|f^{-1}(f(x))\right| - 3\log n - 5 \tag{5}$$

Indeed, fix $i > 2d + n + \log\left|f^{-1}(f(x))\right| + 3\log n$, and let W be the event that

$$\left|\left\{x' \colon h_1(f(x')) = h_1(f(x))\right\}\right| < \left|f^{-1}(f(x))\right| + n^2.$$

It holds that,

$$\begin{aligned}
&\Pr_{(h_1,h_2)\leftarrow\mathcal{H}^2}[(h_1, h_2, x) \in \mathcal{U}_{i,n}] \\
&= 1 - \Pr_{(h_1,h_2)\leftarrow\mathcal{H}^2}\left[\exists x' \neq x, g(h_1, h_2, x)_{<i} = g(h_1, h_2, x)_{<i}\right] \\
&\geq 1 - \Pr_{(h_1,h_2)\leftarrow\mathcal{H}^2}\left[\exists x' \neq x, h_1(f(x)) = h_1(f(x')) \wedge h_2(x)_{<i-n-2d}) = h_2(x')_{<i-n-2d} \mid W\right] \\
&\quad - \Pr_{(h_1,h_2)\leftarrow\mathcal{H}^2}\left[\overline{W}\right] \\
&\geq 1 - 2^{i-n-2d-1} \cdot (\left|f^{-1}(f(x))\right| + n^2) - 2/n \\
&\geq 1 - 4/n,
\end{aligned}$$

where the first equality holds by definition of $\mathcal{U}_{i,n}$, the second inequality by Lemma 3.7 and since $\mathcal{H}$ is two-wise independent, and the last by the bound on i.

Finally, Equation (5) holds since, by the above,

$$\begin{aligned}
\sum_{i>2d+n} \Pr_{(h_1,h_2)\leftarrow\mathcal{H}^2}[(h_1, h_2, x) \in \mathcal{U}_{i,n}] &\geq \sum_{i>2d+n+\log|f^{-1}(f(x))|+3\log n} (1 - 4/n) \\
&\geq (n - \log\left|f^{-1}(f(x))\right| - 3\log n - 1) - 4.
\end{aligned}$$

$\square$

The proof of Theorem 5.6 is now immediate.

Proof of Theorem 5.6. Immediate from Claim 5.7, Claim 5.8 and Theorem 3.12. $\square$

5.3 Next-Bit Unreachable Entropy to Almost-UOWHF

5.3.1 The Construction

We now describe our main construction. We start with some notations.

A *position vector* $p \in [\ell]^q$ is just a vector of indexes from $[\ell]$. For a function $g\colon \{0,1\}^m \to \{0,1\}^\ell$, input vector $w = (x_1, \ldots x_q) \in (\{0,1\}^m)^q$ and a position vector $p = (i_1, \ldots, i_q) \in [\ell]^q$, let $g_p(w) := g(x_1)_{i_1}, \ldots, g(x_q)_{i_q}$. Similarly, define $g_{<p}(w) := g(x_1)_{<i_1}, \ldots, g(x_q)_{<i_q}$, and $g_{\geq p}(w)$ analogously. For a number $k \in \mathbb{N}$, let $p + k := (i_1 + k, \ldots, i_q + k)$. For a number t, let $g^t\colon \{0,1\}^{tm} \to \{0,1\}^{t\ell}$ be the t-fold repetition of g, i.e., $g^t(x_1, \ldots, x_t) = g(x_1), \ldots, g(x_t)$.

We are now ready to present the construction (see Figure 1).

Construction 5.9 (Almost-UOWHF)**.** *Let n be a security parameter, and let $q = q(n), t = t(n)$ and $k = k(n)$ be parameters. Let $g : \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}$ be a function, and let $\mathcal{H}_n = \left\{h : \{0,1\}^{q(n)} \to \{0,1\}^{k(n)}\right\}$ be a 2-universal hash family. Define the function $C\colon \mathcal{H}_n \times [\ell(n)]^{q(n)} \times (\{0,1\}^{m(n)\cdot t(n)})^{q(n)} \to \mathcal{H}_n \times [\ell(n)]^{q(n)} \times \{0,1\}^{\ell(n)\cdot q(n)+(t(n)-1)\ell\cdot k(n)}$ by*

$$C(h, p, z) := h, p, g'_{<p}(z), h(g'_p(z)), h(g'_{p+1}(z)), \ldots, h(g'_{p+(t-1)\ell-1}(z)), g'_{\geq p+(t-1)\ell}(z),$$

for $g' = g^t$.

Our main theorem states that for the right choice of parameters, the above construction is indeed an almost-UOWHF.

Theorem 5.10. *Let $g\colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}$ be an efficient, almost-injective function with next-bit unreachable entropy $v(n) \in \mathbb{N}$. For every $q \in$ poly and $\varepsilon \in 1/\mathrm{poly}$ such that $H(4\varepsilon(n)) \leq 0.1v(n)/\ell(n)$, $q = \omega\left(\log n \cdot \max\left\{\ell, \frac{\ell}{\varepsilon^2(\ell-m-v)}\right\}\right)$ and for $k = q(m - v/3)/\ell$, $t = 3(\ell - m)/v + 2$ the function C as in Construction 5.9 is efficient, shrinking and almost collision resistant on random inputs.*

We prove Theorem 5.10 below, but first we use Theorem 5.6 and Construction 5.9 in order to conclude the parameters of our almost-UOWHF construction.

Corollary 5.11. *Let $s = \omega(1)$. Assuming that one-way functions exist, there exists an almost-UOWHF with key length $O(n^4 \cdot s)$. Moreover, the almost-UOWHF makes $O(n^3 \cdot s)$ non-adaptive calls to the underlying one-way function.*

Proof. By Theorem 5.6, assuming one-way function exists, there is an alomost-injective function $g\colon \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}$ with next-bit unreachable entropy $v(n) = 3\log n$, for $m(n) = O(n)$, $\ell(n) = O(n)$ and $\ell(n) = \Omega(\ell(n) - m(n) - v(n))$. Thus, using Fact 3.15 we can take $\varepsilon = \Omega(1/\ell(n))$, $q = O(s \cdot n^2 \log n)$ and $t = O(n/\log n)$.

The corollary thus follows from Theorem 5.10 and Lemma 5.3, and the fact that Construction 5.9 uses $q \cdot t$ non-adaptive calls to the g. $\square$

In the rest of this section we prove Theorem 5.10. We start with analyzing a simpler function, for which we only consider one hashed column.

5.3.2 Analyzing Single Column

In this part we analyze a simpler function. Let g, q, k and $\mathcal{H}_n$ be as in Construction 5.9. Consider the function $C^1 \colon \mathcal{H}_n \times [\ell(n)]^{q(n)} \times (\{0,1\}^{m(n)})^{q(n)} \to \mathcal{H}_n \times [\ell(n)]^{q(n)} \times \{0,1\}^{k(n)}$ defined by

$$C^1(h,p,w) := h, p, g_{<p}(w), h(g_p(w)),$$

Let $\mathcal{D}_n := \mathcal{H}_n \times [\ell(n)]^{q(n)} \times (\{0,1\}^{m(n)})^{q(n)}$ be the domain of C^1. The next lemma is the crux of the proof of Theorem 5.10, and it is the analog of it with respect to the function C^1.

Lemma 5.12. *Let g, q and k be as in Theorem 5.10. For every $n \in \mathbb{N}$ there exists a set $\mathcal{B}_n \subseteq [\ell(n)]^{q(n)} \times (\{0,1\}^{m(n)})^{q(n)}$ such that $|\mathcal{B}_n| = neg(n) \cdot \left|[\ell(n)]^{q(n)} \times (\{0,1\}^{m(n)})^{q(n)}\right|$, and for every* PPT *algorithm* A*,*

$$\Pr_{\substack{(h,p,w)\leftarrow\mathcal{D}_n\\ w'\leftarrow\mathsf{A}(1^n,h,p,w)}}\left[C^1(h,p,w) = C^1(h,p,w') \wedge (p,w) \notin \mathcal{B}_n \wedge g_p(w) \neq g_p(w')\right] = \operatorname{neg}(n).$$

That is, there is a negligible size set, such that no efficient algorithm can find a collision for C^1 outside this set, unless the value of the last column $g_p(w)$ stays the same.

In order to prove Lemma 5.12, let g, q, k, v and ϵ be as in Theorem 5.10.

Defining $\mathcal{B}_n$. We start with the definition of $\mathcal{B}_n$. Since g has next-bit unreachable entropy $v(n)$, for every $n \in \mathbb{N}$ there is a collection of sets $\{\mathcal{U}_{i,n}\}_{i\in[\ell(n)]}$ that fulfils the definition of Definition 5.4. We further assume without loss of generality that $\mathbf{Pr}_{x\leftarrow\{0,1\}^{m(n)}, i\leftarrow[\ell(n)]}[x \notin \mathcal{U}_{i,n}] = (m(n) - v(n))/\ell(n)$, as we can take v to be larger if the above equality does not holds.

Fix n and omit it from the notation. For $p = (i_1, \ldots, i_q) \in [\ell]^q$ and $w = (x_1, \ldots, x_q) \in (\{0,1\}^m)^q$, let

$$\mathcal{J}_p(w) := \left\{ j \in [q] : x_j \in \mathcal{U}_{i_j} \right\}$$

Note that by definition, the expected size of $\mathcal{J}_p(w)$ is $q(\ell - m + v)/\ell$. Thus, we define

$$\mathcal{B}_n := \left\{ (p,w) \in [\ell(n)]^{q(n)} \times (\{0,1\}^{m(n)})^{q(n)} : |\mathcal{J}_p(w)| < (1-\varepsilon)q \cdot (\ell - m + v)/\ell \right\}.$$

The next claim, proved below by a simple use of Chernoff bound, implies that the size of $\mathcal{B}_n$ is indeed negligible.

Claim 5.13.

$$\Pr_{(\cdot,p,w)\leftarrow\mathcal{D}}[|\mathcal{J}_p(w)| \notin (1 \pm \varepsilon) \cdot q \cdot (\ell - m + v)/\ell] \leq 2^{-\varepsilon^2 \cdot q \cdot (\ell-m+v)/3\ell+1}.$$

Next, fix a PPT collision finder algorithm A for C^1.The proof of Lemma 5.12 now follows by the next two claims. The first shows that all the collisions A can find have a specific structure. For $p \in [\ell]^q$ and $w \in (\{0,1\}^m)^q$, let

$$\mathcal{T}_{p,w} := \left\{ w' = (x'_1, \ldots, x'_q) \colon \mathcal{J}_p(w') \subseteq \mathcal{J}_p(w) \wedge \forall j \in \mathcal{J}_p(w'),\ g(x'_j)_{i_j} = g(x_j)_{i_j} \right\}.$$

Claim 5.14.

$$\Pr_{\substack{(h,p,w)\leftarrow\mathcal{D}_n\\ (h,p,w')\leftarrow\mathsf{A}(1^n,h,p,w)}}\left[w' \notin \mathcal{T}_{p,w}\right] = \operatorname{neg}(n).$$

In the following, for $p \in [\ell(n)]^{q(n)}$, let $\mathcal{B}_p = \{w\colon (w,p) \in \mathcal{B}_n\}$. The second claim states that with all but negligible probability, there is no collision for $C^1(z,p,w)$ in the set $\mathcal{T}_{p,w} \setminus \mathcal{B}_p$.

Claim 5.15.

$$\Pr_{(h,p,w)\leftarrow\mathcal{D}_n}\left[\exists w' \in (\mathcal{T}_{p,w} \setminus \mathcal{B}_p) \textit{ s.t. } C^1(h,p,w) = C^1(h,p,w') \textit{ and } g_p(w) \neq g_p(w')\right] = \operatorname{neg}(n).$$

We prove Claims 5.13 to 5.15 below, but first we use it in order to prove Lemma 5.12.

Proof of Lemma 5.12. For every $n \in \mathbb{N}$, let $\mathcal{B}_n$ be as defined above. By Claim 5.13, the size of $\mathcal{B}_n$ is at most $2^{-\varepsilon^2 q(\ell-m+v)/4\ell} \cdot \left|[\ell(n)]^{q(n)} \times (\{0,1\}^{m(n)})^{q(n)}\right| = \operatorname{neg}(n) \cdot \left|[\ell(n)]^{q(n)} \times (\{0,1\}^{m(n)})^{q(n)}\right|$ by our choice of parameters.

Next, fix a PPT collision finder algorithm A.[13]It follows that,

$$\begin{aligned}
&\Pr_{\substack{(h,p,w)\leftarrow\mathcal{D}_n\\(h,p,w')\leftarrow\mathsf{A}(1^n,h,p,w)}}\left[C^1(h,p,w) = C^1(h,p,w') \wedge (p,w) \notin \mathcal{B}_n \wedge g_p(w) \neq g_p(w')\right]\\
&\leq \Pr_{\substack{(h,p,w)\leftarrow\mathcal{D}_n\\(h,p,w')\leftarrow\mathsf{A}(1^n,h,p,w)}}\left[w' \notin \mathcal{T}_{p,w}\right] + \Pr_{\substack{(h,p,w)\leftarrow\mathcal{D}_n\\(h,p,w')\leftarrow\mathsf{A}(1^n,h,p,w)}}\left[w' \in (\mathcal{T}_{p,w} \setminus \mathcal{B}_p) \text{ and } g_p(w) \neq g_p(w')\right]\\
&\leq \operatorname{neg}(n) + \Pr_{(h,p,w)\leftarrow\mathcal{D}_n}\left[\exists w' \in (\mathcal{T}_{p,w} \setminus \mathcal{B}_p) \text{ s.t. } C^1(h,p,w) = C^1(h,p,w') \text{ and } g_p(w) \neq g_p(w')\right]\\
&\leq \operatorname{neg}(n),
\end{aligned}$$

where the second inequality holds by Claim 5.14 and by our assumption that A is a collision finder, and the last inequality follows by Claim 5.15. □

Proving Claim 5.13.

Proof of Claim 5.13. For each $j \in [q]$, let Z_j be an indicator for the event that $x_j \in \mathcal{U}_{i_j}$. It holds that $\Pr[Z_j = 1] = (\ell - m + v)/\ell$ for every $j \in [q]$, and $Z_1, \ldots, Z_q$ are independent. Thus, by Fact 3.13,

$$\begin{aligned}
\Pr_{(\cdot,p,w)\leftarrow\mathcal{D}}[|\mathcal{J}_p(w)| \notin (1 \pm \varepsilon) \cdot q \cdot (\ell - m + v)/\ell] &= \Pr\left[\left|\sum Z_j - \mathbf{E}\left[\sum Z_j\right]\right| \geq \varepsilon \cdot \mathbf{E}\left[\sum Z_j\right]\right]\\
&\leq 2 \cdot e^{-\varepsilon^2 \cdot \mathbf{E}[\sum Z_j]/3}.
\end{aligned}$$

□

Proving Claim 5.14.

Proof of Claim 5.14. Let A be a PPT collision finder for C^1. Assume toward a contradiction that

$$\Pr_{\substack{(h,p,w)\leftarrow\mathcal{D}_n\\(h,p,w')\leftarrow\mathsf{A}(1^n,h,p,w)}}\left[w' \notin \mathcal{T}_{p,w}\right] \geq 1/n^c \tag{6}$$

for some $c \in \mathbb{N}$ and for infinitely many n's. The following algorithm R breaks the assumed next-bit unreachable entropy of g.

[13] Note that the assumption that A is a collision finder is without loss of generality.

Algorithm 2: The reduction R

Input: 1^n, $x \in \{0,1\}^{m(n)}$,$i \in [\ell(n)]$

Oracle : A

Sample $h \leftarrow \mathcal{H}_n$ and $j \leftarrow [q(n)]$.

Sample $x_1, \ldots x_{j-1}, x_{j+1}, \ldots, x_{q(n)} \leftarrow \{0,1\}^{m(n)}$ and set $x_j = x$. Let $w := (x_1, \ldots, x_{q(n)})$

Sample $p_1, \ldots p_{j-1}, p_{j+1}, \ldots, p_{q(n)} \leftarrow [\ell]^{m(n)}$ and set $p_j = i$. Let $p = (p_1, \ldots, p_{q(n)}$

Compute $(h, p, w') = \mathsf{A}(1^n, h, p, w)$ and output w'_j.

That is, R gets x and i as inputs, and plant it in a random position j inside a random input for A. It then executes A in order to get its output w', and outputs x'_j.

Fix $n \in \mathbb{N}$ such that Equation (6) holds. We next show that for such n, either

$$\alpha_n := \Pr_{\substack{x \leftarrow \{0,1\}^{m(n)}, \\ i \leftarrow [\ell(n)], \\ x' \leftarrow \mathsf{R}^{\mathsf{A}}(1^n, x, i)}} \left[(x' \in \mathcal{U}_{i,n}) \wedge (g(x)_{<i} = g(x')_{<i}) \wedge (x \notin \mathcal{U}_{i,n})\right] \geq 1/2q \cdot 1/n^c$$

or

$$\beta_n := \Pr_{\substack{x \leftarrow \{0,1\}^{m(n)}, \\ i \leftarrow [\ell(n)], \\ x' \leftarrow \mathsf{R}^{\mathsf{A}}(1^n, x, i)}} \left[(x' \in \mathcal{U}_{i,n}) \wedge (g(x)_{<i} = g(x')_{<i}) \wedge (g(x)_i \neq g(x')_i)\right] \geq 1/2q \cdot 1/n^c$$

The above contradicts the assumed next-bit unreachable entropy of g, since $q \in \text{poly}$ and at least one of the above must happen for infinitely many n's.

Indeed,

$$\begin{aligned}
&\alpha_n + \beta_n \\
&\geq \Pr_{\substack{x \leftarrow \{0,1\}^{m(n)}, \\ i \leftarrow [\ell(n)], \\ x' \leftarrow \mathsf{R}^{\mathsf{A}}(1^n, x, i)}} \left[(x' \in \mathcal{U}_{i,n}) \wedge (g(x)_{<i} = g(x')_{<i}) \wedge \left((x \notin \mathcal{U}_{i,n}) \vee (g(x)_i \neq g(x')_i)\right)\right] \\
&\geq \Pr_{\substack{(h,p,w) \leftarrow \mathcal{D}_n, \\ j \leftarrow [q(n)], \\ (h,p,w') \leftarrow \mathsf{A}(1^n, h, p, w)}} \left[(w'_j \in \mathcal{U}_{p_j,n}) \wedge (g(w_j)_{<p_j} = g(w'_j)_{<p_j}) \wedge \left((w_j \notin \mathcal{U}_{p_j,n}) \vee (g(w_j)_{p_j} \neq g(w'_j)_{p_j})\right)\right] \\
&\geq \Pr_{\substack{(h,p,w) \leftarrow \mathcal{D}_n, \\ j \leftarrow [q(n)], \\ (h,p,w') \leftarrow \mathsf{A}(1^n, h, p, w)}} \left[(w'_j \in \mathcal{U}_{p_j,n}) \wedge \left((w_j \notin \mathcal{U}_{p_j,n}) \vee (g(w_j)_{p_j} \neq g(w'_j)_{p_j})\right)\right] \\
&\geq 1/q \cdot \Pr_{\substack{(h,p,w) \leftarrow \mathcal{D}_n, \\ (h,p,w') \leftarrow \mathsf{A}(1^n, h, p, w)}} \left[\exists j \in [q(n)] \text{ s.t. } (w'_j \in \mathcal{U}_{p_j,n}) \wedge \left((w_j \notin \mathcal{U}_{p_j,n}) \vee (g(w_j)_{p_j} \neq g(w'_j)_{p_j})\right)\right] \\
&= 1/q(n) \cdot \Pr_{\substack{(h,p,w) \leftarrow \mathcal{D}_n \\ (h,p,w') \leftarrow \mathsf{A}(1^n, h, p, w)}} \left[w' \notin \mathcal{T}_{p,w}\right] \\
&\geq 1/q(n) \cdot 1/n^c,
\end{aligned}$$

where the first inequality holds by the union bound, the second by the definition of R and the distribution $\mathcal{D}_n$, the third by our assumption that A is a collision finder, and the equality by the definition of $\mathcal{T}_{p,w}$. $\square$

Proving Claim 5.15.

Proof of Claim 5.15. Fix $n \in \mathbb{N}$ and omit it from the notation. By Claim 5.13, With all but negligible probability probability over the choice of (p, w),

$$|\mathcal{J}_p(w)| \in (1 \pm \varepsilon)q(\ell - m + v)/\ell. \tag{7}$$

Fix $p \in [\ell]^q$ and $w \in (\{0,1\}^m)^q$ such that $|\mathcal{J}_p(w)| \in (1 \pm \varepsilon)q(\ell - m + v)/\ell$, and let $J = |\mathcal{J}_p(w)|$. Define

$$\mathcal{S} := \{g_p(w') \colon w' \in (\mathcal{T}_{p,w} \setminus \mathcal{B}_p)\}.$$

That is, $\mathcal{S} \subseteq \{0,1\}^q$ contains all the possible values of $g_p(w')$ where w' is in $(\mathcal{T}_{p,w} \setminus \mathcal{B}_p)$. We start with bounding the size of S. Observe that, by Equation (7) and the definition of $\mathcal{B}_p$, for each $w' \in (\mathcal{T}_{p,w} \setminus \mathcal{B}_p)$, it holds that $\mathcal{J}_p(w') = \mathcal{J}_p(w) \cup \mathcal{J}'$, for some set $\mathcal{J}'$ of size at most $2\varepsilon \cdot q(\ell - m + v)/\ell$. Thus, there are at most $2\varepsilon \cdot q(\ell - m + v)/\ell$ indices $j \in \mathcal{J}_p(w)$ with $g_p(w')_j \neq g_p(w)_j$. The other $(q - J)$ indices outside of $\mathcal{J}_p(w)$ are free to choose. Hence, using Fact 3.14, and the fact that $J > q(\ell - m - v)/2\ell$,

$$\begin{aligned}
\log|\mathcal{S}| &\leq (q - J) + \log \binom{J}{\leq 2\varepsilon \cdot q(\ell - m + v)/\ell} \\
&\leq (q - J) + JH(2\varepsilon \cdot q(\ell - m + v)/(\ell J)) \\
&\leq (q - J) + JH(4\varepsilon) \\
&= q - J(1 - H(4\varepsilon)).
\end{aligned}$$

By our assumptions on J and ϵ, it holds that

$$\begin{aligned}
J(1 - H(4\varepsilon)) &\geq (1 - H(4\varepsilon))(1 - \varepsilon)q(\ell - m + v)/\ell) \\
&\geq (1 - H(4\varepsilon) - \varepsilon)q(\ell - m + v)/\ell) \\
&\geq (1 - 2H(4\varepsilon))q(\ell - m + v)/\ell) \\
&\geq q((\ell - m + v)/\ell - 2H(4\varepsilon)) \\
&\geq q((\ell - m + 0.8v)/\ell),
\end{aligned}$$

where the third inequality holds since $H(\alpha) > \alpha$ for every $\alpha \in [0, 1/2]$. Combining the above, we get that,

$$\log|\mathcal{S}| \leq q(m - 2v/3)/\ell.$$

Thus, by our choice of h, the probability that there is an element $w' \in \mathcal{S}$, with $g_p(w) \neq g_p(w')$, that maps to the same image of $h(g_p(w))$ is at most

$$|\mathcal{S}| \cdot 2^{-k} \leq |\mathcal{S}| \cdot 2^{-q(m-v/3)/\ell} \leq 2^{q(m-2v/3)/\ell} \cdot 2^{-q(m-v/3)/\ell} = 2^{-q \cdot v/(3\ell)} = \text{neg}(n),$$

where the last inequality holds by our choice of q. $\square$

5.3.3 Proving Theorem 5.10

We now prove Theorem 5.10. In the following, let g, m, ℓ, v, q, t and k be as in Theorem 5.10, and let C be as in Construction 5.9. We start with some claims. The first states that the function C is shrinking.

Claim 5.16. *C is length-decreasing.*

Proof of Claim 5.16. Fix $n \in \mathbb{N}$ and omit it from the notation. It is enough to show that $m \cdot t \cdot q > k \cdot \ell \cdot (t-1) + \ell \cdot q$. Indeed,

$$\begin{aligned} m \cdot t \cdot q &= (m - v/3) \cdot t \cdot q + (v/3) \cdot t \cdot q \\ &= (m - v/3) \cdot (t-1) \cdot q + (v/3) \cdot t \cdot q + (m - v/3) \cdot q \\ &= k \cdot \ell \cdot (t-1) + (v/3) \cdot t \cdot q + (m - v/3) \cdot q \\ &> k \cdot \ell \cdot (t-1) + (v/3) \cdot (1 + 3(\ell - m)/v) \cdot q + (m - v/3) \cdot q \\ &= k \cdot \ell \cdot (t-1) + (\ell - m) \cdot q + m \cdot q \\ &= k \cdot \ell \cdot (t-1) + \ell \cdot q. \end{aligned}$$

Where the third equality is by our choice of k, and the inequality is by our choice of t. $\square$

To see the almost collision resistance property, we start with the definition of $\mathcal{B}_n$. Fix $n \in \mathbb{N}$ and omit it from the notation. Let $\mathcal{B}_n^1$ be the set from Lemma 5.12. Informally, the set $\mathcal{B}_n$ contains all the inputs (p, z) such that there exists a column $k \in [\ell \cdot t]$ that is inside $\mathcal{B}_n^1$. That is, the position of the blocks that intersects with the k-th column are untypical. To define it formally, we will need the following definitions.

For a position vector $p \in [\ell \cdot t]^q$ and input $z = ((z_1^1, \ldots, z_1^t), \ldots, (z_q^1, \ldots, z_q^t)) \in (\{0,1\}^{m \cdot t})^q$, let $z_{\overline{p}} := (z_1^{\lceil p_1/\ell \rceil}, \ldots, z_q^{\lceil p_q/\ell \rceil})$. That is, for each row $j \in [q]$, $g((z_{\overline{p}})_j)$ is the ℓ-size block in $g'(z_j)$ that contains the index p_j. Additionally, define the position vector $\widehat{p+k}$ by $(\widehat{p+k})_j := ((p_j + k - 1) \bmod \ell) + 1$. That is, $(\widehat{p+k})_j$ is the relative position of $(p+k)_j$ inside the ℓ-size block.

Let

$$\mathcal{B}_n = \left\{ (h, p, z) \in \mathcal{H} \times [\ell]^q \times (\{0,1\}^m)^{t \cdot q} \colon \exists k \in [t \cdot \ell] \text{ s.t. } (\widehat{p+k}, z_{\overline{p+k}}) \in \mathcal{B}_n^1 \right\}.$$

The next claim bound the size of $\mathcal{B}_n$, using a simple use of the union bound.

Claim 5.17.

$$|\mathcal{B}_n| = \operatorname{neg}(n) \cdot \left| \mathcal{H}_n \times [\ell(n)]^{q(n)} \times (\{0,1\}^{m(n)})^{t(n) \cdot q(n)} \right|.$$

Proof of Claim 5.17. First, observe that for every $k \in [t \cdot \ell]$, the distribution of $(\widehat{p+k}, z_{\overline{p+k}})$, for $p \leftarrow [\ell]^q, z \leftarrow (\{0,1\}^{m \cdot t})^q$, is uniform. Thus, by definition of $\mathcal{B}_n^1$ and Claim 5.13, for every $k \in [t \cdot \ell]$ it holds that

$$\Pr_{(h,p,z) \leftarrow \mathcal{H}_n \times [\ell(n)]^{q(n)} \times (\{0,1\}^{m(n)})^{t(n) \cdot q(n)}} \left[(\widehat{p+k}, z_{\overline{p+k}}) \in \mathcal{B}_n^1 \right] = \operatorname{neg}(n).$$

The claim now follows since $\ell, t \in \operatorname{poly}$ and the union bound.

$\square$

The proof of Theorem 5.10 is now by a simple reduction to the single-column case.

Proof of Theorem 5.10. By Claim 5.16, C is shrinking. Let $\mathcal{B}_n$ be as defined above. By Claim 5.17, the $\mathcal{B}_n$ is of negligible fraction.

We are left to prove that every PPT algorithm can only find collisions inside $\mathcal{B}_n$. This is done by a reduction from the single column case. In the following, we view $z \in (\{0,1\}^{m \cdot t})^q$ as $z = (z_1, \ldots, z_q)$ for $z_j = (z_j^1, \ldots, z_j^t)$, such that $z_j^i \in \{0,1\}^m$. Let A be a collision finder for C, and consider the following algorithm:

Algorithm 3: The reduction R

Input: 1^n, $h \in \mathcal{H}_n$, $p \in [\ell(n)]^{q(n)}$, $w = (w_1, \ldots, w_{q(n)}) \in \{0,1\}^{m(n)q(n)}$
Oracle : A

Choose $k \leftarrow [\ell(n) \cdot t(n)]$.

For each $j \in [q(n)]$, let $p'_j := ((p_j - k - 1) \bmod \ell(n)) + 1$, and $c_j := \left\lceil (p'_j + k)/\ell(n) \right\rceil$.

For each $j \in [q(n)]$, let $z_j^{c_j} = w_j$, and choose $z_j^1, \ldots, z_j^{c_j - 1}, z_j^{c_j + 1}, \ldots, z_j^t \leftarrow \{0,1\}^{m(n)}$. Let $z_j = (z_j^1, \ldots, z_j^t)$ and $z = (z_1, \ldots, z_q)$.

Let $(h, p', z') = \mathsf{A}(1^n, h, p', z)$. Output $({z'}_1^{c_1}, \ldots, {z'}_{q(n)}^{c_{q(n)}})$.

That is, given an input $(h, p, w = (w_1, \ldots, w_q))$ for C^1, the reduction R complete it to be an input $(h, p', ((z_1^1, \ldots z_1^t), \ldots, (z_q^1, \ldots, z_q^t)))$ for C: It chooses a random column $k \in [\ell \cdot t]$, and sets p' and $z = ((z_1^1, \ldots z_1^t), \ldots, (z_q^1, \ldots, z_q^t)))$ such that $g_p(w) = g'_{p'+k}(z)$, and $g'_{<p'+k}(z)$ contains $g_{<p}(w)$. It then executes A on the above inputs in order to get a collision for (h, p, w).

For fixed $n \in \mathbb{N}$, Let K, P', and Z be the random variables representing the values of k, p and z in Algorithm 3 in a random execution with a random inputs $(1^n, H, P, W)$. Let Z' be the value of z', the output of A it this execution, and W' the output of Algorithm 3. Let α be the probability that $Z' \neq Z$ and $Z' \notin \mathcal{B}_n$. We will show that $\alpha \leq \mu(n)$, for some negligible function μ.

We first observe that the distribution (H, P', Z) is independent from the distribution of K. Additionally, since g is almost-injective, if $Z' \neq Z$, with all but negligible probability it holds that $(g'(Z_1), \ldots, g'(Z_q)) \neq (g'(Z'_1), \ldots, g'(Z'_q))$. Thus, there exists k^* such that $g'_{p'+k}(Z) \neq g'_{p'+k}(Z')$. Take the minimal such k^*, such that $g'_{<p'+k^*}(Z) = g'_{<p'+k^*}(Z')$. By construction, if $K = k^*$, the output of R^{A} is a collision for C^1, which happens with probability $1/(\ell(n) \cdot t(n))$.

To conclude, with probability at least $\alpha/(\ell(n) \cdot t(n))$, it holds that $W' \neq W$ and $C^1(H, P, W') = C^1(H, P, W)$. Additionally, since $Z' \notin \mathcal{B}_n$, by construction it holds that $W' \notin \mathcal{B}_n^1$. Using Lemma 5.12 we get that α must be negligible. $\square$

Acknowledgments

We are thankful to Iftach Haitner for very useful discussions.

References

[Agr+19] Rohit Agrawal, Yi-Hsiu Chen, Thibaut Horel, and Salil Vadhan. "Unifying computational entropies via Kullback–Leibler divergence". In: *Annual International Cryptology Conference*. Springer. 2019, pp. 831–858 (cit. on p. 5).

[Ame+12] Scott Ames, Rosario Gennaro, and Muthuramakrishnan Venkitasubramaniam. "The generalized randomized iterate and its application to new efficient constructions of UOWHFs from regular one-way functions". In: *International Conference on the Theory and Application of Cryptology and Information Security*. Springer. 2012, pp. 154–171 (cit. on p. 1).

[App+06] Benny Applebaum, Yuval Ishai, and Eyal Kushilevitz. "Cryptography in NCˆ0". In: *SIAM Journal on Computing* 36.4 (2006), pp. 845–888 (cit. on p. 2).

[AS16] Gilad Asharov and Gil Segev. "Limits on the power of indistinguishability obfuscation and functional encryption". In: *SIAM Journal on Computing* 45.6 (2016), pp. 2117–2176 (cit. on p. 5).

[BH13] Kfir Barhum and Thomas Holenstein. "A cookbook for black-box separations and a recipe for UOWHFs". In: *Theory of Cryptography Conference*. Springer. 2013, pp. 662–679 (cit. on p. 5).

[BM12] Kfir Barhum and Ueli Maurer. "UOWHFs from OWFs: Trading regularity for efficiency". In: *International Conference on Cryptology and Information Security in Latin America*. Springer. 2012, pp. 234–253 (cit. on p. 5).

[BM84] Manuel Blum and Silvio Micali. "How to generate cryptographically strong sequences of pseudorandom bits". In: *SIAM journal on Computing* 13.4 (1984), pp. 850–864 (cit. on p. 1).

[Gal14] David Galvin. "Three tutorial lectures on entropy and counting". In: *arXiv preprint arXiv:1406.7872* (2014) (cit. on p. 14).

[Gen+05] Rosario Gennaro, Yael Gertner, Jonathan Katz, and Luca Trevisan. "Bounds on the efficiency of generic cryptographic constructions". In: *SIAM journal on Computing* 35.1 (2005), pp. 217–246 (cit. on p. 5).

[Hai+06] Iftach Haitner, Danny Harnik, and Omer Reingold. "On the power of the randomized iterate". In: *Annual International Cryptology Conference*. Springer. 2006, pp. 22–40 (cit. on p. 1).

[Hai+09] Iftach Haitner, Omer Reingold, Salil Vadhan, and Hoeteck Wee. "Inaccessible entropy". In: *Proceedings of the forty-first annual ACM symposium on Theory of computing*. 2009, pp. 611–620 (cit. on pp. 2, 5).

[Hai+10] Iftach Haitner, Thomas Holenstein, Omer Reingold, Salil Vadhan, and Hoeteck Wee. "Universal one-way hash functions via inaccessible entropy". In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 2010, pp. 616–637 (cit. on pp. 1–4, 6, 7, 13, 14).

[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. 1–5, 8, 9).

[Hås+99] Johan Håstad, Russell Impagliazzo, Leonid A Levin, and Michael Luby. "A pseudorandom generator from any one-way function". In: *SIAM Journal on Computing* 28.4 (1999), pp. 1364–1396 (cit. on p. 2).

[HS12] Thomas Holenstein and Makrand Sinha. "Constructing a pseudorandom generator requires an almost linear number of calls". In: *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*. IEEE. 2012, pp. 698–707 (cit. on p. 5).

[KK05] Jonathan Katz and Chiu-Yuen Koo. "On constructing universal one-way hash functions from arbitrary one-way functions". In: *Cryptology ePrint Archive* (2005) (cit. on p. 5).

[MZ21] Noam Mazor and Jiapeng Zhang. "Simple Constructions from (Almost) Regular One-Way Functions". In: *Theory of Cryptography Conference*. Springer. 2021, pp. 457–485 (cit. on pp. 1, 3, 6, 7, 14).

[NY89] Moni Naor and Moti Yung. "Universal one-way hash functions and their cryptographic applications". In: *Proceedings of the twenty-first annual ACM symposium on Theory of computing*. 1989, pp. 33–43 (cit. on pp. 1, 5).

[Rom90] John Rompel. "One-way functions are necessary and sufficient for secure signatures". In: *Proceedings of the twenty-second annual ACM symposium on Theory of computing*. 1990, pp. 387–394 (cit. on pp. 1–3, 5).

[VZ12] Salil Vadhan and Colin Jia Zheng. "Characterizing pseudoentropy and simplifying pseudorandom generator constructions". In: *Proceedings of the forty-fourth annual ACM symposium on Theory of computing*. 2012, pp. 817–836 (cit. on pp. 1, 9).

[Yao82] Andrew C Yao. "Theory and application of trapdoor functions". In: *23rd Annual Symposium on Foundations of Computer Science (SFCS 1982)*. IEEE. 1982, pp. 80–91 (cit. on p. 1).

[Yu+15] Yu Yu, Dawu Gu, Xiangxue Li, and Jian Weng. "(Almost) Optimal constructions of UOWHFs from 1-to-1, regular one-way functions and beyond". In: *Annual Cryptology Conference*. Springer. 2015, pp. 209–229 (cit. on pp. 1, 5).

A Missing Proofs

Proving Lemma 3.7.

Lemma A.1 (Lemma 3.7, restated)**.** *Let $f\colon \{0,1\}^n \to \{0,1\}^n$ be a function, and $\mathcal{H} = \{h\colon \{0,1\}^n \to \{0,1\}^n\}$ a two-wise independent family. For every $x \in \{0,1\}^n$ and $c \in \mathbb{N}$ the following holds.*

$$\Pr_{h\leftarrow\mathcal{H}}\left[\left|\left\{x'\colon h(f(x')) = h(f(x))\right\}\right| \geq \left|f^{-1}(f(x))\right| + n^{2c}\right] \leq 2/n^c.$$

Proof. Let $\mathcal{A} = \left\{y \in \mathsf{Im}(f) : \left|f^{-1}(y)\right| \leq n^c\right\}$ and $\mathcal{B} = \left\{y \in \mathsf{Im}(f) : \left|f^{-1}(y)\right| > n^c\right\}$. Notice that $|\mathcal{B}| \leq 2^n/n^c$, and thus,

$$\Pr_{h\leftarrow\mathcal{H}}[\exists y \in \mathcal{B} \setminus \{f(x)\} \text{ s.t. } h(y) = h(f(x))] \leq |\mathcal{B}| \cdot 2^{-n} \leq 1/n^c.$$

Similarly, using Markov inequality, since

$$\mathop{\mathbf{E}}_{h\leftarrow\mathcal{H}}[|\{y \in \mathcal{A} \setminus \{f(x)\}\colon h(y) = h(f(x))\}|] = |\mathcal{A}| \cdot 2^{-n} \leq 1,$$

it holds that

$$\Pr_{h\leftarrow\mathcal{H}}[|\{y \in \mathcal{A} \setminus \{f(x)\}\colon h(y) = h(f(x))\}| \geq n^c] \leq 1/n^c.$$

We get that

$$
\begin{aligned}
&\Pr_{h\leftarrow\mathcal{H}}\left[\left|\left\{x'\colon h(f(x')) = h(f(x))\right\}\right| \geq \left|f^{-1}(f(x))\right| + n^{2c}\right] \\
&\leq \Pr_{h\leftarrow\mathcal{H}}[\exists y \in \mathcal{B} \setminus \{f(x)\} \text{ s.t. } h(y) = h(f(x))] + \Pr_{h\leftarrow\mathcal{H}}[|\{y \in \mathcal{A} \setminus \{f(x)\}\colon h(y) = h(f(x))\}| \geq n^c] \\
&\leq 2/n^c.
\end{aligned}
$$

□

Proving Claim 5.7.

Claim A.2 (Claim 5.7, restated)**.** *g is almost injective function.*

Proof. Fix $n \in \mathbb{N}$ and $x \in \{0,1\}^n$. We will show that

$$\Pr_{h_1,h_2\leftarrow\mathcal{H}}\left[\exists x' \neq x \in \{0,1\}^n \text{ s.t. } (h_1(f(x)), h_2(x)) = (h_1(f(x')), h_2(x'))\right] = neg(n)$$

which implies the claim. Observe that by the pair-wise independence of $\mathcal{H}$, for every x' such that $f(x') \neq f(x)$, it holds that $\mathbf{Pr}_{h_1,h_2\leftarrow\mathcal{H}}[(h_1(f(x)), h_2(x)) = (h_1(f(x')), h_2(x'))] = 2^{-2n}$. Similarly, for every x' with $f(x) = f(x')$ it holds that $\mathbf{Pr}_{h_1,h_2\leftarrow\mathcal{H}}[(h_1(f(x)), h_2(x)) = (h_1(f(x')), h_2(x'))] = 2^{-n}$. Thus, using the union bound we get,

$$
\begin{aligned}
&\Pr_{h_1,h_2\leftarrow\mathcal{H}}\left[\exists x' \neq x \text{ s.t. } (h_1(f(x)), h_2(x)) = (h_1(f(x')), h_2(x'))\right] \\
&\leq 2^n \cdot 2^{-2n} + \left|f^{-1}(f(x))\right| \cdot 2^{-n} \\
&\leq 2^{-n} + neg(n) \\
&= neg(n).
\end{aligned}
$$

Where the last inequality is by Claim 3.2. □

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