On the Correlation Intractability of Obfuscated Pseudorandom Functions

Ran Canetti[1,2], Yilei Chen[1], and Leonid Reyzin[1]

[1] Boston University, USA
{canetti,chenyl,reyzin}@bu.edu
[2] Tel Aviv University, Israel
canetti@tau.ac.il

Abstract. A family of hash functions is called "correlation intractable" if it is hard to find, given a random function in the family, an input-output pair that satisfies any "sparse" relation, namely any relation that is hard to satisfy for truly random functions. Indeed, correlation intractability is a strong and natural random-oracle-like property. However, it was widely considered unobtainable. In fact for some parameter settings, unobtainability has been demonstrated [Canetti, Goldreich, Halevi, J.ACM 04]. We construct a correlation intractable function ensemble that withstands all relations with a priori bounded polynomial complexity. We assume the existence of sub-exponentially secure indistinguishability obfuscators, puncturable pseudorandom functions, and input-hiding obfuscators for evasive circuits. The existence of the latter is implied by Virtual-Grey-Box obfuscation for evasive circuits [Bitansky et al, CRYPTO 14].

1 Introduction

To what extent can we construct efficient function families that "behave like random functions"? This is an intriguing question in cryptography. One of the most elusive properties of random functions is correlation intractability, proposed by Canetti, Goldreich and Halevi [26]. Roughly speaking, correlation intractable functions guarantee that it is infeasible to find input-output pairs that satisfy some "rare" relation. A bit more precisely, a binary relation R is called *sparse*, if for each value x, only a negligible fraction of y values satisfy $(x, y) \in R$. A function family F is *correlation intractable* if, for any sparse relation R, it is infeasible for the adversary to find, given the full description of a random function f in F, a value x such that $(x, f(x))$ is in the relation.

The only known results regarding the existence of correlation intractable functions are negative. Specifically, for some settings of the parameters (e.g. when the key is shorter than the input), correlation intractable functions were shown not to exist. This observation was used in [26] to demonstrate the uninstantiability of the random oracle model [9]. However, whether correlation intractable functions exist for other settings of the parameters, and based on what assumptions, remains open.

Beyond the foundational appeal, correlation intractability is desirable in real world applications. For example, consider the hash function used to build the block chain in the Bitcoin protocol [47]. Its main security property, needed to obtain proofs of work, can be stated as correlation intractability with respect to a specific set of relations, which come from protocol-defined constraints on the input and the output. (Specifically, the input needs to contain appropriate transaction information and the output needs to begin with the correct number of zeros.) It should be noted that we do not claim that our result directly applies to the Bitcoin protocol: in this paper we consider only relations that are negligibly sparse, while for Bitcoin and other proof-of-work applications, it is necessary to consider relations that are moderately sparse and to define a more precise analog of correlation intractability (in which the difficulty of finding $(x, f(x)) \in R$ is closely related to the density of R).

More generally, consider a multi-party game which uses the value returned by a random oracle, applied to the previous moves of players, as a substitute for public randomness. Correlation intractable functions can potentially be used to instantiate the random oracle in such a game without significant change in the properties of the game.

Alternative approaches to obtaining hash functions with random oracle like properties Several alternative notions have been proposed in attempt to capture random-oracle-like properties of hash functions. These notions include entropy preservation [7], seed incompressibility [41], perfect one-wayness [23, 28], non-malleability [16], correlation robustness [43], correlated input security [38], and universal computational extractors [8]. Their relations to correlation intractability will be discussed later in section 1.4. Still, to the best of our knowledge, none of the known results regarding these notions shed light on the question of the existence of correlation intractable functions.

Obfuscated pseudorandom functions. A natural approach to constructing functions with random-oracle-like properties is to obfuscate pseudorandom functions (PRFs). Indeed, if the obfuscation was perfect, then the adversary would be unable to take advantage of the code any more than by merely having oracle access to the function. This would render the function random-oracle-like. Strong security definitions of obfuscation are formalized in the work of Hada [39] and Barak et al. [6], e.g. *Virtual-black-box* (VBB) Obfuscation. However, they also show that VBB obfuscation is impossible for many function families. In particular, Barak et al. [6] explicitly construct a PRF such that given any program (no matter how obfuscated) that computes the PRF, the adversary can find an input which evaluates to a fixed value. This certainly breaks correlation intractability.

We also know that *no* pseudorandom function family can be VBB obfuscated with respect to auxiliary inputs [12, 37]. However, these results do not rule out the possibility that there exist pseudorandom functions whose obfuscated version is correlation intractable.

A reasonable next step may thus be to consider PRFs with additional properties, such as constrained or puncturable PRFs [18, 19, 44]. Indeed, as demonstrated by multiple works, starting with the ingenious work of Sahai and Waters [51], puncturable PRFs are an extremely powerful tool when combined with obfuscation of general programs. In particular, puncturable PRFs have been used together with iO to instantiate some random-oracle-like hash functions, including universal hardcore functions [10], universal computational extractors [22], and functions used for the full-domain-hash construction [42]. Furthermore, the constructions of [10] and [22] are simply obfuscating puncturable PRFs. It is thus natural to ask:

Are obfuscated puncturable PRFs correlation intractable?
If so, under what assumptions?

1.1 Our results

We make progress towards answering the above questions. Specifically, we show that puncturable pseudorandom functions, obfuscated using an indistinguishability obfuscator, satisfy *bounded* correlation intractability. Here "bounded" means that there is a polynomial upper bound on the computational complexity of the sparse relations considered, and the complexity of the function family depends on that bound. (We stress that this bound applies only to the relation. The adversary runs in arbitrary polynomial time.) Bounded correlation intractability is indeed a qualitatively weaker property than full correlation intractability (see definitions in Section 3). Still, even in its bounded form, correlation intractability is a very strong notion that has not been constructed before. In particular, in many specific applications, such as Bitcoin, an upper bound on the complexity of the sparse relation is known.

Our result holds under the assumption of sub-exponentially secure general iO and puncturable PRFs, and also requires the existence of *Input-Hiding Obfuscation* (IHO) for evasive circuit families, which we now explain. Recall that a boolean circuit family is evasive if for any input, only negligibly many circuits in the family evaluate to a non-zero value. An obfuscator on evasive circuits achieves the "input-hiding" property, if it is infeasible for a polytime adversary to find, given an obfuscated version of a random function in the family, a preimage of non-zero output for that function. (Note that no subexponential hardness is assumed here.) Candidate IHOs for general evasive circuits are proposed by Bitansky et al. [13] and Badrinarayanan et al. [3] (see Section 1.3). Our main theorem is thus the following:

Theorem 1 (Bounded correlation intractable function ensembles, informal).

Assume existence of input-hiding obfuscation for evasive circuits, subexponentially secure indistinguishability obfuscation, and subexponentially secure puncturable pseudorandom functions. Then there is a $p(n)$-bounded correlation intractable function ensemble for any polynomial $p(n)$.

Note that if we only consider relations R where for any x, there are only very few y values in the range satisfy $R(x,y)$, and allow the range to be larger than the domain, then correlation intractability becomes easy to obtain. Indeed, for such a R and a 1-universal function f there will with high probability not *exist* inputs x such that $R(x,f(x))$ holds. However, we argue that this case is of less interest. Rather, we are interested in general sparse relations where the "bad inputs" exist, but are hard to find. Our solution is able to handle the general case. For further discussions of the parameters and other special relations, we refer the readers to the end of section 3.

1.2 Our techniques

Our goal is to prove correlation intractability of certain function family. At a high level, our approach is to show, given a relation R, that a function f sampled randomly from the initial function family is indistinguishable from another function, f^R, that is constructed specifically so as to make it hard to find "bad inputs" with respect to the given relation R.

However, the definition of this function f^R, and moreover showing that it is indistinguishable from the original function f, needs to be done with care. In particular, the "naive" methodology of simply puncturing f at all the bad points, so as to obtain a function where no bad points for relation R exist, fails. We start by briefly explaining this failure.

Failure of the "standard" puncturing methodology. Recall that a PRF is puncturable if for any key K and input value x it is possible to generate a key $K\{x\}$ that is "punctured" at x, such that $F_K(x)$ remains pseudorandom even given $K\{x\}$, and yet $K\{x\}$ allows evaluating F_K at all points other than x. To prove security of constructions that use puncturable PRFs obfuscated with iO, the "standard" methodology proceeds in two steps to get an indistinguishable game that an adversary cannot win (thus showing, by indistinguishability, that the adversary also fails in the original game). In the first step (whose indistinguishability is proven via iO), one typically punctures the key at the bad inputs that threaten the security of the scheme, and hardwires the output values for the punctured inputs. In the second step (whose indistinguishability is proven via the puncturable PRF), the output values at the punctured inputs are changed to ensure the adversary can't exploit them.

In our scenario, given a relation R, the "bad" inputs are those x values that satisfy $R(x, F_K(x)) = 1$, where K is randomly sampled after R is fixed. However, it is not clear how puncturing at these bad points helps here, since it is not clear how to argue that changing the output values so as to avoid R is indistinguishable. (In fact, it can be seen from our analysis that such change may well be distinguishable overall.)

Said otherwise, the "standard" puncturing technique is geared toward the case where the bad input values are fixed before the PRF key K is chosen, whereas for correlation intractability, the bad points are determined by K.

A "counterintuitive" puncturing strategy. To get around this difficulty, we start from the following observation: for any sparse relation, the "bad" inputs x (i.e., those for which $R(x, F_K(x)) = 1$) are rare—in fact, they can be recognized by a circuit from an evasive circuit family. All we need to do in order to prove correlation intractability is show an indistinguishable function in which those rare inputs are hidden from the adversary. We do so by decomposing the PRF into two branches: one defined on the bad inputs, which form an evasive set, the other defined on the "innocent" inputs. Then we apply an input-hiding obfuscator to the bad branch. However, the input-hiding obfuscator cannot work in the presence of auxiliary information given by the innocent branch: the value of the function on the innocent inputs may permit the adversary to find the evasive inputs. We therefore puncture the key and change the function at every input that belongs to the innocent branch. To avoid increasing the circuit size beyond polynomial as we puncture at exponentially many points, we build an alternative function family $\mathcal{F}^R$ that is designed to avoid R. The details of the key-switching strategy form the technical heart of the proof.

The proof in a nutshell. To better illustrate the main idea, we present an overview of the proof. The analysis goes through 3 hybrids, as will be presented by the games between the adversary and the challenger. Hybrid 0 represents the original game. Hybrid 1, 2, and 3 are intermediate games that are indistinguishable by the adversary. Finally we will show that the adversary cannot break correlation intractability in hybrid 3, therefore concluding that the adversary also fails in hybrid 0, since hybrids 0 and 3 are indistinguishable.

We note that the circuits being iOed shall be padded to the same size, which is possible in our construction if an a priori bound on the size of the relation is given. Under this limitation, our techniques suffice to prove only a bounded version of correlation intractability. For the simplicity of the overview, we postpone the details of padding to the formal proof and now present the hybrids.

For any sparse relation R that is recognizable by some bounded polynomial sized circuit:

0. The challenger samples a key K of puncturable PRF $\mathcal{F}$ and obfuscates it:
$$h_k^0(\cdot) = \mathsf{iO}(F_K(\cdot))$$
The adversary wins if it outputs x such that $(x, h_k^0(x)) \in R$. This is the original game. The only thing that changes in subsequent games is the circuit obfuscated iO.
1. The challenger samples a key K of puncturable PRF $\mathcal{F}$, and embeds the relation R into the description of the function:
$$h_k^1(x) = \mathsf{iO}\begin{pmatrix} \text{if } R(x, F_K(x)) = 1, & \text{return } F_K(x) & ;\text{the ``bad'' branch} \\ \text{else,} & \text{return } F_K(x) & ;\text{the ``innocent'' branch} \end{pmatrix}$$
Note that h^1 has the same functionality as h^0, and therefore it is indistinguishable from the original function by iO. (Recall that an iO scheme iO

guarantees that $\mathsf{iO}(C) \approx \mathsf{iO}(C')$ for any two circuits C, C' that have the same size and functionality.) This is a preparation step, which enables us to partition the function as described above.

2. Replace the key that is evaluated on the innocent branch with a freshly generated key K' for a different puncturable PRF $\mathcal{F}^R$ parameterized by R:

$$h_k^2(x) = \mathsf{iO}\begin{pmatrix}\text{if } R(x, F_K(x)) = 1, \text{return } F_K(x) & \text{; the "bad" branch} \\ \text{else,} \quad\quad\quad\quad\quad\quad \text{return } F^R_{K'}(x) & \text{; the "innocent" branch}\end{pmatrix}$$

where $\mathcal{F}^R$ is designed such that there is no x such that $(x, F^R_{K'}(x)) \in R$ with high probability. To generate a key K' for $\mathcal{F}^R$, we sample a set of independent puncturable PRF keys K_1, ..., $K_{T(n)}$ from $\mathcal{F}$. The function $F^R_{K'}$ executes in a "rejection sampling" fashion, such that for input x, it goes through the keys K_1, ..., $K_{T(n)}$ one by one, evaluates on the first key K_i for which $(x, F_{K_i}(x))$ is not in the relation. Setting T to be linear in l (in fact, even slightly sublinear) is enough to make sure that x not in the relation is found except with exponentially small probability. A similar construction was proposed in [49] (the results are included in [26]) to achieve "relation-specific" correlation intractable functions.

To prove the indistinguishability of h^1 and h^2, we show that both of them are subexponentially secure puncturable PRFs, based on the subexponential security assumption on the underlying puncturable PRF $\mathcal{F}$. We then use the following lemma (derived from the proof methodology in the work of Canetti et al. [27]) to show that, h^1 and h^2 are indistinguishable after being obfuscated by subexponentially secure iO.

Lemma 1 (Informal). *If h_1 and h_2 are subexponentially secure punctured PRFs and iO is subexponentially secure, then $\mathsf{iO}(h_1)$ and $\mathsf{iO}(h_2)$ are indistinguishable.*

3. Wrap the first "if-trigger", together with the underlying evasive function, by input-hiding obfuscation. The function h_k^3 is then generated as:

$$h_k^3(x) = \mathsf{iO}\begin{pmatrix} y \leftarrow \mathsf{IHO}\begin{pmatrix}\text{if } R(x, F_K(x)) = 1, & \text{return } F_K(x) \\ \text{else,} & \text{return } \perp\end{pmatrix} & \text{; "bad"} \\ \text{if } y = \perp, \ y \leftarrow F^R_{K'}(x) & \text{; "innocent"} \\ \text{return } y & \end{pmatrix}$$

h^3 is indistinguishable from h^2 because they are functionally equivalent and obfuscated by iO.

Finally, we note that finding the x values that trigger the non-zero values on the "input-hiding-box" is hard, given R and an "innocent" function $F^R_{K'}$ generated independently (even if not obfuscated). Since the adversary cannot distinguish whether she is given the original function h^0 or the function h^3, and finding an input on h^3 that satisfies the relation is hard, it should also be infeasible for the adversary to break correlation intractability on the original function.

1.3 More on input-hiding obfuscation for evasive functions

Our result depends on the existence of input-hiding obfuscation (IHO) for evasive circuits. In this section we survey the state of the art regaring the existence of such obfuscation.

IHO for the class NC^1 can be obtained as follows. Start with a primitive called *strong indistinguishability obfuscation* (siO), which guarantees that if two circuits C_0 and C_1 are drawn from two distributions that are *concentrated* on the same function, then $\mathsf{siO}(C_0)$ is indistinguishable from $\mathsf{siO}(C_1)$. We show in section 2.1 that siO for evasive circuit class $\mathcal{C}$ implies input-hiding obfuscation for $\mathcal{C}$. Thus, it is enough get siO for NC^1. Bitansky et al. [13] show that siO is equivalent to worst-case VGB obfuscation, and that siO/VGB for NC^1 circuits can be obtained under the assumptions that certain graded encoding schemes satisfy a strong form of semantic security [50]. Therefore, under the same assumption as made in [13] plus the assumption that puncturable PRFs exist in NC^1 [17], we obtain correlation intractable functions w.r.t. relations recognizable by NC^1 circuits.

IHO for larger circuit classes is currently is not known to follow from simpler primitives. Still, one can simply assume (similarly to [13]) that existing candidate obfuscators for $\mathsf{P/poly}$ are IHO. This assumption is not contradicted by known impossibility results: for evasive (as opposed to general [6]) circuits, there are no impossibility results known even for such a strong notion as average-case VBB [4].

Alternatively, IHO can be built in idealized models. In fact, both VBB obfuscation and IHO for $\mathsf{P/poly}$ were shown possible in a model with idealized graded encodings [2, 5, 20, 54]. Furthermore, IHO for $\mathsf{P/poly}$ was shown possible by Badrinarayanan et al. [3] in a more relaxed idealized model, which avoids the devastating zeroing attack [29] on the candidate graded encodings [30, 34].

Proposing simpler constructions of IHO without going through the full-fledged VGB, or basing IHO on simpler assumptions is an interesting open problem.

1.4 More on related work

Correlation intractability and constant-round public-coin zero-knowledge proofs. Hada and Tanaka show that the existence of correlation intractable hash functions (w.r.t. relations that are not necessarily efficient) implies 3 round public-coin auxiliary-input zero-knowledge proofs exist only for languages in BPP [40]. The key observation is based on the relation $R_{\notin\mathcal{L}}$ defined as

$$(x||\alpha,\beta) \in R_{\notin\mathcal{L}} \Leftrightarrow x \notin \mathcal{L} \wedge \exists\gamma, \Pr[\mathsf{Ver}(x,\alpha,\beta,\gamma) = \mathsf{Accept}] \geq \mathsf{non.negl}.$$

where x is the instance, α,β,γ are the 3 messages in the protocol. The relation is sparse due to the statistical soundness of the underlying proof. Given the fact that the bounded simulator cannot break the correlation intractability, it should be able to decide the membership of the instance.

However, deciding the membership in the relation $R_{\notin \mathcal{L}}$ requires (at least) an auxiliary string γ in addition to the instance x, input α, and output β, whereas the construction of correlation intractable function proposed in this paper can only handle relations that takes exactly one input and one output. An alternative way of describing the relation is proposed by Halevi et al. [41] who define the relation with multiple invocations, and set γ as part of the inputs of the additional invocations. Our construction hasn't been proved to work for relations with multiple invocations.

Entropy-preserving hashing. The notion of "entropy-preserving hashing", formalized by Barak, Lindell and Vadhan [7] as being sufficient to achieve Fiat-Shamir heuristics for proofs [32], is closely related to correlation intractability. Roughly speaking, the definition requires that after the adversary is given the key and chooses the input, the output conditioned on the input has high entropy.

We show (in appendix A) that entropy preservation and correlation intractability implies each other. However, the connections are shown w.r.t. relations that are not necessarily decidable by poly-size circuits. Therefore, our construction is not necessarily entropy-preserving. The existence of entropy-preserving hash functions remains open. In fact Bitansky et al. show that entropy preservation is impossible to prove from black-box reduction to falsifiable assumptions [14]. As a corollary, correlation intractability w.r.t. possibly inefficient relations is impossible to obtain from black-box reduction to falsifiable assumptions. We don't know if the same impossibility holds for CI w.r.t. efficiently recognizable relations.

Alternative approaches to instantiating random oracles. Several alternative definitions have been proposed in order to capture the random-oracle-like properties. These notions include perfect one-wayness [23, 28], non-malleability [16], seed incompressibility (SI) [41], correlation robustness [43], correlated input security (CIH) [38], and universal computational extractors (UCE) [8]. These definitions are quite different from correlation intractability. In particular, SI, CIH and UCE model the security game in two stages, where the adversary in the first stage doesn't get full access to the description of the function, to avoid the impossibility results in [26]. It turns out that one can separate correlation intractability and each of these notions. An example is given in appendix A that separates CIH/UCE and correlation intractability.

Separations, of course, do not show incompatibility: indeed, a construction may naturally satisfy many security definitions simultaneously. For example, essentially the same construction as in this paper (obfuscated puncturable PRFs) was shown to also satisfy a subclass of UCE by Brzuska and Mittelbach [22]. Further exploring constructions that satisfy multiple definitions simultaneously (and, in particular, gaining a better understanding of puncturable PRFs) is an interesting future direction.

Additional related work. A canonical construction of a PRF from a pseudorandom generator (PRG), now known as the GGM PRF, was given by Goldreich, Goldwasser and Micali [36]. Suppose we simply publish a GGM PRF seed in the clear to allow public evaluation, without any obfuscation. Is such a function correlation intractable? This questions was posed in the 1990s and answered negatively by Goldreich [35]. He constructed a specialized PRG, such that the GGM PRF built on this PRG is not correlation intractable. In fact one can find a preimage of $0^{m(n)}$ with non-negligible probability.

Correlation intractability is a natural criterion for designing efficient ciphers and hash functions. For example, it is used by Mandal et al. [46] to analyze the 6-round Feistel construction. In particular, they show that the 6-round Feistel construction is sequentially indifferentiable from a random invertible permutation, which implies that it is correlation intractable under an idealized assumption on the Feistel round function.

2 Preliminaries

Many experiments and probability statements in this paper contain randomized algorithms (such as obfuscators or adversaries) within them. The probability of success of an experiment is always taken over the random coins used by the relevant randomized algorithms; therefore, we do not mention these coins explicitly.

A function ensemble $\mathcal{F}$ has a key generation function $g : S \to K$; on seeds s of length $\sigma(n)$, g produces a key k of length $\kappa(n)$ for a function with input length $l(n)$ and output length $m(n)$:

$$\mathcal{F} = \{f_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}, k = g(s), s \in \{0,1\}^{\sigma(n)}\}_{n\in\mathbb{N}}$$

By default we denote $k \overset{\$}{\leftarrow} \mathcal{F}_n$ (sometimes abbreviated as k in the equations) as sampling a key k uniformly random from $\mathcal{F}_n$.

For any definition based on computational indistinguishability, we will say that the relevant security notion is *subexponential* if for every distinguisher there exists $\epsilon > 0$ such that the distinguisher's advantage is $2^{-n^{\epsilon}}$, where n is the security parameter.

2.1 Obfuscation

In this work we use indistinguishability obfuscation for all circuits, and input-hiding obfuscation for all evasive circuit collections. Both obfuscators considered in this paper perfectly preserve the functionality, and cause a polynomial blow-up on the size of the function description. To be precise, for the circuit family $\mathcal{F} = \{f : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{f\in\mathcal{F}_n}$, a probabilistic algorithm Obf is an obfuscator, if

1. The string $\mathsf{Obf}(f)$ describes a circuit that computes the same function as f;
2. There is a polynomial $B(\cdot)$ such that $|\mathsf{Obf}(f)| \leq B(|f|)$.

The difference lies in the security properties: indistinguishability obfuscation guarantees that the obfuscation of any functionally equivalent circuits cannot be distinguished; whereas input-hiding obfuscation only applies on evasive circuits, and promises to hide all the inputs which lead to non-zero outputs.

Definition 1 (Indistinguishability Obfuscation [6]). Obf *is an indistinguishability Obfuscator (iO) for $\mathcal{F}$ if for any feasible adversary A, there is a negligible function $\mathsf{negl}(\cdot)$ such that for all circuits f_0 and f_1 that have identical functionalities, and are of the same size, it holds that*

$$|\Pr[A(\mathsf{iO}(f_0)) = 1] - \Pr[A(\mathsf{iO}(f_1)) = 1]| \leq \mathsf{negl}(n)$$

Definition 2 (Evasive circuit collections). *Let $\mathcal{F} = \{f_k : \{0,1\}^{l(n)} \rightarrow \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$ be a circuit collection, we say $\mathcal{F}_n$ is* evasive *if there is a negligible function $\mathsf{negl}(\cdot)$ such that for all $x \in \{0,1\}^{l(n)}$:*

$$\Pr_k[f_k(x) \neq 0^{m(n)}] \leq \mathsf{negl}(n)$$

Definition 3 (Input-hiding Obfuscation for evasive circuits [4]). *An obfuscator for a evasive circuit collection $\mathcal{F}$ is* input-hiding *(IHO) if for every p.p.t. adversary A there exist a negligible function $\mathsf{negl}(\cdot)$ s.t. for every auxiliary input $z \in \{0,1\}^{\mathsf{poly}(n)}$:*

$$\Pr_k[f_k(A(\mathsf{IHO}(f_k), z)) \neq 0^{m(n)}] \leq \mathsf{negl}(n)$$

The notion of IHO (unlike iO) is inherently average-case, i.e., the function f_k is random and independent of the auxiliary input z (see [4, Section 2] for a discussion of this issue). In particular, impossibility results, such as [21], for notions of obfuscation that allow a related auxiliary input, do not apply.

Remark 1. The original definitions of evasive circuit collections and the corresponding obfuscators proposed by Barak et al. [4] are stated for circuits with 1-bit output; whereas our definition of evasive circuit collections is for multi-bit output. For the case of input-hiding obfuscation, the existence of IHO for *all* evasive circuits with 1-bit output implies the existence of IHO for *all* evasive circuits with multi-bit output: for circuit $C(x)$ with m-bit output, we can obfuscate the circuit $C(x;i) = C(x)^{(i)}$ that returns the i-th output bit, and run $\mathsf{IHO}(C(x;i))$ with $i \in [m]$. This transformation is mentioned by Bitansky et al. [13] for VGB obfuscation for all circuits. We note that the transformation also works for certain restricted circuit classes including NC^1.

Throughout this paper, we will assume the existence of IHO for all evasive circuits with 1-bit output, and use IHO for evasive circuits with possibly multi-bit output without loss of generality.

Input-hiding obfuscation from VGB obfuscation We introduce one of the known approaches to designing input-hiding obfuscation for evasive circuits. As a corollary of the result from [13], IHO is implied by Virtual-Grey-Box (VGB) obfuscation, or equivalently, strong indistinguishability obfuscation (siO).

Definition 4 (Concentrated / Evasive function distribution). *Let $\mathcal{F} = \{f_k : \{0,1\}^{l(n)} \to \{0,1\}\}_{n\in\mathbb{N}}$ be a function ensemble, $\tilde{\mathcal{F}}_n$ be a distribution on $\mathcal{F}_n$. Let $\mathsf{maj}_{\tilde{F}_n}(x) = \mathbb{E}_{f\leftarrow\tilde{\mathcal{F}}_n} f(x)$ be the common output on x for functions drawn from $\tilde{\mathcal{F}}_n$.*

1. *$\tilde{\mathcal{F}}_n$ is concentrated if there is a negligible function $\mathsf{negl}(\cdot)$ that*

$$\max_{x\in\{0,1\}^{l(n)}} \Pr_{f\leftarrow\tilde{\mathcal{F}}_n}[f(x) \neq \mathsf{maj}_{\tilde{\mathcal{F}}_n}(x)] \leq \mathsf{negl}(n)$$

2. *(Rephrasing definition 2 for 1-bit output) $\tilde{\mathcal{F}}_n$ is evasive if it is concentrated, and $\forall x \in \{0,1\}^{l(n)}$, $\mathsf{maj}_{\tilde{F}_n}(x) = 0$*

Definition 5 (Strong indistinguishability Obfuscator [13]). *An obfuscator is a strong indistinguishability Obfuscator (siO) for $\mathcal{F}$ if for any two concentrated distribution ensembles $\tilde{\mathcal{F}}_n^0$, $\tilde{\mathcal{F}}_n^1$ on $\mathcal{F}_n$ s.t. $\mathsf{maj}_{\tilde{\mathcal{F}}_n^0} \equiv \mathsf{maj}_{\tilde{\mathcal{F}}_n^1}$, and for any p.p.t. adversary A, there is a negligible function $\mathsf{negl}(\cdot)$:*

$$\left| \Pr_{f_0\leftarrow\tilde{\mathcal{F}}_n^0}[A(\mathsf{siO}(f_0)) = 1] - \Pr_{f_1\leftarrow\tilde{\mathcal{F}}_n^1}[A(\mathsf{siO}(f_1)) = 1] \right| \leq \mathsf{negl}(n)$$

Definition 6 (Virtual-Grey-Box Obfuscation [11]). Obf *is a Virtual-Grey-Box (VGB) Obfuscator for $\mathcal{F}$ if for any feasible adversary A, there is a simulator S, and a negligible function $\mathsf{negl}(\cdot)$ such that for all $f \in \mathcal{F}$:*

$$|\Pr[A(\mathsf{Obf}(f)) = 1] - \Pr[S^f(1^{|f|}) = 1]| \leq \mathsf{negl}(|f|)$$

where the running time of S is computationally unbounded, but only sends polynomially many queries to f (such a simulator is usually called "semi-bounded").

Theorem 2 ([13]). *An obfuscator is* siO *for $\mathcal{F}$ iff it is worst-case VGB obfuscator for $\mathcal{F}$.*

Theorem 3 (SiO implies IHO for evasive functions). *Let $\mathcal{F} = \{f_k : \{0,1\}^{l(n)} \to \{0,1\}\}_{n\in\mathbb{N}}$ be an evasive function ensemble,* Obf *be a strong iO for $\mathcal{F}$, then* Obf *is an input-hiding obfuscator for $\mathcal{F}$.*

Proof. Let $\tilde{\mathcal{F}}_n^0$ be the uniform distribution on $\mathcal{F}$ and $\tilde{\mathcal{F}}_n^1$ be the one-element distribution consisting of the zero function. Then $\mathsf{maj}_{\tilde{\mathcal{F}}_n^0} \equiv \mathsf{maj}_{\tilde{\mathcal{F}}_n^1} \equiv 0$. Therefore

$$\Pr_{f_0\leftarrow\tilde{\mathcal{F}}_n^0}[f_0(A(\mathsf{siO}(f_0), z)) = 1] \leq \Pr_{f_1\leftarrow\tilde{\mathcal{F}}_n^1}[f_1(A(\mathsf{siO}(f_1), z)) = 1] + \mathsf{negl}(n) = \mathsf{negl}(n)$$

2.2 Puncturable pseudorandom functions

Definition 7 (Puncturable PRF [18, 19, 44, 51]). *Let $l(n)$ and $m(n)$ be the input and output lengths. A family of puncturable pseudorandom functions $\mathcal{F} = \{F_K\}$ is given by a triple of efficient functions (Gen, Eval, $\mathsf{Puncture}$), where $\mathsf{Gen}(1^n)$ generates the key K, such that F_K maps from $\{0,1\}^{l(n)}$ to $\{0,1\}^{m(n)}$; $\mathsf{Eval}(K, x)$ takes a key K, an input x, outputs $F_K(x)$; $\mathsf{Puncture}(K, x^*)$ takes a key and an input x^*, outputs a punctured key $K\{x^*\}$.*

It satisfies the following conditions:

Functionality preserved over unpunctured points: *For all x^* and keys K, if $K\{x^*\} = \mathsf{Puncture}(K, x^*)$, then for all $x \neq x^*$, $\mathsf{Eval}(K, x) = \mathsf{Eval}(K\{x^*\}, x)$.*

Pseudorandom on the punctured points: *For every input x^*, the value of F on x^* is indistinguishable from random in the presence of the key punctured at x^*. That is, the following two distributions are indistinguishable for every x^*:*

$$(x^*, K\{x^*\}, F_K(x^*)) \text{ and } (x^*, K\{x^*\}, r^*)\,,$$

where K is output by $\mathsf{Gen}(1^n)$, $K\{x^*\}$ is output by $\mathsf{Puncture}(K, x^*)$, and r^* is uniform in $\{0,1\}^{m(n)}$.

Theorem 4 ([18, 19, 36, 44]). *If one-way function exists, then for all length parameters $l(n)$, $m(n)$, there is a puncturable PRF family that maps from $l(n)$ bits to $m(n)$ bits.*

3 Correlation Intractability

We recall the definitions of correlation intractability, initially proposed in [25,26].

Definition 8 (Sparse relations[3]). *A binary relation R is sparse with respect to length parameters $l(n)$, $m(n)$, if there is a negligible function $\delta(\cdot)$ such that for every $x \in \{0,1\}^{l(n)}$:*

$$\Pr_{y \in \{0,1\}^{m(n)}}[R(x, y) = 1] \leq \delta(n)$$

[3] This is called $(l(n), m(n))$-restricted sparse relation in [26], as opposed to the "unrestricted" version where the input length is not prescribed. In this paper we remove the "restriction" in the term, since the case where the input length is unbounded is shown to be impossible (cf. claim 3), and the "restricted" definition is indeed a natural and interesting setting. Also, in [26] and subsequently in [40,41,46], they also define "evasive" relations, which is equivalent to sparse for relations with 1-invocation, and with non-uniform adversaries. Throughout this paper, we only define and use "sparse" relations, since we focus on 1-invocation relations. The term "evasive" only serves the definition of "evasive circuit collections" [4] (cf. def. 2) to avoid confusion.

In some cases, we quantitatively describes the relations as $\delta(n)$-sparse, and even more precisely, $\delta_x(n)$-sparse when specifying the density on the input x.

Definition 9 (Correlation intractability). *A family of functions $\mathcal{H} = \{h_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$ is correlation intractable (CI) if for all (nonuniform, p.p.t.) adversary A, for all sparse relations R, there's a negligible function $\mathsf{negl}(\cdot)$ such that:*

$$\Pr_{k \xleftarrow{\$} \mathcal{H}_n} [x \leftarrow A(k) : R(x, h_k(x)) = 1] < \mathsf{negl}(n)$$

In the definition above, the sparse relations may not be efficiently recognizable. A reasonable weakening on definition 9 is to restrict the relations to be recognizable by poly-size circuits:

Definition 10 (CI-$\mathsf{P/poly}$[4]). *The definition is same as definition 9 except that we restrict the relations to be recognizable by poly-size circuits*

$$C : \{0,1\}^{l(n)+m(n)} \to \{0,1\}$$

s.t. $C(x, y) = 1$ iff $R(x, y) = 1$.

This definition can be further weakened by giving an a priori bound $p(n)$ on the size of the circuit that defines the relation, instead of allowing circuits of arbitrary polynomial size.

Definition 11 (Bounded correlation intractability). *Given a polynomial $p(\cdot)$. A family of functions $\mathcal{H} = \{h_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$ is $p(n)$-bounded correlation intractable (bounded CI, or $p(\cdot)$-CI) if for all (non-uniform, p.p.t.) adversary A, for all sparse relations R that can be recorgnized by a circuit of size smaller or equal to $p(n)$, there's a negligible function $\mathsf{negl}(\cdot)$ such that:*

$$\Pr_{k \xleftarrow{\$} \mathcal{H}_n} [x \leftarrow A(k) : R(x, h_k(x)) = 1] < \mathsf{negl}(n)$$

On the length parameters It is shown in [26] that a function family cannot be correlation intractable when the key length $\kappa(n)$ of the function is short compared to the input length $l(n)$:

Claim ([26]). $\mathcal{H}_n$ is not correlation intractable w.r.t. poly-size relations when $\kappa(n) \leq l(n)$.

Proof. Consider the diagonalization relation $R = \{(k, h_k(k)) | k \in K\}$ (pad k with 0s to get length $l(n)$ if $\kappa(n) < l(n)$). The attacker outputs k (padded with 0s to length $l(n)$ as the x).

[4] This notion is called "weak correlation intractability" in [26].

If $\kappa(n) > l(n)$, then there is no way to pad k to get x. However, some extensions of the impossibility result are still possible; we refer the readers to [26] for the details.

As opposed to the relation between input and key lengths, the relation between input and outputs lengths is not restricted. The only requirement is that the output length $m(n)$ shall be super-logarithmic, i.e. $m(n) \geq \omega(\log(n))$. Although CI is meant to model cryptographic hash functions (which have short outputs), the definition of CI is also meaningful for the functions whose output is longer than their input. In fact, our construction works for both cases.

We note that a function family that is correlation intractable against a more general class of sparse relations captures an essential feature of random oracles better. However, if one is interested in defending against certain restricted types of sparse relations, we may have simpler constructions based on standard cryptographic assumptions. For example, Ajtai's function [1], based on the hardness of approximating the Short Independent Vector Problem for Lattice in the worst case, suffices to prevent the adversary from finding the preimage of any fixed output. We also note that any 1-universal hash function family is correlation intractable, if one only considers very sparse relations — more specifically relations where, for any x, the number of y's that stand in the relation with x is at most a negligible fraction of the ratio between the size of the range and the size of the domain of functions in the family. Indeed, in this case with high probability a random function from the 1-universal hashing family has no input-output pairs in the relation. (We note that in this case the output is inherently longer than the input.)

4 Bounded Correlation Intractability from Obfuscating Puncturable PRF

In this section we give the construction of correlation intractable function ensembles with respect to all the sparse relations recognizable by circuits of size up to a given polynomial $p(\cdot)$.

Construction 5 (Bounded CI) *Let $\mathcal{F} = \{F_K : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$ be a puncturable pseudorandom function. Let the function ensemble $\mathcal{H} = \{h_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$ be constructed as*

$$h_k(\cdot) = \mathsf{iO}(F_K(\cdot), \mathsf{padding}(n))$$

where $K \xleftarrow{\$} \mathcal{F}_n$, for some length of $\mathsf{padding}$.

Theorem 6 (Bounded CI). *Let $p(n)$ be a polynomial in the security parameter n. Assuming the existence of input-hiding obfuscation for all evasive circuits, sub-exponentially secure indistinguishability obfuscation for $\mathsf{P/poly}$, and sub-exponentially secure puncturable PRF, there is an appropriate polynomial size of $\mathsf{padding}$ such that the family $\mathcal{H}$ is $p(n)$-bounded correlation intractable.*

The size of padding (which represents arbitrary gates that do not change the functionality of the circuit) will be discussed at the end of the proof (see remark 2). In short, it depends on p and the blow-up due to input-hiding obfuscation. In the proof below, we drop the explicit mention of padding from the construction in order to simplify notation.

Proof of Theorem 6: The proof in this section follows the outline presented in Section 1.2. The proof goes through 3 hybrids. From the original game which captures the security definition of correlation intractability, we move to intermediate games 1, 2, and 3 that are indistinguishable by the adversary. Finally we will show that the adversary cannot win in game 3 except for negligible probability. We conclude that the adversary also fails in game 0, since the adversary cannot distinguish game 0 and game 3.

More specifically, fix an adversary and a $\delta(n)$-sparse relation R. Then:

Game 0: The original game. The adversary receives the key of the function h_k^0 constructed by the challenger:

$$h_k^0(\cdot) = \mathsf{iO}(F_K(\cdot)) \tag{0}$$

The adversary wins if he outputs an x such that $R(x, h_k^0(x)) = 1$. The winning condition is the same in each subsequent game; what changes is that h^0 is replaced by h^1, h^2, and h^3, which are computed as obfuscations of different circuits, each described in the corresponding game below.

Game 1: Embed the relation into the description without changing the functionality. The challenger samples a puncturable key K, then generates h_k^1 which has the relation R embedded:

$$h_k^1(x) = \mathsf{iO}\begin{pmatrix}\text{if } R(x, F_K(x)) = 1, & \text{return } F_K(x) \\ \text{else,} & \text{return } F_K(x)\end{pmatrix} \tag{1}$$

The hybrids h_k^0 and h_k^1 have identical functionality. Therefore, because both h_k^0 and h_k^1 are obfuscated by iO, they are indistinguishable for any p.p.t. adversary.

Game 2: Switch to a function where the "innocent" branch is generated independently from the "bad" branch and avoids R. The challenger constructs a new function family $\mathcal{F}^R$ that always avoids R, as described below, and generates h_k^2 as:

$$h_k^2(x) = \mathsf{iO}\begin{pmatrix}\text{if } R(x, F_K(x)) = 1, & \text{return } F_K(x) \\ \text{else,} & \text{return } F_{K'}^R(x)\end{pmatrix} \tag{2}$$

where $F_K \xleftarrow{\$} \mathcal{F}_n$ and $F_{K'}^R \xleftarrow{\$} \mathcal{F}^R$. The function family $\mathcal{F}^R$ is constructed as follows:

Construction 7 **($\mathcal{F}^R$)** *Let $\mathcal{F}^R = \{F^R_{K'} : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_n$ be a function family, where each $F^R_{K'}$ is constructed as follows:*

$$F^R_{K'}(x) = \left(\begin{array}{l} \underline{K' = (K_1, K_2, \ldots, K_{T(n)})} \\ \text{for } i = 1 \text{ to } T(n): \\ \quad \text{if } R(x, F_{K_i}(x)) = 0, \text{ return } F_{K_i}(x) \\ \text{return } \perp \end{array} \right) \tag{2.else}$$

where $T(n) = \frac{l(n)}{\log(n)}$. The functions F_{K_1}, ..., $F_{K_{T(n)}}$ are sampled independently from any puncturable PRF family $\mathcal{F}$.

The functionality of $F^R_{K'}$ is to output, given an input x, the pseudorandom value $F_{K_i}(x)$, where K_i is the first key among $K_1, ..., K_{T(n)}$ s.t. $R(x, F_{K_i}(x)) = 0$ (if no such K_i exists, output $\perp$). The iteration bound $T(n)$ is set large enough to make sure that $F^R_{K'}$ outputs $\perp$ with probability less than $2^{-l(n)} \cdot \mathsf{negl}(n)$ (we prove and use this fact in Lemma 2).

To prove that h^2_k is indistinguishable from h^1_k, let g^2_k be the same as h^2_k but without the iO:

$$g^2_k(x) = \begin{cases} \text{if } R(x, F_K(x)) = 1, \text{return } F_K(x) \\ \text{else,} \qquad\qquad\qquad\quad \text{return } F^R_{K'}(x) \end{cases} \tag{2.inner}$$

First, using subexponential security of F_K, we show in Lemma 2 that the g^2_k is also a subexponentially secure puncturable PRF. Then, in Lemma 3 (whose proof methodology is derived from the work of Canetti et al. [27]), we show that any two subexponentially secure puncturable PRFs are indistinguishable after being obfuscated by subexponentially secure iO. This makes $h^k_2 = \mathsf{iO}(g^2_k)$ indistinguishable from $h^k_0 = \mathsf{iO}(F_K)$, and therefore also indistinguishable from h^k_1. (Note that technically h^k_1 is not needed at all—we can move directly from h^k_0 to h^k_2; but we believe that moving to h^k_1 first clarifies presentation.)

Lemma 2 and Lemma 3 below are based on the sub-exponential hardness of puncturability and iO, respectively. Let $\epsilon_{\mathsf{Puncture}}$ be the adversary's advantage of winning the puncturability game of $\mathcal{F}$ and ϵ_{iO} be the advantage of distinguishing the iO of two identical functions. We need to set

$$\epsilon_{\mathsf{Puncture}} = \epsilon_{\mathsf{iO}} = 2^{-l(n)} \cdot \mathsf{negl}(n)$$

This level of security can always be achieved from subexponential hardness by setting the security parameter λ for the puncturable PRF and for iO sufficiently high, but still polynomial in n: if the security of these two objects is $2^{-\lambda^\epsilon}$ for security parameter λ, then setting $\lambda = (2l(n))^{1/\epsilon}$ is sufficient.

Lemma 2. *Assume that $\mathcal{F}$ is a subexponentially secure puncturable PRF with the advantage of distinguishing being $\epsilon_{\mathsf{Puncture}} = 2^{-l(n)} \cdot \mathsf{negl}(n)$. Then the function g^2_k (i.e., the function being obfuscated in h^2_k) is also a subexponentially secure puncturable PRF with the advantage of distinguishing at most $2^{-l(n)} \cdot \mathsf{negl}(n)$.*

Proof. To puncture g_k^2 on input x^*, we puncture all the inner PRF keys K, K_1, ..., $K_{T(n)}$ on x^*, and construct the punctured function as follows:

$$\begin{array}{l}\underline{k\{x^*\} = (R, K\{x^*\}, K'\{x^*\} = (K_1\{x^*\}, \ldots, K_{T(n)}\{x^*\}))} \\ g_{k\{x^*\}}(x) = \begin{pmatrix} \text{if } R(x, F_{K\{x^*\}}(x)) = 1, & \text{return } F_{K\{x^*\}}(x) \\ \text{else,} & \text{return } F^R_{K'\{x^*\}}(x) \end{pmatrix}\end{array} \tag{2.p}$$

where $F^R_{K'\{x^*\}}$ is constructed as:

$$F^R_{K'\{x^*\}}(x) = \begin{pmatrix} \underline{K'\{x^*\} = (K_1\{x^*\}, \ldots, K_{T(n)}\{x^*\})} \\ \text{for } i = 1 \text{ to } T(n): \\ \quad \text{if } R(x, F_{K_i\{x^*\}}(x)) = 0,\ \text{return } F_{K_i\{x^*\}}(x) \\ \text{return } \perp \end{pmatrix} \tag{2.else.p}$$

By the puncturability of $\mathcal{F}$, the outputs of $F_{K\{x^*\}}$ and $F_{K_i\{x^*\}}$ on the punctured points are indistinguishable from random even given $k\{x^*\}$. More precisely,

$$\big(k\{x^*\}, F_K(x^*), F_{K_1}(x^*), ..., F_{K_{T(n)}}(x^*)\big) \approx \big(k\{x^*\}, U_0, U_1, \ldots, U_{T(n)}\big)$$

(where $(U_0, U_1, ..., U_{T(n)}) \xleftarrow{\$} \{0,1\}^{(T(n)+1)\cdot m(n)}$). The advantage of any p.p.t. adversary to distinguish these two tuples is

$$(T(n)+1)\cdot \epsilon_{\mathsf{Puncture}} = (T(n)+1)\cdot 2^{-l(n)} \cdot \mathsf{negl}(n) = 2^{-l(n)} \cdot \mathsf{negl}(n)$$

Construct the distribution V_{x^*} by sampling random $U_0, \ldots, U_{T(n)}$ and computing

$$V_{x^*} = \begin{pmatrix} \text{if } R(x^*, U_0) = 1,\ \text{return } U_0 \\ \text{else}:\ \text{for } i = 1 \text{ to } T(n): \\ \qquad\qquad \text{if } R(x^*, U_i) = 0,\ \text{return } U_i \\ \text{return } \perp \end{pmatrix}$$

From the indistinguishability of $F_K(x^*)$ and $F_{K_i}(x^*)$ from uniform, it follows that V_{x^*} is indistinguishable from $g_k^2(x^*)$:

$$\big(k\{x^*\}, g_k^2(x^*)\big) \approx (k\{x^*\}, V_{x^*})$$

and the advantage of any p.p.t. adversary to distinguish these two pairs is $2^{-l(n)}\cdot \mathsf{negl}(n)$. To complete the proof, we will show that V_{x^*} is very close to uniform over $\{0,1\}^{m(n)}$: it differs from uniform by the probability that $V_{x^*} = \perp$. Indeed,

- For all $y \in \{0,1\}^{m(n)}$ such that $R(x^*, y) = 1$,

$$\Pr[V_{x^*} = y] = \Pr[U_0 = y] = 2^{-m(n)}$$

- $\Pr[V_{x^*} = \perp] = (1 - \delta_{x^*}(n))\delta_{x^*}(n)^{T(n)}$

– For all $y \in \{0,1\}^{m(n)}$ such that $R(x^*, y) = 0$ (note that there are $2^{m(n)}(1 - \delta_{x^*}(n))$ such values)

$$\begin{aligned}
&\Pr[V_{x^*} = y] \\
=\ &\Pr[V_{x^*} = y | R(x^*, V_{x^*}) \neq 1 \wedge V_{x^*} \neq \perp] \Pr[R(x^*, V_{x^*}) \neq 1 \wedge V_{x^*} \neq \perp] \\
=\ &\frac{1}{2^{m(n)}(1 - \delta_{x^*}(n))}(1 - \Pr[V_{x^*} \neq \perp \wedge R(x^*, V_{x^*}) = 1] - \Pr[V_{x^*} = \perp]) \\
=\ &\frac{1}{2^{m(n)}(1 - \delta_{x^*}(n))}(1 - \delta_{x^*}(n) - (1 - \delta_{x^*}(n))\delta_{x^*}(n)^{T(n)}) \\
=\ &2^{-m(n)} \cdot \left(1 - \frac{(1 - \delta_{x^*}(n))\delta_{x^*}(n)^{T(n)}}{1 - \delta_{x^*}(n)}\right) = 2^{-m(n)} \cdot \left(1 - \delta_{x^*}(n)^{T(n)}\right)
\end{aligned}$$

Thus, the statistical difference between V_{x^*} and the uniform distribution on $\{0,1\}^{m(n)}$ (which is a bound on any distinguisher's advantage) is

$$\begin{aligned}
&\frac{1}{2} \sum_{y \in \{\perp\} \cup \{0,1\}^n} |\Pr[V_{x^*} = y] - \Pr[U = y]| \quad (U \text{ is uniform over } \{0,1\}^{m(n)}) \\
= &\frac{1}{2}\Bigg((1 - \delta_{x^*}(n))\delta_{x^*}(n)^{T(n)} \\
&\qquad + \sum_{y \text{ s.t. } R(x^*,y)=0} \left(2^{-m(n)} - 2^{-m(n)} \cdot \left(1 - \delta_{x^*}(n)^{T(n)}\right)\right)\Bigg) \\
= &(1 - \delta_{x^*}(n))\delta_{x^*}(n)^{T(n)} \leq \delta_{x^*}(n)^{T(n)}
\end{aligned}$$

We thus obtain that V_{x^*} can be distinguished from uniform with advantage at most $\delta_{x^*}(n)^{T(n)} = 2^{-l(n)} \cdot \mathsf{negl}(n)$, because $T(n) = \frac{l(n)}{\log(n)}$ and $\delta_x(n)$ is a negligible function.

V_{x^*} is independent of $k\{x^*\}$. Therefore, the advantage of any adversary in distinguishing $(k\{x^*\}, V_{x^*})$ from $(k\{x^*\}, U)$ is $2^{-l(n)} \cdot \mathsf{negl}(n)$. And we already know the same is true for distinguishing $(k\{x^*\}, g_k^2(x^*))$ from $(k\{x^*\}, V_{x^*})$. Thus, even given $k\{x^*\}$, g_k^2 cannot be distinguished from uniform with advantage better than $2^{-l(n)} \cdot \mathsf{negl}(n)$, which concludes the proof.

Next we show that for arbitrary puncturable PRF families $\mathcal{F}_1, \mathcal{F}_2 : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}$ that are $2^{-l(n)} \cdot \mathsf{negl}(n)$-secure, the pseudorandom functions sampled independently from these families are indistinguishable after being obfuscated by $2^{-l(n)} \cdot \mathsf{negl}(n)$-secure indistinguishability obfuscation. The following lemma is derived from the "piO" proof methodology developed in the work of Canetti et al. [27].

Lemma 3. *Let $\mathcal{F}_1, \mathcal{F}_2 : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}$ be $2^{-l(n)} \cdot \mathsf{negl}(n)$-secure puncturable PRF families, iO be $\epsilon_{\mathsf{iO}} = 2^{-l(n)} \cdot \mathsf{negl}(n)$-secure indistinguishability obfuscation. Let $F_{K_1} \xleftarrow{\$} \mathcal{F}_1$, $F_{K_2} \xleftarrow{\$} \mathcal{F}_2$, then $\mathsf{iO}(F_{K_1})$ and $\mathsf{iO}(F_{K_2})$ are indistinguishable.*

Proof. We prove the indistinguishability via $2^{l(n)} + 1$ intermediate hybrids, one for each input. More precisely, for $z^* \in \{0, 1, ..., 2^{l(n)} - 1, 2^{l(n)}\}$, we construct f_{z^*} as

$$f_{z^*}(x) = \mathsf{iO}\begin{pmatrix} \text{if } x = z^*, \text{ return } F_{K_1}(x) & \\ \text{else,} \qquad \text{return} \begin{pmatrix} \text{if } x > z^*, \text{ return } F_{K_1}(x) \\ \text{else,} \qquad \text{return } F_{K_2}(x) \end{pmatrix} \end{pmatrix}$$

Note that f_0 is functionally equivalent to F_{K_1}, therefore, they are $2^{-l(n)} \cdot \mathsf{negl}(n)$ indistinguishable after being obfuscated by iO. Likewise, $f_{2^{l(n)}}$ is functionally equivalent to F_{K_2}, hence being $2^{-l(n)} \cdot \mathsf{negl}(n)$-indistinguishable following iO.

Next we show that each intermediate pairs f_{z^*} and f_{z^*+1}, $z^* \in \{0, 1, ..., 2^{l(n)} - 1\}$, are $2^{-l(n)} \cdot \mathsf{negl}(n)$-indistinguishable. We introduce 3 more sub-hybrids:

$$f_{z^*,y^*}(x) = \mathsf{iO}\begin{pmatrix} \text{if } x = z^*, \text{ return } y^* & \\ \text{else,} \qquad \text{return} \begin{pmatrix} \text{if } x > z^*, \text{ return } F_{K_1\{z^*\}}(x) \\ \text{else,} \qquad \text{return } F_{K_2\{z^*\}}(x) \end{pmatrix} \end{pmatrix}$$

where y^* equals to $F_{K_1}(z^*)$, $U \xleftarrow{\$} \{0,1\}^{m(n)}$, and $F_{K_2}(z^*)$ respectively.

Note that $f_{z^*,F_{K_1}(z^*)}$ is functionally equivalent to f_{z^*}; $f_{z^*,F_{K_2}(z^*)}$ is functionally equivalent to f_{z^*+1}. They are $2^{-l(n)} \cdot \mathsf{negl}(n)$-indistinguishable following iO. In between, $f_{z^*,F_{K_1}(z^*)}$ is indistinguishable from $f_{z^*,U}$ and $f_{z^*,U}$ is indistinguishable from $f_{z^*,F_{K_2}(z^*)}$, following the $2^{-l(n)} \cdot \mathsf{negl}(n)$-puncturability of K_1 and K_2.

To conclude, f_{z^*} and f_{z^*+1} are $4 \cdot 2^{-l(n)} \cdot \mathsf{negl}(n)$-indistinguishable following the $2^{-l(n)} \cdot \mathsf{negl}(n)$ security of $\mathcal{F}_1$, $\mathcal{F}_2$, and iO. Summing up all the $2^{l(n)} + 1$ intermediate hybrids, the total advantage of distinguishing $\mathsf{iO}(F_{K_1})$ and $\mathsf{iO}(F_{K_2})$ is negligible.

Combining lemma 2 and lemma 3, h_k^1 is indistinguishable from h_k^2.

Game 3: Wrap the "bad" branch by input-hiding obfuscation, without changing the functionality. The challenger generates h_k^3 that is functionally equivalent to h_k^2 but is computed differently. The difference is that in game 3, the challenger first wraps the if statement together with the true branch with input-hiding obfuscation (the challenger also applies iO to the entire function, just like in the previous games, which ensures that h_k^2 is indistinguishable from h_k^3):

$$h_k^3(x) = \mathsf{iO}\begin{pmatrix} y \leftarrow \mathsf{IHO}\begin{pmatrix} \text{if } R(x, F_K(x)) = 1, \text{ return } F_K(x) \\ \text{else,} \qquad\qquad\qquad \text{return } \bot \end{pmatrix} \\ \text{if } y = \bot \text{ , } y \leftarrow F^R_{K'}(x) \\ \text{return } y \end{pmatrix} \tag{3}$$

Let $E^R_K(x)$ denote $\begin{pmatrix} \text{if } R(x, F_K(x)) = 1, & \text{return } F_K(x) \\ \text{else,} & \text{return } \perp \end{pmatrix}$.

Proposition 1. *$\mathcal{E}^R = \{E^R_K : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$ is an evasive circuit family.*

Proof. Assume, for contradiction, that there is an input $x' \in \{0,1\}^{l(n)}$ on which there are non-negligibly many keys that evaluate to a value other than $\perp$. We can then build a (non-uniform) adversary that distinguishes the PRF $F_K(x)$ from a truly random function with non-neglible advantage. The adversary simply queries input x' to the function and checks if the output y satisfies $R(x', y)$.

Note that h^2_k and h^3_k are functionally equivalent. Therefore, by indistinguishability obfuscation, the adversary cannot distinguish game 2 and game 3.

Finally, in Game 3: Suppose that there is a p.p.t. adversary A who gets h^3_k, finds an input x such that $R(x, h^3_k(x)) = 1$ with non-negligible probability $\eta(n)$, we build an adversary A' that breaks IHO for evasive circuit family $\mathcal{E}^R$: A' gets $\mathsf{IHO}(E^R_K(\cdot))$, samples $F^R_{K'}$ independently, and creates h^3_k as described in construction (3), sends it to A. For adversary A, finding an input x to h^3 such that $R(x, h^3_k(x)) = 1$ is equivalent to finding such an input to $\mathsf{IHO}(E^R_K(\cdot))$ that evaluates to an non-bottom value, because $F^R_{K'}$ is independently generated and always avoids R ($F^R_{K'}$ outputs $\perp$ rather than hit R).

The advantage of adversary A' is thus the following:

$$\begin{aligned} &\Pr_K[A'(\mathsf{IHO}(E_{R,K}(\cdot))) \to x : \; E_{R,K}(x) \neq \perp] \\ =&\Pr_{K,K'}[A(\mathsf{IHO}(E_{R,K}(\cdot)), R, F^R_{K'}) \to x : \; E_{R,K}(x) \neq \perp] \\ \geq&\Pr_k[A(h^3_k(\cdot)) \to x : R(x, h^3_k(x)) = 1] \geq \eta(n) \end{aligned}$$

which forms the contradiction.

If a p.p.t. adversary could find x such $R(x, h^0_k(x)) = 1$, then she could distinguish h^0 from h^3 (because testing R is polynomial-time). Thus, we complete the proof that $\mathcal{H}$ is correlation intractable. □

Remark 2 (The size of padding). Let $\kappa_{\mathcal{F}}(n)$ be the key size of $\mathcal{F}_n$, $\kappa^*_{\mathcal{F}}(n)$ be the punctured key size of $\mathcal{F}_n$, $B(\cdot)$ be the maximum blow-up of the input-hiding obfuscation. The size of $F^R_{K'}$ is $T(n) \cdot (p(n) + 2 \cdot \kappa_{\mathcal{F}}(n))$. The maximum size of $\mathsf{IHO}(E_{R,K})$ is $B(p(n) + 2 \cdot \kappa_{\mathcal{F}}(n))$. The size of padding is bounded by

$$\begin{aligned} &|\mathsf{padding}(n)| \\ \leq\; & B(p(n) + 2 \cdot \kappa_{\mathcal{F}}(n)) + T(n) \cdot (p(n) + 2 \cdot \kappa_{\mathcal{F}}(n)) + (T(n) + 2) \cdot \kappa^*_{\mathcal{F}}(n) \\ =\; & \mathsf{poly}(n) \end{aligned}$$

As the analysis suggests, the key size of the function inherently exceeds the maximum size of R. The existence of correlation intractable functions with a prescribed description size that works for all poly-size relations (i.e. CI-P/poly) remains an open problem.

Acknowledgments

We are grateful to Nir Bitansky, Cheng Chen, Omer Paneth, and Oxana Poburinnaya for their enlightening discussions in the early stage of this work. We thank Ethan Heilman for discussions on the Bitcoin protocol, and Stefano Tessaro for pointing out "the piO proof methodology", which simplified the proof compared to earlier versions of this paper. We also thank the anonymous reviewers for their helpful comments.

This work is supported by US NSF grants 1012798, 1012910, 1218461, 1413920, 1422965, ISF grant 1523/14, and the Check Point Institute for Information Security. Part of the research by Y.C. was conducted while at Tel Aviv University funded by the Check Point Institute for Information Security.

References

1. Miklós Ajtai. Generating hard instances of lattice problems (extended abstract). In *STOC*, pages 99–108, 1996.
2. Benny Applebaum and Zvika Brakerski. Obfuscating circuits via composite-order graded encoding. In Dodis and Nielsen [31], pages 528–556.
3. Saikrishna Badrinarayanan, Eric Miles, Amit Sahai, and Mark Zhandry. Post-zeroizing obfuscation: The case of evasive circuits. Cryptology ePrint Archive, Report 2015/167, 2015. http://eprint.iacr.org/.
4. Boaz Barak, Nir Bitansky, Ran Canetti, Yael Tauman Kalai, Omer Paneth, and Amit Sahai. Obfuscation for evasive functions. In Lindell [45], pages 26–51.
5. Boaz Barak, Sanjam Garg, Yael Tauman Kalai, Omer Paneth, and Amit Sahai. Protecting obfuscation against algebraic attacks. In Nguyen and Oswald [48], pages 221–238.
6. Boaz Barak, Oded Goldreich, Russell Impagliazzo, Steven Rudich, Amit Sahai, Salil P. Vadhan, and Ke Yang. On the (im)possibility of obfuscating programs. *J. ACM*, 59(2):6, 2012.
7. Boaz Barak, Yehuda Lindell, and Salil P. Vadhan. Lower bounds for non-black-box zero knowledge. *J. Comput. Syst. Sci.*, 72(2):321–391, 2006.
8. Mihir Bellare, Viet Tung Hoang, and Sriram Keelveedhi. Instantiating random oracles via uces. In Ran Canetti and Juan A. Garay, editors, *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part II*, volume 8043 of *Lecture Notes in Computer Science*, pages 398–415. Springer, 2013.
9. Mihir Bellare and Phillip Rogaway. Random oracles are practical: A paradigm for designing efficient protocols. In *ACM Conference on Computer and Communications Security*, pages 62–73, 1993.

10. Mihir Bellare, Igors Stepanovs, and Stefano Tessaro. Poly-many hardcore bits for any one-way function and a framework for differing-inputs obfuscation. In Sarkar and Iwata [52], pages 102–121.
11. Nir Bitansky and Ran Canetti. On strong simulation and composable point obfuscation. In Tal Rabin, editor, *Advances in Cryptology - CRYPTO 2010, 30th Annual Cryptology Conference, Santa Barbara, CA, USA, August 15-19, 2010. Proceedings*, volume 6223 of *Lecture Notes in Computer Science*, pages 520–537. Springer, 2010.
12. Nir Bitansky, Ran Canetti, Henry Cohn, Shafi Goldwasser, Yael Tauman Kalai, Omer Paneth, and Alon Rosen. The impossibility of obfuscation with auxiliary input or a universal simulator. In Garay and Gennaro [33], pages 71–89.
13. Nir Bitansky, Ran Canetti, Yael Tauman Kalai, and Omer Paneth. On virtual grey box obfuscation for general circuits. In Garay and Gennaro [33], pages 108–125.
14. Nir Bitansky, Dana Dachman-Soled, Sanjam Garg, Abhishek Jain, Yael Tauman Kalai, Adriana López-Alt, and Daniel Wichs. Why ”fiat-shamir for proofs” lacks a proof. In *TCC*, pages 182–201, 2013.
15. Manuel Blum. How to prove a theorem so no one else can claim it, August 1986. Invited 45 minute address to the International Congress of Mathematicians, 1986. To appear in the Proceedings of ICM 86.
16. Alexandra Boldyreva, David Cash, Marc Fischlin, and Bogdan Warinschi. Foundations of non-malleable hash and one-way functions. In Mitsuru Matsui, editor, *ASIACRYPT*, volume 5912 of *Lecture Notes in Computer Science*, pages 524–541. Springer, 2009.
17. Dan Boneh, Kevin Lewi, Hart William Montgomery, and Ananth Raghunathan. Key homomorphic prfs and their applications. In Canetti and Garay [24], pages 410–428.
18. Dan Boneh and Brent Waters. Constrained pseudorandom functions and their applications. In Kazue Sako and Palash Sarkar, editors, *ASIACRYPT (2)*, volume 8270 of *Lecture Notes in Computer Science*, pages 280–300. Springer, 2013.
19. Elette Boyle, Shafi Goldwasser, and Ioana Ivan. Functional signatures and pseudorandom functions. In Hugo Krawczyk, editor, *Public Key Cryptography*, volume 8383 of *Lecture Notes in Computer Science*, pages 501–519. Springer, 2014.
20. Zvika Brakerski and Guy N. Rothblum. Virtual black-box obfuscation for all circuits via generic graded encoding. In Lindell [45], pages 1–25.
21. Christina Brzuska and Arno Mittelbach. Indistinguishability obfuscation versus multi-bit point obfuscation with auxiliary input. In Sarkar and Iwata [52], pages 142–161.
22. Christina Brzuska and Arno Mittelbach. Using indistinguishability obfuscation via uces. In Sarkar and Iwata [52], pages 122–141.
23. Ran Canetti. Towards realizing random oracles: Hash functions that hide all partial information. In Burton S. Kaliski Jr., editor, *Advances in Cryptology - CRYPTO '97, 17th Annual International Cryptology Conference, Santa Barbara, California, USA, August 17-21, 1997, Proceedings*, volume 1294 of *Lecture Notes in Computer Science*, pages 455–469. Springer, 1997.
24. Ran Canetti and Juan A. Garay, editors. *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part I*, volume 8042 of *Lecture Notes in Computer Science*. Springer, 2013.
25. Ran Canetti, Oded Goldreich, and Shai Halevi. The random oracle methodology, revisited (preliminary version). In Vitter [53], pages 209–218.

26. Ran Canetti, Oded Goldreich, and Shai Halevi. The random oracle methodology, revisited. *J. ACM*, 51(4):557–594, 2004.
27. Ran Canetti, Huijia Lin, Stefano Tessaro, and Vinod Vaikuntanathan. Obfuscation of probabilistic circuits and applications. In Dodis and Nielsen [31], pages 468–497.
28. Ran Canetti, Daniele Micciancio, and Omer Reingold. Perfectly one-way probabilistic hash functions (preliminary version). In Vitter [53], pages 131–140.
29. Jung Hee Cheon, Kyoohyung Han, Changmin Lee, Hansol Ryu, and Damien Stehlé. Cryptanalysis of the multilinear map over the integers. In Elisabeth Oswald and Marc Fischlin, editors, *Advances in Cryptology - EUROCRYPT 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part I*, volume 9056 of *Lecture Notes in Computer Science*, pages 3–12. Springer, 2015.
30. Jean-Sébastien Coron, Tancrède Lepoint, and Mehdi Tibouchi. Practical multilinear maps over the integers. In Canetti and Garay [24], pages 476–493.
31. Yevgeniy Dodis and Jesper Buus Nielsen, editors. *Theory of Cryptography - 12th Theory of Cryptography Conference, TCC 2015, Warsaw, Poland, March 23-25, 2015, Proceedings, Part II*, volume 9015 of *Lecture Notes in Computer Science*. Springer, 2015.
32. Amos Fiat and Adi Shamir. How to prove yourself: Practical solutions to identification and signature problems. In *CRYPTO*, pages 186–194, 1986.
33. Juan A. Garay and Rosario Gennaro, editors. *Advances in Cryptology - CRYPTO 2014 - 34th Annual Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2014, Proceedings, Part II*, volume 8617 of *Lecture Notes in Computer Science*. Springer, 2014.
34. Sanjam Garg, Craig Gentry, and Shai Halevi. Candidate multilinear maps from ideal lattices. In Thomas Johansson and Phong Q. Nguyen, editors, *EUROCRYPT*, volume 7881 of *Lecture Notes in Computer Science*, pages 1–17. Springer, 2013.
35. Oded Goldreich. The ggm construction does not yield correlation intractable function ensembles. *IACR Cryptology ePrint Archive*, 2002:110, 2002.
36. Oded Goldreich, Shafi Goldwasser, and Silvio Micali. How to construct random functions. *J. ACM*, 33(4):792–807, 1986.
37. Shafi Goldwasser and Yael Tauman Kalai. On the impossibility of obfuscation with auxiliary input. In *FOCS*, pages 553–562. IEEE Computer Society, 2005.
38. Vipul Goyal, Adam O'Neill, and Vanishree Rao. Correlated-input secure hash functions. In Yuval Ishai, editor, *Theory of Cryptography - 8th Theory of Cryptography Conference, TCC 2011, Providence, RI, USA, March 28-30, 2011. Proceedings*, volume 6597 of *Lecture Notes in Computer Science*, pages 182–200. Springer, 2011.
39. Satoshi Hada. Zero-knowledge and code obfuscation. In Tatsuaki Okamoto, editor, *ASIACRYPT*, volume 1976 of *Lecture Notes in Computer Science*, pages 443–457. Springer, 2000.
40. Satoshi Hada and Toshiaki Tanaka. Zero-knowledge and correlation intractability. *IEICE Transactions*, 89-A(10):2894–2905, 2006.
41. Shai Halevi, Steven Myers, and Charles Rackoff. On seed-incompressible functions. In Ran Canetti, editor, *TCC*, volume 4948 of *Lecture Notes in Computer Science*, pages 19–36. Springer, 2008.
42. Susan Hohenberger, Amit Sahai, and Brent Waters. Replacing a random oracle: Full domain hash from indistinguishability obfuscation. In Nguyen and Oswald [48], pages 201–220.
43. Yuval Ishai, Joe Kilian, Kobbi Nissim, and Erez Petrank. Extending oblivious transfers efficiently. In Dan Boneh, editor, *Advances in Cryptology - CRYPTO*

2003, 23rd Annual International Cryptology Conference, Santa Barbara, California, USA, August 17-21, 2003, Proceedings, volume 2729 of *Lecture Notes in Computer Science*, pages 145–161. Springer, 2003.

44. Aggelos Kiayias, Stavros Papadopoulos, Nikos Triandopoulos, and Thomas Zacharias. Delegatable pseudorandom functions and applications. In Ahmad-Reza Sadeghi, Virgil D. Gligor, and Moti Yung, editors, *ACM Conference on Computer and Communications Security*, pages 669–684. ACM, 2013.
45. Yehuda Lindell, editor. *Theory of Cryptography - 11th Theory of Cryptography Conference, TCC 2014, San Diego, CA, USA, February 24-26, 2014. Proceedings*, volume 8349 of *Lecture Notes in Computer Science*. Springer, 2014.
46. Avradip Mandal, Jacques Patarin, and Yannick Seurin. On the public indifferentiability and correlation intractability of the 6-round feistel construction. In Ronald Cramer, editor, *Theory of Cryptography - 9th Theory of Cryptography Conference, TCC 2012, Taormina, Sicily, Italy, March 19-21, 2012. Proceedings*, volume 7194 of *Lecture Notes in Computer Science*, pages 285–302. Springer, 2012.
47. Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. *Consulted*, 1(2012):28, 2008.
48. Phong Q. Nguyen and Elisabeth Oswald, editors. *Advances in Cryptology - EUROCRYPT 2014 - 33rd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Copenhagen, Denmark, May 11-15, 2014. Proceedings*, volume 8441 of *Lecture Notes in Computer Science*. Springer, 2014.
49. Kobbi Nissim. Two results regarding correlation intractability. *Manuscript*, 1999.
50. Rafael Pass, Karn Seth, and Sidharth Telang. Indistinguishability obfuscation from semantically-secure multilinear encodings. In Juan A. Garay and Rosario Gennaro, editors, *CRYPTO (1)*, volume 8616 of *Lecture Notes in Computer Science*, pages 500–517. Springer, 2014.
51. Amit Sahai and Brent Waters. How to use indistinguishability obfuscation: deniable encryption, and more. In David B. Shmoys, editor, *STOC*, pages 475–484. ACM, 2014.
52. Palash Sarkar and Tetsu Iwata, editors. *Advances in Cryptology - ASIACRYPT 2014 - 20th International Conference on the Theory and Application of Cryptology and Information Security, Kaoshiung, Taiwan, R.O.C., December 7-11, 2014, Proceedings, Part II*, volume 8874 of *Lecture Notes in Computer Science*. Springer, 2014.
53. Jeffrey Scott Vitter, editor. *Proceedings of the Thirtieth Annual ACM Symposium on the Theory of Computing, Dallas, Texas, USA, May 23-26, 1998*. ACM, 1998.
54. Joe Zimmerman. How to obfuscate programs directly. In Elisabeth Oswald and Marc Fischlin, editors, *Advances in Cryptology - EUROCRYPT 2015 - 34th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Sofia, Bulgaria, April 26-30, 2015, Proceedings, Part II*, volume 9057 of *Lecture Notes in Computer Science*, pages 439–467. Springer, 2015.

Appendices

A Correlation intractability versus other notions

We explore the relation between correlation intractability and other security definitions for cryptographic hash functions.

A.1 Relations with entropy-preserving hashing

Recall the definition of *Entropy Preserving* (EP) from [7]:

Definition 12 (Entropy preservation). *A family of hash function* $\mathcal{H} = \{h_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}, k = g(s), s \in \{0,1\}^{\sigma(n)}\}_{n\in\mathbb{N}}$ *ensures conditional entropy*[5] *greater than* $\delta(n)$ *if for all (non-uniform, p.p.t.) adversary* A*:*

$$H(h_k(A(k))|A(k)) > \delta(n)$$

Equivalently:

$$\mathbb{E}_{k,A}[H(h_k(X)|_{X=A(k)})] > \delta(n)$$

Notice that in order to get meaningful (i.e. non-zero) conditional entropy, the length of the key $\kappa(n)$ must be bigger then the length of the input $l(n)$, otherwise the adversary could always output the key (i.e. $A(k) \to k$) so that the conditional entropy will be zero (same to the diagonalization attack of correlation intractability [26]). In other words, we hope that there are multiple choices of keys that could lead the adversary to return the same input, and $h_k(x)$ on these candidate seeds and fixed input has different values.

[7] proposed 3 bounds for $\delta(n)$, each being interested on its own:

- (Best possible) $\delta(n) > m(n) - O(\log n)$. If achievable, would imply that constant-round public-coin auxiliary-input zero-knowledge proofs exist only for languages in BPP.
- (Somewhat) $\delta(n) > 1/\mathsf{poly}(n)$, also interesting. If achievable, would imply that 3-round public-coin auxiliary-input optimally sound zero-knowledge proofs exist only for languages in BPP.
- (Minimum/Weakest) $\delta(n) > 0$, still interesting. Even the existance of the weakest entropy-preserving hash functions implies that the parallel composition of some classic protocols (e.g. Blum's protocol [15]) is not auxiliary-input zero-knowledge.

An equivalent formalization of the minimum/weakest notion:

Conjecture 1 ([7]). There is a polynomial $p(\cdot)$ such that the following holds: For every *non-uniform deterministic* polynomial-time algorithm A and all sufficiently large n, there are circuits C_1, C_2 of size at most $p(n)$ such that $\alpha = A(C_1) = A(C_2)$ but $C_1(\alpha) \neq C_2(\alpha)$.

Note that even the construction of the weakest notion of entropy-preservation is unknown. In fact it is shown by Bitansky et al. to be impossible to obtain from black-box reduction to falsifiable assumptions [14].

[5] The entropy of a random variable X is defined as $H(X) = \mathbb{E}_{x \xleftarrow{\$} X}[\log \frac{1}{\Pr[X=x]}]$. For jointly distributed random variables (X, Y), the conditional entropy of X given Y is defined to be $\mathbb{E}_{y \xleftarrow{\$} Y}[H(X|_{Y=y})]$, where $X|_{Y=y}$ denotes the conditional distribution of X given that $Y = y$.

Connections with CI. We show that correlation intractability (where the sparse relations are not necessarily efficiently recognizable) impies entropy preservation; and entropy preservation implies a weaker variant of correlation intractability in which if the adversary exists, it breaks correlation intractability with probability 1.

Theorem 8. *If a function family $\mathcal{H}$ is correlation intractable, then it is also entropy-preserving, i.e. for all p.p.t. adversary A:*

$$H(h_k(A(k))|A(k)) > m(n) - O(\log(n))$$

Proof. Assume by contradiction that $\mathcal{H}$ is not entropy-preserving, then there's an Adv A, such that

$$H(h_k(A(k))|A(k)) < m(n) - \omega(\log(n))$$

We define a relation by enumerating the keys, and query A on each key to get x, and the corresponding $y = h_k(x)$, then adding (x, y) into the relation. Formally, let R be:

$$R = \{(x, h_k(x)) \mid x = A(k),\ k = g(s),\ s \in \{0,1\}^{\sigma(n)}\}$$

R is sparse since the adversary can always break entropy-preservation, which means the portion of the possible outputs conditioned on the adversary's choice of the input is negligible.

Notice that this relation is not likely to be efficiently recognizable, which means our construction of bounded correlation intractable functions is not necessarily entropy-preserving.

Definition 13 (Weak correlation intractability[6]). *A family of functions $\mathcal{H} = \{h_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$ is weak correlation intractable (wCI) if for all (non-uniform, p.p.t.) adversary A, for all sparse relations R, there's a non-negligible function $\mathsf{non.negl}(\cdot)$ such that:*

$$\Pr_{k \xleftarrow{\$} \mathcal{H}_n}[x \leftarrow A(k) : R(x, h_k(x)) = 1] < 1 - \mathsf{non.negl}(n)$$

Theorem 9. *If a function family $\mathcal{H}$ guarantees the best possible entropy preservation, i.e. for all p.p.t. adversary A:*

$$H(h_k(A(k))|A(k)) > m(n) - O(\log(n))$$

then it is weakly correlation intractable.

[6] This notion is different from the "weak correlation intractability" in [26]. The "weak correlation intractability" in [26] is redefined as CI-$\mathsf{P/poly}$ in this article, cf. definition 10.

Proof. If $\mathcal{H}$ is not weakly correlation intractable, which means there is a sparse relation R, an adversary A that:

$$\Pr_k[x \leftarrow A(k) : (x, h_k(x)) \in R] = 1$$

Since R is sparse, which means for all x, the possible y values form a negligibly small subset of the range. Therefore the conditional entropy is:

$$H(h_k(A(k))|A(k)) < m(n) - \omega(\log(n))$$

which forms a contradiction.

A.2 Separations between correlation intractability and other notions

Several random-oracle-like notions are defined in an "indistinguishability" fashion. These definitions attempt to capture the intuition that, given only limited access to or partial information from the function, it is hard for the adversary to distinguish whether the information is obtained from the hash function or a truly random function. The notions defined in this way include correlation robustness[7] [43], seed-incompressibility[8] [41], correlated input security (CIH) [38], and universal computational extractor (UCE) [8].

These notions are quite different from correlation intractability. In the next few paragraphs, we demonstrate the difference by showing that a simple version of correlated-input hash function (defined by [38], rephrased by [8] as a subclass of UCE and by [22] as q-CIH) is separated from correlation intractability. We emphasize that the purpose of showing separations is to demonstrate the properties of these definitions on their own, rather than showing incompatibility. In fact, there is evidence that these notions are compatible with correlation intractability: the same construction that we show to be correlation intractable (iO of puncturable PRFs with appropriate padding) was shown to satisfy a subclass of UCE by Brzuska and Mittelbach [22].

Definition 14. *(q-CIH [8, 22, 38]) Let q be a polynomial. For a hash function family $\mathcal{H} = \{h_k : \{0,1\}^{l(n)} \rightarrow \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$, consider the following game between the p.p.t. adversary $A = (A_1, A_2)$ and the challenger:*

1. *The challenger samples a hash function from the family $h_k \overset{\$}{\leftarrow} \mathcal{H}_n$.*
2. *A_1 samples $q(n)$ (possibly correlated) inputs x_i, $i \in [q(n)]$.*

[7] Correlation robustness is defined for keyless hash functions, unlike the other notions in this article.

[8] [41] discussed both indistinguishability-style and correlation intractability-style definitions, when the adversary is only given partial information of the key (e.g. with an a priori bound on the length).

3. *The challenger tosses a coin b. If $b = 0$, then let $y_i = h_k(x_i)$, $i \in [q(n)]$; if $b = 1$, then let $y_i \xleftarrow{\$} \{0,1\}^{m(n)}$, $i \in [q(n)]$.*
4. *A_2 gets h_k, y_i, $i \in [q(n)]$, outputs $b' \in \{0,1\}$, and wins if $b' = b$.*

$\mathcal{H}$ is called q-CIH if any p.p.t. adversary $A = (A_1, A_2)$ wins with probability less than $1/2 + \mathsf{negl}(n)$.

Theorem 10. *If q-CIH exists, then there is a function ensemble that is q-CIH but not correlation intractable. If correlation intractable function ensemble exists, then there is a function ensemble that is correlation intractable but not q-CIH.*

Proof. The constructions that demonstrate the separation of CIH and correlation intractability are very similar to the ones in [8], section 4.4, where they are used to separate UCE from other notions including collision resistance.

Consider the following constructions:

Construction 11 *Let $\mathcal{H} = \{h_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)}\}_{n\in\mathbb{N}}$ be q-CIH. We construct $\mathcal{H}'$ by adding a uniformly random string $u \in \{0,1\}^{l(n)}$ as the prefix of the key, and define $h'_{k'} = h'_{u||k}$ as:*

$$h'_{u||k}(x) = \begin{cases} \text{if } x = u, \text{ return } 0^{m(n)}\,; \\ \text{else, } \qquad \text{return } h_k(x)\,. \end{cases}$$

Lemma 4. *$\mathcal{H}'$ is q-CIH but not correlation intractable.*

Proof. To break correlation intractability, the adversary outputs u which is a preimage of $0^{m(n)}$.

To show $\mathcal{H}'$ is q-CIH, assume by contradiction that there is an adversary $A' = (A'_1, A'_2)$ that wins the q-CIH game with probability $1/2 + \eta(n)$ where η is non-negligible. We use the exact same adversary to break the q-CIH of $\mathcal{H}$: note that with probability $(1 - 2^{-l(n)})^{q(n)}$, A'_1 won't sample an input that equals to u, beyond which the view of A'_2 will be exactly the same for $\mathcal{H}$ and $\mathcal{H}'$. Therefore, A' wins the q-CIH game for $\mathcal{H}$ with probability no less than

$$(1 - 2^{-l(n)})^{q(n)} \cdot (1/2 + \eta(n)) \geq 1/2 + \eta(n) - q(n) \cdot 2^{-l(n)}$$

where $\eta(n) - q(n) \cdot 2^{-l(n)}$ is non-negligible, thus forming a contradiction.

Construction 12 *Let $\mathcal{H} = \{h_k : \{0,1\}^{l(n)} \to \{0,1\}^{m(n)-1}, k = g(s), s \in \{0,1\}^{\sigma(n)}\}_{n\in\mathbb{N}}$ be a correlation intractable function ensemble, we construct $\mathcal{H}'$ by padding an 1-bit at the end of the output:*

$$h'_{k'}(x) = h_k(x)||1$$

Lemma 5. *$\mathcal{H}'$ is correlation intractable but not q-CIH.*

Proof. To break q-CIH, the adversary outputs 0 if all the y_i, $i \in [q(n)]$ end with 1; otherwise, the adversary outputs 1.

To show $\mathcal{H}'$ is correlation intractable, assume by contradiction that there is an attacker A', a sparse relation $R' : \{0,1\}^{l(n)+m(n)} \to \{0,1\}$, and a non-negligible function $\eta(\cdot)$ such that

$$\Pr_{k'}[x \leftarrow A'(k') : R'(x, h'_{k'}(x)) = 1] > \eta(n)$$

Then we build an adversary A and a sparse relation $R : \{0,1\}^{l(n)+m(n)-1} \to \{0,1\}$ against $\mathcal{H}$: the relation R is defined as

$$R = \{(x,y) \mid R'(x, y||1) = 1,\ x \in \{0,1\}^{l(n)},\ y \in \{0,1\}^{m(n)-1}\}$$

The density of R is at most twice as much as the density of R', so it is sparse. Given the key k, A constructs $h'_{k'}$ by padding a bit '1' at the end of the output of h_k, then sends $h'_{k'}$ to A' and outputs the answer of A'. The probability that A breaks R is exactly the probability that A' breaks R', which contradicts the assumption that $\mathcal{H}$ is correlation intractable.

Note that this transformation works regardless of the efficiency of checking the relation.

The proof completes by combining construction 11, 12 and lemma 4, 5.