Amplification of Non-Interactive Zero Knowledge, Revisited

Nir Bitansky* Nathan Geier†

Abstract

In an $(\varepsilon_s, \varepsilon_z)$-weak non-interactive zero knowledge (NIZK), the soundness error is at most ε_s and the zero-knowledge error is at most ε_z. Goyal, Jain, and Sahai (CRYPTO 2019) stated that if $\varepsilon_s + \varepsilon_z < 1$ for some constants $\varepsilon_s, \varepsilon_z$, then $(\varepsilon_s, \varepsilon_z)$-weak NIZK can be turned into fully-secure NIZK, assuming *sub-exponentially-secure public-key encryption*. However, they have since discovered a gap in their proof.

We revisit the problem of NIZK amplification:

- We amplify NIZK arguments assuming only *polynomially-secure* public-key encryption, for any constants $\varepsilon_s + \varepsilon_z < 1$.
- We amplify NIZK proofs assuming only *one-way functions*, for any constants $\varepsilon_s + \varepsilon_z < 1$.
- When the soundness error ε_s is negligible to begin with, we can also amplify NIZK arguments assuming only one-way functions.

Our results take a different route than that of Goyal, Jain, and Sahai. They are based on the *hidden-bits paradigm*, and can be viewed as a reduction from NIZK amplification to the better understood problem of pseudorandomness amplification.

*New York University and Tel Aviv University. nbitansky@gmail.com.

†Tel Aviv University. nathangeier.cs@gmail.com

Contents

1 Introduction **3**
- 1.1 Our Results . . . 3
- 1.2 Technical Overview . . . 4

2 Preliminaries **8**
- 2.1 Non-Interactive Zero Knowledge and the Hidden-Bits Model . . . 9
- 2.2 Useful Lemmas . . . 12

3 The Amplifiers **13**
- 3.1 AND/OR Easy-Subset Lemmas . . . 13
- 3.2 Zero-Knowledge Amplification . . . 15
- 3.3 Soundness* Amplification . . . 23

4 Simultaneous Amplification **26**
- 4.1 Soundness to Soundness* via Public-Key Encryption . . . 26

5 Open Questions **28**

Acknowledgements **29**

References **29**

A Appendix **31**
- A.1 Hybrid Indistinguishability Lemma . . . 31
- A.2 Useful Facts . . . 33

1 Introduction

Security amplification is a foundational problem in cryptography: given a cryptographic primitive that is *weakly secure* in some sense, we would like to make it *strongly secure*. One central motivation for studying such amplification is that weak primitives are often easier to construct. Another motivation comes from *cryptographic combiners* [HKN+05, Her05] where amplification may allow to combine several constructions, only some of which are secure, into one construction that is guaranteed to be secure. Intuitively, by picking one of the constructions at random we obtain a weakly secure construction, which we can then amplify.

Given its basic importance, the problem of security amplification has been extensively studied. Perhaps the most famous examples are Yao's amplification of weak one-way functions and Yao's XOR lemma for amplifying unpredictability [Yao82]. Other examples include public-key encryption [DNR04, LT13] and key-agreement [Hol05], oblivious transfer (OT) [CK90, DKS99, Wul07], commitments [DKS99, HR08], cryptographic arguments [Hai09], and many more.

Amplifying Non-Interactive Zero Knowledge. We consider amplification of *non-interactive zero knowledge* (NIZK) [BFM88], a central cryptographic object. NIZKs allow the prover to generate a proof of validity of an NP statement without revealing any information about the witness, in a model where both the prover and verifier have access to a trusted *common reference string* (CRS). This is captured by the existence of an efficient simulator that can generate from the statement alone a CRS and proof that are indistinguishable from real ones.

An $(\varepsilon_s, \varepsilon_z)$-weak NIZK is such that a cheating prover can cause the verifier to accept a false statement, adaptively chosen based on the CRS, with probability at most ε_s, and the simulated proof and CRS are distinguishable from real ones with advantage at most ε_z. Here the focus is on the non-trivial case where $\varepsilon_s + \varepsilon_z < 1$.[1] A challenge in the context of NIZK is coming up with *simultaneous amplifiers*, which eventually reduce both the soundness and ZK errors to negligible. The difficulty in simultaneous amplification, also expressed in primitives such as *oblivious transfer* [DKS99, Wul07], is that amplifying one property would typically degrade the other property.[2]

The problem of simultaneous amplification for NIZKs was studied by Goyal, Jain and Sahai [GJS19]. Their main statement is that for any constants ε_s and ε_z such that $\varepsilon_s + \varepsilon_z < 1$, $(\varepsilon_s, \varepsilon_z)$-weak NIZK can be amplified to full-fledged NIZK assuming *sub-exponentially-secure public-key encryption* (PKE). However, later on they discovered a gap in their proof [Jai24].

1.1 Our Results

We revisit the problem of simultaneous amplification of NIZKs, and present new and improved simultaneous amplification results, as detailed next.

In the setting of arguments, we rely on polynomially (rather than sub-exponentially) secure PKE.

[1] Note that when $\varepsilon_s + \varepsilon_z = 1$, we can include in the CRS an ε_s-biased bit which determines whether to use the trivially sound system where the witness is sent in the clear, or the trivially ZK system where the prover sends nothing, and the verifier accepts.

[2] We remark that we focus on security properties and do not deal with weak completeness; namely, we assume a negligible completeness error.

Theorem 1.1 (Informal)**.** *For any constants $\varepsilon_s, \varepsilon_z$ such that $\varepsilon_s + \varepsilon_z < 1$, $(\varepsilon_s, \varepsilon_z)$-weak NIZK arguments for NP can be amplified to fully-secure NIZK arguments for NP, assuming polynomially-secure PKE.*

In the setting of proofs, we further reduce the assumption to one-way functions (OWFs).

Theorem 1.2 (Informal)**.** *For any constants $\varepsilon_s, \varepsilon_z$ such that $\varepsilon_s + \varepsilon_z < 1$, $(\varepsilon_s, \varepsilon_z)$-weak NIZK proofs for NP can be amplified to fully-secure NIZK proofs for NP, assuming OWFs.*

In the case that the soundness error ε_s is negligible to begin with (which arises for instance in the construction of NIZK from batch arguments [BKP+24]), we can rely only on OWFs (rather than PKE) also in the case of arguments, and even for $\varepsilon_z = 1 - o(1)$.

Theorem 1.3 (Informal)**.** *For any constant $\delta < 1$, $(n^{-\omega(1)}, 1 - n^{-\delta})$-weak NIZK arguments for NP can be amplified to fully-secure NIZK arguments for NP, assuming OWFs.*

In the last two theorems (where PKE is not assumed), we also preserve a random CRS if the original weak NIZK has a random CRS, whereas in the first theorem we collapse to the so called structured common reference string model.

A Reduction to Pseudorandomness and Soundness*. The main component behind our results is a zero-knowledge amplifier that diverges from the approach of [GJS19], which was based on the MPC-in-the-head paradigm [IKOS07]. The amplifier is based on the hidden-bits paradigm, and specifically on amplifying the pseudorandomness of hidden-bits generators.

For soundness amplification, we use basic parallel repetition as in [GJS19]. We introduce a natural soundness notion that we call soundness*, for which amplification can be proven without any additional computational assumptions. While in certain cases (such as proofs) this notion is equivalent to plain soundness, in the case of general arguments, to upgrade plain (weak) soundness to soundness*, we rely on PKE.

To achieve simultaneous amplification we carefully combine the two, using security degradation theorems from the literature (and generalization thereof). We next elaborate on our techniques.

1.2 Technical Overview

We first focus on the case that the zero-knowledge error is large, say $\varepsilon_z \approx 0.99$, but the soundness error is negligible. This will already convey the main technical components behind our amplifiers. We will then discuss the general case of simultaneous amplification where the zero-knowledge and soundness errors are some constants $\varepsilon_s, \varepsilon_z$ such that $\varepsilon_s + \varepsilon_z < 1$.

Hidden-Bits Amplification. The zero-knowledge amplifier we construct is based on the hidden-bits model [FLS99]. Recall that in the hidden-bits model, a trusted party generates a random string and shares it with the prover, who then picks a subset of bits to reveal to the verifier. Feige, Lapidot and Shamir [FLS99] show how to construct a NIZK proof in the hidden-bits model (HBM) without any computational assumptions. Thus, in order to construct NIZK, it suffices to implement a so called *hidden-bits generator* (HBG) [QRW19, KMY20].

Roughly speaking, an HBG is a pseudorandom generator $G : \{0,1\}^n \to \{0,1\}^t$, accompanied with the ability to prove, for any $I \subseteq [t]$, membership in the image of G_I (the restriction of G to the bits in I). Furthermore, a proof π_I for $G_I(s)$, does not compromise the pseudorandomness of the

rest of the bits $G_{\overline{I}}(s)$. Combining an HBG and an HBM proof system with (statistical) soundness 2^{-2n} and $t(n) = \mathsf{poly}(n)$ hidden bits, yields a NIZK with a common random string (CRS) $r \in \{0,1\}^t$ as follows. The hidden bit string is set to be $r^* := G(s) \oplus r$ for a random seed s, and the prover reveals r^*_I, by revealing $y = G_I(s)$ along with a corresponding proof π_I that y is indeed in the image of G_I. The fact that $G(s)$ can be described by a short seed $s \in \{0,1\}^n$, implies that the HBM soundness error will only increase from $\approx 2^{-2n}$ to $\approx 2^{-n}$.

Clearly, if we had a NIZK to begin with, then an HBG could be easily constructed from any pseudorandom generator G, as the required proofs π_I can be realized using the NIZK. Indeed, the fact that the proof π_I is zero knowledge implies that no information is leaked on the seed s, and accordingly the non-revealed bits $G_{\overline{I}}(s)$ remain pseudorandom. The soundness of proofs follows from the adaptive soundness of the underlying NIZK.[3] If our NIZK is only weakly ZK, however, then we obtain an HBG with weak pseudorandomness. This suggests a reduction to pseudorandomness amplification, which has been thoroughly studied starting from the work of Yao [Yao82].

Perhaps the most natural first attempt is to apply Yao's XOR Lemma [Yao82, GNW95]. Namely, rather than using a single pseudorandom string $G(s)$, consider the XOR over several independent seeds $G(s_1) \oplus \cdots \oplus G(s_k)$ with independent proofs $\pi_I(s_1), \ldots, \pi_I(s_k)$ for the revealed $G_I(s_1), \ldots, G_I(s_k)$. If we start with an HBM proof with sufficiently small soundness error $\ll 2^{-nk}$, we would obtain a sound NIZK as before. In terms of ZK, if each proof is ε_z-ZK, then each $G_{\overline{I}}(s_i)$ is $2\varepsilon_z$-indistinguishable from random (as the hybrid argument goes through the simulator twice), which in turn implies that each bit is $4\varepsilon_z$-unpredictable. Accordingly, the XOR Lemma implies that $G_{\overline{I}}(s_1) \oplus \cdots \oplus G_{\overline{I}}(s_k)$ is $\approx t \cdot (4\varepsilon_z)^k$-pseudorandom. In fact, a more careful analysis shows that the error can be reduced to $t \cdot (2\varepsilon_z)^k$. But still, this approach is limited to $\varepsilon_z < 1/2$.

Tighter Amplification via Extraction. To be able to deal with $\varepsilon_z \approx 0.99$, we turn to a different approach taken by Maurer and Tessaro in the context of weak PRG amplification [MT10]. They rely on the concatenate and extract approach: Output $\mathsf{Ext}(G(s_1), \ldots, G(s_k); r), r$ for a strong randomness extractor Ext (where r denotes the seed). They show that this approach could amplify ε-weak PRGs for an arbitrary $\varepsilon < 1 - 1/\mathsf{poly}(n)$. However, the above construction is not applicable in our setting as is. Specifically, the output $\mathsf{Ext}_I(G(s_1), \ldots, G(s_k); r)$ may depend on all underlying inputs bits $G(s_1), \ldots, G(s_k)$, rather than just the restricted $G_I(s_1), \ldots, G_I(s_k)$ in the case of the XOR construction.

Instead, we use each s_i as the seed of an n-bit output PRF, to generate t blocks $F_{s_i}(1), \ldots, F_{s_i}(t)$.[4] Then, we apply the extractor to each block separately. That is, each bit $j \in [t]$ of the amplified HBG is taken to be $\mathsf{Ext}(F_{s_1}(j), \ldots, F_{s_k}(j); r)$. This way, to reveal a set I, we exhibit $F_{s_1}(I), \ldots, F_{s_k}(I)$, along with independent proofs $\pi_I(s_1), \ldots, \pi_I(s_k)$. We show that for each of the remaining indices $j \notin I$, the string $(F_{s_1}(j), \ldots, F_{s_k}(j))$ has certain computational entropy, which is extracted by Ext.

In more detail, to show that the bits in $\overline{I}$ remain hidden, we build on the analysis of [MT10]. We rely on their *indistinguishability hard-core lemma* (Lemma 2.14) to derive a *hybrid indistinguishability lemma*, which intuitively says that if X is weakly-computationally-indistinguishable from Y, then it is strongly-computationally-indistinguishable from a hybrid distribution X', that is weakly-statistically-indistinguishable from Y.

[3] Note that a malicious prover has the power to choose the set I and values $G_I(s)$ adaptively based on the CRS; hence, adaptive soundness is required.

[4] The transition to PRFs is for ease of notation, instead of addressing blocks of output of G.

Lemma 1.4 (Informal, see formal Lemma 2.15 in the body)**.** *If X is ε-indistinguishable from Y for some $\varepsilon < 1 - 1/\mathsf{poly}(n)$, then for every $\varepsilon' = 1/\mathsf{poly}(n)$, there exists X', such that:*

- *X is ε'-indistinguishable from X',*
- *X' is ε-**statistically**-indistinguishable from Y.*

Thus, we can use the fact that $F_{s_i}(\overline{I})$ is ε_z-pseudorandom for all $i \in [k]$, and reduce our analysis to the information-theoretic setting where $F'_{s_i}(\overline{I})$ is ε_z-statistically-close to random. To be more accurate, we are not guaranteed that $F_{s_i}(\overline{I})$ is ε_z-pseudorandom given the real transcript, but we are only guaranteed that

$$\mathrm{Real}_i, F_{s_i}(\overline{I}) \overset{\mathrm{c}}{\approx}_{\varepsilon_z} \mathrm{Sim}_i, F_{s_i}(\overline{I}) \overset{\mathrm{c}}{\approx} \mathrm{Sim}_i, U^{n\cdot|\overline{I}|}\,,$$

where Real_i denotes the revealed bits $F_{s_i}(I)$ along with their proofs $\pi_I(s_i)$, and Sim_i denotes the distribution where the proofs are generated by the ZK simulator from $F_{s_i}(I)$ with no direct leakage on the witness s_i.

Now, invoking the above hybrid indistinguishability lemma, we deduce:

$$\mathrm{Real}_i, F_{s_i}(\overline{I}) \overset{\mathrm{c}}{\approx}_{\varepsilon'} \mathrm{Real}'_i, F'_{s_i}(\overline{I}) \overset{\mathrm{s}}{\approx}_{\varepsilon_z} \mathrm{Sim}_i, U^{n\cdot|\overline{I}|}\,.$$

Given the above statistical guarantee, we prove by statistical coupling that except with probability ε_z^k, for every $j \in \overline{I}$, the (average) min-entropy of $F'_{s_1}(j)\ldots F'_{s_k}(j)$, given $\left\{\mathrm{Real}'_i, F'_{s_i}(\overline{I}\setminus j)\right\}_i$, is at least $n - \log(1/(1-\varepsilon_z))$, which suffices for extraction.[5] Finally, the construction also works when using k samples from $\mathrm{Real}_i, F_{s_i}(\overline{I})$ instead of $\mathrm{Real}'_i, F'_{s_i}(\overline{I})$, as they are computationally indistinguishable.

Simultaneous Amplification and Soundness*. So far, we have assumed that the soundness error ε_s is negligible. Considering the general case where ε_s could be any constant such that $\varepsilon_s + \varepsilon_z < 1$, we need to take into account how the soundness of the above transformation degrades.

Going back to the described amplifier, a naive bound shows that using k weak proofs, soundness would degrade from ε_s to at most $k \cdot \varepsilon_s$. While this is good enough when ε_s is negligible, in the general case, we aim to prove a tighter bound of $1 - (1-\varepsilon_s)^k \leq k \cdot \varepsilon_s$.

To do this we prove a simple *OR easy-subset lemma* (see Lemma 3.2), which roughly states that if an adversary breaks at least one out of k independent challenges with probability $1-(1-\varepsilon)^k$, then for some coordinate $i \in [k]$, at least ε-fraction of challenges are broken with noticeable probability over the choice of the other challenges.[6] This suggests a standard reduction to breaking the ε_s-soundness of a single proof. Given a challenge CRS, we generate polynomially many candidates by iteratively embedding the CRS in every coordinate $i \in [k]$ and running the adversary polynomially many times where the other coordinates are sampled at random.

[5] Formally, in the body we use a two-universal hash function as the extractor and the generalized left-over hash lemma [DORS08].

[6] This lemma can be viewed as a parallel version of an *AND easy-subset lemma*, commonly used in hardness amplification (for instance, in Yao's product amplification of OWFs). We also state this lemma and use it below.

Soundness*. The reduction just described does not quite work as is. The problem is that this reduction given a CRS generates (with the required probability ε_s) a list of statements and proofs, only some of which are guaranteed to be actually false. So in order to break plain soundness, we would have to be able to recognize false statements, which cannot necessarily be done efficiently. Instead, we work with a more robust notion of weak soundness that we term soundness*. In ε_s-soundness*, except with probability ε_s over the CRS, the adversary cannot even generate a list of statements and accepting proofs, such that at least one of the statements is false. We later observe that plain ε_s-soundness can always be upgraded to ε_s-soundness* assuming PKE, and in certain cases arises naturally on its own, or follows automatically from plain soundness (see further discussion below).

Soundness* Amplification. Having deduced that the zero-knowledge amplifier degrades ε_s-soundness* to $1-(1-\varepsilon_s)^k$-soundness*, and assuming k is not too large, we then aim to amplify soundness*. We consider the standard parallel repetition amplifier. Here the proof is analogous to Yao's product amplification [Yao82], and again relies on the robustness of the soundness* notion. Specifically, we rely on an *AND easy-subset lemma* (Lemma 3.1), stating that if the adversary breaks all k independent challenges with probability ε^k, then for some coordinate $i \in [k]$, at least ε-fraction of challenges are broken with noticeable probability over the choice of the other challenges. We then use a similar embedding reduction to the one described above to breaking soundness* of a single instance.

Now, we need to take into account how this parallel repetition degrades ZK. We build on the analysis of [HR08, Gei22], and apply an indistinguishability degradation bound for product distributions (Lemma 2.17), which roughly states that if $X \overset{c}{\approx}_{\varepsilon} Y$, then $X^k \overset{c}{\approx}_{1-(1-\varepsilon)^k} Y^k$.

Combining the Amplifiers. Overall, we obtain a pair of ZK and soundness* amplifiers where each amplifies one parameter $\varepsilon \in \{\varepsilon_s, \varepsilon_z\}$ by $\varepsilon \to \varepsilon^k$, but degrades the other by $\overline{\varepsilon} \to 1-(1-\overline{\varepsilon})^k$. Hence, they must be carefully combined in order to amplify both ZK and soundness* simultaneously. It is important to note here that to maintain efficiency, we can only interchangeably apply the two amplifiers a constant number of times, as the ZK amplifier may polynomially increase the size of the instance. When $\varepsilon_s, \varepsilon_z$ are constants, three applications are enough.

Let us demonstrate here concretely for

$$(\varepsilon_s, \varepsilon_z) = (0.8, 0.1) = (1 - 2^{-\log 5}, 2^{-\log 10}) .$$

Apply the ZK amplifier with $k = \log n$, then we get $(1-n^{-\log 5}, n^{-\log 10})$. Now apply the soundness* amplifier with $k = n^3$, noting that $1 - x \le 2^{-x}$ and $1-(1-x)^k \le k \cdot x$, to obtain

$$\left(\left(1 - n^{-\log 5}\right)^k, 1 - \left(1 - n^{-\log 10}\right)^k\right) \le \left(2^{-k \cdot n^{-\log 5}}, k \cdot n^{-\log 10}\right) \le (2^{-\sqrt{n}}, n^{-0.3}) .$$

Finally, apply the ZK amplifier again with $k = \omega(1)$ to obtain $(\mathrm{negl}(n), \mathrm{negl}(n))$.[7]

[7] Here we essentially used the fact that $-\log(1-0.8) < 3 < -\log(0.1)$, which generalizes for any $\varepsilon_s + \varepsilon_z < 1$, replacing 3 with $c \in (-\log(1-\varepsilon_s), -\log(\varepsilon_z))$.

More on Soundness*. We note that the notion of ε_s-soundness* is equivalent to plain soundness when ε_s is negligible; in this case, the reduction can simply guess which statement is false and gain noticeable advantage. Furthermore, ε_s-soundness* is equivalent to plain soundness in the case of proofs, since an inefficient prover can test on its own whether any given statement is false.

Beyond the above, soundness* may arise naturally in certain amplification scenarios, such as constructing combiners. For instance, consider the case of 2-*out-of*-3 *combiners*, where we have three NIZK candidates, two of which are fully secure, and one which is completely insecure. Then choosing one of them at random yields a $(1/3, 1/3)$-NIZK that in fact has $1/3$-soundness* (rather than just $1/3$ plain soundness). Indeed, provided that we have chosen one of the two valid constructions, we are guaranteed a negligible soundness error.

Finally, we observe that one can always upgrade plain soundness to soundness* using PKE, by guaranteeing witness extraction [DP92]. Specifically, the prover uses a designated public key in the CRS to encrypt its witness and then uses the NIZK to prove that the encryption is to a valid witness. In this case the reduction can efficiently test which statement is false by decrypting the witness using the corresponding secret key. This transformation is also used in [GJS19] to directly amplify plain soundness.

2 Preliminaries

For $n \in \mathbb{N}$, we denote by $[n]$ the set $\{1, \ldots, n\}$. For a function $f : [n] \to \mathcal{X}$ and a subset $I \subseteq [n]$, we denote by $f(I)$ the concatenation of $f(i)$ over $i \in I$. For a distribution X over a set Ω, we use $x \leftarrow X$ to denote the result of sampling according to X, and $x \leftarrow \Omega$ to denote a uniformly random sample from the set. For a set of events $\{A_i\}_{i\in[k]}$ and $I \subseteq [k]$, we denote by $\left(A_I, \overline{A}_{\overline{I}}\right)$ the event where A_i occurred for every $i \in I$ and did not occur otherwise. Bernoulli's inequality states that $(1+x)^k \geq 1 + kx$ for $k \in \mathbb{N}$ and $x \geq -1$, which implies $(\alpha - \beta)^k \geq \alpha^k - k\beta$, for any $1 \geq \alpha \geq \beta \geq 0$. For random variables X and Y, we denote by $\widetilde{H}_\infty(X \mid Y)$ the average min-entropy of X given Y [DORS08]:

$$\widetilde{H}_\infty(X \mid Y) = -\log\left(\mathop{\mathbb{E}}_{y\leftarrow Y}\left[\max_x \Pr\left[X = x \mid Y = y\right]\right]\right),$$

and for an event A, we denote by $\widetilde{H}_\infty(X \mid Y, A) := \widetilde{H}_\infty(X_A \mid Y_A)$ the average min-entropy of X given Y under the conditional distribution of A.

We rely on standard computational concepts and notation:

- We say that a function $f : \mathbb{N} \to \mathbb{R}$ is *negligible* if for all constants $c > 0$, there exists $N \in \mathbb{N}$ such that for all $n > N$, $f(n) < n^{-c}$. We sometimes denote negligible functions by negl. We say that a function $f : \mathbb{N} \to \mathbb{R}$ is *overwhelming* if $1 - f$ is negligible. We say that a function $f : \mathbb{N} \to \mathbb{R}$ is *noticeable* if for some constant $c > 0$ and $N \in \mathbb{N}$, for all $n > N$, $f(n) > n^{-c}$.

- A PPT algorithm is a probabilistic polynomial-time algorithm. A family of circuits $\mathsf{A} = \{\mathsf{A}_n\}_{n\in\mathbb{N}}$ is $s(n)$-*sized* if $|\mathsf{A}_n| \leq s(n)$. It is *polynomial-sized* if $s(n) \leq \mathsf{poly}(n)$. We follow the common practice of modeling any efficient adversary as a family of polynomial-size circuits $\mathsf{A} = \{\mathsf{A}_n\}_{n\in\mathbb{N}}$. We also say that such an A runs in *non-uniform polynomial time*.

- We denote *statistical distance* by SD. For two random variables X, Y and $\varepsilon \in [0, 1]$, we write $X \overset{\text{s}}{\approx}_\varepsilon Y$ to denote the fact that $\mathsf{SD}(X, Y) \leq \varepsilon$ and say that X is ε-*statistically indistinguishable* from Y. For two ensembles $\mathcal{X} = \{X_n\}_{n\in\mathbb{N}}$ and $\mathcal{Y} = \{Y_n\}_{n\in\mathbb{N}}$ and function ε, we write $\mathcal{X} \overset{\text{s}}{\approx}_\varepsilon \mathcal{Y}$

if for all large enough n, $X_n \overset{s}{\approx}_{\varepsilon(n)} Y_n$. For $\mathcal{X}$, $\mathcal{Y}$ and ε as above, we write that $\mathcal{X} \overset{c}{\approx}_{\varepsilon} \mathcal{Y}$ against $s(n)$-sized circuits if the *computational distance* against $s(n)$-sized circuits is at most ε. That is, for every $s(n)$-sized family of distinguishers $\mathsf{D} = \{\mathsf{D}_n\}_{n\in\mathbb{N}}$ and large enough $n \in \mathbb{N}$:

$$\left| \Pr_{x\leftarrow X_n} [\mathsf{D}_n(x) = 1] - \Pr_{y\leftarrow Y_n} [\mathsf{D}_n(y) = 1] \right| \leq \varepsilon(n) \,.$$

We omit the bound on the size when referring to a polynomial-sized family of distinguishers. In addition, we may drop ε from the subscript when it is a negligible function.

Pseudorandom Functions. A PPT seeded function $F_s : \{0,1\}^* \to \{0,1\}^{m(n)}$ with $s \in \{0,1\}^n$, is an m-bit-output PRF, if for every oracle-aided polynomial-sized circuit family of distinguishers $\mathsf{D} = \{\mathsf{D}_n\}_{n\in\mathbb{N}}$:

$$\left| \Pr_{s\leftarrow\{0,1\}^n} \left[\mathsf{D}_n^{F_s} = 1\right] - \Pr_{R_m} \left[\mathsf{D}_n^{R_m} = 1\right] \right| \leq \mathrm{negl}(n) \,,$$

where $R_m(\cdot)$ is a random oracle to uniform m-bit strings.

Public-Key Encryption. A PKE scheme is a PPT triplet $(\mathsf{Gen}, \mathsf{Enc}, \mathsf{Dec})$ such that:

- Correctness: With overwhelming probability over $(pk, sk) \leftarrow \mathsf{Gen}(1^n)$, for all $m \in \{0,1\}^*$ we have
$$\Pr\left[\mathsf{Dec}_{sk}(\mathsf{Enc}_{pk}(m)) = m\right] = 1 \,,$$
where the probability is taken over the random coins of the encryption algorithm.
- Security: For every ensemble of polynomial-size messages $m = \{m_n\}_{n\in\mathbb{N}}$:
$$pk, \mathsf{Enc}_{sk}(m_n) \overset{c}{\approx} pk, \mathsf{Enc}_{sk}(0^{|m_n|}) \,,$$
where $(pk, sk) \leftarrow \mathsf{Gen}(1^n)$.

Remark 2.1. We assume w.l.o.g. almost-all-keys perfect-correctness, as public-key encryption schemes can be *immunized*, see [DNR04].

We note that the existence of OWFs is equivalent to that of PRFs, and we will be assuming OWFs whenever we need a PRF. Further, note that PKE implies OWFs, which is why we do not need to additionally assume OWFs when already assuming PKE. We also remark that the existence of NIZK for a hard-on-average language implies the existence of non-uniform OWFs [OW93, Ps05].

2.1 Non-Interactive Zero Knowledge and the Hidden-Bits Model

In this paper, we only consider *doubly-efficient* non-interactive protocols in the *common reference string* (CRS) model. Below we define this primitive and the different notions of soundness and privacy that we need. We always refer to n as the size of the statement $|x| = n$.

Definition 2.2 (Non-Interactive Protocol in the CRS Model)**.** *A* non-interactive protocol (NIP) Π in the CRS model *for an* NP *relation* $\mathcal{R}$ *is a triplet of PPT algorithms* $(\mathsf{Gen}, \mathsf{P}, \mathsf{V})$*, with the following syntax:*

- $crs \leftarrow \mathsf{Gen}(1^n)$*: Given the instance size* n*, the randomized* set-up *algorithm* Gen *outputs a CRS* crs*.*
- $\pi \leftarrow \mathsf{P}(crs, x, w)$*: Given CRS* crs*, instance* x *and witness* w*, the randomized* prover *outputs a proof* π*.*
- $b := \mathsf{V}(crs, x, \pi)$*: Given CRS* crs*, instance* x*, and proof* π*, the deterministic* verifier *returns a bit* b *representing accept or reject.*

In the following definitions, Π is a non-interactive protocol in the CRS model for an NP relation $\mathcal{R}$.

Definition 2.3 (Completeness)**.** *NIP* Π *satisfies* completeness *if for every ensemble* $\{(x, w) \in \mathcal{R}_n\}_{n\in\mathbb{N}}$*, the following probability is overwhelming:*

$$\Pr_{\substack{crs\leftarrow\mathsf{Gen}(1^n)\\ \pi\leftarrow\mathsf{P}(crs,x,w)}}[\mathsf{V}(crs, x, \pi) = 1] \,.$$

Definition 2.4 (Soundness (Adaptive))**.** *NIP* Π *is* ε_s-computationally-sound *if, for every polynomial-sized circuit family of malicious provers* $\mathsf{P}^* = \{\mathsf{P}^*_n\}_{n\in\mathbb{N}}$*, we have that*

$$\Pr_{\substack{crs\leftarrow\mathsf{Gen}(1^n)\\ (x,\pi)\leftarrow\mathsf{P}^*_n(crs)}}[|x| = n \wedge x \notin \mathcal{L}(\mathcal{R}_n) \wedge \mathsf{V}(crs, x, \pi) = 1] \leq \varepsilon_s(n) \,.$$

If the above holds also for unbounded circuit families, we say that Π *is* ε_s-statistically-sound*. We may omit* ε_s *when it is negligible.*

Definition 2.5 (Soundness*)**.** *NIP* Π *is* ε_s-computationally-sound* *if, for every polynomial-sized circuit family of malicious provers* $\mathsf{P}^* = \{\mathsf{P}^*_n\}_{n\in\mathbb{N}}$*, we have that*

$$\Pr_{\substack{crs\leftarrow\mathsf{Gen}(1^n)\\ (x_1,\pi_1),\ldots,(x_t,\pi_t)\leftarrow\mathsf{P}^*_n(crs)}}[\exists i \in [t] : |x_i| = n \wedge x_i \notin \mathcal{L}(\mathcal{R}_n) \wedge \mathsf{V}(crs, x_i, \pi_i) = 1] \leq \varepsilon_s(n) \,.$$

If the above holds also for unbounded circuit families, we say that Π *is* ε_s-statistically-sound**. We may omit* ε_s *when it is negligible.*

Remark 2.6. Note that ε_s-statistical-soundness is equivalent to ε_s-statistical-soundness*, since an unbounded adversary can always test whether $x \notin \mathcal{L}(\mathcal{R}_n)$. In the computational setting, soundness* implies soundness, so it is a stronger notion. However, note that negligible-computational-soundness is equivalent to negligible-computational-soundness*, since we can preserve non-negligible success probability by picking an instance-proof pair at random.

Definition 2.7 (Zero Knowledge)**.** *NIP* Π *is* ε_z-zero knowledge *if there exists a PPT simulator* Sim *such that for all* $\{(x, w) \in \mathcal{R}_n\}_{n\in\mathbb{N}}$*, we have*

$$(crs, \pi) \overset{c}{\approx}_{\varepsilon_z(n)} \mathsf{Sim}(x) \,,$$

where $crs \leftarrow \mathsf{Gen}(1^n)$ *and* $\pi \leftarrow \mathsf{P}(crs, x, w)$*. We may omit* ε_z *when it is negligible.*

Remark 2.8. Throughout the paper, we refer to non-adaptive zero knowledge. There is a general transformation from NIZK with non-adaptive to adaptive ZK [DN00, KMY20] (where both have a negligible adaptive-soundness error). Also, we restrict our attention to single-instance ZK, noting that there is a general transformation that allows to handle multiple instances, assuming OWFs [FLS90].

Definition 2.9 (NIZK)**.** *Let* Π *be a NIP that satisfies completeness. It is an* $(\varepsilon_s, \varepsilon_z)$*-weak NIZK if it satisfies* ε_s*-soundness and* ε_z*-ZK. It is an* $(\varepsilon_s, \varepsilon_z)$*-weak** *NIZK if it satisfies* ε_s*-soundness** *and* ε_z*-ZK. It is a standard NIZK if it satisfies* negl*-soundness and* negl*-ZK. It is an* $(\varepsilon_s, \varepsilon_z)$*-weak statistically-sound NIZK if it satisfies* ε_s*-statistical-soundness and* ε_z*-ZK. It is a statistically-sound NIZK if it satisfies* negl*-statistical-soundness and* negl*-ZK.*

Definition 2.10 (NIZK in the Hidden-Bits Model)**.** *An* HBM NIZK Π^{hbm} *with* $t(n) = \mathsf{poly}(n)$ *hidden bits for an* NP *relation* $\mathcal{R}$*, is a pair of PPT algorithms* $(\mathsf{P}^{\mathsf{hbm}}, \mathsf{V}^{\mathsf{hbm}})$*, with the following syntax:*

- $(I, \pi^{\mathsf{hbm}}) \leftarrow \mathsf{P}^{\mathsf{hbm}}(r, x, w)$*: Given hidden bits* $r \in \{0,1\}^t$*, instance* x *and witness* w*, the randomized* prover *outputs a subset* $I \subseteq [t]$ *and a proof* π^{hbm}*.*
- $b := \mathsf{V}^{\mathsf{hbm}}(I, r_I, x, \pi^{\mathsf{hbm}})$*: Given subset* $I \subseteq [t]$*, string* $r_I \in \{0,1\}^{|I|}$*, instance* x *and a proof* π^{hbm}*, the deterministic* verifier *returns a bit* b *representing accept or reject.*

The following properties should be satisfied by it:

- *Completeness: For all* $\{(x,w) \in \mathcal{R}_n\}_{n\in\mathbb{N}}$*, the following probability is overwhelming:*

$$\Pr_{\substack{r\leftarrow\{0,1\}^t \\ (I,\pi^{\mathsf{hbm}})\leftarrow\mathsf{P}^{\mathsf{hbm}}(r,x,w)}}\left[\mathsf{V}^{\mathsf{hbm}}(I, r_I, x, \pi^{\mathsf{hbm}}) = 1\right].$$

- *Statistical Soundness: For every unbounded circuit family of malicious provers* $\mathsf{P}^* = \{\mathsf{P}^*_n\}_{n\in\mathbb{N}}$*:*

$$\Pr_{\substack{r\leftarrow\{0,1\}^t \\ (x,I,\pi^{\mathsf{hbm}})\leftarrow\mathsf{P}^*_n(r)}}\left[|x| = n \wedge x \notin \mathcal{L}(\mathcal{R}_n) \wedge \mathsf{V}^{\mathsf{hbm}}(I, r_I, x, \pi) = 1\right] \leq \mathrm{negl}(n).$$

- *Zero Knowledge: There exists a PPT simulator* $\mathsf{Sim}^{\mathsf{hbm}}$ *such that for all* $\{(x,w) \in \mathcal{R}_n\}_{n\in\mathbb{N}}$*, we have*

$$(I, r_I, \pi^{\mathsf{hbm}}) \overset{s}{\approx} \mathsf{Sim}^{\mathsf{hbm}}(x),$$

  where $r \leftarrow \{0,1\}^t$ *and* $(I, \pi^{\mathsf{hbm}}) \leftarrow \mathsf{P}^{\mathsf{hbm}}(r, x, w)$*.*

Remark 2.11. We can always amplify soundness to $2^{-k(n)} \cdot \mathrm{negl}(n)$, using $k(n)$ parallel repetitions.

Theorem 2.12 ([FLS99])**.** *There exists HBM NIZK for all* NP*.*

2.2 Useful Lemmas

Lemma 2.13 (Generalized Leftover Hash Lemma [DORS08])**.** *Let X be a random variable over $\mathcal{X}$, and Z be some jointly distributed random variable. Further, let $\mathcal{H} = \left\{h : \mathcal{X} \to \{0,1\}^\ell\right\}$ be a universal hash family for some $\ell \in \mathbb{N}$. Then we have that*

$$Z, h, h(X) \overset{s}{\approx}_\varepsilon Z, h, U^\ell \,,$$

over $h \leftarrow \mathcal{H}$, where $\varepsilon \leq 0.5\sqrt{2^\ell / 2^{\widetilde{H}_\infty(X|Z)}}$.

Lemma 2.14 (Indistinguishability Hard-Core Lemma [MT10], Reinterpreted)**.** *Let X and Y be random variables, and let $\delta, \varepsilon \in (0,1)$ and $s \in \mathbb{N}$ be given. Assume that*

$$X \overset{c}{\approx}_\delta Y \,,$$

against s-sized circuits. Then, there exists a pair of events A and B, with probability $1-\delta$ each, such that

$$X \mid A \overset{c}{\approx}_\varepsilon Y \mid B \,,$$

against s'-sized circuits with $s' := \frac{s \cdot \varepsilon^2}{128(\log|\mathrm{Im}(X)| + \log|\mathrm{Im}(Y)| + 1)}$.

The above indistinguishability hard-core lemma implies the following:

Lemma 2.15 (Hybrid Indistinguishability Lemma)**.** *Let X and Y be random variables, and let $\delta, \varepsilon \in (0,1)$ and $s \in \mathbb{N}$ be given. Assume that*

$$X \overset{c}{\approx}_\delta Y \,,$$

against s-sized circuits. Then there exists a hybrid distribution X' such that

$$X \overset{c}{\approx}_\varepsilon X' \overset{s}{\approx}_\delta Y \,,$$

where the first indistinguishability is computational against s'-sized circuits with $s' := \frac{s \cdot (\varepsilon/(1-\delta))^2}{128(\log|\mathrm{Im}(X)| + \log|\mathrm{Im}(Y)| + 1)}$, and the second is statistical.

Proof. Let $X' = (1-\delta) \cdot (Y \mid B) + \delta \cdot (X \mid \overline{A})$, namely, we sample from $Y \mid B$ w.p. $1-\delta$ and otherwise from $X \mid \overline{A}$. Then the computational distance between X and X' is equal to $(1-\delta)$ times the computational distance between $X \mid A$ and $Y \mid B$, and the statistical distance from Y is at most δ. □

We also give in the appendix (Appendix A.1) a proof sketch based on the original statement of [MT10].

Remark 2.16. There is also a uniform version of the indistinguishability hard-core lemma, where X and Y are efficiently samplable and δ, ε are noticeable and efficiently computable. In this setting, the poly-time distinguishers get oracle access to the conditional distributions $A \mid X$ and $B \mid Y$, for input-independent queries. That is, the oracle-aided distinguisher may pick queries $x_1, \ldots, x_q$ and $y_1, \ldots, y_q$, depending only on its internal randomness and previous queries, and receive answers $\{\Pr[A \mid X = x_i]\}_{i \in [q]}$ and $\{\Pr[B \mid Y = y_i]\}_{i \in [q]}$. So in particular, it can generate samples from

$X \mid A$, $X \mid \overline{A}$, $Y \mid B$, $Y \mid \overline{B}$, by rejection sampling and deciding when to accept based on the oracle.

We can thus also generalize the hybrid indistinguishability lemma to the uniform setting, where we allow distinguishers access to samples from X', as we can use the oracle access to $A \mid X$ and $B \mid Y$ in order to sample from X'.

Lemma 2.17 (Indistinguishability Bound for Product Distributions [HR08, Gei22])**.** *Let X and Y be distributions over n bits such that*

$$X \overset{c}{\approx}_{\delta} Y\,,$$

against s-sized circuits. Then, for every $m \in \mathbb{N}$ and $\varepsilon \in (0,1)$, we have that

$$X^m \overset{c}{\approx}_{1-(1-\delta)^m+\varepsilon} Y^m\,,$$

against s'-sized circuits, with $s' := \frac{(s-1)\cdot((1-\delta)^m-\varepsilon)}{\log(1/\delta)+\log(1/\varepsilon)} - 5m \cdot n - 1$.

Remark 2.18. We note that a uniform version also exists for the indistinguishability bound for product distributions, although the dependency on the slackness $1/\varepsilon$ is then polynomial instead of logarithmic, this is still good enough for us. We also remark that the above bound could be shown using [MT10], which is the approach taken by [GJS19].

3 The Amplifiers

In this section, we present and analyze our pair of amplifiers. We start by presenting two general information-theoretic lemmas that will be used to analyze the soundness degradation and amplification of our amplifiers. Then, we present the zero-knowledge amplifier based on the hidden-bits paradigm, which is the main component behind our results. Finally, we show soundness amplification by parallel repetition of the base protocol. Throughout the section we work with soundness* as the security notion, and in Section 4.1 we derive corresponding corollaries for plain soundness (assuming PKE).

3.1 AND/OR Easy-Subset Lemmas

In this section, we present two general information-theoretic lemmas, that are used in our analysis of soundness degradation and amplification. Roughly speaking, we show that if k independent challenges can all be solved (AND) w.p. noticeably larger than ε^k, or if the probability to solve at least one challenge (OR) is noticeably larger than $1-(1-\varepsilon)^k$, then for some coordinate $i \in k$ we can solve ε-fraction of challenges with noticeable probability, over the choice of other challenges and the internal solver's randomness.

Lemma 3.1 (AND Easy-Subset Lemma, adapted from Yao's OWF amplification [Yao82])**.** *Let $f : \mathcal{X}^k \to \{0,1\}$ be any boolean mapping (possibly randomized), where $\mathcal{X}$ is some finite set and $k \in \mathbb{N}$. Also, let $D : \mathcal{X} \to [0,1]$ be any distribution over $\mathcal{X}$. For any $q \in \mathbb{N}$ and $\varepsilon \in (0,1)$ such that*

$$\Pr_{\forall i \in [k]\colon x_i \leftarrow D}[f(x_1,\ldots,x_k)=1] \geq \varepsilon^k + k/q\,,$$

there exists some i and $G_i \subseteq \mathcal{X}$ with $D(G_i) \geq \varepsilon$, such that for every $x_i \in G_i$:

$$\Pr_{\forall j \in [k]\setminus i\colon x_j \leftarrow D}[f(x_1,\ldots,x_k)=1] \geq 1/q\,.$$

Proof. For every i we let G_i be the set of all x_i such that

$$\Pr_{\forall j\in[k]\setminus i:\, x_j\leftarrow D}[f(x_1,\ldots,x_k)=1]\geq 1/q\,,$$

and assume toward contradiction that $D(G_i)<\varepsilon$ for all i. Since the events $\{x_i\notin G_i\}_i$ and $\forall i: x_i\in G_i$ are covering (not necessarily disjoint) we have that

$$\begin{aligned}&\Pr_{\forall i:\, x_i\leftarrow D}[f(x_1,\ldots,x_k)=1]\leq\\&\sum_{i=1}^{k}\Pr[x_i\notin G_i]\Pr[f(x_1,\ldots,x_k)=1\mid x_i\notin G_i]+\\&\Pr[\forall i: x_i\in G_i]\Pr[f(x_1,\ldots,x_k)=1\mid \forall i: x_i\in G_i]<\\&\varepsilon^k\cdot 1+k\cdot 1\cdot 1/q=\varepsilon^k+k/q\,,\end{aligned}$$

which is a contradiction. □

Lemma 3.2 (OR Easy-Subset Lemma)**.** *Let $f:\mathcal{X}^k\to\{0,\ldots,k\}$ be any mapping (possibly randomized), where $\mathcal{X}$ is some finite set and $k\in\mathbb{N}$. Also, let $D:\mathcal{X}\to[0,1]$ be any distribution over $\mathcal{X}$. For any $q\in\mathbb{N}$ and $\varepsilon\in(0,1)$ such that*

$$\Pr_{\forall i\in[k]:\, x_i\leftarrow D}[f(x_1,\ldots,x_k)=0]<(1-\varepsilon)^k-k/q\,,$$

there exists some i and $G_i\subseteq\mathcal{X}$ with $D(G_i)\geq\varepsilon$, such that for every $x_i\in G_i$:

$$\Pr_{\forall j\in[k]\setminus i:\, x_j\leftarrow D}[f(x_1,\ldots,x_k)=i]\geq 1/q\,.$$

Proof. For every i we let G_i be the set of all x_i such that

$$\Pr_{\forall j\in[k]\setminus i:\, x_j\leftarrow D}[f(x_1,\ldots,x_k)=i]\geq 1/q\,,$$

and assume toward contradiction that $\forall i:\ D(G_i)<\varepsilon$. Note that for every i we have that

$$\begin{aligned}1/q>&\Pr[f(x_1,\ldots,x_k)=i\mid x_i\notin G_i]\geq\\&\Pr[\forall j\in[k]\setminus i: x_j\notin G_j]\Pr[f(x_1,\ldots,x_k)=i\mid\forall j: x_j\notin G_j]\geq\\&(1-\varepsilon)^{k-1}\cdot\Pr[f(x_1,\ldots,x_k)=i\mid\forall j: x_j\notin G_j]\,.\end{aligned}$$

This inequality is used in the third transition below

$$\begin{aligned}&\Pr_{\forall i:\, x_i\leftarrow D}[f(x_1,\ldots,x_k)=0]\geq\Pr[\forall j: x_j\notin G_j]\Pr[f(x_1,\ldots,x_k)=0\mid\forall j: x_j\notin G_j]=\\&\Pr[\forall j: x_j\notin G_j]\left(1-\sum_{i=1}^{k}\Pr[f(x_1,\ldots,x_k)=i\mid\forall j: x_j\notin G_j]\right)\geq\\&(1-\varepsilon)^k\cdot\left(1-\frac{k}{q\,(1-\varepsilon)^{k-1}}\right)=(1-\varepsilon)^k-\frac{k(1-\varepsilon)}{q}\geq(1-\varepsilon)^k-k/q\,,\end{aligned}$$

and we arrive at a contradiction. □

3.2 Zero-Knowledge Amplification

In this section, we present and analyze our zero-knowledge amplifier that is based on the hidden-bits paradigm.

3.2.1 Statistical Tools

Lemma 3.3 (Useful facts)**.** *We defer the proofs of the following useful facts to the appendix (Appendix A.2).*

1. *Let $(X_1, Z_1), (X_2, Z_2)$ be two independent pairs of jointly distributed random variables. Then*
$$\widetilde{H}_\infty(X_1, X_2 \mid Z_1, Z_2) = \widetilde{H}_\infty(X_1 \mid Z_1) + \widetilde{H}_\infty(X_2 \mid Z_2)\,.$$

2. *Let X, Z be a pair of jointly distributed random variables and A be an event, then*
$$\widetilde{H}_\infty(X \mid Z, A) \geq \widetilde{H}_\infty(X \mid Z) - \log(1/\Pr[A])\,.$$

3. *Let $Y := (Z, X_1, \ldots, X_t)$ be a random variable and denote by Y^{-j} the r.v. with X_j omitted. Further, let $f(\cdot\,; r)$ be a randomized function. Assume that for all $j \in [t]$ we have*
$$Y^{-j}, R, f(X_j; R) \stackrel{s}{\approx}_\delta Y^{-j}, R, U\,,$$
then we also have
$$Z, R, f(X_1; R), \ldots, f(X_t; R) \stackrel{s}{\approx}_{t\cdot\delta} Z, R, U^t\,.$$

4. *Let X, Y be a pair of jointly distributed random variables and $\{A_i\}_{i\in[k]}$ be a set of events, then*
$$\mathsf{SD}(X, Y) \leq \sum_{\substack{I\subseteq[k]\\ \Pr[A_I, \overline{A}_{\overline{I}}]>0}} \Pr\left[A_I, \overline{A}_{\overline{I}}\right] \cdot \mathsf{SD}(X, Y \mid A_I, \overline{A}_{\overline{I}})\,.$$

Lemma 3.4 (Statistical Extraction Lemma)**.** *Let $(Z, X_1, \ldots, X_t) \stackrel{s}{\approx}_\delta (\widetilde{Z}, U_1^n, \ldots, U_t^n)$ be distributions over $\mathcal{Z} \times \{0,1\}^{t\cdot n}$ for some $t, n \in \mathbb{N}$ and $\delta \in (0,1)$, where each U_i^n is uniform and independent. Further, let $\mathcal{H} = \left\{h : \{0,1\}^{k\cdot n} \to \{0,1\}\right\}$ be a universal hash family and consider k independent copies of $\{(Z_i, X_{i1}, \ldots, X_{it})\}_{i\in[k]}$, for some $k \in \mathbb{N}$. Then we have that*
$$Z_1, \ldots, Z_k, h, \{h(X_{1j}, \ldots, X_{kj})\}_{j\in[t]} \stackrel{s}{\approx}_{\delta'} Z_1, \ldots, Z_k, h, U^t\,,$$
over $h \leftarrow \mathcal{H}$, where $\delta' \leq \delta^k + t/\sqrt{2^{n-\log(1/(1-\delta))}}$.

Proof. Using the coupling method, there exist events $A, \widetilde{A}$ with probability $1-\delta$ each, such that
$$(Z, X_1, \ldots, X_t) \mid A \equiv (\widetilde{Z}, U_1^n, \ldots, U_t^n) \mid \widetilde{A}\,.$$
Let us denote by Y and $\widetilde{Y}$ the r.v.s $(Z, X_1, \ldots, X_t)$ and $(\widetilde{Z}, U_1^n, \ldots, U_t^n)$ respectively. Further, for $j \in [t]$, denote by Y^{-j} (resp. $\widetilde{Y}^{-j}$) the r.v. where X_j (resp. U_j^n) is omitted. The k pairs

$(Y_1, A_1), \ldots, (Y_k, A_k)$ are mutually independent. For every non-empty subset $\emptyset \neq I \subseteq [k]$ and $j \in [t]$, we can choose some $i^* \in I$ and have that

$$\widetilde{H}_\infty\left(X_{1j}, \ldots, X_{kj} \mid Y_1^{-j}, \ldots, Y_k^{-j}, A_I, \overline{A}_{\overline{I}}\right) = \sum_{i \in [k]} \widetilde{H}_\infty\left(X_{ij} \mid Y_i^{-j}, A_I, \overline{A}_{\overline{I}}\right) \geq$$
$$\widetilde{H}_\infty\left(X_{i^*j} \mid Y_{i^*}^{-j}, A_I, \overline{A}_{\overline{I}}\right) = \widetilde{H}_\infty\left(X_{i^*j} \mid Y_{i^*}^{-j}, A_{i^*}\right) = \widetilde{H}_\infty\left(U_j^n \mid \widetilde{Y}^{-j}, \widetilde{A}\right) \geq$$
$$\widetilde{H}_\infty\left(U_j^n \mid \widetilde{Y}^{-j}\right) - \log\left(1/\Pr\left[\widetilde{A}\right]\right) = n - \log(1/(1-\delta))\,,$$

where we used Fact 1 in the first equality, and Fact 2 in the second inequality. Specifically, for Fact 1, we used that the k pairs $(X_{1j}, Y_1^{-j}), \ldots, (X_{kj}, Y_k^{-j})$ remain mutually independent conditioned on any $A_I, \overline{A}_{\overline{I}}$. Then, using the generalized leftover hash lemma (Lemma 2.13), we conclude that

$$Y_1^{-j}, \ldots, Y_k^{-j}, h, h(X_{1j}, \ldots, X_{kj}) \mid A_I, \overline{A}_{\overline{I}} \stackrel{s}{\approx}_{\delta^*} Y_1^{-j}, \ldots, Y_k^{-j}, h, U \mid A_I, \overline{A}_{\overline{I}}\,,$$

over $h \leftarrow \mathcal{H}$, and for $\delta^* \leq 2^{\frac{-n+\log(1/(1-\delta))}{2}}$. Now we can apply a hybrid argument over $j \in [t]$ as in Fact 3 to conclude that

$$Z_1, \ldots, Z_k, h, \{h(X_{1j}, \ldots, X_{kj})\}_{j \in [t]} \mid A_I, \overline{A}_{\overline{I}} \stackrel{s}{\approx}_{t \cdot \delta^*} Z_1, \ldots, Z_k, h, U^t \mid A_I, \overline{A}_{\overline{I}}\,,$$

over $h \leftarrow \mathcal{H}$. Finally, using Fact 4 and over $h \leftarrow \mathcal{H}$, it holds that

$$\mathsf{SD}\left((Z_1, \ldots, Z_k, h, \{h(X_{1j}, \ldots, X_{kj})\}_{j \in [t]}), (Z_1, \ldots, Z_k, h, U^t)\right) \leq$$
$$\sum_{I \subseteq [k]} \Pr\left[A_I, \overline{A}_{\overline{I}}\right] \cdot \mathsf{SD}\left((Z_1, \ldots, Z_k, h, \{h(X_{1j}, \ldots, X_{kj})\}), (Z_1, \ldots, Z_k, h, U^t) \mid A_I, \overline{A}_{\overline{I}}\right) \leq$$
$$\Pr\left[A_\emptyset, \overline{A}_{[k]}\right] + \sum_{\emptyset \neq I \subseteq [k]} \Pr\left[A_I, \overline{A}_{\overline{I}}\right] \cdot t \cdot \delta^* \leq \delta^k + t \cdot \delta^*\,.$$

□

3.2.2 The Zero-Knowledge Amplifier

Theorem 3.5. *Consider the protocol Π_z described in Fig. 3.1, where Π_{wk} is an $(\varepsilon_s(n), \varepsilon_z(n))$-weak* NIZK for NP, the amplification parameter is set to k, and Π^{hbm} is an $(2^{-k \cdot n} \cdot \mathrm{negl}(n))$-sound HBM NIZK. Then, Π_z is an $(1 - (1 - \varepsilon_s(n'))^k + \mathrm{negl}, \varepsilon_z^k(n') + \mathrm{negl})$-weak* NIZK for NP, for some fixed polynomial $n' = k \cdot \mathsf{poly}(n)$.*

Note that usually we think of $\varepsilon_s(n)$ and $\varepsilon_z(n)$ as non-increasing, so depending on $n' > n$ is not worse. However, if for example $\varepsilon_z(n) = 1 - 1/n$, then $\varepsilon_z(n') > \varepsilon(n)$.

Proof. Correctness follows readily from the correctness of Π_{wk} and Π^{hbm}. We focus on proving soundness and zero knowledge.

Given a weak* NIZK Π_{wk}, amplification parameter k, HBM NIZK Π^{hbm} with $t = k \cdot \mathsf{poly}(n)$ hidden bits, n-bit-output PRF F_s and a universal hash family $\mathcal{H} = \left\{h : \{0,1\}^{k \cdot n} \to \{0,1\}\right\}$, we construct Π_z as follows:

$\underline{crs_z \leftarrow \mathsf{Gen}_z(1^n)}$

1. Sample a hash function $h \leftarrow \mathcal{H}$.
2. Sample $r \leftarrow \{0,1\}^t$.
3. For all $i \in [k]$, sample $crs_i \leftarrow \mathsf{Gen}_{wk}(1^{n'})$, for some fixed polynomial $n' = k \cdot \mathsf{poly}(n)$ corresponding to the size of the statement below.
4. Output $crs_z := (h, r, crs_1, crs_2, \ldots, crs_k)$ as the CRS.

$\underline{\pi_z \leftarrow \mathsf{P}_z(crs_z, x, w)}$

1. For all $i \in [k]$, sample a PRF seed $s_i \leftarrow \{0,1\}^n$.
2. Compute the hidden bit-string r^* where $\forall j \in [t] :\ r^*_j = r_j \oplus h(F_{s_1}(j), \ldots, F_{s_k}(j))$.
3. Compute the HBM proof $(J, \pi^{\mathsf{hbm}}) \leftarrow \mathsf{P}^{\mathsf{hbm}}(r^*, x, w)$.
4. For all $i \in [k]$, generate a (weak ZK) proof that a consistent seed w.r.t. the revealed indices $F_{s_i}(J)$ exists, namely $\pi_i \leftarrow \mathsf{P}_{wk}\left(crs_i,\ \exists s \in \{0,1\}^n :\ F_s(J) = F_{s_i}(J),\ s_i\right)$.
5. Output $\pi_z := \left(J,\ \pi^{\mathsf{hbm}},\ \{F_{s_i}(J)\}_{i \in [k]},\ \{\pi_i\}_{i \in [k]}\right)$ as the proof.

$\underline{b_z := \mathsf{V}_z(crs_z, x, \pi_z)}$

1. Parse $\pi_z := \left(J,\ \pi^{\mathsf{hbm}},\ \{y_{ij}\}_{i \in [k],\, j \in J},\ \{\pi_i\}_{i \in [k]}\right)$ and $crs_z := (h, r, crs_1, crs_2, \ldots, crs_k)$.
2. Check that $\mathsf{V}_{wk}(crs_i,\ \exists s \in \{0,1\}^n :\ F_s(J) = y_i(J),\ \pi_i) = 1$, for all $i \in [k]$.
3. Compute the revealed bit-string r^*_J where $\forall j \in J :\ r^*_j = r_j \oplus h(y_{1j}, \ldots, y_{kj})$.
4. Check that $\mathsf{V}^{\mathsf{hbm}}(J, r^*_J, x, \pi^{\mathsf{hbm}}) = 1$.
5. If all checks passed output $b_z := 1$ (accept π_z), otherwise output $b_z := 0$ (reject π_z).

Figure 3.1: ZK Amplifier

Soundness*. Let $A_z(crs_z)$ be an adversary that breaks the soundness* of Π_z with probability $\varepsilon = \varepsilon(n)$, and assume toward contradiction that $\varepsilon(n) > 1-(1-\varepsilon_s(n'))^k + 3k/q$ for some polynomial $q = q(n)$ and infinitely many n's. We construct an adversary $A_{wk}(crs)$ that breaks the soundness* of Π_{wk} as follows:

1. Initialize OUT as the empty string.
2. For every $i \in [k]$, repeat $q \log q$ times:
   (a) Sample $\forall j \in [k] \setminus i :\ crs_j \leftarrow \mathsf{Gen}_{wk}(1^{n'})$, with $h \leftarrow \mathcal{H}$ and $r \leftarrow \{0,1\}^t$.
   (b) Set $crs_i := crs$ and $crs_z := (h, r, crs_1, crs_2, \ldots, crs_k)$.
   (c) Run $(x_1, \pi_1), \ldots, (x_p, \pi_p) \leftarrow A_z(crs_z)$, where each π_ℓ is parsed as
   $$\left(J_\ell,\ \pi_\ell^{\mathsf{hbm}},\ \{y_{\ell i}(J_\ell)\}_{i\in[k]}\ ,\ \{\pi_{\ell i}\}_{i\in[k]}\right).$$
   Recall that each π_ℓ induces k statement-proof pairs $\{(x_{\ell i}, \pi_{\ell i})\}_{i\in[k]}$, where
   $$x_{\ell i} := \exists s \in \{0,1\}^n :\ F_s(J_\ell) = y_{\ell i}(J_\ell)\,.$$
   (d) For every $\ell \in [p]$, concatenate the i'th internal statement-proof pair of π_ℓ to OUT, that is, concatenate the tuple $(x_{\ell i}, \pi_{\ell i})$ taken from π_ℓ.
3. Output OUT.

The running time of A_{wk} increases to $k \cdot q \log q \cdot (\mathrm{TIME}(A_z) + \mathsf{poly}(n,k))$ so it remains polynomial.

To analyze the success probability of A_{wk}, we define a randomized and possibly inefficient function $T(crs_1, crs_2, \ldots, crs_k) \to \{0, \ldots, k\}$, that runs the adversary A_z and tests which crs_i was broken. That is, T samples $h \leftarrow \mathcal{H}$ and $r \leftarrow \{0,1\}^t$, runs $(x_1, \pi_1), \ldots, (x_p, \pi_p) \leftarrow A_z(h, r, crs_1, crs_2, \ldots, crs_k)$, and outputs an $i \in [k]$ such that the i'th internal statement-proof pair $(x_{\ell i}, \pi_{\ell i})$ of some π_ℓ, is a false statement with an accepting proof w.r.t. crs_i. If for all $\ell \in [p]$ no such $i \in [k]$ exists T outputs 0, and if multiple such i's exist simply pick one at random or pick the minimal.

By the $\left(2^{-k\cdot n} \cdot \mathrm{negl}(n)\right)$-statistical-soundness of Π^{hbm}, except for negligible probability over $r \in \{0,1\}^t$, there does not exist an accepting HBM proof of a no-instance for all choices of the PRF seeds. Specifically, fixing any $h \in \mathcal{H}$, and taking a union bound over all choices of $s_1, \ldots, s_k \in \{0,1\}^n$, this probability (over $r \in \{0,1\}^t$) is bounded by:

$$2^{k\cdot n} \cdot \left(2^{-k\cdot n} \cdot \mathrm{negl}(n)\right) = \mathrm{negl}(n)\,.$$

Therefore, with probability at least $\varepsilon(n) - \mathrm{negl}(n)$, we have that A_z breaks soundness* against a "good" r (no accepting HBM proof of a no-instance for all seed choices), then for every false instance with an accepting proof (x, π_z) it cannot be that all $\{y_i(J)\}_{i\in[k]}$ are consistent with PRF seeds. Thus, for at least one ℓ and one i, we have that $(x_{\ell i}, \pi_{\ell i})$ is a no-instance with an accepting proof w.r.t. crs_i. We conclude that

$$\begin{aligned}
&\Pr_{\forall i:\ crs_i \leftarrow \mathsf{Gen}_{wk}(1^{n'})}[T(crs_1, crs_2, \ldots, crs_k) = 0] \le 1 - \varepsilon + \mathrm{negl} < \\
&(1-\varepsilon_s)^k - 3k/q + \mathrm{negl} \le (1 - \varepsilon_s - 1/q)^k + k/q - 3k/q + \mathrm{negl} < \\
&(1 - (\varepsilon_s + 1/q))^k - k/q\,,
\end{aligned}$$

The simulator $\mathsf{Sim}_z = \mathsf{Sim}_z(x)$ is defined below:

$\underline{(crs_z, \pi_z) \leftarrow \mathsf{Sim}_z(x)}$

1. Sample a hash function $h \leftarrow \mathcal{H}$.
2. Simulate the Π^{hbm} proof $(J, r^*_J, \pi^{\mathsf{hbm}}) \leftarrow \mathsf{Sim}^{\mathsf{hbm}}(x)$.
3. For $i \in [k]$:
   (a) Sample $crs_i \leftarrow \mathsf{Gen}_{wk}(1^{n'})$.
   (b) Sample a PRF seed $s_i \leftarrow \{0,1\}^n$.
   (c) Generate a (weak ZK) proof that a consistent seed w.r.t. the revealed indices $F_{s_i}(J)$ exists, namely $\pi_i \leftarrow \mathsf{P}_{wk}\left(crs_i, \exists s \in \{0,1\}^n : F_s(J) = F_{s_i}(J), s_i\right)$.
4. Compute the random CRS bit-string at revealed indices r_J where $\forall j \in J : r_j = r^*_j \oplus h(F_{s_1}(j), \ldots, F_{s_k}(j))$.
5. Sample the random CRS bit-string at hidden indices $r_{\overline{J}} \leftarrow \{0,1\}^{|\overline{J}|}$.
6. Set $crs_z := (h, r, crs_1, crs_2, \ldots, crs_k)$ as the CRS.
7. Set $\pi_z := \left(J, \pi^{\mathsf{hbm}}, \{F_{s_i}(J)\}_{i\in[k]}, \{\pi_i\}_{i\in[k]}\right)$ as the proof.
8. Output (crs_z, π_z).

Figure 3.2: Simulator Sim_z

where we used that $\alpha^k \leq (\alpha - \beta)^k + \beta \cdot k$ for all $0 \leq \beta \leq \alpha \leq 1$. Now we apply the OR easy-subset lemma (Lemma 3.2) and get that there exists some i and a set of CRS's S with

$$\Pr_{crs \leftarrow \mathsf{Gen}_{wk}(1^{n'})}[crs \in S] \geq \varepsilon_s + 1/q\,,$$

such that for every $crs_i \in S$:

$$\Pr_{\forall j \in [k]\setminus i:\ crs_j \leftarrow \mathsf{Gen}_{wk}(1^{n'})}[T(crs_1, crs_2, \ldots, crs_k) = i] \geq 1/q\,.$$

So if $crs_i \in S$, the probability to never get $T = i$ during $q \log q$ repetitions is at most $(1 - 1/q)^{q \log q}$. Note that $A_{wk}(crs)$ goes over all possible values of $i \in [k]$ so in particular it hits the correct value, hence its probability to succeed breaking the soundness* of Π_{wk}, over the choice of $crs \leftarrow \mathsf{Gen}_{wk}(1^{n'})$ and its internal randomness, is at least $\varepsilon_s + 1/q - (1 - 1/q)^{q \log q} > \varepsilon_s$ in contradiction to the security of Π_{wk}.

Zero Knowledge. We start by describing the simulator in Fig. 3.2. Then, we prove by a hybrid argument that the real distribution of CRS and proof (crs_z, π_z) given by an honest execution of Π_z

The distribution $H_1 = H_1(x, w)$ is defined below:

$\underline{(crs_z, \pi_z) \leftarrow H_1(x, w)}$

1. Sample a hash function $h \leftarrow \mathcal{H}$.
2. Sample $r^* \leftarrow \{0,1\}^t$.
3. Compute the HBM proof $(J, \pi^{\mathsf{hbm}}) \leftarrow \mathsf{P}^{\mathsf{hbm}}(r^*, x, w)$.
4. For $i \in [k]$:
   (a) Sample $crs_i \leftarrow \mathsf{Gen}_{wk}(1^{n'})$.
   (b) Sample a PRF seed $s_i \leftarrow \{0,1\}^n$.
   (c) Generate a (weak ZK) proof that a consistent seed w.r.t. the revealed indices $F_{s_i}(J)$ exists, namely $\pi_i \leftarrow \mathsf{P}_{wk}\left(crs_i, \exists s \in \{0,1\}^n : \; F_s(J) = F_{s_i}(J), \; s_i\right)$.
5. Compute the random CRS bit-string r where $\forall j \in [t] : \; r_j = r^*_j \oplus h(F_{s_1}(j), \ldots, F_{s_k}(j))$.
6. Set $crs_z := (h, r, crs_1, crs_2, \ldots, crs_k)$ as the CRS.
7. Set $\pi_z := \left(J, \pi^{\mathsf{hbm}}, \{F_{s_i}(J)\}_{i\in[k]}, \{\pi_i\}_{i\in[k]}\right)$ as the proof.
8. Output (crs_z, π_z).

Figure 3.3: Hybrid H_1

The distribution $H_2 = H_2(x, w)$ is defined below:

$\underline{(crs_z, \pi_z) \leftarrow H_2(x, w)}$

1. Sample a hash function $h \leftarrow \mathcal{H}$.
2. Sample $r^* \leftarrow \{0,1\}^t$.
3. Compute the HBM proof $(J, \pi^{\mathsf{hbm}}) \leftarrow \mathsf{P}^{\mathsf{hbm}}(r^*, x, w)$.
4. For $i \in [k]$:
   (a) Sample $crs_i \leftarrow \mathsf{Gen}_{wk}(1^{n'})$.
   (b) Sample a PRF seed $s_i \leftarrow \{0,1\}^n$.
   (c) Generate a (weak ZK) proof that a consistent seed w.r.t. the revealed indices $F_{s_i}(J)$ exists, namely $\pi_i \leftarrow \mathsf{P}_{wk}\left(crs_i, \exists s \in \{0,1\}^n : F_s(J) = F_{s_i}(J), s_i\right)$.
5. Compute the random CRS bit-string at revealed indices r_J where $\forall j \in J : r_j = r^*_j \oplus h(F_{s_1}(j), \ldots, F_{s_k}(j))$.
6. Sample the random CRS bit-string at hidden indices $r_{\overline{J}} \leftarrow \{0,1\}^{|\overline{J}|}$.
7. Set $crs_z := (h, r, crs_1, crs_2, \ldots, crs_k)$ as the CRS.
8. Set $\pi_z := \left(J, \pi^{\mathsf{hbm}}, \{F_{s_i}(J)\}_{i\in[k]}, \{\pi_i\}_{i\in[k]}\right)$ as the proof.
9. Output (crs_z, π_z).

Figure 3.4: Hybrid H_2

over (x, w), which we refer to as H_0, is indistinguishable from the simulated distribution of CRS and proof, which we refer to as H_3.

We first observe that the hybrid H_1 described in Fig. 3.3, has exactly the same distribution of H_0: We simply changed the order of sampling, as r is sampled uniformly and independently of $\{s_i\}_i$, it is equivalent whether we sample $r \leftarrow \{0,1\}^t$ and let $r^* = r \oplus g(s_1, \ldots, s_k)$, or sample $r^* \leftarrow \{0,1\}^t$ and let $r = r^* \oplus g(s_1, \ldots, s_k)$, where $g(s_1, \ldots, s_k) \in \{0,1\}^t$ is defined by $\forall j \in [t] : g_j = h(F_{s_1}(j), \ldots, F_{s_k}(j))$.

Next, in the hybrid H_2 described in Fig. 3.4, we randomly sample $r_{\overline{J}} \leftarrow \{0,1\}^{|\overline{J}|}$ instead of computing $r_{\overline{J}} = r^*_{\overline{J}} \oplus g_{\overline{J}}(s_1, \ldots, s_k)$. For the indistinguishability analysis, fix some choice of $(r^*, J, \pi^{\mathsf{hbm}})$ in the second and third lines. We observe that the k random variables $\left\{\left(F_{s_i}(J), crs_i, \pi_i\right), F_{s_i}(\overline{J})\right\}_{i\in[k]}$ are i.i.d., and denote by $\left(F_s(J), crs, \pi\right), F_s(\overline{J})$ their shared distribution. We have that

$$\begin{aligned}&\left(F_s(J), crs, \pi\right), F_s(\overline{J}) \overset{\mathrm{c}}{\approx}_{\varepsilon_z} \left(F_s(J), \mathsf{Sim}_{wk}(F_s(J))\right), F_s(\overline{J}) \overset{\mathrm{c}}{\approx} \\ &\left(F_s(J), \mathsf{Sim}_{wk}(F_s(J))\right), U^{n\cdot|\overline{J}|},\end{aligned}$$

where the first transition is based on the weak-ZK of Π_{wk}, and the second is based on the security of the PRF. Also, we used $\mathsf{Sim}_{wk}(F_s(J))$ as shorthand for $\mathsf{Sim}_{wk}\left(\exists s' \in \{0,1\}^n : F_{s'}(J) = F_s(J)\right)$. Now, using the hybrid indistinguishability lemma (Lemma 2.15), for every $\varepsilon > 0$ there exists some hybrid distribution $\left(F'_s(J), crs', \pi'\right), F'_s(\overline{J})$ such that

$$\begin{aligned}&\left(F_s(J), crs, \pi\right), F_s(\overline{J}) \overset{\mathrm{c}}{\approx}_{\varepsilon} \left(F'_s(J), crs', \pi'\right), F'_s(\overline{J}) \overset{\mathrm{s}}{\approx}_{\varepsilon_z+\mathrm{negl}} \\ &\left(F_s(J), \mathsf{Sim}_{wk}(F_s(J))\right), U^{n\cdot|\overline{J}|},\end{aligned}$$

with respect to circuits smaller by $\mathsf{poly}(\varepsilon/n)$. Therefore, we have that

$$\left\{\left(F_{s_i}(J), crs_i, \pi_i\right), F_{s_i}(\overline{J})\right\}_{i\in[k]} \overset{\mathrm{c}}{\approx}_{k\cdot\varepsilon} \left\{\left(F'_{s_i}(J), crs'_i, \pi'_i\right), F'_{s_i}(\overline{J})\right\}_{i\in[k]}.$$

We next apply the statistical extraction lemma (Lemma 3.4), to conclude that

$$\begin{aligned}&\left\{\left(F'_{s_i}(J), crs'_i, \pi'_i\right)\right\}_{i\in[k]}, h, \left\{h(F'_{s_1}(j), \ldots, F'_{s_k}(j))\right\}_{j\in\overline{J}} \overset{\mathrm{s}}{\approx}_{\varepsilon_z^k+\mathrm{negl}} \\ &\left\{\left(F'_{s_i}(J), crs'_i, \pi'_i\right)\right\}_{i\in[k]}, h, U^{|\overline{J}|},\end{aligned}$$

where $h \leftarrow \mathcal{H}$. Going back to the original distributions, we arrive at

$$\begin{aligned}&\{(F_{s_i}(J), crs_i, \pi_i)\}_{i\in[k]}, h, \{h(F_{s_1}(j), \ldots, F_{s_k}(j))\}_{j\in\overline{J}} \overset{\mathrm{c}}{\approx}_{\varepsilon_z^k+2k\varepsilon+\mathrm{negl}} \\ &\{(F_{s_i}(J), crs_i, \pi_i)\}_{i\in[k]}, h, U^{|\overline{J}|},\end{aligned}$$

where $h \leftarrow \mathcal{H}$. Since the above holds for every fixed choice of $(r^*, J, \pi^{\mathsf{hbm}})$, it also holds for a random sample. Hence, the hybrid H_2 is at most $\varepsilon_z^k + 2k\varepsilon + \mathrm{negl}$ computationally indistinguishable from H_1, seeing that we replaced $\forall j \in \overline{J} : r_j = r^*_j \oplus h(F_{s_1}(j), \ldots, F_{s_k}(j))$ by $\forall j \in \overline{J} : r_j \leftarrow U \equiv r^*_j \oplus U$.

Finally, in the hybrid H_3 induced by the simulator Sim_z, we simulate $(J, r^*_j, \pi^{\mathsf{hbm}}) \leftarrow \mathsf{Sim}^{\mathsf{hbm}}(x)$ using the simulator of Π^{hbm} instead of generating $(J, r^*, \pi^{\mathsf{hbm}})$ honestly, which suffices since we are not using $r^*_{\overline{J}}$. By the ZK of Π^{hbm}, hybrid H_3 is indistinguishable from H_2.

Overall, we conclude that $\mathsf{Sim}_z(x)$ is at most $\varepsilon_z^k + 2k\varepsilon + \mathrm{negl}$ computationally indistinguishable from (crs_z, π_z) given by an honest execution of Π_z over (x, w), and since this holds for any $\varepsilon > 0$ with the running time of the reduction being polynomial in $1/\varepsilon$, we have $\varepsilon_z^k + \mathrm{negl}$ indistinguishability. □

Remark 3.6. We can only compose the ZK amplifier a constant number of times to maintain efficiency, as the statement size grows polynomially. We also remark that it works for adaptive soundness with looser degradation $k \cdot \varepsilon_s$, but not for non-adaptive soundness as the attacked instances are induced by the adversary. Finally, we remark that since the PRF is only applied to $[t]$, a PRG with an output length of $t \cdot n$ bits parsed as t blocks of size n would suffice, and we use a PRF strictly for notational ease.

Remark 3.7. Note that the above reduction for zero-knowledge amplification is non-uniform. Specifically, we used non-uniformity to fix some choice of $(r^*, J, \pi^{\mathsf{hbm}})$ and obtain samples from the induced hybrid distribution $(F'_s(J), crs', \pi'), F'_s(\overline{J})$, for the hybrid argument

$$\left\{(F_{s_i}(J), crs_i, \pi_i), F_{s_i}(\overline{J})\right\}_{i\in[k]} \overset{c}{\approx}_{k\cdot\varepsilon} \left\{\left(F'_{s_i}(J), crs'_i, \pi'_i\right), F'_{s_i}(\overline{J})\right\}_{i\in[k]} .$$

Even though the uniform version of the hybrid indistinguishability lemma allows to sample from the hybrid X', it is not immediately applicable here, because of the auxiliary parameter $(r^*, J, \pi^{\mathsf{hbm}})$. We suppose that a more careful consideration of Holenstein's uniform hard-core lemma [Hol05] should suffice for this generalization, but leave it open for the curious reader.

The above zero-knowledge amplifier directly implies the following:

Corollary 3.8 (Weak Zero-Knowledge Amplification)**.** *Assuming OWFs and an* $(\mathrm{negl}, 1 - 1/n^{\varepsilon})$*-weak NIZK for some constant* $\varepsilon < 1$*, there also exists standard NIZK.*

Proof. Recall that negl-soundness is equivalent to negl-soundness*. Using the zero-knowledge amplifier (Theorem 3.5), noting that $1 - (1-\varepsilon)^k \leq \varepsilon \cdot k$ and $1 - x \leq e^{-x}$, yields (up to negl)

$$\left(k \cdot \mathrm{negl},\ e^{-k/(k\cdot n^c)^{\varepsilon}}\right) = \left(\mathrm{negl},\ e^{-k^{1-\varepsilon}/n^{c\cdot\varepsilon}}\right) ,$$

where we used that $n' = k \cdot n^c$ for some constant $c > 0$. We want to have $k^{1-\varepsilon}/n^{c\cdot\varepsilon} \gg \log n$, so we can pick for example $k = n^{c/(1-\varepsilon)} = \mathsf{poly}(n)$. □

In more general, the above corollary also applies to the case where the ZK error may be $1 - 1/\mathsf{poly}(\lambda)$ in the security parameter λ, but not where it arbitrarily grows with the instance size n.

3.3 Soundness* Amplification

In this section, we explicitly present and analyze the parallel repetition soundness* amplifier.

Theorem 3.9. *Consider the direct-product protocol* Π_s *described in Fig. 3.5, where* Π_{wk} *is an* $(\varepsilon_s, \varepsilon_z)$*-weak* NIZK and the amplification parameter is set to* k*. Then,* Π_s *is an* $(\varepsilon_s^k + \mathrm{negl}, 1 - (1-\varepsilon_z)^k + \mathrm{negl})$*-weak* NIZK.*

Proof. Correctness follows readily from the correctness of Π_{wk}.

Soundness*. Let A_s be an adversary that breaks the soundness* of Π_s with probability $\varepsilon = \varepsilon(n)$, and assume toward contradiction that $\varepsilon > \varepsilon_s^k + 2k/q$ for some polynomial $q = q(n)$ and infinitely many n's. We construct an adversary $A_{wk}(crs)$ against Π_{wk} as follows:

1. Initialize OUT as the empty string.

Given a weak* NIZK Π_{wk} and amplification parameter k, we construct Π_s as follows:

$\underline{crs_s \leftarrow \mathsf{Gen}_s(1^n)}$

1. For all $i \in [k]$, sample $crs_i \leftarrow \mathsf{Gen}_{wk}(1^n)$.
2. Output $crs_s := (crs_1, \ldots, crs_k)$ as the CRS.

$\underline{\pi_s \leftarrow \mathsf{P}_s(crs_s, x, w)}$

1. For all $i \in [k]$, generate a (weak ZK) proof $\pi_i \leftarrow \mathsf{P}_{wk}(crs_i, x, w)$.
2. Output $\pi_s := (\pi_1, \ldots, \pi_k)$ as the proof.

$\underline{b_s := \mathsf{V}_s(crs_s, x, \pi_s)}$

1. Check that $\mathsf{V}_{wk}(crs_i, x, \pi_i) = 1$, for all $i \in [k]$.
2. If all checks passed output $b_s := 1$ (accept π_s), otherwise output $b_s := 0$ (reject π_s).

$\underline{(crs_s, \pi_s) \leftarrow \mathsf{Sim}_s(x)}$

1. For all $i \in [k]$ simulate $(crs_i, \pi_i) \leftarrow \mathsf{Sim}_{wk}(x)$.
2. Set $crs_s := (crs_1, \ldots, crs_k)$ as the CRS.
3. Set $\pi_s := (\pi_1, \ldots, \pi_k)$ as the proof.
4. Output (crs_s, π_s).

Figure 3.5: Soundness* Amplifier (Direct-Product)

2. For every $i \in [k]$, repeat $q \log q$ times:
   (a) Sample $crs_j \leftarrow \mathsf{Gen}_{wk}(1^n)$ for all $j \in [k] \setminus i$.
   (b) Set $crs_i := crs$ and $crs_s := (crs_1, crs_2, \ldots, crs_k)$.
   (c) Run $(x_1, \pi_1), \ldots, (x_p, \pi_p) \leftarrow A_s(crs_s)$, where each π_ℓ is parsed as $(\pi_{\ell 1}, \ldots, \pi_{\ell k})$.
   (d) For every $\ell \in [p]$, concatenate the i'th internal proof of π_ℓ to OUT, namely, the tuple $(x_\ell, \pi_{\ell i})$.
3. Output OUT.

The running time of A_{wk} increases to $k \cdot q \log q \cdot (\text{TIME}(A_s) + \mathsf{poly}(n, k))$ so it remains polynomial.

For the analysis of the success probability, we define a randomized and possibly inefficient function $T(crs_1, crs_2, \ldots, crs_k) \to \{0, 1\}$, that runs the adversary $A_s(crs_1, crs_2, \ldots, crs_k)$ and tests whether it succeeds breaking soundness*. We have that

$$\Pr_{\forall i:\, crs_i \leftarrow \mathsf{Gen}_{wk}(1^n)} [T(crs_1, crs_2, \ldots, crs_k) = 1] = \varepsilon > \varepsilon_s^k + 2k/q \geq (\varepsilon_s + 1/q)^k + k/q\,,$$

where we used that $(\alpha - \beta)^k + \beta \cdot k \geq \alpha^k$ for all $0 \leq \beta \leq \alpha \leq 1$. Now we apply the AND easy-subset lemma (Lemma 3.1) and get that there exists some i and a set of CRS's S with

$$\Pr_{crs \leftarrow \mathsf{Gen}_{wk}(1^n)} [crs \in S] \geq \varepsilon_s + 1/q\,,$$

such that for every $crs_i \in S$:

$$\Pr_{\forall j \in [k] \setminus i:\, crs_j \leftarrow \mathsf{Gen}_{wk}(1^n)} [T(crs_1, crs_2, \ldots, crs_k) = 1] \geq 1/q\,.$$

So if $crs_i \in S$, the probability to never get $T = 1$ during $q \log q$ repetitions is at most $(1 - 1/q)^{q \log q}$. Note that $A_{wk}(crs)$ goes over all possible values of $i \in [k]$ so in particular it hits the correct value, hence its probability to succeed breaking the soundness* of Π_{wk}, over the choice of $crs \leftarrow \mathsf{Gen}_{wk}(1^n)$ and its internal randomness, is at least $\varepsilon_s + 1/q - (1 - 1/q)^{q \log q} > \varepsilon_s$ in contradiction to the security of Π_{wk}.

Zero Knowledge. We have that $\mathsf{Sim}_{wk}(x) \overset{c}{\approx}_{\varepsilon_z} (crs, \pi)$ where $crs \leftarrow \mathsf{Gen}_s(1^n)$, $\pi \leftarrow \mathsf{P}_{wk}(crs, x, w)$, and we are now switching to k independent copies, so we can apply the indistinguishability bound for product distributions (Lemma 2.17). As long as $1/(1 - \varepsilon_z)^k$ is polynomial, we maintain efficiency for every inverse-polynomial slackness error ε (and even beyond that).

□

It is important to note here, that all of the above soundness* security reductions are oblivious to statistical-soundness. That is, if the underlying protocol is (weakly) statistically-sound*, the reduction preserves this property (with the same amplification and degradation parameters).

4 Simultaneous Amplification

In this section, we show how to combine our pair of amplifiers in order to amplify weak to fully secure NIZK. We state a general theorem for the notion of soundness*, and derive corollaries for standard notions of soundness.

Theorem 4.1 (Weak* NIZK Amplification)**.** *Assuming OWFs and an (α, β)-weak* NIZK for NP, for some constants α, β such that $\alpha + \beta < 1$, there also exists standard NIZK for NP.*

Proof. We start from $(\alpha, \beta) = (1 - 1/2^{-\log(1-\alpha)}, 1/2^{-\log\beta})$. Apply the ZK amplifier with $k = \log n$, to get

$$(1 - 1/n^{-\log(1-\alpha)} + \mathrm{negl}, 1/n^{-\log\beta} + \mathrm{negl}) \leq (1 - 1/n^{c_1}, 1/n^{c_2}),$$

for some constants $c_1 < c_2$, since $-\log(1-\alpha) < -\log\beta$. It is important to note we do not want to choose k to be much larger, so that $2^{-k\log(1-\alpha)}$ remains polynomial. Now apply the soundness* amplifier with $k = n^{(c_1+c_2)/2}$, noting that $1 - x \leq e^{-x}$ and $1 - (1-\varepsilon)^k \leq \varepsilon \cdot k$, we get

$$(e^{-k/n^{c_1}} + \mathrm{negl}, k/n^{c_2} + \mathrm{negl}) = (\mathrm{negl}, 1/n^{c}),$$

for some positive constant c. Finally, apply the ZK amplifier again with $k = \log n$, to obtain $(\mathrm{negl}, \mathrm{negl})$. □

We remark that we can also do soundness* → ZK → soundness*, with a slightly more careful argument taking into account the growth in statement size for the increasing function $1 - 1/n^c$.

Corollary 4.2 (Weak Statistically-Sound NIZK Amplification)**.** *Assuming OWFs and an (α, β)-weak statistically-sound NIZK for NP, for some constants α, β such that $\alpha + \beta < 1$, there also exists a statistically-sound NIZK for NP.*

Proof. Recall that α-statistical-soundness is equivalent to α-statistical-soundness*, since a computationally unbounded adversary can always pick the false statement with an accepting proof out of several candidates. Then, we can apply the above weak* NIZK amplifier (Theorem 4.1). □

4.1 Soundness to Soundness* via Public-Key Encryption

In this section, we construct weak* NIZK from weak NIZK and public-key encryption.

Theorem 4.3. *Consider the protocol Π_{wk^*} described in Fig. 4.1, where Π_{wk} is an $(\varepsilon_s(n), \varepsilon_z(n))$-weak NIZK for NP, and PKE is a public-key encryption scheme. Then, Π_{wk^*} is an $(\varepsilon_s(n') + \mathrm{negl}, \varepsilon_z(n') + \mathrm{negl})$-weak* NIZK for NP, for some fixed polynomial $n' = \mathsf{poly}(n)$.*

We note that up to this point, all our constructions preserve a random CRS. That is, if the underlying weak NIZK has a random CRS, the amplification protocol preserves this property. This is no longer necessarily true for the transformation given by the above theorem.

Proof. Consider NP relation $\mathcal{R}$ with $m(n)$-sized witnesses. Throughout the proof, we denote by $x(pk, c)$ the statement

$$\exists(w, r) : \; c = \mathsf{PKE}.E(pk, w; r) \wedge (x, w) \in \mathcal{R}.$$

The correctness of Π_{wk^*} follows readily from the correctness of Π_{wk}, so let us focus on proving soundness* and ZK.

Given a weak NIZK Π_{wk}, and an almost-all-keys perfectly-correct public-key encryption scheme PKE, we construct a weak* NIZK Π_{wk^*} as follows:

$\underline{crs_{wk^*} \leftarrow \mathsf{Gen}_{wk^*}(1^n)}$

1. Sample $(pk, sk) \leftarrow \mathsf{PKE}.G(1^n)$.
2. Sample $crs_{wk} \leftarrow \mathsf{Gen}_{wk}(1^{n'})$, for some fixed polynomial $n' = \mathsf{poly}(n)$ corresponding to the size of the statement below.
3. Output $crs_{wk^*} := (pk, crs_{wk})$ as the CRS.

$\underline{\pi_{wk^*} \leftarrow \mathsf{P}_{wk^*}(crs_{wk^*}, x, w)}$

1. Encrypt $c_w \leftarrow \mathsf{PKE}.E(pk, w; r)$.
2. Generate a weak proof that a valid witness was encrypted, namely

$$\pi_{wk} \leftarrow \mathsf{P}_{wk}\left(crs_{wk},\ \exists(w, r) :\ c_w = \mathsf{PKE}.E(pk, w; r) \wedge (x, w) \in \mathcal{R},\ (w, r)\right) .$$

3. Output $\pi_{wk^*} := (c_w, \pi_{wk})$ as the proof.

$\underline{b_{wk^*} := \mathsf{V}_{wk^*}(crs_{wk^*}, x, \pi_{wk^*})}$

1. Parse $crs_{wk^*} := (pk, crs_{wk})$ and $\pi_{wk^*} := (c, \pi_{wk})$.
2. Return $\mathsf{V}_{wk}(crs_{wk},\ \exists(w, r) :\ c = \mathsf{PKE}.E(pk, w; r) \wedge (x, w) \in \mathcal{R},\ \pi_{wk})$.

$\underline{(crs_{wk^*}, \pi_{wk^*}) \leftarrow \mathsf{Sim}_{wk^*}(x)}$

1. Sample $(pk, sk) \leftarrow \mathsf{PKE}.G(1^n)$.
2. Encrypt $c_0 \leftarrow \mathsf{PKE}.E(pk, 0^{m(n)}; r)$, where $m(n)$ is the size of the witness.
3. Simulate $(crs_{wk}, \pi_{wk}) \leftarrow \mathsf{Sim}_{wk}(\exists(w, r) :\ c_0 = \mathsf{PKE}.E(pk, w; r) \wedge (x, w) \in \mathcal{R})$.
4. Set $crs_{wk^*} := (pk, crs_{wk})$ as the CRS.
5. Set $\pi_{wk^*} := (c_0, \pi_{wk})$ as the proof.
6. Output (crs_{wk^*}, π_{wk^*}).

Figure 4.1: Soundness* From Soundness and PKE

Soundness*. Let $A_{wk^*}(crs_{wk^*})$ be an adversary breaking the soundness* of Π_{wk^*} with probability at least $\varepsilon = \varepsilon(n)$. We construct an adversary $A_{wk}(crs_{wk})$ breaking the (adaptive) soundness of Π_{wk} with probability $\varepsilon - \text{negl}$ as follows: given $crs_{wk} \leftarrow \mathsf{Gen}_{wk}(1^{n'})$ we sample $(pk, sk) \leftarrow \mathsf{PKE}.G(1^n)$ and set $crs_{wk^*} := (pk, crs_{wk})$. Then, we run $(x_1, \pi_1), \ldots, (x_p, \pi_p) \leftarrow A_{wk^*}(crs_{wk^*})$ where each proof π_ℓ is of the form $(c_\ell, \pi_{wk,\ell})$. We return $(x_\ell(pk, c_\ell), \pi_{wk,\ell})$ for some $\ell \in [p]$ such that π_ℓ is accepting, but $w_\ell := \mathsf{PKE}.D(sk, c_\ell)$ is not a valid witness for x_ℓ. If no such ℓ exists, we return $\perp$.

With probability at least $\varepsilon - \text{negl}$, soundness* is broken while (pk, sk) is perfectly-correct. Since soundness* is broken, there exists a false instance with an accepting proof, and in particular no witness is valid for it, so we can efficiently find an appropriate ℓ as described above, by decrypting every ciphertext and testing whether it contains a valid witness. Because of perfect-correctness, there cannot exist (w', r') with $w' \neq w_\ell$ such that $c_\ell = \mathsf{PKE}.E(pk, w'; r')$. Hence, the statement $x_\ell(pk, c_\ell)$ must be a false statement, and we know that $\pi_{wk,\ell}$ is an accepting proof for it w.r.t. crs_{wk}, since π_ℓ is accepting.

Zero Knowledge. For any given $(x, w) \in \mathcal{R}_n$, the security of PKE implies that $pk, c_w \overset{c}{\approx} pk, c_0$, where

$$(pk, sk) \leftarrow \mathsf{PKE}.G(1^n),\ c_w \leftarrow \mathsf{PKE}.E(pk, w; r),\ c_0 \leftarrow \mathsf{PKE}.E(pk, 0^{m(n)}; r)\,.$$

Then, we have that

$$\mathsf{Sim}_{wk^*}(x) = pk, c_0, \mathsf{Sim}_{wk}(x(pk, c_0)) \overset{c}{\approx} pk, c_w, \mathsf{Sim}_{wk}(x(pk, c_w))\,,$$

where $(pk, sk) \leftarrow \mathsf{PKE}.G(1^n)$ and r is chosen at random, as we applied an efficient transformation over the encryption. From the ZK of Π_{wk}, for all (pk, r), and in particular over $(pk, sk) \leftarrow \mathsf{PKE}.G(1^n)$ and a random r, we have that

$$pk, c_w, \mathsf{Sim}_{wk}(x(pk, c_w)) \overset{c}{\approx}_{\varepsilon_z} pk, c_w, crs_{wk}, \pi_{wk}\,,$$

where $crs_{wk} \leftarrow \mathsf{Gen}_{wk}(1^{n'})$ and $\pi_{wk} \leftarrow \mathsf{P}_{wk}\left(crs_{wk},\ x(pk, c_w),\ (w, r)\right)$. The latter is equivalent to an honest execution of Π_{wk^*} over (x, w). □

Corollary 4.4 (Weak NIZK Amplification)**.** *Assuming PKE and an (α, β)-weak NIZK for NP, for some constants α, β such that $\alpha + \beta < 1$, there also exists standard NIZK for NP.*

Proof. Follows by combining Theorem 4.3 and Theorem 4.1. □

5 Open Questions

In this section, we discuss a few open directions.

Noticeable Gap. Ideally, we would like to amplify for any $\alpha(n) + \beta(n) < 1 - 1/\mathsf{poly}(n)$, for example a $(1/2, 1/2 - 1/n)$-weak NIZK. For comparison, in the setting of weak-OT amplification, we can indeed amplify whenever the gap from 1 is noticeable - this is done by alternating between the amplifiers for a logarithmic number of times. However, in our setting the ZK amplifier increases the size of the statement polynomially even when k is constant, so this would result an extremely inefficient protocol.

Still, there are some non-constant parameters we can capture. For example, given a $(1 - 2/n, 1/n)$-weak NIZK, we can apply the soundness amplifier with $k = n$ to get roughly $(1/e^2, 1 - 1/e)$, then use the constants amplifier. A similar in spirit argument could also be made for the ZK amplifier, although we have to be more careful because of the statement size growth.

Non-adaptive Soundness. Our ZK amplifier requires an adaptively-sound base protocol. It is an interesting open question how to amplify ZK when only non-adaptive soundness is guaranteed.

Removing PKE. Even though we were able to remove the assumption of public-key encryption in some settings, it remains open whether we can amplify the general case of (α, β)-weak NIZK without this assumption.

Although, we do mention that even without PKE we can still show adaptive-soundness degradation of $k \cdot \varepsilon$ in the ZK amplifier, and non-adaptive soundness amplification in the parallel repetition soundness amplifier. Hence, we can amplify $(1/\log n, 1/\log n)$-weak NIZK to fully secure non-adaptively-sound NIZK without PKE, by first using the ZK amplifier with $k = \log n/2$, then the soundness amplifier with $k = n$.

Acknowledgements

This project was supported in part by the European Research Council (ERC) under the European Union's Horizon Europe research and innovation programme (grant agreement No. 101042417, acronym SPP).

References

[BFM88] Manuel Blum, Paul Feldman, and Silvio Micali. Non-interactive zero-knowledge and its applications (extended abstract). In *20th ACM STOC*, pages 103–112. ACM Press, May 1988. 3

[BKP+24] Nir Bitansky, Chethan Kamath, Omer Paneth, Ron Rothblum, and Prashant Nalini Vasudevan. Batch proofs are statistically hiding. In *56th ACM STOC*. ACM Press, June 2024. 4

[CK90] Claude Crépeau and Joe Kilian. Weakening security assumptions and oblivious transfer (abstract). In Shafi Goldwasser, editor, *CRYPTO'88*, volume 403 of *LNCS*, pages 2–7. Springer, Heidelberg, August 1990. 3

[DKS99] Ivan Damgård, Joe Kilian, and Louis Salvail. On the (im)possibility of basing oblivious transfer and bit commitment on weakened security assumptions. In Jacques Stern, editor, *EUROCRYPT'99*, volume 1592 of *LNCS*, pages 56–73. Springer, Heidelberg, May 1999. 3

[DN00] Cynthia Dwork and Moni Naor. Zaps and their applications. In *41st FOCS*, pages 283–293. IEEE Computer Society Press, November 2000. 11

[DNR04] Cynthia Dwork, Moni Naor, and Omer Reingold. Immunizing encryption schemes from decryption errors. In Christian Cachin and Jan Camenisch, editors, *EUROCRYPT 2004*, volume 3027 of *LNCS*, pages 342–360. Springer, Heidelberg, May 2004. 3, 9

[DORS08] Yevgeniy Dodis, Rafail Ostrovsky, Leonid Reyzin, and Adam D. Smith. Fuzzy extractors: How to generate strong keys from biometrics and other noisy data. *SIAM J. Comput.*, 38(1):97–139, 2008. 6, 8, 12

[DP92] Alfredo De Santis and Giuseppe Persiano. Zero-knowledge proofs of knowledge without interaction (extended abstract). In *33rd FOCS*, pages 427–436. IEEE Computer Society Press, October 1992. 8

[FLS90] Uriel Feige, Dror Lapidot, and Adi Shamir. Multiple non-interactive zero knowledge proofs based on a single random string (extended abstract). In *31st FOCS*, pages 308–317. IEEE Computer Society Press, October 1990. 11

[FLS99] Uriel Feige, Dror Lapidot, and Adi Shamir. Multiple noninteractive zero knowledge proofs under general assumptions. *SIAM J. Comput.*, 29(1):1–28, 1999. 4, 11

[Gei22] Nathan Geier. A tight computational indistinguishability bound for product distributions. In Eike Kiltz and Vinod Vaikuntanathan, editors, *TCC 2022, Part II*, volume 13748 of *LNCS*, pages 333–347. Springer, Heidelberg, November 2022. 7, 13

[GJS19] Vipul Goyal, Aayush Jain, and Amit Sahai. Simultaneous amplification: The case of non-interactive zero-knowledge. In Alexandra Boldyreva and Daniele Micciancio, editors, *CRYPTO 2019, Part II*, volume 11693 of *LNCS*, pages 608–637. Springer, Heidelberg, August 2019. 3, 4, 8, 13

[GNW95] Oded Goldreich, Noam Nisan, and Avi Wigderson. On yao's xor-lemma. *Electron. Colloquium Comput. Complex.*, TR95-050, 1995. 5

[Hai09] Iftach Haitner. A parallel repetition theorem for any interactive argument. In *50th FOCS*, pages 241–250. IEEE Computer Society Press, October 2009. 3

[Her05] Amir Herzberg. On tolerant cryptographic constructions. In Alfred Menezes, editor, *CT-RSA 2005*, volume 3376 of *LNCS*, pages 172–190. Springer, Heidelberg, February 2005. 3

[HKN+05] Danny Harnik, Joe Kilian, Moni Naor, Omer Reingold, and Alon Rosen. On robust combiners for oblivious transfer and other primitives. In Ronald Cramer, editor, *EUROCRYPT 2005*, volume 3494 of *LNCS*, pages 96–113. Springer, Heidelberg, May 2005. 3

[Hol05] Thomas Holenstein. Key agreement from weak bit agreement. In Harold N. Gabow and Ronald Fagin, editors, *37th ACM STOC*, pages 664–673. ACM Press, May 2005. 3, 23

[HR08] Shai Halevi and Tal Rabin. Degradation and amplification of computational hardness. In Ran Canetti, editor, *TCC 2008*, volume 4948 of *LNCS*, pages 626–643. Springer, Heidelberg, March 2008. 3, 7, 13

[IKOS07] Yuval Ishai, Eyal Kushilevitz, Rafail Ostrovsky, and Amit Sahai. Zero-knowledge from secure multiparty computation. In David S. Johnson and Uriel Feige, editors, *39th ACM STOC*, pages 21–30. ACM Press, June 2007. 4

[Jai24] 2024. Aayush Jain, personal communication. 3

[KMY20] Fuyuki Kitagawa, Takahiro Matsuda, and Takashi Yamakawa. NIZK from SNARG. In Rafael Pass and Krzysztof Pietrzak, editors, *TCC 2020, Part I*, volume 12550 of *LNCS*, pages 567–595. Springer, Heidelberg, November 2020. 4, 11

[LT13] Huijia Lin and Stefano Tessaro. Amplification of chosen-ciphertext security. In Thomas Johansson and Phong Q. Nguyen, editors, *EUROCRYPT 2013*, volume 7881 of *LNCS*, pages 503–519. Springer, Heidelberg, May 2013. 3

[MT10] Ueli M. Maurer and Stefano Tessaro. A hardcore lemma for computational indistinguishability: Security amplification for arbitrarily weak PRGs with optimal stretch. In Daniele Micciancio, editor, *TCC 2010*, volume 5978 of *LNCS*, pages 237–254. Springer, Heidelberg, February 2010. 5, 12, 13, 31, 32

[OW93] Rafail Ostrovsky and Avi Wigderson. One-way fuctions are essential for non-trivial zero-knowledge. In *Second Israel Symposium on Theory of Computing Systems, ISTCS 1993, Natanya, Israel, June 7-9, 1993, Proceedings*, pages 3–17. IEEE Computer Society, 1993. 9

[Ps05] Rafael Pass and Abhi shelat. Unconditional characterizations of non-interactive zero-knowledge. In Victor Shoup, editor, *CRYPTO 2005*, volume 3621 of *LNCS*, pages 118–134. Springer, Heidelberg, August 2005. 9

[QRW19] Willy Quach, Ron D. Rothblum, and Daniel Wichs. Reusable designated-verifier NIZKs for all NP from CDH. In Yuval Ishai and Vincent Rijmen, editors, *EUROCRYPT 2019, Part II*, volume 11477 of *LNCS*, pages 593–621. Springer, Heidelberg, May 2019. 4

[Wul07] Jürg Wullschleger. Oblivious-transfer amplification. In Moni Naor, editor, *EUROCRYPT 2007*, volume 4515 of *LNCS*, pages 555–572. Springer, Heidelberg, May 2007. 3

[Yao82] Andrew Chi-Chih Yao. Theory and applications of trapdoor functions (extended abstract). In *23rd FOCS*, pages 80–91. IEEE Computer Society Press, November 1982. 3, 5, 7, 13

A Appendix

A.1 Hybrid Indistinguishability Lemma

Below, we give a proof sketch for the hybrid indistinguishability lemma (Lemma 2.15), from the original statement of the indistinguishability hard-core lemma, as it appears in [MT10].

First, some definitions: A measure $\mathcal{M}$ over a set $\mathcal{X}$ is a function $\mathcal{M} : \mathcal{X} \to [0, 1]$, and its density is $\mu(\mathcal{M}) := \mathbb{E}_{x \leftarrow \mathcal{X}}[\mathcal{M}(x)]$. We denote by $\mathsf{P}_{\mathcal{M}}$ sampling according to $\mathcal{M}$, that is, $\mathsf{P}_{\mathcal{M}}(x) = \frac{\mathcal{M}(x)}{\mu(\mathcal{M}) \cdot |\mathcal{X}|}$.

Lemma A.1 (Indistinguishability Hard-Core Lemma [MT10])**.** *Let $E : \mathcal{U} \to \mathcal{X}$ and $F : \mathcal{V} \to \mathcal{X}$ be functions, and let $\delta, \varepsilon \in (0,1)$ and $s \in \mathbb{N}$ be given. Assume that*

$$E(U) \overset{c}{\approx}_{\delta} F(V)\,,$$

against s-sized circuits, where $U \leftarrow \mathcal{U}$, $V \leftarrow \mathcal{V}$. Then there exist measures $\mathcal{M}$ on $\mathcal{U}$ and $\mathcal{N}$ on $\mathcal{V}$, each with density at least $1-\delta$, such that

$$E(U') \overset{c}{\approx}_{\varepsilon} F(V')\,,$$

against s'-sized circuits, where $U' \leftarrow \mathsf{P}_{\mathcal{M}}$, $V' \leftarrow \mathsf{P}_{\mathcal{N}}$, and $s' := \frac{s \cdot \varepsilon^2}{128(\log|\mathcal{U}|+\log|\mathcal{V}|+1)}$.

Proof Sketch of Lemma 2.15. Given a distribution X we let $N = |\mathrm{Im}(X)|$ and $n = \log N$, then we can consider its probability vector $(p_1, \ldots, p_N)$ and round each probability down to an $n + \log(1/\varepsilon^*)$-bits representation. The statistical distance from X when sampling according to the new probabilities (after normalizing) is at most ε^*. Now we can define an inefficient function $E : \{0,1\}^{n+\log(1/\varepsilon^*)} \to \mathrm{Im}(X)$ using inverse transform sampling on the rounded probabilities, such that $E(U^{n+\log(1/\varepsilon^*)}) \overset{s}{\approx}_{\varepsilon^*} X$, and do the same for $F(U^{m+\log(1/\varepsilon^*)}) \overset{s}{\approx}_{\varepsilon^*} Y$. We have that

$$E(U^{n+\log(1/\varepsilon^*)}) \overset{c}{\approx}_{\delta+2\varepsilon^*} F(U^{m+\log(1/\varepsilon^*)})\,,$$

against s-sized circuits, and we can use the indistinguishability hard-core lemma with ε' to conclude that there exist measures $\mathcal{N}$ on $\{0,1\}^{\log|\mathrm{Im}(X)|+\log(1/\varepsilon^*)}$ and $\mathcal{M}$ on $\{0,1\}^{\log|\mathrm{Im}(Y)|+\log(1/\varepsilon^*)}$, each with density at least $1-\delta-2\varepsilon^*$, such that

$$E(\mathsf{P}_{\mathcal{N}}) \overset{c}{\approx}_{\varepsilon'} F(\mathsf{P}_{\mathcal{M}})\,,$$

against s'-sized circuits where $s' := \frac{s \cdot \varepsilon'^2}{128(\log|\mathrm{Im}(X)|+\log(1/\varepsilon^*)+\log|\mathrm{Im}(X)|+\log(1/\varepsilon^*)+1)}$. We can assume w.l.o.g. the measures have density exactly $1-\delta-2\varepsilon^*$, because we can always scale them down. Note that the following are distributionally equivalent

$$\begin{aligned} U^{n+\log(1/\varepsilon^*)} &\equiv (1-\delta-2\varepsilon^*) \cdot \mathsf{P}_{\mathcal{N}} + (\delta+2\varepsilon^*) \cdot \mathsf{P}_{\overline{\mathcal{N}}}\,, \\ U^{m+\log(1/\varepsilon^*)} &\equiv (1-\delta-2\varepsilon^*) \cdot \mathsf{P}_{\mathcal{M}} + (\delta+2\varepsilon^*) \cdot \mathsf{P}_{\overline{\mathcal{M}}}\,. \end{aligned}$$

Now we can define the hybrid distribution X' using

$$X' := (1-\delta-2\varepsilon^*) \cdot F(\mathsf{P}_{\mathcal{M}}) + 3\varepsilon^* \cdot F(\mathsf{P}_{\overline{\mathcal{M}}}) + (\delta-\varepsilon^*) \cdot E(\mathsf{P}_{\overline{\mathcal{N}}})\,,$$

then we have that the computational distance between X' and $E(U^{n+\log(1/\varepsilon^*)})$ is at most

$$(1-\delta-2\varepsilon^*) \cdot \varepsilon' + 3\varepsilon^* \cdot 1 + (\delta-\varepsilon^*) \cdot 0 \leq (1-\delta) \cdot \varepsilon' + 3\varepsilon^*\,,$$

and the statistical distance from $F(U^{m+\log(1/\varepsilon^*)})$ is at most $(\delta - \varepsilon^*)$. Finally, we switch back to the original distributions to conclude that

$$X \overset{c}{\approx}_{(1-\delta)\cdot\varepsilon'+4\varepsilon^*} X' \overset{s}{\approx}_{\delta} Y\,,$$

and by setting $\varepsilon' = \varepsilon(1-\varepsilon)/(1-\delta)$ with $\varepsilon^* = \varepsilon^2/4$, we get what we wanted w.r.t. ε. □

A.2 Useful Facts

Here, we prove the useful facts appearing in Section 3.2.1, for the proof of the statistical extraction lemma (Lemma 3.4).

Proof of Lemma 3.3. Below are the proofs for the useful facts

1. Let X_1, Z_1 be a pair of jointly distributed random variables and X_2, Z_2 be an independent pair. Then

$$\widetilde{H}_\infty(X_1, X_2 \mid Z_1, Z_2) = -\log\left(\mathop{\mathbb{E}}_{z_1,z_2\leftarrow Z_1,Z_2}\left[\max_{x_1,x_2}\Pr\left[X_1, X_2 = x_1, x_2 \mid Z_1, Z_2 = z_1, z_2\right]\right]\right) =$$
$$-\log\left(\mathop{\mathbb{E}}_{z_1,z_2\leftarrow Z_1,Z_2}\left[\max_{x_1,x_2}\Pr\left[X_1 = x_1 \mid Z_1 = z_1\right]\Pr\left[X_2 = x_2 \mid Z_2 = z_2\right]\right]\right) =$$
$$-\log\left(\mathop{\mathbb{E}}_{z_1\leftarrow Z_1}\left[\mathop{\mathbb{E}}_{z_2\leftarrow Z_2}\left[\max_{x_1}\Pr\left[X_1 = x_1 \mid Z_1 = z_1\right]\max_{x_2}\Pr\left[X_2 = x_2 \mid Z_2 = z_2\right]\right]\right]\right) =$$
$$-\log\left(\mathop{\mathbb{E}}_{z_1\leftarrow Z_1}\left[\max_{x_1}\Pr\left[X_1 = x_1 \mid Z_1 = z_1\right]\right]\mathop{\mathbb{E}}_{z_2\leftarrow Z_2}\left[\max_{x_2}\Pr\left[X_2 = x_2 \mid Z_2 = z_2\right]\right]\right) =$$
$$-\log\left(\mathop{\mathbb{E}}_{z_1\leftarrow Z_1}\left[\max_{x_1}\Pr\left[X_1 = x_1 \mid Z_1 = z_1\right]\right]\right) - \log\left(\mathop{\mathbb{E}}_{z_2\leftarrow Z_2}\left[\max_{x_2}\Pr\left[X_2 = x_2 \mid Z_2 = z_2\right]\right]\right) =$$
$$\widetilde{H}_\infty(X_1 \mid Z_1) + \widetilde{H}_\infty(X_2 \mid Z_2)\,.$$

2. Let X, Z be a pair of jointly distributed random variables and A be an event. We have that

$$\mathop{\mathbb{E}}_{z\leftarrow Z|A}\left[\max_x \Pr\left[X = x \mid Z = z, A\right]\right] = \sum_z \Pr\left[Z = z \mid A\right]\max_x \Pr\left[X = x \mid Z = z, A\right] =$$
$$\sum_z \max_x \Pr\left[X = x, Z = z \mid A\right] = \frac{1}{\Pr[A]}\sum_z \max_x \Pr\left[X = x, Z = z, A\right] \le$$
$$\frac{1}{\Pr[A]}\sum_z \max_x \Pr\left[X = x, Z = z\right] = \frac{1}{\Pr[A]}\mathop{\mathbb{E}}_{z\leftarrow Z}\left[\max_x \Pr\left[X = x \mid Z = z\right]\right]\,.$$

   By applying the decreasing function $-\log(\cdot)$ on both sides, we conclude that

$$\widetilde{H}_\infty(X \mid Z, A) \ge \widetilde{H}_\infty(X \mid Z) - \log(1/\Pr[A])\,.$$

3. Let $Y := (Z, X_1, \ldots, X_t)$ be a random variable and denote by Y^{-j} the r.v. with X_j omitted. Further, let $f(\cdot\,; r)$ be a randomized function. Assume that for all $j \in [t]$ we have

$$Y^{-j}, R, f(X_j; R) \overset{\text{s}}{\approx}_\delta Y^{-j}, R, U\,,$$

   then we also have

$$Z, R, f(X_1; R), \ldots, f(X_j; R), U^{t-j} \overset{\text{s}}{\approx}_\delta Z, R, f(X_1; R), \ldots, f(X_{j-1}; R), U^{t-j+1}\,,$$

   as we applied some function on both sides. Using a hybrid argument over $j \in [t]$ we conclude that

$$Z, R, f(X_1; R), \ldots, f(X_t; R) \overset{\text{s}}{\approx}_{t\cdot\delta} Z, R, U^t\,.$$

4. Let X, Y be a pair of jointly distributed random variables and $\{A_i\}_{i\in[k]}$ be a set of events, then

$$2 \cdot \mathsf{SD}(X,Y) = \sum_z |\Pr[X=z] - \Pr[Y=z]| =$$

$$\sum_z \left| \sum_{\substack{I\subseteq[k] \\ \Pr\left[A_I, \overline{A}_{\overline{I}}\right]>0}} \Pr\left[A_I, \overline{A}_{\overline{I}}\right] \left(\Pr[X=z \mid A_I, \overline{A}_{\overline{I}}] - \Pr[Y=z \mid A_I, \overline{A}_{\overline{I}}]\right) \right| \leq$$

$$\sum_z \sum_{\substack{I\subseteq[k] \\ \Pr\left[A_I, \overline{A}_{\overline{I}}\right]>0}} \Pr\left[A_I, \overline{A}_{\overline{I}}\right] \left|\Pr[X=z \mid A_I, \overline{A}_{\overline{I}}] - \Pr[Y=z \mid A_I, \overline{A}_{\overline{I}}]\right| =$$

$$\sum_{\substack{I\subseteq[k] \\ \Pr\left[A_I, \overline{A}_{\overline{I}}\right]>0}} \Pr\left[A_I, \overline{A}_{\overline{I}}\right] \sum_z \left|\Pr[X=z \mid A_I, \overline{A}_{\overline{I}}] - \Pr[Y=z \mid A_I, \overline{A}_{\overline{I}}]\right| =$$

$$2 \sum_{\substack{I\subseteq[k] \\ \Pr\left[A_I, \overline{A}_{\overline{I}}\right]>0}} \Pr\left[A_I, \overline{A}_{\overline{I}}\right] \mathsf{SD}(X, Y \mid A_I, \overline{A}_{\overline{I}}) \,.$$

□