Does Witness Indistinguishability Require Cryptography?

Gal Arnon[1], Noam Mazor[2], Rafael Pass[3], and Jad Silbak[4]

[1]NTT Research*, galarnon42@gmail.com
[2]New York University, noammaz@gmail.com
[3]Technion, Cornell Tech, and Tel Aviv University, rafael@cs.cornell.edu
[4]MIT, jadsilbak@gmail.com

November 19, 2026

Abstract

An interactive proof is witness indistinguishable (WI) if the verifier cannot tell which of several valid witnesses the prover used. WI is a fundamental privacy guarantee that is implied by a number of central primitives, including zero knowledge (ZK), batch arguments, and succinct arguments.

Ostrovsky and Wigderson [OW93] showed that, assuming the worst-case hardness of NP, ZK proofs for NP imply auxiliary-input one-way functions. This was recently strengthened by Hirahara and Nanashima [HN24], who showed that, under the same assumption, ZK proofs for NP imply standard one-way functions. For WI, no analogous result was known; in fact, we do not known if cryptography is necessary for WI at all.

One source of difficulty in proving a cryptographic implication from WI is the possible unconditional existence of witness canonicalization. More formally, an ε-witness canonicalization (ε-WC) is a deterministic polynomial-time map that sends any valid witness w for x, together with public randomness r, to an efficiently verifiable string that depends only on x and r (and not on the specific witness w), except with probability ε. Completeness and soundness hold perfectly (only canonicalization is allowed to fail). Indeed, if a valid witnesses for an NP statement can be efficiently mapped to a single canonical representation, then WI is trivial and requires no cryptography: the prover sends the canonical witness, and so reveals nothing about its own.

We prove the following:

1. Let $\mathcal{L}$ be an NP language that admits no ε-WC for some inverse-polynomial ε. Then, under a derandomization assumption, every constant-round (honest-verifier) WI proof for $\mathcal{L}$ implies infinitely-often auxiliary-input one-way functions. The derandomization assumption can be avoided if one settles for canonicalization with negligible completeness and soundness errors.
2. Combining (a variant of) the above with a theorem of Bitansky, Kamath, Paneth, Rothblum, and Vasudevan [BKPRV24] yields analogous consequences for sufficiently compressing batch arguments and succinct non-interactive arguments.

Since it is unlikely that SAT has a ε-WC, even for constant ε, the above results establish that cryptography is necessary for designing WI proofs (or sufficiently compressing BARG/SNARGs) for SAT. As far as we are aware, these are the first results establishing the necessity of cryptographic hardness for witness indistinguishability, BARGs, and SNARGs.

*The majority of this work was done while the author was a postdoc at Bocconi University.

Contents

1 Introduction **1**
- 1.1 Our Contributions . . . 2
- 1.2 Technical Overview . . . 4

2 Preliminaries **11**
- 2.1 Complexity Classes and NP Relations . . . 12
- 2.2 Auxiliary-Input One-Way Functions . . . 12
- 2.3 Interactive Protocols . . . 13

3 Deterministic-Prover Witness Indistinguishability and Witness Canonicalization **16**
- 3.1 Extensions to Batch Arguments and Succinct Arguments. . . . 18
- 3.2 Next-bit Predictors, Samplers, and Simulators . . . 20
- 3.3 Canonical Simulation . . . 22
- 3.4 Proof of Theorem 3.1 . . . 22
- 3.5 Canonical Simulation: Proving Lemma 3.11 . . . 26
- 3.6 Round Collapse for Constant-Round Deterministic-Prover SWI . . . 28
- 3.7 Error Elimination via Derandomization . . . 29

A Application to Hard Oracles **37**
- A.1 Definitions and Notions for Application to Hard Oracles . . . 37
- A.2 Application to Hard Functions: Theorem Statement . . . 39
- A.3 Proof of Theorem A.9 . . . 41

B Deferred Proofs **44**
- B.1 Proving Lemma 2.4 . . . 44
- B.2 Proving Claim 3.9 . . . 45
- B.3 Proving Lemma 3.10 . . . 46
- B.4 Proving Claim 3.14 . . . 48

1 Introduction

Witness indistinguishability (WI), introduced by Feige and Shamir [FS90], is a basic privacy guarantee for interactive proof systems. Informally, an interactive protocol is WI if no efficient verifier can determine which of two witnesses was used by the prover. More precisely, for every instance x and witnesses w_0 and w_1, the verifier's views in interactions with $P(x, w_0)$ and with $P(x, w_1)$ are indistinguishable.

WI provides a weaker privacy guarantee than zero knowledge (ZK), which requires that the verifier's entire view can be efficiently simulated from x alone [GMR89]. Nevertheless, WI enjoys properties that make it especially useful. Most notably, WI is preserved under arbitrary composition, including parallel composition [FS90]. It also admits low-round realizations: two-message WI protocols (ZAPs) in the plain model [DN07], or single-message WI protocols under cryptographic assumptions [GOS06]. This stands in contrast to ZK, for which two-message protocols for languages outside BPP are impossible [GO94a]. Consequently, WI has become a fundamental component in the construction and analysis of zero-knowledge protocols, non-interactive proof systems, batch arguments, and succinct arguments. Recent work further shows that batch arguments with sufficient compression yields statistical forms of witness indistinguishability [BKPRV24].

Is cryptography necessary for proof systems? A foundational goal in cryptography is to characterize the minimal computational assumptions required for basic primitives. Most basic cryptographic objects (e.g., pseudorandom generators [BM84; HILL99], pseudorandom functions [GGM86], private-key encryption [GM84], digital signatures [NY89; Rom90], commitment schemes [Nao91; HR07], identification protocols [FS86], coin-flipping protocols [Blu83], and more), rely on the existence of one-way functions (OWF) [DH76], which are functions that are easy to compute but hard to invert. Such results can be seen as a lower bound: showing that a primitive implies one-way functions rules out constructing it solely from complexity assumptions that may hold in a world without cryptography.

For zero knowledge, this question is by now well understood. Classical constructions show that one-way functions suffice for ZK proofs for all of NP [GMW91; Nao91; HILL99; FS90]. In the reverse direction, Ostrovsky and Wigderson [Ost91; OW93] show that auxiliary-input one-way functions are necessary for zero-knowledge assuming worst-case hardness of NP.[1] An auxiliary-input one-way function is a function *family* $\{f_a\}$ so that there is no efficient adversary A that inverts *all* functions in the family simultaneously. This result was an important milestone, as it was the first to establish the necessity of any from of cryptographic hardness for the construction of zero-knowledge proofs for NP.

Recently, Hirahara and Nanashima [HN24] extend these results to show that, assuming worst-case hardness of NP, zero-knowledge proofs imply (standard) one-way functions. Thus, for ZK, the picture of cryptographic necessity is complete.

Despite the central role of witness indistinguishability, no analogous characterization was known for WI itself. In particular, prior work did not determine whether a non-trivial WI protocol must entail any form of cryptographic hardness.

Cryptography and WI. One may hope to extend the complexity-theoretic lower bounds for ZK to WI. Indeed, WI can equivalently be viewed as ZK with an unbounded simulator [Pas03; Pas04]. This equivalence, however, also explains why the known approach for ZK does not directly apply. Roughly speaking, the one-way function in the construction of Ostrovsky and Wigderson [OW93] is the ZK simulator itself. Replacing the efficient simulator by an unbounded one therefore yields an inefficient candidate function.

There is also evidence that this difficulty is inherent to the known proof strategy. Baril and Haitner [BH26] give a fully black-box barrier: in their setting, a reduction deriving one-way functions from WI proofs together with an additional hardness assumption that can be formalized as a *game* (such as the worst-case hardness of NP) would already yield a black-box reduction from that additional assumption alone. In particular, their result rules out a direct black-box analogue of the known ZK implication. Thus, understanding the cryptographic complexity of WI appears to require substantially different ideas.

[1] We mention that [OW93] additionally show that, assuming the *average-case* hardness of NP, *standard* one-way functions are necessary for zero-knowledge.

Witness canonicalization. There is a second, more conceptual, reason why an unrestricted implication from WI to one-way functions cannot hold. Consider a relation in which every instance has a unique witness. For such a relation, WI is vacuous: there are no two witnesses whose use must be hidden. More generally, WI is easy to achieve whenever all valid witnesses can be efficiently mapped to the same canonical representation.

This motivates the notion of *witness canonicalization*:

Definition 1.1 (ε-WC, informal)**.** *An ε-witness canonicalization (ε-WC) for a language $\mathcal{L} \in$ NP consists a deterministic polynomial-time mapping algorithm* map *and deterministic polynomial-time NP verifier V,*[2] *both taking in 3 inputs, with the following properties:*

- Completeness. *For every $x \in \mathcal{L}$, valid witness w, and r: $V(r, x, \pi) = 1$ where $\pi = \mathsf{map}(r, x, w)$.*
- Soundness. *For every $x \notin \mathcal{L}$, r, and π: $V(r, x, \pi) = 0$.*
- Canonicalization. *There exists an ensemble $\{\pi_{x,r}\}$ of canonical witnesses such that for any $x \in \mathcal{L}$ and valid witness w,*
$$\Pr_{r \leftarrow \{0,1\}^{\ell(|x|)}} \left[\mathsf{map}(r, x, w) = \pi_{x,r}\right] \geq 1 - \varepsilon.$$

We denote by $\mathcal{WC}[\varepsilon]$ the class of all languages with an ε-WC.

Note that the shared randomness r affects only the choice of canonical witness, and not completeness and soundness, which hold perfectly. Thus, an ε-WC can be seen as a shared-randomness reduction of NP witnesses into a canonical representation that depends only on the instance and the public randomness, and not on the original witness.

It seems unlikely that $\mathsf{NP} \subseteq \mathcal{WC}[\varepsilon]$, *even for constant ε*. Indeed, map would have to coordinate among potentially unrelated witnesses: for any two valid witnesses w_0, w_1 for the same instance x,

$$\Pr_r \left[\mathsf{map}(r, x, w_0) = \mathsf{map}(r, x, w_1)\right] \geq 1 - 2\varepsilon,$$

even though NP-complete languages provide no structure relating two separate witnesses. This intuition is consistent with the classical result of Hemaspaandra, Naik, Ogihara and Selman [HNOS96], who showed that the existence of a nondeterministic polynomial-time procedure selecting a unique satisfying assignment for every satisfiable formula would collapse the polynomial hierarchy.

Witness canonicalization captures a natural barrier to deriving cryptographic hardness from WI. Indeed, any relation with an ε-WC (for a negligible ε) has an immediate (non-interactive) WI protocol:[3] the prover samples and sends r, together with $\pi = \mathsf{map}(r, x, w)$, and the verifier checks that $V(r, x, \pi) = 1$. The protocol hides the witness, not through cryptography, but because the witness has been canonicalized.

1.1 Our Contributions

Our main technical contribution is to prove that if $\mathsf{NP} \not\subseteq \mathcal{WC}[\varepsilon]$ for inverse-polynomial ε then (under a derandomization assumption) witness-indistinguishability implies auxiliary-input one-way functions. Specifically, we make use of the following derandomization assumption by Liu and Pass [LP25] following the result of Chen and Tell [CT21]: There exists a constant $\gamma > 0$ such that, for every $k \in \mathbb{N}$,

$$\mathsf{E} \not\subseteq \mathsf{ioNTIME}[2^{kn}]/2^{\gamma n}. \tag{1}$$

where $\mathsf{E} = \mathsf{DTIME}[2^{O(n)}]$.

As far as we know, the following theorem is the first result to establish, under any assumption, the necessity of cryptographic hardness for witness indistinguishability:

Theorem 1.2 (informal, Theorem 3.25)**.** *Let $\mathcal{L} \in \mathsf{NP} \setminus \mathcal{WC}[\varepsilon]$ be a language for $\varepsilon \in 1/\mathrm{poly}$ and assume Eq. (1) holds. If there is a constant-round honest-verifier WI proof for $\mathcal{L}$ then infinitely-often auxiliary-input one-way functions exist.*

[2] In particular, V is an NP verifier for a relation for $\mathcal{L}$ whose witnesses are tuples (r, π).

[3] Note that if the WC has a soundness error, we only get honest-verifier WI, in which the honest-verifier chooses r.

Equivalently, in a world with no cryptography, any (constant-round) witness-indistinguishable proof yields a procedure to canonicalize NP witnesses. We further remark that Theorem 1.2 can be altered in the following ways,

- The assumption that Eq. (1) holds can be replaced with the assumption that $\mathcal{L} \in \mathsf{NP} \setminus \mathcal{WC}_{\mathsf{negl}}[\varepsilon]$, where $\mathcal{WC}_{\mathsf{negl}}[\varepsilon]$ is the class of languages with an ε-WC for which completeness and soundness hold except with negligible probability rather than perfectly.
- Theorem 1.2 can be extended to show that computationally-sound WI implies infiniteley-often auxiliary-input one-way functions under the assumption that $\mathcal{L} \notin \mathsf{Arg}\mathcal{WC}[\varepsilon]$ where $\mathsf{Arg}\mathcal{WC}[\varepsilon]$, is the set of all languages with a ε-WC where soundness holds against computationally-bounded provers (and both completeness and computational soundness hold except with negligible probability). As in the previous item, we do not need to assume that Eq. (1) holds.

Implications to BARGs. Bitansky, Kamath, Paneth, Rothblum, and Vasudevan [BKPRV24] show that batch arguments (BARGs) and succinct non-interactive arguments (SNARGs) imply non-uniform (honest-verifier) statistical WI. Combining their result with (a variant of) Theorem 1.2 gives a corresponding consequence for batch and succinct arguments, indicating that the construction of BARGs and SNARGs necessitates cryptography.

Theorem 1.3 (informal, Corollary 3.5)**.** *Assume there is no non-uniform ε-WC with computational soundness for SAT for some $\varepsilon \in 1/\mathrm{poly}$. If there is a constant-round batch-argument for SAT with small enough prover-to-verifier communication,*[4] *then there exists an infinitely-often auxiliary-input one-way function.*

As succinct arguments almost trivially imply BARGs, Theorem 1.3 applies to succinct arguments as well. Thus, under the same non-canonicalizability premise, sufficiently strong batch compression or succinctness entails cryptographic hardness.

A round-collapse theorem for WI. Our results are proven by combining a *canonical simulation* procedure with a new round-collapse theorem for honest-verifier WI in the abscence of one-way functions. We believe this round reduction theorem to be of independent interest:

Corollary 1.4. *Assume there is a constant-round WI argument for a language $\mathcal{L}$. Either infinitely-often auxiliary-input one-way functions exist or there is a* two-round *honest-verifier statistical WI argument for $\mathcal{L}$ with inverse-polynomial WI error.*

Previsouly, Babai and Moran [BM88] gave the classical round collapse for information-theoretically sound public-coin protocols. Pass and Venkitasubramaniam [PV20] developed an efficient analogue for computationally sound protocols in the absence of one-way functions, but their transformation does not preserve WI for a general randomized prover.

Open questions. Our work fills in a gap in our understanding of WI proof systems that has been open since their inception: cryptographic hardness, and not simply complexity assumptions, is (likely to be) necessary for the construction of WI proofs. However, we leave open a few interesting questions to be tackled in future work:

- *Auxiliary-input one-way functions.* The hardness implied by Theorem 1.2 that of "auxiliary-input" one-way functions (AI-OWFs), rather than the more standard notion of one-way functions. AI-OWFs functions are natural to consider when working with proof systems, with the auxiliary-input being some hard instance for the language in question. Indeed, the OWF of Ostrovsky and Wigderson [OW93] is (roughly) the simulator $S(x; r)$ with x being a fixed worst-case hard instance. Similarly, our one-way function is (roughly) the algorithm that, given x and w generates a random interaction transcript (here x and w are the auxiliary inputs). Given the recent breakthrough of Hirahara and Nananshima [HN24], who upgraded [OW93] to achieve full one-way functions, one could ask whether their techniques could be used to achieve a parallel strengthening

[4] More precisely, there is a universal constant $c > 1$ such that the implication holds if there is a k-round batch argument for SAT, for a constant k such that, when proving the satisfiablility of t instances, the prover-to-verifier communication is bounded by $(t\varepsilon/k)^{1/ck}$.

for WI. Unfortunately, their techniques do not seem to apply to our case. As such, we leave the gap between auxiliary-input one-way functions and standard one-way functions open.

- *Constant rounds.* WI proofs are a primitive primarily used for their extremely small round complexity. Indeed, for over 3 rounds one could usually simply use a ZK proof with its upgraded security notion instead. Thus, our result captures the most interesting case for WI proofs. We remark that improving our techniques to achieve a full characterization of WI proofs seems especially challenging as transforming a many-round WI protocol into WC seems to require a significantly stronger round-collapse theorem than is currently known even for non-WI protocols.

1.2 Technical Overview

In this section, we discuss the main ideas used to derive Theorem 1.2. For clarity and simplicity, in this overview we will be imprecise about when properties hold in the standard way compared to holding "infinitely-often".

Our goal is to show that in the absence of auxiliary-input one-way functions (AI-OWFs), constant-round WI implies WC.

First note that by the recent result of Chakraborty, Hulett, Khurana, and Tomer [CHKT26], if auxiliary-input one-way functions don't exist then any private-coin WI protocol can be transformation into a public-coin WI protocol. Thus, without loss of generality, we can assume that the protocol is public-coin.

Our proof proceeds by successively applying the following transformations:

1. **Deterministic-prover WI.** We show that, if auxiliary-input one-way functions don't exist, then any WI protocol can be turned into a *deterministic-prover* statistical WI with inverse-polynomial error. We sketch this transformation and its proof in Section 1.2.1.

2. **Round collapse.** We next show that the Babai–Moran [BM88]/Pass–Venkitasubramaniam [PV20] round-collapse preserves honest-verifier statistical WI when the prover is deterministic. This yields a two-message public-coin protocol with a deterministic prover. Such a protocol directly defines a witness canonicalization scheme (with a small soundness and completeness error) since the verifier's message can be viewed as the public randomness, and statistical WI (with a deterministic prover) implies that the prover's responses under any two valid witnesses agree with high probability. We explain this step in Section 1.2.2.

3. **Derandomization.** Finally, in the statistically sound setting and under a derandomization assumption, we eliminate the remaining completeness and soundness errors while preserving inverse-polynomial uniqueness. This yields a statistically sound witness canonicalization scheme with no completeness or soundness errors. We explain this step in Section 1.2.3.

In the following sections, we give further details about each transformation, thereby sketching the proof of Theorem 1.2.

1.2.1 Deterministic-prover WI

We describe our deterministic-prover transformation, proving the following theorem:

Theorem 1.5 (informal, Theorem 3.1)**.** *Assume that auxiliary-input one-way functions do not exists, and let $\mathcal{L}$ be a language with a k-round honest-verifier WI proof. Then $\mathcal{L}$ has a k-round* deterministic-prover *honest-verifier statistical WI proof with inverse-polynomial WI error.*

That is, our goal is to show that every WI interactive proof (P, V) for a relation $\mathcal{R}$ either already yields a distributional auxiliary-input one-way function,[5] or can be compiled into an equivalent SWI interactive proof with a *deterministic* prover.

From WI to SWI for the transcript function. We start by considering the transcript function defined as follows: For auxiliary-input $a = (x, w_0, w_1)$ where $(x, w_0) \in \mathcal{R}$ and $(x, w_1) \in \mathcal{R}$, we define

$$f_a(i, r_P, r_V, b) \;=\; \big(\langle P(x, w_b; r_P), V(x; r_V)\rangle\big)_{\le i}$$

[5] A distributional auxiliary-input one-way function is a family $\{f_a\}_a$ such that, given a and $y = f_a(X)$ for uniformly random X, it is computationally hard to sample a value distributed like $X \mid f_a(X) = y$, see formal definition Definition 2.6. Impagliazzo and Luby [IL89] showed that distributional one-way functions imply one-way functions.

where $i \leq |\langle P(x, w_b; r_P), V(x; r_V)\rangle|$ is an index of a bit in the transcript, $b \in \{0, 1\}$, and r_P and r_V are the randomness used by the prover and verifier. That is, on random input, f simply outputs the prefix of the transcript on a random invocation of the protocol using a random choice of witness.

It is well known that, assuming that auxiliary-input one-way functions do not exist, any two efficiently samplable distributions which are computationally indistinguishable are, in fact, statistically close [Gol90]. The same argument yields that if $\{f_a\}_a$ defined above is not a distributional AI-OWF then (P, V) is not only computational WI but is, in fact, statistical WI (SWI).[6]

Thus, for the remainder of the proof, we can assume that the protocol is SWI. We will next show that if $\{f_a\}$ is not a distributional auxiliary-input one-way then we can turn the protocol into a new SWI proof $(\widehat{P}, \widehat{V})$ in which the prover $\widehat{P}$ is deterministic.

A naive attempt. Recall that we are aiming to construct (honest-verifier) WI with a deterministic prover. A natural idea is to consider a protocol $(\widehat{P}, \widehat{V})$ where the verifier chooses the prover's randomness: In the first round $\widehat{V}$ uniformly samples R_P and sends it to the prover, after which the parties run the original protocol with the prover using R_P as its randomness. It is immediate that this transformation preserves completeness and soundness as the honest prover and verifier have essentially has the same behavior. However, it does not preserve even honest-verifier WI. For $(x, w) \in \mathcal{R}$, the transcript in a random execution on input x and witness w is

$$\langle \widehat{P}(x, w), \widehat{V}(x; R_V, R_P)\rangle := (R_P, \langle P(x, w; R_P), V(x; R_V)\rangle).$$

Observe that, unlike in the original protocol, the verifier's view contains R_P. To see why this does not preserve honest-verifier WI suppose, for example, that in the original protocol P sends as a message a one-time pad of the witness $m = w \oplus R_P$. In the original protocol, where the verifier cannot see R_P, the witness is completely hidden, but is can be easily derived in the new protocol.

Canonical simulation. Unlike zero-knowledge, that has an efficient simulator that samples a random transcript of the protocol without knowing the witness, for WI this is not the case. Of course, given the instance x and the witness w, it is trivial to sample such a transcript. However, as mentioned above, the mapping from the randomness used for this sampling to the output could be highly *correlated* with the witness. Our main observation is that if f_a is not a distributional AI-OWF then we can sample such a transcript in a manner that is (almost) *independent* of the witness (even given the randomness used for the sampling).

In more detail, we show an efficient (randomized) algorithm Sim that for every $(x, w) \in \mathcal{R}$ simulates correctly the interaction between $P(x, w)$ and V. Namely,

$$\mathrm{Sim}_{x,w}(R_{\mathrm{Sim}}) \approx \langle P(x, w; R_P), V(x; R_V)\rangle, \tag{2}$$

where $\mathrm{Sim}_{x,w}$ denotes the algorithm Sim with additional auxiliary-input (x, w).

More interestingly, Sim "canonicalizes" the protocol: for any (x, w_0, w_1) and (small enough) $\delta \in 1/\mathrm{poly}$,

$$\mathop{\mathbf{Pr}}_{r_{\mathrm{Sim}} \leftarrow \{0,1\}^{\ell_{\mathrm{Sim}}}} [\mathrm{Sim}_{x,w_0}(r_{\mathrm{Sim}}) \neq \mathrm{Sim}_{x,w_1}(r_{\mathrm{Sim}})] \leq \delta. \tag{3}$$

Notice that the above implies that, given a sample from $\mathrm{Sim}_{x,w_b}(r_{\mathrm{Sim}})$, no distinguisher is likely to predict the value of the bit b, *even given the leakage* r_{Sim} *as an additional input.*

In fact, Sim will be a *next-bit simulator* for the transcript distribution (in fact, next-block simulator where each block corresponds to a message). That is, $\mathrm{Sim}_{x,w}$ can be described as generating its output z in an online fashion: $z_1 = \mathrm{Sim}_{x,w}(r_{\mathrm{Sim},1})$, and for every i set $z_i = \mathrm{Sim}_{x,w}(z_{<i}, r_{\mathrm{Sim},i})$, where $z_{<i} = (z_1, \ldots, z_{i-1})$. Notice, crucially, that in order to output the i-th block, Sim only needs x, w, the transcript so far, and fresh randomness (i.e., it does not need the randomness used to output previous blocks).

We first show how to use such a simulator for deterministic-prover WI, and then discuss why one exists.

[6] See Lemma 3.10 for formal statement and proof.

The new protocol. The canonical simulator allows us to overcome the problem encountered in the naive attempt. Fix (x, w_0, w_1), and let $\mathrm{Sim}^b = \mathrm{Sim}_{x,w_b}$ We define a new protocol $(\widehat{P}, \widehat{V})$ as follows. On input x, the honest verifier $\widehat{V}$ uniformly samples $r_{\mathrm{Sim}} = (r_{\mathrm{Sim},1}, \ldots, r_{\mathrm{Sim},k})$ and sends r_{Sim} together with its first message. The parties then execute the original protocol, with $\widehat{V}$ emulating the original verifier V with the following key distinction: In round $i \in [k]$, instead of generating the next message according to $P(x, w_b)$, the prover $\widehat{P}(x, w_b)$ generates this message using the online simulator Sim^b, the transcript so far, and the fresh randomness $r_{\mathrm{Sim},i}$. Since all the randomness used by $\widehat{P}$ is given by the verifier, the new prover is deterministic.

It is important that Sim is a *canonical* simulator, and not merely an online simulator. It is well known that, in the absence of one-way functions, an efficiently samplable distribution can be simulated in an online, next-bit manner. Indeed, [Ost91; OW93] used this fact to construct an online prover from a zero-knowledge simulator. However, in our setting, such an online simulation on its own, does not guarantee that the new deterministic protocol is SWI. In fact, just like the naive attempt, the simulator's randomness could be highly correlated with the witness. The additional guarantee in Eq. (3) is precisely required to break this correlation (that is, whenever the simulators corresponding to two valid witnesses are invoked on the same transcript prefix and with the same randomness, they produce the same *canonical* next message except with probability δ. Thus, revealing r_{Sim} does not reveal which witness is being used).

We now briefly argue that the resulting protocol has the desired properties. Soundness is preserved since $\widehat{V}$ uses the same acceptance predicate as V. Completeness follows from Eq. (2) since the transcript generated using $\mathrm{Sim}_{x,w}$ is statistically close to the transcript of a genuine execution of $\langle P(x, w; R_P), V(x; R_V)\rangle$ it follows that the acceptance probability decreases only by the corresponding simulation error.[7]

Finally, the SWI property follows from Eq. (3). Consider the two executions with witnesses w_0 and w_1, using the same verifier randomness and the same simulation randomness r_{Sim}. If the transcripts agree through round $i-1$, then the verifier sends the same next message in both executions, and Eq. (3) guarantees that the corresponding prover messages differ with probability at most δ. Thus, by a union bound over the k rounds, the two verifier views differ with probability at most $\delta' := k\delta$. It follows that $(\widehat{P}, \widehat{V})$ is a deterministic-prover honest-verifier δ'-SWI protocol, where δ' is inverse-polynomial by taking δ to be sufficiently small.

This proves that the existence of the canonical simulator suffices for Theorem 1.5. It remains to construct Sim with the properties described in Eqs. (2) and (3). We do this below.

Constructing the canonical simulator. We now construct our canonical simulator. For this, we first construct a *next-bit predictor* to the transcript distribution:

$$T \equiv \langle P(x, w; R_P), V(x; R_V)\rangle.$$

For $\mu \in [0, 1]$, a μ-next-bit predictor for T (over $\{0,1\}^n$) is a function Pred such that, for every $i \in [n]$,

$$\mathop{\mathbf{Pr}}_{\substack{t_{<i} \leftarrow T_{<i} \\ r_{\mathsf{Pred}} \leftarrow \{0,1\}^{\ell_{\mathsf{Pred}}}}} [|\mathsf{Pred}(t_{<i}; r_{\mathsf{Pred}}) - p(t_{<i})| > \mu] \leq \mu \tag{4}$$

where $p(t_{<i}) = \mathbf{Pr}[T_i = 1 \mid T_{<i} = t_{<i}]$. That is, the function Pred can predict on average the bias of the next-bit in the distribution given a prefix of a sample from the distribution. By our choice of function family f it holds that if f is not a distributional auxiliary-input one-way function, then, for any $\mu \in 1/\mathrm{poly}$, there exists a μ-next-bit predictor $\mathsf{Pred}_{x,w}$ for the distribution T of the interaction between the prover and the verifier on input x and witness w.

Indeed, given (x, w) and a partial transcript $T_{<i}$, the predictor $\mathsf{Pred}_{x,w}$ can (distributionally) invert the function $f_{x,w,w}$ to get candidate randomnesses r_P, r_V that are consistent with the transcript so far. Then, it can compute the next bit sent in the interaction using these strings as the parties' randomness. Since r_P and r_V are distributed according to the right distribution, $\mathsf{Pred}_{x,w}$ can get (an estimation) for the probability of the next bit to be 1. In the following, we fix a triplet (x, w_0, w_1), and let $\mathsf{Pred}^b = \mathsf{Pred}_{x,w_b}$.

Given such a next-bit predictor, we are ready to define our canonical simulator Sim^b which uses Pred^b to iteratively sample each bit in the distribution according to the probability indicated by the

[7] This completeness error can then be made negligible by simple parallel repetition.

predictor Pred^b. That is, for randomness $r_{\mathrm{Sim}} = (r_{\mathrm{Sim},1}, \cdots, r_{\mathrm{Sim},n})$, $\mathrm{Sim}^b(r_{\mathrm{Sim}})$ outputs a sequence of bits $(z_1, \cdots, z_n)$, defined as follows:

For every $i \in [n]$, Sim^b interprets $r_{\mathrm{Sim},i} = (r_{\mathsf{Pred},i}, r_{\mathrm{Samp},i})$, for $r_{\mathrm{Samp},i} \leftarrow [0,1]$, and outputs

$$z_i = \begin{cases} 1 & \text{if } r_{\mathrm{Samp},i} \leq \mathsf{Pred}(z_{<i}, r_{\mathsf{Pred},i}), \\ 0 & \text{else} \end{cases}$$

Since Pred^b is a μ-next-bit predictor, a standard argument yields that

$$\mathrm{Sim}^b(R_{\mathrm{Sim}}) \approx_\nu \langle P(x, w_b; R_P), V(x; R_V) \rangle \tag{5}$$

for $\nu := 2n\mu$, which satisfies Equation (2). Thus, in order to conclude the proof for the canonical simulation and prove Eq. (3) it suffices to show:

Claim 1.6. *If $\langle P(x, w_0; R_P), V(x; R_V) \rangle$ and $\langle P(x, w_1; R_P), V(x; R_V) \rangle$) are ϵ-close, then*

$$\Pr_{r_{\mathrm{Sim}} \leftarrow \{0,1\}^{\ell_{\mathrm{Sim}}}} \left[\mathrm{Sim}^0(r_{\mathrm{Sim}}) \neq \mathrm{Sim}^1(r_{\mathrm{Sim}}) \right] \leq \delta$$

for $\delta = 3\epsilon n + 1/\mathrm{poly}$.

Proof sketch. Let $T^b = \langle P(x, w_b; R_P), V(x; R_V) \rangle$. For ease of notation we drop the subscript Sim and write r and ℓ instead of r_{Sim} and ℓ_{Sim}. For every $i \in [n]$ define the indicator function $E_i(r)$ that is 1 if i is the first index for which $\mathrm{Sim}^0(r)_i \neq \mathrm{Sim}^1(r)_i$. That is,

$$E_i(r) = \begin{cases} 1 & \mathrm{Sim}^0(r)_i \neq \mathrm{Sim}^1(r)_i \ \wedge \ \mathrm{Sim}^0(r)_{<i} = \mathrm{Sim}^1(r)_{<i}, \\ 0 & \text{else} \end{cases}$$

Then we have that $\mathbf{Pr}_{r \leftarrow \{0,1\}^{\ell \cdot n}} \left[\mathrm{Sim}^0(r) \neq \mathrm{Sim}^1(r) \right] = \sum_{i=1}^n \mathbf{Pr}_{r \leftarrow \{0,1\}^{\ell \cdot n}} \left[E_i(r) = 1 \right]$. Thus, it suffices to show, for every $i \in [n]$, that

$$\Pr_r[E_i(r) = 1] \leq 3\varepsilon + 4\mu + \nu.$$

Recall that for every $b \in \{0,1\}$, $\mathrm{Sim}^b(r_{\leq i})_i = \mathrm{Sim}^b(z, r_i)$ where $z = \mathrm{Sim}^b(r_{<i})$. Thus, we get that for every $i \in [n]$

$$\begin{aligned} \Pr_{r \leftarrow \{0,1\}^{\ell \cdot n}} [E_i(r) = 1] &\leq \Pr_{\substack{r_{<i} \leftarrow \{0,1\}^{\ell \cdot (i-1)} \\ z = \mathrm{Sim}^0(r_{<i}) \\ r_i \leftarrow \{0,1\}^\ell}} \left[\mathrm{Sim}^0(z, r_i) \neq \mathrm{Sim}^1(z, r_i) \right] \\ &\leq \Pr_{\substack{z \leftarrow T^0_{<i} \\ r_i \leftarrow \{0,1\}^\ell}} \left[\mathrm{Sim}^0(z, r_i) \neq \mathrm{Sim}^1(z, r_i) \right] + \nu \end{aligned} \tag{6}$$

where the last inequality follows from Equation (5) since $\mathrm{Sim}^0(r_{<i})_{<i}$ is statistically close to $T^0_{<i}$. Letting $p^b(z) = \mathbf{Pr}[T^b_i = 1 \mid T^b_{<i} = z]$, again since $T^0 \approx_\varepsilon T^1$, it holds that $\mathbf{E}_{z \leftarrow T^0_{<i}} \left[|p^0(z) - p^1(z)| \right] \leq 2\varepsilon$. Thus, by the μ-next bit predictor guarantee (Equation (4)) and the triangle inequality

$$\mathop{\mathbf{E}}_{\substack{z_{<i} \leftarrow T^0_{<i} \\ r_{\mathsf{Pred}} \leftarrow \{0,1\}^{\ell_{\mathsf{Pred}}}}} \left[|\mathsf{Pred}^0(z_{<i}; r_{\mathsf{Pred}}) - \mathsf{Pred}^1(z_{<i}; r_{\mathsf{Pred}})| \right] \leq 3\varepsilon + 4\mu. \tag{7}$$

Recall that $r = (r_{\mathsf{Pred}}, r_{\mathrm{Samp}})$ and that by our definition of Sim it holds that $\mathrm{Sim}^0(z_{<i}, r_i) \neq \mathrm{Sim}^1(z_{<i}, r_i)$ with probability at most $|\mathsf{Pred}^0(z_{<i}; r_{\mathsf{Pred}}) - \mathsf{Pred}^1(z_{<i}; r_{\mathsf{Pred}})|$ where the probability is taken over r_{Samp} (that is interpreted as a random element in $[0,1]$). Putting the bounds together, we obtain

$$\Pr_r[E_i(r) = 1] \leq 3\varepsilon + 4\mu + \nu.$$

Summing over i gives

$$\Pr_r[\mathrm{Sim}^0(r) \neq \mathrm{Sim}^1(r)] \leq 3\varepsilon n + n(4\mu + \nu).$$

Since $\nu = 2n\mu$, we can choose μ small enough that the second term is the desired inverse polynomial. □

1.2.2 Round-collapse for honest-verifier WI

We now describe how to collapse constant-round honest-verifier WI protocols into 2-message ones in the absence of one-way functions. Our main observation is that the classical Babai–Moran [BM88] round-collapse, which was shown by Pass and Venkitasubramaniam [PV20] to also hold (in the absence of one-way functions) for computationally-sound protocols, preserves witness intistinguishability provided that the prover is deterministic. Fortunately, this is exactly what the transformation described in the previous section achieves! Putting this together gives us the Corollary 1.4 (and the round-collapse needed in the process of proving Theorem 1.2).

We start by explaining the Babai–Moran transformation. For simplicity, suppose the protocol has 3 messages with the prover going first, and we would like to transform it into a 2-message one with the verifier going first (the many-round version follows by applying the transformation on the final 3 rounds of the protocol and then repeating this process on the derived protocol sufficiently many times):

1. First, amplify soundness of the protocol by parallel repetition of the last two rounds of the protocol. That is, the protocol is identical up to the final round. In the final round, the verifier sends many different challenges and the prover replies to all of these challenges. The verifier accepts if all of the replies to challenges are accepting.

2. Now, we switch the order of the last two messages of the protocol. That is, the verifier begins by sending its random message ρ, and the prover replies with both of its messages m_1 and m_2 together.

3. Amplify soundness of the protocol using standard parallel repetition.

Babai and Moran show that the protocol above is sound, which roughly stems from the fact that the protocol has been amplified to a degree such that (when the prover is malicious) the probability over ρ that there *exists* an m_2 such that the verifier would accept (m_1, ρ, m_2) is small.

Honest-verifier WI is preserved by the transformations in steps 2 and 3. However, this is not true for step 1. Indeed, suppose that if the first bit of ρ is 0 then $m_2 = u$ for an initially chosen random u and if ρ starts with 1 then $m_2 = w \oplus u$ where w is the witness. While standard parallel repetition would resample u, thus leaking nothing, repeating only the last rounds allows the verifier to recover w.

Thankfully, if the prover is *deterministic*, then step 1 does preserve honest-verifier WI. Indeed, for every choice of ρ, the transcript (m_1, ρ, m_2) generated by the protocol is a fixed deterministic function f_ρ that depends only on the witness used by the prover. Consider two witnesses w_0 and w_1. Whenever $f_\rho(w_0) \neq f_\rho(w_1)$, a distinguisher can easily distinguish whether the prover used w_0 or w_1 from the verifier's view. Hence, by (honest-verifier) WI, $f_\rho(w_0) = f_\rho(w_1)$ with overwhelming probability over ρ. Thus, even seeing m_2 for many choices of ρ will not leak the witness used, and so step 1 preserves honest-verifier WI.

1.2.3 WC with Perfect Completeness and Soundness Via Derandomization

Following the steps outlined in previous sections, we have a WC scheme with nonzero completeness and soundness errors. We now explain how to eliminate these errors from our WC, while preserving a inverse-polynomial canonization error.

The approach of Barak, Ong, and Vadhan. Our starting point is the approach of Barak, Ong, and Vadhan [BOV07] which showed how to derandomize ZAPs, which are two-message WI protocols. Specifically, suppose that we are given a ZAP in which the verifier sends ℓ uniformly random bits to the prover. Barak, Ong, and Vadhan use a complexity-theoretic PRG $G\colon \{0,1\}^d \to \{0,1\}^\ell$ to remove the verifier's randomness while retaining the WI property. Specifically, they use the PRG of Impagliazzo and Wigderson [IW97] and its extension by Klivans and van Melkebeek [KM02] to nondeterministic, non-uniform adversaries.

The key idea is to enumerate all seeds of G, shifted according to a random uniform offset $u \in \{0,1\}^\ell$. That is, for every $s \in \{0,1\}^d$, execute the original scheme using randomness $u \oplus G(s)$. The new WI outputs the vector of all resulting proofs, and the new verifier accepts if a strict majority of the coordinates are accepted.

Loosely speaking, the random shift is added to preserve the WI property, while the enumerating over the PRG seeds enables us to remove the randomness of the verifier. More formally, let $V(x, r, \pi)$ denote the verifier's decision, where x is the instance, and r and π are the verifier and prover messages in the original ZAP. For a fixed shift u, define $r_s(u) = u \oplus G(s)$. WI, completeness, and soundness are preserved as follows:

- *WI.* For every fixed seed s, if u is uniform, then $r_s(u) = u \oplus G(s)$ is uniform. Hence, the s-th coordinate is distributed as an honest execution of the original protocol. A hybrid argument over the 2^d seeds shows that, if the original protocol has WI error ε, then the resulting vector has WI error at most $2^d\varepsilon$. In particular, this remains negligible when 2^d is polynomial.

- *Completeness.* For fixed $x \in L$, valid witness w, and shift u, consider the deterministic test

$$C_{x,w,u}(r) := V\big(x, u \oplus r, P(w; u \oplus r)\big).$$

  Since $u \oplus r$ is uniform when r is uniform, if G fools this test, then completeness implies that, for every shift u, a strict majority of the coordinates contain accepting honest-prover messages, as desired.

- *Soundness.* For fixed $x \notin L$ and shift u, consider the nondeterministic test

$$S_{x,u}(r) = 1 \quad \Longleftrightarrow \quad \exists \pi : \ V(x, u \oplus r, \pi) = 1.$$

  Again, since $u \oplus r$ is uniform when r is uniform, if G fools this nondeterministic test, then statistical soundness implies that, for every shift u, fewer than half of the coordinates admit an accepting prover message, so no vector of messages passes the majority check.

The problem in our setting. By viewing the deterministic prover as the WC mapping algorithm, Theorem 1.5 and our round-collapse from Section 1.2.2 yields a n^{-c}-WC for some constant c. We would like to apply the [BOV07] approach to this WC to remove the randomness. The issue is that the classical PRGs used above may have a seed length whose constant depends on the running time of the completeness test. Since this test runs the mapping algorithm and the verifier, the number of seeds may grow with the running time of the WC. Thus, we cannot afford to union bound over all seeds, since it might be the case that $n^{-c} \cdot 2^d > 1$.[8]

This creates a circularity. We would like to choose c large enough that the error n^{-c} remains small even after taking a union bound over all seeds. However, increasing c changes the starting WC and may increase the running time of its mapping algorithm, thereby increasing the number of seeds in the same union bound.

Breaking the circularity using PRGs for fast derandomization. We overcome this problem using the recent derandomization result of Liu and Pass [LP25]. We rely on their stronger derandomization assumption, which states that there exists a constant $\gamma > 0$ such that, for every $k \in \mathbb{N}$,

$$\mathsf{E} \not\subseteq \mathsf{ioNTIME}[2^{kn}]/2^{\gamma n}.$$

Under this assumption, Liu and Pass construct a constant-error PRG with seed length

$$\hat{a} \log N,$$

where the crucial point is that the constant $\hat{a}$ is independent of the running time of the computation being fooled. More precisely, for every polynomial t, their PRG fools $\mathsf{NTIME}[t(N)]/N$. The generator and its running time may depend on t, but the number of seeds does not.

We would now like to apply the transformation of [BOV07] using this new PRG. However, one issue remains: N must be at least the length ℓ of the public randomness, and ℓ again may depend on c.[9] Indeed, since the shift u is hardwired into the test, we need $N \geq |u| = \ell$. If $\ell = n^{v_c}$ for an exponent

[8] Observe that in our construction, to get n^{-c} WI error, we need to invert distributional one-way function with error smaller than n^{-c}. Thus, the seed length of the PRG we need depends on the running time of the one-way function inverter.

[9] In our construction, ℓ depends on the length of the randomness used by the one-way function inverter.

v_c depending on c, then the Liu Pass generator has $n^{v_c\hat{a}}$ seeds. Thus, the number of seeds may again depend on c. We avoid this by first compressing the public randomness.

We first reduce the public randomness to logarithmic length. Namely, we use a standard inverse-polynomial-error PRG

$$G_1 : \{0,1\}^d \to \{0,1\}^\ell, \qquad d = O_c(\log n),$$

which fools the completeness, soundness, canonization tests with error at most n^{-c}. We do not enumerate the seeds of G_1, and therefore it does not matter that the constant hidden in $O_c(\log n)$ depends on c. The purpose of G_1 is to compress the public randomness while preserving an inverse polynomial canonization error.

The seed $u \in \{0,1\}^d$ of G_1 becomes the public randomness of the new scheme.

Thus, we can now apply the Liu Pass generator

$$G_2 : \{0,1\}^{\lceil \hat{a} \log N \rceil} \to \{0,1\}^d.$$

For every seed s of G_2, we use $r_s(u) = G_1\big(u \oplus G_2(s)\big)$ as the public randomness of the original scheme. The new mapping outputs all the resulting proofs, and the new checker accepts if a strict majority of them are accepted.

For every fixed u, the pseudorandomness of G_2 and the majority rule give perfect completeness and soundness. Moreover, for each seed s, the shift $u \oplus G_2(s)_{\le d}$ is uniform, so the disagreement probability is $O(n^{-c})$. Since G_2 has $O(n^{v\hat{a}})$ seeds, where $v\hat{a}$ is independent of c, a union bound gives error $O(n^{v\hat{a}-c})$, which is inverse polynomial for sufficiently large c.

Additional Related Works

As mentioned in above, [OW93; HN24] prove that one-way functions are necessary for zero-knowledge arguments for NP, and [LMP24] extended this result to the uniform settings. [MPS25] defined and study witness pseudo-canonicalization (WPC), a notion closely related to Non-Interactive WI proofs.

Faonio, Nielsen, and Venturi [FNV17] defined the notion of predictive arguments and showed it implies WE. [FS90] defined the notion of WI (and a related notion of witness hiding) and showed that together with one-way functions, WI for NP implies zero-knowledge. Using the OR trick [FLS99], [CPW23] showed that WI arguments with a unique proof for NP implies WE for NP, assuming one-way functions. [Kiy24] showed a similar result starting from statistical WI proofs that are *resettable* (that is, that are secure when the prover uses the same randomness multiple times).

[BOV07] showed that under assumptions from the literature on derandomization [NW94], it is possible to derandomize the verifier of a ZAP and construct Non-Interactive WI (NIWI). Goldreich and Oren [GO94b] showed that languages outside BPP do not have two-message ZK arguments, or that auxiliary-input ZK with a deterministic prover. [CLMP12] consider two-message arguments with super-polynomial simulator (SPS), and showed a barrier to prove soundness of such arguments using black-box reductions. [BC20] showed that without auxiliary-input, ZK with deterministic prover for NP can be constructed from strong cryptographic assumptions. Additionally, [BC20] showed that such protocols imply witness encryption. [Wee06] presented an oracle with respect to which there exists a succinct argument for a hard language in NP, but one-way functions do not exist.

Unique witness maps. [CPW20; CPW23] consider a number of types of "unique witness maps" (UWM). Unique witness maps can be thought of as identical to WC except that the common reference string is allowed to be structured (as opposed to be uniformly random). [CPW20] construct UWM for NP from the combined assumptions of one-way functions and indistinguishability obfuscation via a Sahai–Waters style construction, where the "canonical witness" is essentially a signature applied to the instance. As such, the structured nature of the CRS is crucial for uniqueness as it encodes within it the signing key. The closest object to ϵ-WC considered in prior works is pseudo-UWM, defined in [CPW23], which are ϵ-WC where a structured reference string is allowed. The same paper shows that a version of ultra-compact UWM implies pseudo-UWM.

OWFs from WI of knowledge. In a related work to this paper, [AMPS26] show that, assuming NP is hard, the existence auxiliary-input one-way functions is implied by (constant-round) WI arguments for NP that satisfy *knowledge soundness*, roughly meaning that an extractor, given black-box access to a

highly-convincing prover, can derive a valid witness. We note that this is incomparable to Theorem 1.2 as (a) the assumption that NP has no ϵ-WC does not seem to be implied by the worst-case hardness of NP, (b) it assumes additional structure on the protocol in the form of the extractor.

2 Preliminaries

General notation. We use the following conventions throughout the paper:

- All logarithms are taken in base 2.
- Calligraphic letters denote sets and distributions. Uppercase letters denote random variables; lowercase letters denote values and functions.
- poly denotes the set of polynomially bounded functions. PPT denotes probabilistic polynomial-time algorithms. A non-uniform polynomial-time algorithm A is equipped with a fixed polynomial-size advice sequence $\{z_n\}_{n\in\mathbb{N}}$ (typically omitted from notation), and we write A_n for A with advice z_n on inputs of length n.
- For a randomized algorithm A, we write $A(\cdot;r)$ for A with randomness fixed to $r \in \{0,1\}^*$. We write negl for a negligible function.
- For $n \in \mathbb{N}$, we let $[n] \coloneqq \{1,\ldots,n\}$. For $i \le j$, we let $[i:j] \coloneqq \{i,\ldots,j\}$.
- For a vector $v \in \Sigma^n$, we write v_i for its i-th entry, $v_{<i} = (v_1,\ldots,v_{i-1})$, and $v_{\le i} = (v_1,\ldots,v_i)$. For $\mathcal{I} \subseteq [n]$, we write $v_{\mathcal{I}}$ for the ordered tuple $(v_i)_{i\in\mathcal{I}}$.
- For strings $x, y \in \{0,1\}^*$, we write $x||y$ for their concatenation.

Probability notation. When unambiguous, we will naturally identify a random variable with its distribution. The support of a finite distribution $\mathcal{P}$ is defined by $\mathrm{Supp}(\mathcal{P}) \coloneqq \{x\colon \mathbf{Pr}_{\mathcal{P}}[x] > 0\}$. For a (discrete) distribution $\mathcal{P}$, let $x \leftarrow \mathcal{P}$ denote that x was sampled according to $\mathcal{P}$. Similarly, for a set $\mathcal{S}$, let $x \leftarrow \mathcal{S}$ denote that x is drawn uniformly from $\mathcal{S}$. For $m \in \mathbb{N}$, we use $\mathbf{U}_m$ to denote a uniform random variable over $\{0,1\}^m$ (that is independent from other random variables in consideration). For $p \in [0,1]$ we use U_p to denote the random variable that takes 1 with probability p and 0 otherwise.

A collection of distributions $\{\mathcal{D}_a\}_{a\in\{0,1\}^{\lambda(n)},n\in\mathbb{N}}$ is samplable if there exists a poly(n)-time randomized algorithm S such that for every $n \in \mathbb{N}$ and every $a \in \{0,1\}^{\lambda(n)}$ $S(a) \equiv \mathcal{D}_a$.

Statistical distance. The statistical distance of two distributions $\mathcal{P}$ and $\mathcal{Q}$ over a discrete domain $\mathcal{X}$ is defined by $\mathrm{SD}(\mathcal{P},\mathcal{Q}) \coloneqq \max_{\mathcal{S}\subseteq\mathcal{X}} |\mathcal{P}(\mathcal{S}) - \mathcal{Q}(\mathcal{S})| = \frac{1}{2}\sum_{x\in\mathcal{X}} |\mathcal{P}(x) - \mathcal{Q}(x)|$.

Entropy. For a random variable X, let $H(X) = \mathbf{E}_{x\leftarrow X}\left[\log \frac{1}{\mathbf{Pr}[X=x]}\right]$ denote the **(Shannon) entropy** of X, and let $\mathrm{H}_\infty(X) = \min_{x\in Supp(X)}\left\{\log \frac{1}{\mathbf{Pr}[X=x]}\right\}$ denote the **min-entropy** of X.

We will make use of the following facts in probability.

Lemma 2.1 (Coupling)**.** *Let $\mathcal{D}_A, \mathcal{D}_B$ be two distributions. Then there exists a triplet of jointly distributed random variables (A,B,E), such that the following hold:*

- *The marginal distribution of A is $\mathcal{D}_A$ and the marginal distribution of B is $\mathcal{D}_B$.*
- $\mathbf{Pr}[E \neq 1] = \mathrm{SD}(\mathcal{D}_A,\mathcal{D}_B)$*, and,*
- $\mathbf{Pr}[A = B | E = 1] = 1$.

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

Lemma 2.2 (Chernoff bound)**.** *Let $A_1,\ldots,A_n$ be independent random variables s.t. $A_i \in \{0,1\}$. Let $\widehat{A} = \Sigma_{i=1}^n A_i$ and $\mu = \mathbf{E}\left[\widehat{A}\right]$. For every $\epsilon \in [0,1]$ It holds that:*

$$\mathbf{Pr}\left[|\widehat{A} - \mu| \ge \epsilon\cdot\mu\right] \le 2\cdot e^{-\epsilon^2\cdot\mu/3}.$$

The above lemmas is a standard facts in information theory.

2.1 Complexity Classes and NP Relations

For a language $\mathcal{L} \subseteq \{0,1\}^*$, let $\mathcal{L}_n = \mathcal{L} \cap \{0,1\}^n$, and for $\mathcal{L} \in \mathsf{NP}$ let $\mathcal{R}(\mathcal{L}) \subseteq \mathcal{L} \times \{0,1\}^*$ be a canonical NP relation. Similarly, for an NP relation $\mathcal{R} \subseteq \{0,1\}^* \times \{0,1\}^*$, we let $\mathcal{L}(\mathcal{R}) = \{x | \ \exists w : (x,w) \in \mathcal{R}\}$. For a string x, we let $\mathcal{L}(x)$ be the indicator function that outputs 1 if $x \in \mathcal{L}$ and 0 otherwise.

The SAT language and relation. We define the canonical relation for the NP complete language SAT to be $\mathcal{R}_{\mathsf{SAT}} = \{(\varphi, v) \colon \varphi$ is a Boolean formula and v is an assignment such that $\varphi[v] = 1\}$. The language OR-SAT consists of all instances of the form $\varphi := (\varphi_0 \vee \varphi_1)$ where $|\varphi_0| = |\varphi_1|$ and φ is a satisfiable SAT formula. The canonical relation $\mathcal{R}_{\mathsf{OR\text{-}SAT}}$ is defined similarly to $\mathcal{R}_{\mathsf{SAT}}$.

Definition 2.3 (AvgBPP/poly)**.** *A pair $(\mathcal{L}, \mathcal{D})$ where $\mathcal{L}$ is a language and $\mathcal{D} = \{\mathcal{D}_n\}_{n \in \mathbb{N}}$ is a samplable distribution is in* AvgBPP/poly *if there exists a non-uniform PPT A such that the following holds for infinitely many $n \in \mathbb{N}$:*

- *For every $x \in \mathsf{Supp}(\mathcal{D}_n)$, $\mathbf{Pr}\left[A(x) \in \{\perp, \mathcal{L}(x)\}\right] \geq 0.9$*
- $\mathbf{Pr}_{x \leftarrow \mathcal{D}_n}\left[A(x) = \perp\right] \leq 1/4$.

The following lemma is proven in Section B.1

Lemma 2.4. *Assume that $\mathcal{L} \notin \mathsf{ioSIZE}[T(n)]$. Then, there exists a pair of distributions ensemble $\mathcal{D}^Y = \{\mathcal{D}^Y_n\}_{n \in \mathbb{N}}, \mathcal{D}^N = \{\mathcal{D}^N_n\}_{n \in \mathbb{N}}$ with $\mathcal{D}^Y_n \subseteq \mathcal{L}_n$ and $\mathcal{D}^N_n \subseteq \{0,1\}^n \setminus \mathcal{L}_n$, such that*

$$\left| \mathop{\mathbf{Pr}}_{x \leftarrow \mathcal{D}^Y}\left[A(x) = 1\right] - \mathop{\mathbf{Pr}}_{x \leftarrow \mathcal{D}^N}\left[A(x) = 1\right] \right| \leq \epsilon(n)$$

for any $A \in \mathsf{SIZE}[\frac{\epsilon^2(n)(T(n) - \mathrm{poly}(n, 1/\epsilon(n)))}{160n}]$ and any large enough n.

2.2 Auxiliary-Input One-Way Functions

Definition 2.5 (AI-OWF)**.** *An efficiently computable function family $\{f_a \colon \{0,1\}^{m(n)} \to \{0,1\}^{m(n)}\}_{a \in \{0,1\}^{\lambda(n)}}$ is an* auxiliary-input one-way function (AI-OWF) *if for every PPT $\mathcal{A}$ there exists a negligible function μ such that for every $n \in \mathbb{N}$, there exists some $a \in \{0,1\}^{\lambda(n)}$ such that*

$$\mathop{\mathbf{Pr}}_{x \leftarrow \{0,1\}^{m(n)}}\left[\mathcal{A}(a, f_a(x)) \in f_a^{-1}(f_a(x))\right] \leq \mu(n) \tag{8}$$

If a is simply $0^{\lambda(n)}$, we refer to the family as simply a one-way function (OWF)*.*

We say that $\{f_a\}$ is non-uniform-secure *if the above holds for every non-uniform PPT $\mathcal{A}$. We say that $\{f_a\}$ is* infinitely often auxiliary-input one-way function (ioAI-OWF) *if for every $\mathcal{A}$ the above holds for infinitely many $n \in \mathbb{N}$.*

We now define the notion of distributional auxiliary-input one-way functions.

Definition 2.6 (Distributional Auxiliary-input One-way functions, [IL89])**.** *An efficiently computable function family $\{f_a \colon \{0,1\}^{m(n)} \to \{0,1\}^{m(n)}\}_{a \in \{0,1\}^{\lambda(n)}}$ is an ε-*distributional auxiliary-input one-way function (distributional AI-OWF) *if for every* PPT $\mathcal{A}$ *and for every $n \in \mathbb{N}$, there exists some $a \in \{0,1\}^{\lambda(n)}$ such that*

$$\mathrm{SD}\Big(\big(\mathcal{A}(a, f_a(X)), f_a(X)\big), \big(X, f_a(X)\big)\Big) \geq \varepsilon$$

where X is uniformly sampled from $\{0,1\}^{m(n)}$. The family $\{f_a\}$ is a simply a distributional auxiliary-input one-way function *if there exists a constant $0 < c$ such that $\{f_a\}$ is $1/n^c$-distributional AI-OWF.*

We say that $\{f_a\}$ is infinitely often (ϵ-)distributional auxiliary-input one-way function *if for every $\mathcal{A}$ the above holds for infinitely many $n \in \mathbb{N}$.*

Theorem 2.7 (Distributional one-way functions imply one-way functions, [IL89])**.** *If there exists an auxiliary-input distributional one-way function then there exists an auxiliary-input one-way function.*

2.3 Interactive Protocols

Interactive Protocol. Let $P = (P_1, \ldots, P_k)$ and $V = (V_1, \ldots, V_k, V^*)$ be interactive algorithms (the prover and verifier, respectively). We assume $k = k(|x|) \in \text{poly}$.

For inputs x, y and random coins r_P, r_V, we denote by

$$\mathsf{tr} \leftarrow \langle P(x; r_P), V(y; r_V) \rangle$$

the transcript of their interaction, and let

$$\mathsf{View}_V^P(x, y) := (r_V, \langle P(x; r_P), V(y; r_V) \rangle)$$

denote the view of V in the interaction. When we omit r_P, r_V, they are sampled uniformly and independently.

The transcript has the form $\mathsf{tr} = (\rho_1, m_1, \ldots, \rho_k, m_k)$, where for $i = 1, \ldots, k$:

$$\begin{aligned}(\rho_i, \mathsf{st}_{V,i}) &= V_i(y, \mathsf{st}_{V,i-1}, m_{i-1}; r_{V,i}),\\ (m_i, \mathsf{st}_{P,i}) &= P_i(x, \mathsf{st}_{P,i-1}, \rho_i; r_{P,i}),\end{aligned}$$

with $\mathsf{st}_{V,0} = \mathsf{st}_{P,0} = \emptyset$ and $m_0 = \emptyset$.

We define the verifier's output as

$$\mathsf{Out}(\langle P(x), V(y) \rangle) \coloneqq V^*(y, \mathsf{st}_{V,k}, m_k).$$

We write $\mathsf{tr}_{<i} = (\rho_1, m_1, \ldots, \rho_{i-1}, m_{i-1})$ and $\mathsf{tr}_{\leq i} = (\rho_1, m_1, \ldots, \rho_i, m_i)$.

Structural properties.

- The protocol is **public-coin** if for every i, $\rho_i = r_{V,i}$ and $\mathsf{st}_{V,i} = \emptyset$.
- The protocol is $(1-\varepsilon)$**-unique** for $\mathcal{R}$ if for every x, w_0, w_1 with $(x, w_0), (x, w_1) \in \mathcal{R}$,

$$\mathop{\mathbf{Pr}}_{r_P, r_V}\left[\langle P(x, w_0; r_P), V(x; r_V) \rangle = \langle P(x, w_1; r_P), V(x; r_V) \rangle\right] \geq 1 - \varepsilon.$$

- The protocol is γ**-succinct** if for every $(x, w) \in \mathcal{R}$,

$$|\langle P(x, w), V(x) \rangle| \leq (|x| + |w|)^\gamma.$$

Definition 2.8 (Interactive Proofs and Arguments, Completeness and Soundness)**.** *An* interactive protocol *for a relation $\mathcal{R} \subseteq \{0,1\}^* \times \{0,1\}^*$ is a pair of PPT interactive machines (P, V) such that:*

- ***Completeness:*** *For every $(x, w) \in \mathcal{R}$,*

$$\mathbf{Pr}\left[\mathsf{Out}(\langle P(x, w), V(x) \rangle) = 1\right] \geq 1 - \alpha(|x|).$$

- ***Computational Soundness:*** *For every $x \notin L_\mathcal{R}$ and every non-uniform PPT prover $\widehat{P}$,*

$$\mathbf{Pr}\left[\mathsf{Out}(\langle \widehat{P}, V(x) \rangle) = 1\right] \leq \beta(|x|).$$

- ***Statistical Soundness:*** *For every $x \notin L_\mathcal{R}$ and every (possibly inefficient) prover $\widehat{P}$,*

$$\mathbf{Pr}\left[\mathsf{Out}(\langle \widehat{P}, V(x) \rangle) = 1\right] \leq \beta(|x|).$$

We say that (P, V) is an interactive argument (IA) *for $\mathcal{R}$ if it satisfies completeness and computational soundness.*

We say that (P, V) is an interactive proof (IP) *for $\mathcal{R}$ if it satisfies completeness and statistical soundness.*

For an oracle $\mathcal{O}$, we say that a scheme is an IA with respect to $\mathcal{O}$ *(resp., an* IP with respect to $\mathcal{O}$*) if the above holds when (P, V) and the relevant cheating prover have oracle access to samples from $\mathcal{O}$. When α and β are not specified, they are assumed to be negligible functions.*

2.3.1 Witness Indistinguishability

As discussed in the introduction, witness indistinguishability is equivalent to the existence of an unbounded simulator. However, for ease of use, we define the notion of witness indistinguishability explicitly below.

Definition 2.9. *An interactive protocol (P, V) for a relation $\mathcal{R}$ satisfies* (computational) witness indistinguishability *with distinguishing advantage δ (δ-WI) if for every x and every w_0, w_1 with $(x, w_0), (x, w_1) \in \mathcal{R}$, and every non-uniform PPT verifier $\widehat{V}$, and every non-uniform PPT distinguisher D:*

$$\left|\mathbf{Pr}\left[D\left(\mathsf{View}^P_{\widehat{V}}(x, w_0)\right) = 1\right] - \mathbf{Pr}\left[D\left(\mathsf{View}^P_{\widehat{V}}(x, w_1)\right) = 1\right]\right| < \delta(|x|),$$

We further define:

- *A protocol satisfies* statistical witness indistinguishability *δ-SWI as above, except it must hold against any (possibly inefficient) distinguisher (equivalently, the two distributions above have statistical distance at most δ).*
- *A protocol is δ-honest-WI (resp. δ-honest-SWI) if indistinguishability holds when $\widehat{V}$ is the honest verifier V (so that δ-WI implies δ-honest-WI, but the opposite is not necessarily true).*
- *When δ is a negligible function, we sometimes omit δ from all of the above notation (e.g., δ-WI becomes WI).*

Remark 2.10 (Views vs. Transcripts in Witness Indistinguishability)**.** *Witness indistinguishability is formally defined with respect to the* view *of the verifier as is the case in Definition 2.9. However, we note that our results hold even when the indistinguishability is only guaranteed with respect to the transcript of the interaction. The transcript-indistinguishability variant is a strictly weaker (i.e., more easily achievable) notion as any distinguisher can always restrict itself to looking only at the transcript. For simplicity we state our results with respect to that standard notion (though we do not make use of this stringer guaranty in the proofs).*

We additionally note that for SWI the two notions are equivalent: since the verifier's randomness is independent of the witness used by the prover, its view can be post-sampled given the transcript.

The following theorem shows that WI arguments with constant rounds can be transform to public-coin. Chakraborty, Hulett, Khurana, and Tomer [CHKT26] proved this theorem for ZK, but the transformation does not use the efficiency of the simulator.

Theorem 2.11 (Private to public-coin, [CHKT26])**.** *Assume that infinity often AI-OWFs do not exist. Let (P, V) be a β-sound, $(1-\alpha)$-complete honest-verifier ϵ-WI argument for $\mathcal{R}$ with k rounds for constant k. Then for any $q \in$ poly there exists $(\beta + 1/q)$-sound, $(1 - \alpha - 1/q)$-complete, k-rounds honest-verifier $(\epsilon + 1/q)$-WI public-coin argument for $\mathcal{R}$. If the original protocol is statistically sound, then the resulting public-coin protocol is also $(\beta + 1/q)$-statistically sound.*

Moreover, the prover-to-verifier communication stays the same as in the original protocol.

We note that the result of [CHKT26] was stated only for arguments (computational soundness) and not for proofs (statistical soundness), but a careful examination of the transformation presented [CHKT26] shows that the result also holds for proofs. This holds since the verifier in the new protocol is essentially unchanged and it retains its original accepting condition.

We will also make use of the following standard completeness amplification claim.

Claim 2.12 (OR-parallel repetition)**.** *Let (P, V) be an δ-SWI interactive protocol with $(1-\alpha)$-completeness and β-soundness. It holds that the t-fold independent parallel repetition with OR accepting condition is an $t\delta$-SWI interactive protocol with $(1-\alpha^t)$-completeness and $t\beta$-soundness*

Proof of Claim 2.12. The completeness bound follows simple independence. The soundness bound follows by a union bound (or by embedding a base challenge in a random coordinate), and the SWI bound follows by a standard hybrid argument over the t instances. □

2.3.2 Witness Canonicalization

Definition 2.13 (Witness Canonicalization)**.** *An ε-witness canonicalization scheme ($\varepsilon - \mathsf{WC}$) for an NP relation $\mathcal{R}$ is a pair (map, V) of two deterministic polynomial time algorithms such that there exists polynomial ℓ for which the following holds:*

- ***Completeness:*** *For every $(x, w) \in \mathcal{R}$,*

$$\Pr_{r \leftarrow \{0,1\}^{\ell(|x|)}} [V(r, x, \mathsf{map}(r, x, w)) = 1] \geq 1 - \mathsf{negl}(|x|).$$

- ***(Computational) Soundness:*** *For any non-uniform PPT adversary A and every $x \notin L_{\mathcal{R}}$*

$$\Pr_{\substack{r \leftarrow \{0,1\}^{\ell(|x|)} \\ \pi \leftarrow A(r)}} [V(r, x, \pi) = 1] \leq \mathsf{negl}(|x|).$$

  *We say that the protocol is **statistically sound** if the above holds even when A is inefficient. In such a case we say that the scheme is s statistically sound.*

- ***Uniqueness:*** *For every x, there exists a family $\{\pi_{x,r}\}_{r \in \{0,1\}^{\ell(|x|)}}$ such that for every w with $(x, w) \in \mathcal{R}$,*

$$\Pr_{r \leftarrow \{0,1\}^{\ell(|x|)}} [\mathsf{map}(r, x, w) = \pi_{x,r}] \geq 1 - \varepsilon.$$

*Moreover, we say that a WC has **perfect completeness** and **perfect soundness** if the negligible functions in completeness and soundness items above is the zero constant function. Finally, we say that a WC is **non-uniform** if both the map map and the verifier V have access to polynomial-length non-uniform advice.*

Unique interactive arguments/proofs: Note that the above definition is equivalent to a public coin, one-round interactive argument/proof with an honest verifier with $(1 - \varepsilon)$-*uniqueness*, and negligible completeness and soundness errors.

We will make use of the following standard claim.

2.3.3 Batch Protocols

We formalize batch verification for relations.

Definition 2.14 (Batch Protocols)**.** *A batch protocol for a relation $\mathcal{R}$ is an interactive protocol for $\bigcup_{t \in \mathbb{N}} \mathcal{R}^{\otimes t}$, where*

$$\mathcal{R}^{\otimes t} := \{((x_1, \ldots, x_t), (w_1, \ldots, w_t)) : |x_1| = \cdots = |x_t| \text{ and } (x_i, w_i) \in \mathcal{R} \text{ for all } i \in [t]\}.$$

The protocol has soundness and completeness as defined in Definition 2.8. The protocol has compression rate $\rho = \rho(n, t)$ (for instance length n and batch size t) if the number of bits sent by the prover across the entire interaction is at most $\rho \cdot t$.

We will make use of the following theorem by Bitansky, Kamath, Paneth, Rothblum, and Vasudevan [BKPRV24]. We make light changes in the theorem statement to match our usage (we explain this change below).

Theorem 2.15. *[Batch protocols imply SWI, [BKPRV24]] Let $\mathcal{R}$ be an NP relation. Suppose there exists a batch protocol $\Pi = (P, V)$ for $\mathcal{R}$ such that, when run on some polynomial $t = t(n)$ instances of length n, it has compression rate $\rho = \rho(n, t) < 1$. Then there exists a protocol (P'_a, V'_a) for $\mathcal{R}$, where P' and V' are uniform polynomial time algorithms with advice, such that there exists a polynomial-size advice string a for which the following hold:*

- *(P'_a, V'_a) is an* honest-verifier *SWI protocol with statistical distance $O(\sqrt{\rho})$.*
- *(P'_a, V'_a) has the same completeness error as Π run on t instances.*
- *If Π is statistically sound, then for any advice a' of size $|a|$, $(P'_a, V'_{a'})$ is statistically sound with the same soundness error as Π on t instances.*

- *If* Π *is computationally sound, then for any advice* a' *of size* $|a|$, $(P'_a, V'_{a'})$ *is computationally sound.*
- *The communication and round complexity of* (P'_a, V'_a) *match those of* Π*, except that in the first round* V'_a *sends an additional uniform random string of some fixed polysize.*

We make two changes in the statement of the above theorem, when compared to [BKPRV24]. First, unlike [BKPRV24] we assume also that the verifier is non-uniform. The result stated in [BKPRV24] only require that the prover is non-uniform. This change is made so that we can reduce the amount of information sent by the prover. Specifically, it enables us to make the second change made explicit in the last item of Theorem 2.15, we now explain this change and explain why it follows form the result of [BKPRV24].

In the result stated in [BKPRV24], when running the protocol $\Pi'_a = (P'_a, V'_a)$ on instance x and witness w: The prover P'_a gets a string a as advice, where a describes a polynomial-size multiset $S = \{((x_1, y_1, z_1), \cdots, (x_t, y_t, z_t)\}$ such that for every $i \in [t]$, both $(x_i, y_i) \in \mathcal{R}$ and $(x_i, z_i) \in \mathcal{R}$. On execution of protocol, the prover samples at random $((x_1, y_1, z_1), \cdots, (x_t, y_t, z_t)) \leftarrow S$ and an index $i \leftarrow [t]$ and sends $(x_1, \cdots, x_t)$ and the index i to the prover. The prover and verifier then proceed to run the batch protocol for the instance $(x_1, \cdots, x_{i-1}, x, x_{i+1}, \cdots, x_t)$ where the verifier uses witness $(w_1, \cdots, w_{i-1}, w, w_{i+1}, \cdots, w_t)$, where for each $w_i \in [t]/\{i\}$ the prover samples a bit $b \leftarrow \{0,1\}$ and sets $w_i = y_i$ if $b = 0$ and $w_i = z_i$ if $b = 1$, and $w_i = w$ is the original witness given as input to the prover.

Observe that in the above execution of the protocol since the protocol gives a statistical guaranty and we are aiming for an honest-verifier, instead of just sending $(x_1, \cdots, x_t)$ the random instances to the verifier, the prover and verifier could both get the advice a that describes the entire multiset S and then have the verifier V_a sample at random instancies and random index i used in the execution. The prover can then on its own sample a random witness for each of the instances from the set S, and they can now jointly run the batch protocol using these instances and witness as explained above. Notice that this change, matches our description in the last item in the statement of Theorem 2.15, namely, the amount of communication required from the prover consists only of the running the batch protocol.

3 Deterministic-Prover Witness Indistinguishability and Witness Canonicalization

In this section we show that, assuming auxiliary-input one-way functions do not exist, any honest-verifier WI protocol can be transformed into a new honest-verifier SWI protocol with a deterministic prover. We then use this transformation to derive our main consequences, including the connection to witness canonicalization and extensions to batch and succinct arguments.

WI protocol + $\nexists$ AI-OWF $\Rightarrow$ deterministic-prover WI protocol. The following theorem states that given an WI protocol either AI-OWFs exist or there exits a new protocol that is (1/poly)-SWI and has a deterministic prover.

Theorem 3.1. *Let* (P, V) *be an honest verifier* ε*-witness indistinguishable argument [resp., proof] for a relation* $\mathcal{R}$ *with* $(1-\alpha)$*-completeness and* β*-soundness. Let* ℓ *denote a polynomial upper bound on the length of the transcript, let* ℓ_P *denote a polynomial upper bound on the total communication sent by* P *in each round, and let* k *denote the number of rounds. Then at least one of the following holds:*

1. *The function family* $\{f_a\}_{a \in \{0,1\}^{m(n)}}$ *is an infinitely often distributional auxiliary-input one-way function, where for* $a = (x, w_0, w_1) \subseteq \{0,1\}^{m(n)}$ *with* $(x, w_0) \in \mathcal{R}_n$ *and* $(x, w_1) \in \mathcal{R}_n$*, and for every bit* $b \in \{0,1\}$*, index* $i \in [\ell]$*, and prover and verifier randomness* r_P *and* r_V*:*

$$f_a(i, b, r_P, r_V) \stackrel{\text{def}}{=} \langle P(x, w_b; r_P), V(x; r_V)\rangle_{\leq i}$$

2. *For every polynomial* q*, there exists a protocol* $(\widehat{P}, \widehat{V})$*, such that* $\widehat{P}$ *is a* deterministic *polynomial-time algorithm, and* $(\widehat{P}, \widehat{V})$ *is an honest verifier* γ*-statistical witness indistinguishable argument [resp., proof] for* $\mathcal{R}$ *with* $(1 - \alpha - 1/q)$*-completeness and* β*-soundness for* $\gamma \leq k \cdot \ell_P \cdot 3\varepsilon + 1/q$*.*

Moreover, $\widehat{P}$ has the same per-round communication upper bound as P and if (P, V) is public-coin, then so is $(\widehat{P}, \widehat{V})$.

That is, either the function that, on input the index i, the bit b, and the randomness of P and V, outputs the first i bits of the transcript of the protocol execution on instance (x, w_b) is an ioAI-OWF, or the protocol can be transformed into a new WI protocol with a deterministic prover. Note that by standard OR-parallel repetition (see Claim 2.12) we can make the completeness error in Theorem 3.1 negligible. In particular, if we start with a WI protocol with negligible WI, completeness and soundness errors, then (if ioAI-OWF don't exist) for any $\gamma \in$ poly there exists a *deterministic prover* γ-SWI with negligible soundness and completeness errors.

The proof of Theorem 3.1 is given in Section 3.4.

Deterministic-prover WI to WC. We now relate deterministic-prover WI to the notion of witness canonicalization (WC). We will make use of the following lemma, proved in Section 3.6

Lemma 3.2 (Round Collapse)**.** *Assume there is a k-round honest-verifier deterministic-prover δ-SWI public-coin interactive argument for $\mathcal{R}$ with $(1-\alpha)$-completeness and ϵ-soundness errors, prover-to-verifier communication $\ell_P \in$ poly and total communication complexity ℓ where k is a constant and $\epsilon < \frac{1}{64}$. Then either infinitely-often auxiliary-input OWFs exist, or there exists a two-message (AM) honest-verifier deterministic-prover $(\delta(n)\cdot(\text{poly}(\ell_P, \log(n)))^k)$-SWI public-coin interactive argument for $\mathcal{R}$ with negligible soundness error, $(1-\alpha\cdot\text{poly}(\ell_P, \log(n))^k)$-completeness and (total) communication complexity $\ell\cdot(\text{poly}(\ell_P, \log(n)))^k$.*

Moreover, if the argument is ϵ statistically *sound, then the resulting protocol is statistically sound with negligible error.*

The construction of WC from WI is by first applying Theorem 2.11 to obtain a public-coin protocol, then applying Theorem 3.1, and finally collapsing to two rounds using Lemma 3.2, we obtain the following corollary. The last step uses the fact that every two-round deterministic-prover honest-verifier SWI protocol yields a WC.

Theorem 3.3 (WI with no WC implies ioAI-OWF)**.** *Let $\mathcal{R}$ be an NP relation for which there is no ε-WC for some $\varepsilon \in 1/\text{poly}$ [resp., statistically sound ε-WC]. If there exists a constant-round honest-verifier witness indistinguishable interactive argument [resp., interactive proof] for $\mathcal{R}$, then infinitely-often auxiliary-input one-way functions exist.*

Proof of Theorem 3.3. Let k be a constant, and let $\mathcal{R}$ be an NP relation with $(1-\alpha)$-complete, β-sound, k-round, honest-verifier δ-WI indistinguishable argument [resp., proof] (P, V), for negligible α, β and δ. Assume that infinitely-often auxiliary-input one-way functions do not exists. We will construct ε-WC [resp., statistically sound ε-WC] for $\mathcal{R}$.

For concreteness and simplicity the proof below is stated for computational soundness but the same argument holds for statistical soundness as well.

Let ℓ_P be the total prover-to-verifier communication in (P, V), and let $\epsilon' \in 1/\text{poly}$ to be chosen later. First, we use Theorem 2.11 to construct a *public-coins* constant-round honest-verifier witness indistinguishable argument (P', V'). By Theorem 2.11 we get $(1-\alpha-\epsilon') \geq (1-2\epsilon')$-complete, $\beta+\epsilon' \leq 2\epsilon'$-sound honest-verifier k-round $\delta+\epsilon' \leq 2\epsilon'$-WI argument for $\mathcal{R}$, (P', V'), with the same prover-to-verifier communication ℓ_P.

We next make the prover a *deterministic* prover. By Theorem 3.1, there exists a $(1-3\epsilon')$-complete, $2\epsilon'$-sound honest-verifier k-round $8(k\ell_P)\epsilon'$-SWI argument (P'', V'') with deterministic prover for $\mathcal{R}$, with the same prover-to-verifier communication ℓ_P.

Next, we reduce the number of messages to 2. By Lemma 3.2, we get that assuming $2\epsilon' \leq 1/64$, there exists a $(1-3\epsilon'\cdot T)$-complete, negl-sound honest-verifier 2-message, public-coin, $8T(k\ell_P)\epsilon'$-SWI argument for $\mathcal{R}$, (P''', V'''), for $T = \text{poly}(\ell_P, \log n)^k \in$ poly.

Lastly, set $t = \lceil \log n \rceil$ and take $\epsilon' := \frac{\varepsilon}{128tTk\ell_P}$. By Claim 2.12, $(\widehat{P}, \widehat{V})$ defined to be the t-fold OR-parallel repetition of (P''', V''') has $((3T\epsilon')^t = \text{negl})$-completeness, negl-soundness and has SWI error at most $8tT(k\ell_P)\epsilon' \leq \varepsilon$. Moreover, $(\widehat{P}, \widehat{V})$ preserves the number of rounds (two-rounds), public-coin, and the determinism of the prover.

We finally construct a WC (map, V) for $\mathcal{R}$. Let r denote the uniform randomness chosen as the first random message of the verifier $\widehat{V}$ and not that the length of r is polynomial. Let $\mathsf{map}(r, x, w) :=$

$\widehat{P}(x, w, r)$ be the message generated by the prover, and let $V(r, x, \pi)$ be the function computed by the verifier once it receives the provers message. Completeness follows directly from the completeness of $(\widehat{P}, \widehat{V})$, and the same holds for soundness, with the appropriate variant inherited from $(\widehat{P}, \widehat{V})$: computational soundness in the computational case, and statistical soundness in the statistical case. Regarding uniqueness, for each $x \in \mathcal{L}(\mathcal{R})$, let w_0 be some canonical witness, and let $\pi_{x,r} = \mathsf{map}(r, x, w_0)$. Then for any w such that $(x, w) \in \mathcal{R}$, we have that $\mathbf{Pr}_r[\mathsf{map}(r, x, w) \neq \pi_{x,r}] = \mathbf{Pr}_r[\mathsf{map}(r, x, w) \neq \mathsf{map}(r, x, w_0)] \leq \epsilon$. □

The rest of this section is organized as follows.

- In Section 3.1 we show how to extend the above stated results for batch and succinct arguments.
- In Section 3.2 we give some additional preliminaries needed for the proof of Theorem 3.1.
- In Section 3.3 we state our main lemma needed for the proof of Theorem 3.1, regarding canonical simulation. This lemma is proven in Section 3.5.
- In Section 3.4 we finally prove Theorem 3.1.
- Lastly, in Section 3.6 we prove our round collapse result (Lemma 3.2).

3.1 Extensions to Batch Arguments and Succinct Arguments.

We next extend the above implications to batch arguments and succinct arguments. The result of Bitansky, Kamath, Paneth, Rothblum, and Vasudevan, stated in Theorem 2.15, showed that sufficiently strong compression in batch arguments yields witness indistinguishability. Using this connection together with Theorem 3.1 we conclude that sufficiently strong compression or succinctness implies the existence of auxiliary-input one-way functions whenever deterministic-prover SWI arguments do not exist.

Corollary 3.4 (Batch arguments to deterministic-prover)**.** *There exists a universal constant $c > 0$ such that the following holds. Let $\mathcal{R}$ be an NP relation for which, for some $\varepsilon \in 1/\mathrm{poly}$, there is no non-uniform honest-verifier ε-SWI argument [resp., proof] with a deterministic prover. Both parties may receive the same fixed polynomial-length advice a_n. If one of the following exists:*

- *A k-round batch argument [resp., proof] for $\mathcal{R}$ with $\rho(t, n)$-compression rate for $\rho(t, n) < (\frac{\varepsilon}{c \cdot k \cdot t})^{\frac{2}{3}}$.*
- *A γ-succinct interactive argument [resp., proof] for the relation $\mathcal{R}^{\otimes t}$ for $\gamma < 1/3$ and a sufficiently large $t \in \mathrm{poly}(n)$.*

Then there exists an infinitely often auxiliary-input one-way function (AI-OWF).

Proof of Corollary 3.4. For simplicity we prove the result for interactive arguments however the same proof below works for interactive proofs. Assume towards contradiction that there is no infinitely often auxiliary-input one-way function. We show that each of the two items implies a non-uniform honest-verifier ε-SWI argument for $\mathcal{R}$ with a deterministic prover, contradicting the assumption of the corollary.

We start with the first item. Let $\Pi = (P, V)$ be a k-round batch argument for $\mathcal{R}$ with compression rate $\rho = \rho(t, n)$ satisfying

$$\rho(t, n) < \left(\frac{\varepsilon}{c \cdot k \cdot t}\right)^{2/3}$$

for some $\varepsilon \in 1/\mathrm{poly}$.

Apply Theorem 2.15 to Π. We obtain an honest-verifier SWI protocol (P'_a, V'_a) (where a is a the non-uniform advice) for $\mathcal{R}$ with statistical distance

$$\varepsilon_{\mathsf{WI}} \leq C\sqrt{\rho(t, n)}$$

for some absolute constant C, and prover communication is still at most

$$\ell_P \leq \rho(t, n) \cdot t,$$

and number of rounds k.

We apply the proof of Theorem 3.1, including the advice in the auxiliary input of its inversion family. More precisely, use the auxiliary input (a, x, w_0, w_1), with the transcript computed by (P'_a, V'_a). Evaluation is uniform given this input. The absence of ioAI-OWFs supplies inversion guarantees uniformly over these auxiliary inputs. We then fix the good advice sequence a_n supplied by Theorem 2.15. Consequently, for every polynomial q there is a non-uniform deterministic-prover honest-verifier δ-SWI protocol for $\mathcal{R}$, where

$$\delta \leq 3k\ell_P\varepsilon_{\mathsf{WI}} + \frac{1}{q}.$$

Substituting the bounds on ℓ_P and $\varepsilon_{\mathsf{WI}}$, we get

$$\delta \leq 3k \cdot (\rho t) \cdot C\sqrt{\rho} + \frac{1}{q} = 3Ckt\rho^{3/2} + \frac{1}{q}.$$

By the assumption on ρ, it holds that $\rho^{3/2} < \frac{\varepsilon}{c \cdot k \cdot t}$, and therefore

$$\delta < \frac{3C}{c}\varepsilon + \frac{1}{q} \leq \varepsilon.$$

where the last inequality follows by choosing the constant c in the statement to be sufficiently large, and taking q to be a sufficiently large polynomial.

Both parties retain the fixed advice a_n. Let α be the original completeness error, so the compiled protocol has completeness error at most $\alpha + 1/q$. To obtain perfect completeness, append the following round:

1. The verifier sends its original acceptance decision.
2. On rejection, the prover sends its witness; on acceptance, it sends a fixed dummy response. The verifier accepts if it originally accepted or the supplied witness verifies.

The prover remains deterministic with advice, and soundness is unchanged. For each witness, the honest-verifier view differs from the compiled view extended by the decision and a dummy response with probability at most $\alpha + 1/q$. Thus the new honest-verifier SWI error is at most

$$\delta + 2(\alpha + 1/q) < \frac{3C}{c}\varepsilon + 2\alpha + \frac{3}{q}.$$

Taking $c \geq 12C$ and $q \geq 12/\varepsilon$, this is at most ε for all sufficiently large input lengths, since α is negligible. Finitely many exceptional lengths can be handled by exhaustive canonical-witness search. We have therefore obtained a non-uniform honest-verifier ε-SWI argument with a deterministic prover, contradicting the premise.

We next consider the second item. Let (P, V) be a k-round γ-succinct interactive argument for $\mathcal{R}^{\otimes t}$, where $\gamma < 1/3$. Since for instances of $\mathcal{R}$ we have $|x| = n$ and $|w| = \mathrm{poly}(n)$, there exists a constant $d > 0$ such that $|x| + |w| \leq n^d$ for all relevant instance-witness pairs. Hence, for a batch of t instances, the total size of the input and witness is at most tn^d, and by γ-succinctness the total interaction length is at most

$$(tn^d)^\gamma.$$

Viewing (P, V) as a batch protocol for $\mathcal{R}$, the induced compression rate is therefore

$$\rho(n, t) \leq \frac{(tn^d)^\gamma}{t} = n^{d\gamma}t^{\gamma-1}.$$

Fix any inverse polynomial $\varepsilon(n) = n^{-a}$. To satisfy the first item, it is enough that $n^{d\gamma}t^{\gamma-1} < \left(\frac{\varepsilon(n)}{c \cdot k \cdot t}\right)^{2/3}$, or equivalently,

$$n^{d\gamma}t^{\gamma-1/3} < \left(\frac{\varepsilon(n)}{c \cdot k}\right)^{2/3}.$$

Now choose $t = n^{c_t}$ for a sufficiently large constant c_t. Since $\varepsilon(n) = n^{-a}$, and writing $k(n) \leq n^b$ for some constant $b \geq 0$, the above inequality is implied by $d\gamma + c_t(\gamma - 1/3) < -\frac{2(a+b)}{3}$, or equivalently

$$c_t\left(\frac{1}{3} - \gamma\right) > d\gamma + \frac{2(a+b)}{3}.$$

Since $\gamma < 1/3$, the coefficient $(1/3 - \gamma)$ is positive, and therefore this holds for all sufficiently large constants c_t. Thus, for sufficiently large polynomial $t = n^{c_t}$, the succinct protocol induces a batch protocol satisfying the first item. □

Corollary 3.5 (Batch arguments to witness canonicalization)**.** *There exists a universal constant $c > 1$ such that the following holds. If a relation $\mathcal{R}$ does not have a non-uniform ε-WC argument for some $\varepsilon \in 1/\mathrm{poly}$, then, if one of the following exists:*

- *A k-round batch argument for $\mathcal{R}$, for a constant k, such that when batch proving a sufficiently large number of instances $t \in \mathrm{poly}(n)$, the prover-to-verifier communication satisfies $\ell_P \leq t^\gamma$, for some constant $\gamma < 1/ck$.*
- *A k-round γ-succinct interactive argument for $\mathcal{R}^{\otimes t}$, for a constant k, $\gamma < 1/ck$, and a sufficiently large $t \in \mathrm{poly}(n)$.*

Then there exists an infinitely often auxiliary-input one-way function (AI-OWF).

Proof. The proof is similar to the proofs of Theorem 3.3 and Corollary 3.4. In particular, we first use Theorem 2.15 to construct an SWI argument with a non-uniform verifier. We retain the polynomial-size advice a_n for both parties, fixed for each instance length n, throughout the transformations below. We then use Theorem 2.11 to get a public-coin SWI protocol, and then continue similarly to the proof of Corollary 3.4 to get a public-coin SWI protocol with a deterministic prover and non-uniform verifier.

Next, we apply the round-reduction lemma, similarly to the proof of Theorem 3.3, to get a 2-message public-coin SWI protocol with a deterministic prover and non-uniform verifier. For $s := \lceil \log n \rceil$, we choose the parameters so that this protocol has (ε/s)-SWI error. We then apply the s-fold OR-parallel repetition from Claim 2.12. The final protocol we get has negligible completeness and soundness errors and is ε-SWI.

Similar to Theorem 3.3, we interpret this final protocol as a non-uniform ε-WC, where both map and V retain the advice a_n.

Thus, by the proof Theorem 3.3 to obtain a non-uniform ε-WC after the s-fold repetition, we need to start with an ε'-SWI protocol with k rounds and prover-to-verifier communication ℓ_P, where $\varepsilon' = O\left(\frac{\varepsilon}{s \cdot T \cdot k\ell_P}\right)$ and $T = \left(\mathrm{poly}(\ell_P, \log n)\right)^k$. By Theorem 2.15, this is achievable if

$$\sqrt{\frac{\ell_P}{t}} \leq O\left(\frac{\varepsilon}{s \cdot T \cdot k\ell_P}\right).$$

Equivalently, it suffices that $\ell_P^3 T^2 \leq O\left(\frac{\varepsilon^2 t}{s^2 k^2}\right)$. Since $T \leq \left(\ell_P \log n\right)^{dk}$ for some universal constant d and sufficiently large n, it suffices that

$$\ell_P^{2dk+3} (\log n)^{2dk+2} \leq O\left(\frac{\varepsilon^2 t}{k^2}\right). \tag{9}$$

Since $\varepsilon \in 1/\mathrm{poly}$, by taking c to be a sufficiently large constant it follows that if $\ell_P \leq t^\gamma$ for $\gamma < 1/ck$ then Eq. (9) holds.

The second item follows as in the proof of Corollary 3.4. Indeed, fix a constant γ' such that $\gamma < \gamma' < 1/(ck)$. If the protocol is γ-succinct, then

$$\ell_P \leq (tn^a)^\gamma$$

for some constant a. For a sufficiently large polynomial choice of t, we have $(tn^a)^\gamma \leq t^{\gamma'}$, and the conclusion follows from the first item. □

3.2 Next-bit Predictors, Samplers, and Simulators

In this section, we introduce next-bit predictors which we will use to define canonical samplers and simulators for distributions over binary strings. In particular, we will use the simulator to establish our results on witness indistinguishability. We start with definitions and notations we need for the proof.

3.2.1 Next-bit Predictors and OWFs

Informally, a next-bit predictor approximates the conditional probability of the next bit given the current prefix.

Definition 3.6 (Next-Bit Predictor)**.** *For a distribution Z over $\{0,1\}^m$, we say that a function $\mathsf{Pred}\colon \{0,1\}^* \times \{0,1\}^{\ell_{\mathsf{Pred}}} \to [0,1]$, whose outputs are finite binary fractions, is a* next-bit ε-predictor *for Z, if for every $i \in [m]$*

$$\Pr_{\substack{z_{<i} \leftarrow Z_{<i} \\ r_{\mathsf{Pred}} \leftarrow \{0,1\}^{\ell_{\mathsf{Pred}}}}} \left[|\mathsf{Pred}(z_{<i}; r_{\mathsf{Pred}}) - p(z_{<i})| \leq \varepsilon \right] \geq 1 - \varepsilon$$

for $p(z_{<i}) = \mathbf{Pr}\left[Z_i = 1|_{Z_{<i}=z_{<i}}\right]$, with $p(z_{<i}) = 0$ when $\mathbf{Pr}\left[Z_{<i} = z_{<i}\right] = 0$.

For a collection of distributions $Z = \{Z_a\}_{a\in\{0,1\}^{\lambda(n)},n\in\mathbb{N}}$, we say that Pred *is a next-bit ε-predictor for the collection Z if there exists a $q \in$ poly such that for every sequence $\{a_n \in \{0,1\}^{\lambda(n)}\}_n$ and every sufficiently large $n \in \mathbb{N}$, the function $\mathsf{Pred}_{a_n}(\cdot) \coloneqq \mathsf{Pred}(a_n, \cdot)$ is a next-bit $\varepsilon(n)$-predictor for Z_{a_n} and the runtime of Pred_{a_n} is at most $q(n)$.*

We relate next-bit predictors to distributional one-way functions. Informally, if a function family is not a distributional auxiliary-input one-way function, then the induced output distributions admit efficient next-bit predictors. This connection, originated in the work of Impagliazzo and Luby [IL89].

Theorem 3.7 (Distributional AI-OWF and next-bit predictor, [IL89])**.** *Let $\{f_a : \{0,1\}^{m(n)} \to \{0,1\}^{\ell(n)}\}_{a\in\{0,1\}^{\lambda(n)}}$ be an efficiently computable function family and define the function family $\{f'_a\}_{a\in\{0,1\}^{\lambda(n)}}$ so that for every $i \in [\ell(n)]$ and $x \in \{0,1\}^{m(n)}$ it holds that $f'_a(i,x) \coloneqq f_a(x)_{\leq i}$. If $\{f'_a\}_{a\in\{0,1\}^{\lambda(n)}}$ is not an infinity often distributional one way function then for any $\mu \in 1/\mathrm{poly}$ there exists an μ-next bit predictor for the collection of distributions $\{f_a(X_n)\}_{a\in\{0,1\}^{\lambda(n)},n\in\mathbb{N}}$ where X_n is uniformly sampled at random from $\{0,1\}^{m(n)}$.*

3.2.2 Samplers and Simulators from Next-bit Predictors

We now define the notion of samplers and simulators using next-bit predictors. In particular, These predictors suffice to generate a distribution that is statistically close to the original one, by sampling the bits sequentially according to the estimated conditional probabilities.

Definition 3.8 (Samplers and Simulators)**.** *Given a next-bit predictor* Pred *for a distribution $Z \in \{0,1\}^m$ the* next-bit sampler $\mathrm{Samp}^{\mathsf{Pred}}\colon \{0,1\}^* \times \{0,1\}^{\ell_{\mathsf{Pred}}+\ell_S} \to \{0,1\}$ *is a function defined as follows: On input $z_{<i} \in \{0,1\}^{i-1}$ and $r = (r_{\mathsf{Pred}}, r_S) \in \{0,1\}^{\ell_{\mathsf{Pred}}+\ell_S}$, $\mathrm{Samp}^{\mathsf{Pred}}(z_{<i}, r) = 1$ if r_S belongs to the first $\mathsf{Pred}(z_{<i}, r_{\mathsf{Pred}}) \cdot 2^{\ell_S}$ stings in $\{0,1\}^{\ell_S}$ (by lexicographic order) and 0 otherwise.*[10]

The simulator $\mathrm{Sim}^{\mathsf{Pred}}$ *is a function that on input random coins $r = (r_1, \cdots, r_m)$ and initial state $\tilde{z}_0 \in \{0,1\}^*$, generates*

$$\tilde{z} = (\tilde{z}_1, \ldots, \tilde{z}_m)$$

where for every $i \in [m]$,

$$\tilde{z}_i = \mathrm{Samp}^{\mathsf{Pred}}(\tilde{z}_{<i}; r_i).$$

We refer to the distribution $\tilde{Z}$ generated by applying $\mathrm{Sim}^{\mathsf{Pred}}$ on a random input, as the simulated distribution *induced by* Pred *on initial state $\tilde{z}_0$.*

The corresponding simulator $\mathrm{Sim}^{\mathsf{Pred}}$ and sampler $\mathrm{Samp}^{\mathsf{Pred}}$ for a collection of distributions $Z = \{Z_a\}_{a\in\{0,1\}^{\lambda(n)},n\in\mathbb{N}}$ are defined analogously.

The next claim, proven in Section B.2, is standard and simply states the statistical distance between Z and the simulated distribution of Z (created from the next-bit predictor of Z).

Claim 3.9 (Statistical distance from simulation)**.** *Let $X = (Z, Y)$ be a distribution over $\{0,1\}^{n_0} \times \{0,1\}^n$, and let* Pred *be a next-bit μ-predictor for X. Let $\tilde{Y} \equiv \mathrm{Sim}^{\mathsf{Pred}}(Z, R)$ be the random variable generated by the simulated distribution of* Pred *on initial state Z and independent uniform stings $R = (R_1, \ldots, R_n)$. Then for every $i \in [n+1]$,*

$$\mathrm{SD}((Z, Y_{<i}), (Z, \tilde{Y}_{<i})) \leq 2 \cdot (i-1) \cdot \mu.$$

[10] We assume without lose of generality that ℓ_S is sufficiently large and in particular larger than the runtime of Pred so that $\mathsf{Pred}(z_{<i}, r_{\mathsf{Pred}}) \cdot 2^{\ell_S} \in \mathbb{N}$, and there are no rounding errors. Note that given a bound on the runtime of Pred, finding such an ℓ_S can be done in polytime.

3.2.3 From Computational to Statistical WI

A classical well know fact is that if one-way functions do not exists then computational indistinguishability of samplable distributions implies statistical indistinguishability. In the following corollary, we show that if the specific function family associated with the transcript distribution of a WI protocol is not a distributional auxiliary-input one-way function, then the protocol is in fact *statistically* witness indistinguishable (SWI).

Lemma 3.10 (From computational to statistical WI, distributional OWFs)**.** *Let (P, V) be a computational honest-verifier ϵ-WI for a relation $\mathcal{R}$ with an efficient prover (given a witness), and let $\ell(n)$ denote the polynomial that bounds the length of the transcript of the protocol. Consider the function family $\{f_a\}_{a\in\{0,1\}^{m(n)}}$ defined for $a = (x, w_0, w_1) \subseteq \{0,1\}^{m(n)}$ with $(x, w_0) \in \mathcal{R}_n$ and $(x, w_1) \in \mathcal{R}_n$, and for every bit $b \in \{0,1\}$, index $i \in [\ell(n)]$ and prover and verifier randomness r_P and r_V:*

$$f_a(i, b, r_P, r_V) \stackrel{\text{def}}{=} \langle P(x, w_b; r_P), V(x; r_V)\rangle_{\leq i}.$$

If $\{f_a\}_a$ is not an infinitely often distributional auxiliary-input one-way function, then the protocol (P, V) is a honest-verifier ϵ'–SWI, for $\epsilon' = \epsilon + \mathsf{negl}$.

Lemma 3.10 is proven in Section B.3.

3.3 Canonical Simulation

Recall that for a distribution $Z \in \{0,1\}^n$ the next-bit predictor Pred for Z guaranties that for every $i \in [n]$ on average Pred can predict the probability that $Z_i = 1$ given $Z_{<i}$ as an input. Recall that every next-bit predictor Pred defines also a simulator $\mathrm{Sim}^{\mathsf{Pred}}$ (see Definition 3.8 for the formal statement). The next lemma is our main observation for this section.

Lemma 3.11 (Canonical simulation)**.** *Let $X^0 = (Z, Y^0)$ and $X^1 = (Z', Y^1)$ be two distribution over $\{0,1\}^{n_0} \times \{0,1\}^n$, and let Pred^0 and Pred^1 be a next-bit μ-predictor such that for $b \in \{0,1\}$, Pred^b is a predictor for X^b, and let Sim^b denote $\mathrm{Sim}^{\mathsf{Pred}^b}$. If $\mathrm{SD}(X^0, X^1) \leq \varepsilon$ then it holds that,*

$$\Pr_{\substack{z\leftarrow Z \\ r\leftarrow(\{0,1\}^\ell)^n}}\left[\mathrm{Sim}^0(z, r) \neq \mathrm{Sim}^1(z, r)\right] \leq n \cdot \beta + n \cdot (n-1) \cdot \mu$$

for $\beta = 3\varepsilon + 4\mu$.

We restate and prove Lemma 3.11 in Section 3.5. Loosely speaking, the above lemma states that for any two statistically close distributions Z^0 and Z^1 with next-bit predictors Pred^0 and Pred^1, the simulator Sim we defined in Definition 3.8 specified by an oracle access to a next-bit predictor is a *canonically simulator* in the following sense: When defining the two simulators $\mathrm{Sim}^0 := \mathrm{Sim}^{\mathsf{Pred}^0}$ and $\mathrm{Sim}^1 := \mathrm{Sim}^{\mathsf{Pred}^1}$ it holds that for a uniform choice of randomness r, $\mathrm{Sim}^b(r)$ is statistically close to Z^b, and furthermore $\mathrm{Sim}^0(r) = \mathrm{Sim}^1(r)$ with high probability over the choice of r. Which means that with high probability over r, given a sample $\mathrm{Sim}^b(r)$ it is not likely to identify the bit $b \in \{0,1\}$ even if r is also revealed. We use this property to prove our result for a deterministic prover.

3.4 Proof of Theorem 3.1

In this section, we show that if AI-OWFs do not exists then WI protocols could be turned into a SWI protocols with a deterministic provers. Specifically, we prove Theorem 3.1.

Proof of Theorem 3.1. Recall that $\langle P(x, w_b; r_P), V(x; r_V)\rangle$ denotes the transcript of the interaction between P and V on instance x witness w and where P uses randomness r_P and V uses randomness r_V and let ℓ denote the polynomial upper bound on the length of the transcript. Consider the function defined as follows: for every $a = (x, w_0, w_1)$,

$$f_a(i, b, r_V, r_P) = \langle P(x, w_b; r_P), V(x; r_V)\rangle_{\leq i}$$

First note that since we are dealing with an efficient prover, by Lemma 3.10 we can assume without lose of generality that the protocol we are given is in fact SWI (and not just WI).

Thus, we now assume that the WI we are given is in fact SWI. Since f_a is not infinitely often distributional AI-OWF by Theorem 3.7 for every $\mu \in 1/\text{poly}$, there exists a PPT algorithms Pred, such that $\mathsf{Pred}(x, w, \cdot)$ is a next-bit μ-predictor for the distribution induced by

$$T = \langle P(x, w; R_P), V(x; R_V) \rangle$$

where R_V and R_P are uniformly and independently sampled. This holds since we can consider the auxiliary-input $a = (x, w, w)$ with the *same* witness w, which implies that for random bit B, $f_{(x,w,w)}(I, B, R_P, R_V) \equiv T_{\leq I}$.

We now define the new protocol $(\widehat{P}, \widehat{V})$ as follows:

- **Settings and parameters:** Recall that (P, V) is a k-round protocol. Let ℓ_P denote an upper bound (in bits) on the total communication sent by the prover in each round, and let ℓ_{r_V} be a polynomial bound on the randomness used by V.

  We construct a new protocol $(\widehat{P}, \widehat{V})$ in which the prover $\widehat{P}$ is a deterministic polynomial-time algorithm.

  Fix a parameter $\mu \in 1/\text{poly}$ to be specified later, and let Pred be the next-bit μ-predictor guaranteed as above.

  Let ℓ_{Sim} be a polynomial bounding the randomness required to simulate the relevant transcript distributions using Pred. Concretely, for every (x, w) define

  $$\text{Sim}^{x,w} \coloneqq \text{Sim}^{\mathsf{Pred}(x,w,\cdot)}$$

  as in Definition 3.8.

- **The new verifier $\widehat{V}$:** On input randomness $r_{\widehat{V}} = (r_V, r_{\text{Sim}}) \in \{0,1\}^{\ell_{r_V} + \ell_{\text{Sim}}}$ such that $r_V = (r_{V,1}, \cdots, r_{V,k})$. The new verifier mimics the original verifier with the exception that on the first round it also sends the simulation randomness r_{Sim}.

  That is, in round $i = 1$, $\widehat{V}_1(x, \emptyset; r_{V,1})$ sends $(r_{\text{Sim}}, \hat{\rho}_1 \coloneqq \rho_1)$ where $(\rho_1, \mathsf{st}_{V,1}) = V_1(x, \emptyset; r_{V,1})$. On round $i \in [k]$ for $i > 1$, $\widehat{V}_i$ sends $\hat{\rho}_i = \rho_i$ where

  $$(\rho_i, \mathsf{st}_{V,i}) = V_i(x, \mathsf{st}_{\widehat{V},i-1}, \hat{m}_{i-1}; r_{V,i})$$

  where $\hat{m}_{i-1}$ is the messaged received from $\widehat{P}$ on round $i-1$, the new verifier then updates its state according to the original verifier V, that is, $\mathsf{st}_{\widehat{V},i} = \mathsf{st}_{V,i}$. The answer of $\widehat{V}$ is defined to be the same answer that the original verifier V would have given (that is, $V^*(x, \mathsf{st}_{\widehat{V},k}, \hat{m}_k)$).

- **The new prover $\widehat{P}$:** Recall that the prover is deterministic, and as such also stateless (and only depends on the transcript so far, and x and w). Let $r_{\text{Sim}} = (r_{\text{Sim},1}, \cdots, r_{\text{Sim},k})$ be the simulation randomness sent by $\widehat{V}$ on the first round, and let $\hat{\rho}_i$ be the message sent by $\widehat{V}$ on round i.

  Let

  $$\hat{\mathsf{tr}}_{<i} = (r_{\text{Sim}}, \hat{\rho}_1, \hat{m}_1, \cdots, \hat{m}_{i-1})$$

  denote the transcript until round $i-1$. The prover $\widehat{P}$, computes $\mathsf{tr}'_{<i} = (\hat{\rho}_1, \hat{m}_1, \cdots, \hat{m}_{i-1})$, (that is, that transcript without the simulation randomness) and applies $\text{Sim}^{x,w}$ with initial state $z = (\mathsf{tr}'_{<i}, \hat{\rho}_i)$ and randomness $r_{\text{Sim},i}$ to generate the simulated version of the next message output according to the original prover. That is, $\widehat{P}_i$ sends $\hat{m}_i = \text{Sim}^{x,w}(z, r_{\text{Sim},i})$.

We now argue about the properties of the new protocol.

Public-coin: First note that by construction if the original protocol is public-coin then the new protocol is public-coin as well.

Prover communication: By construction, in each round i, the prover $\widehat{P}$ outputs a simulated version of the ith message of the original prover P. Hence, $\widehat{P}$ sends a message of the same format and length as P would send in round i. It follows that $\widehat{P}$ satisfies the same per-round communication upper bounds as P, and in particular the total communication sent by $\widehat{P}$ is at most ℓ_P in each round.

Soundness: The new protocol has the same soundness as the original protocol. Moreover, the construction preserves the relevant notion of soundness: if the original protocol is computationally sound, then the resulting protocol is computationally sound, and if the original protocol is statistically sound, then the resulting protocol is statistically sound.

To see why it holds, suppose that (P, V) has soundness β and assume for contradiction that there exists a cheating prover P^* and $x \notin L_{\mathcal{R}}$ such that

$$\mathbf{Pr}\left[\mathsf{Out}(\langle P^*, \widehat{V}(x)\rangle) = 1\right] > \beta.$$

Define a cheating prover P' for the original verifier V as follows: on the first round when it receives ρ_1 it samples r_{Sim} uniformly by itself, and sends (r_{Sim}, ρ_1) to P^*, and relay P^*'s reply to V. And for any other round $i \in [k]$ when it receives ρ_i from V, it just send ρ_i to P^* and relay P^*'s reply to V. At the end, the projected transcript tr' seen by V is exactly the projection used by $\widehat{V}$, and the acceptance predicate is the same. Therefore

$$\mathbf{Pr}[\mathsf{Out}(\langle P', V(x)\rangle) = 1] \;=\; \mathbf{Pr}\left[\mathsf{Out}(\langle P^*, \widehat{V}(x)\rangle) = 1\right] \;>\; \beta,$$

contradicting the soundness of (P, V). Hence, $(\widehat{P}, \widehat{V})$ has the same soundness as (P, V).

Completeness. For a fixed $(x, w) \in \mathcal{R}$, let R_V be a random variable that denotes the randomness used by the verifier; we drop the subscript V and just write $R_V = R = (R_1, \cdots, R_k)$ to avoid clutter. Let $T = (Q_1, M_1, \cdots Q_k, M_k) = \langle P(x, w; R_P), V(x; R)\rangle$ be the distribution over the transcripts induces by an execution of the original protocol. Let $R_{\text{Sim}} = (R_{\text{Sim},1}, \cdots, R_{\text{Sim},k})$ be a random variable that denotes the random coins used by the simulator and let $\hat{T} = (\hat{Q}_1, \hat{M}_1, \cdots, \hat{Q}_k, \hat{M}_k) = \langle \widehat{P}(x, w), \widehat{V}(x; (R, R_{\text{Sim}}))\rangle$ be the distribution over the transcripts induces by an execution of the new protocol **without** including the simulation randomness. Recall also that $\hat{M}_i = \text{Sim}^{x,w}((\hat{T}_{<i}, \hat{Q}_i), R_{\text{Sim},i})$ and that μ is the error parameter in the next-bit predictor used to define the simulator.

First note that since the accepting predicate in the new protocol is the same as the original protocol, if we can show that

Claim 3.12. *For every $i \in [k]$, $\text{SD}((R_{\le i}, T_{\le i}), (R_{\le i}, \hat{T}_{\le i})) \le 2 \cdot \mu \cdot i \cdot \ell_P$.*

Then it follows that

$$\begin{aligned} &\mathbf{Pr}\left[\mathsf{Out}(\langle \widehat{P}(x, w), \widehat{V}(x; (R_{\text{Sim}}, R))\rangle) = 1\right] \\ &\ge \mathbf{Pr}\left[\mathsf{Out}(\langle P(x, w; R_P), V(x; R)\rangle) = 1\right] - \text{SD}((R, T), (R, \hat{T})) \\ &\ge 1 - \alpha - 2\mu \cdot k \cdot \ell_P. \end{aligned} \tag{10}$$

Where the last equality follows since ℓ_P is a bound on the total communication sent by the prover and k is the number of rounds. Finally for any fixed $q \in \text{poly}$ by taking μ to be a sufficiently small polynomial, we can show that

$$\mathbf{Pr}\left[\mathsf{Out}(\langle \widehat{P}(x, w), \widehat{V}(x)\rangle) = 1\right] \ge 1 - \alpha - 1/q.$$

Thus it remains to prove Claim 3.12.

Proof of Claim 3.12. We will prove the claim by induction.

We will show that for every $i \in [k]$, it holds that

$$\text{SD}((R_{\le i}, T_{\le i}), (R_{\le i}, \hat{T}_{\le i})) \le 2 \cdot \mu \cdot i \cdot \ell_P.$$

Note that for $i = 1$, $(R_{\le 1}, T_{\le 1}) = (R_1, Q_1, M_1)$ and $(R_{\le 1}, \hat{T}_{\le 1}) = (R_1, \hat{Q}_1, \hat{M}_1)$, and since $Q_1 \equiv \hat{Q}_1$ by Claim 3.9 it follows that $\text{SD}((Q_1, M_1), (\hat{Q}_1, \hat{M}_1)) \le 2\mu \cdot \ell_P$. For some $i \ge 1$ assume $\text{SD}((R_{<i}, T_{<i}), (R_{<i}, \hat{T}_{<i}) \le 2\mu \cdot (i-1) \cdot \ell_P$. Recall also that $(\hat{Q}_i, ST_{V,i}) = V_i(x, ST_{\widehat{V},i-1}, Q_{i-1}; R_i)$, and that this implies that there exists a function g_i such that $\hat{Q}_i = g_i(\hat{T}_{<i}, R_{\le i})$ and $Q_i = g_i(T_{<i}, R_{\le i})$ which by post-processing implies that

$$\text{SD}((R_{\le i}, T_{<i}, Q_i), (R_{\le i}, \hat{T}_{<i}, \hat{Q}_i)) \le 2\mu \cdot (i-1) \cdot \ell_P.$$

Let $Z \coloneqq (T_{<i}, Q_i)$ and note that by coupling Lemma 2.1, we can assume without lose of generality that $Z = (T_{<i}, Q_i) = (\hat{T}_{<i}, \hat{Q}_i)$ with probability $1 - 2\mu \cdot (i-1) \cdot \ell_P$. Moreover, since by Claim 3.9

$$\mathrm{SD}\Big((Z, M_i)\,,\ (Z, \hat{M}_i \coloneqq \mathrm{Sim}^{x,w}((Z; R_{\mathrm{Sim}_i})) \Big) \leq 2\mu \cdot \ell_P,$$

we conclude that $\mathrm{SD}((R_{\leq i}, T_{\leq i}), (R_{\leq i}, \hat{T}_{\leq i})) \leq 2\mu \cdot (i-1) \cdot \ell_P + 2\mu \cdot \ell_P \leq 2\mu \cdot i \cdot \ell_P$. □

Witness Indistinguishability. We argue that the protocol is also γ-SWI for $\gamma = k \cdot \ell_P \cdot 3\varepsilon + 1/q$. If (P, V) is an ε-SWI then by definition for every x, w_0 and w_1 such that $(x, w_0) \in \mathcal{R}$ and $(x, w_1) \in \mathcal{R}$ it holds that

$$\mathrm{SD}(T^0, T^1) \leq \varepsilon$$

where for every $b \in \{0,1\}$, we define $T^b = \langle P(x, w_b; R_P), V(x; R_V) \rangle$ where R_V and R_P are random uniformly and independently sampled.

Note that by construction for every $b \in \{0,1\}$:

$$\langle \widehat{P}(x, w_b), \widehat{V}(x; (R^b_{\mathrm{Sim}}, R^b)) \rangle = (R^b_{\mathrm{Sim}}, \hat{T}^b)$$

where $\hat{T}^b = (\hat{Q}^b_1, \hat{M}^b_1, \cdots, \hat{Q}_k, \hat{M}^b_k)$ denotes the transcript of the interaction when using randomness $(R^b_{\mathrm{Sim}}, R^b)$. Consider now the event $E \coloneqq \{(R^0_{\mathrm{Sim}}, R^0) = (R^1_{\mathrm{Sim}}, R^1)\}$ where the verifier tapes are sampled independently and uniformly.

To prove the protocol is γ-SWI it suffices to prove that,

Claim 3.13. $\mathbf{Pr}\left[\hat{T}^0 \neq \hat{T}^1 \mid E\right] \leq \gamma$.

That is, it suffices to show that the protocol is γ-unique (and note that this is a necessary condition, since the prover is deterministic). Which by coupling implies that

$$\mathrm{SD}(((R^0_{\mathrm{Sim}}, R^0), \hat{T}^0), ((R^1_{\mathrm{Sim}}, R^1), \hat{T}^1)) \leq \gamma.$$

Thus to conclude the proof, it suffices to prove Claim 3.13 which we do below.

Proof of Claim 3.13. For every $i \in [k]$ and $b \in \{0,1\}$ let $\hat{Z}^b_{<i}$ denote the transcript $\hat{T}^b$ till round $i-1$ with addition to $\hat{Q}_i$ the ith message of the verifier. That is $\hat{Z}^b_{<i} \coloneqq (\hat{T}^b_{<i}, \hat{Q}_i)$, and similarly let $Z^b_{<i} \coloneqq (T^b_{<i}, Q_i)$.

Recall that for every $i \in [k]$, $\hat{M}^b_i = \mathrm{Sim}^{x,w_b}(\hat{Z}^b_{<i}, R^b_{\mathrm{Sim},i})$. Note that conditioned on the event E it follows that, $\hat{Q}^0_i = \hat{Q}^1_i$ if $\hat{T}^0_{<i} = \hat{T}^1_{<i}$. Moreover by Claim 3.12, it holds that for every $i \in [k]$, $\mathrm{SD}(\hat{Z}^0_{<i}, Z^0_{<i}) \leq 2\mu \cdot (i-1) \cdot \ell_P$. Thus,

$$\begin{aligned}
\mathbf{Pr}\left[\hat{T}^0 \neq \hat{T}^1 | E\right] &\leq \mathbf{Pr}\left[\exists i \in [k] \text{ s.t. } \hat{T}^0_i \neq \hat{T}^1_i \ \cap\ \hat{T}^0_{<i} = \hat{T}^1_{<i} | E\right] \qquad (11)\\
&\leq \sum_{i=1}^{k} \mathbf{Pr}\left[\hat{M}^0_i \neq \hat{M}^1_i \ \cap\ \hat{Z}^0_{<i} = \hat{Z}^1_{<i} | E\right]\\
&\leq \sum_{i=1}^{k} \mathbf{Pr}\left[\mathrm{Sim}^{x,w_0}(\hat{Z}^0_{<i}, R^0_{\mathrm{Sim},i}) \neq \mathrm{Sim}^{x,w_1}(\hat{Z}^1_{<i}, R^0_{\mathrm{Sim},i}) \ \cap\ \hat{Z}^0_{<i} = \hat{Z}^1_{<i} | E\right]\\
&\leq \sum_{i=1}^{k} \mathbf{Pr}\left[\mathrm{Sim}^{x,w_0}(\hat{Z}^0_{<i}, R^0_{\mathrm{Sim},i}) \neq \mathrm{Sim}^{x,w_1}(\hat{Z}^0_{<i}, R^0_{\mathrm{Sim},i}) | E\right]\\
&\leq \sum_{i=1}^{k} \mathbf{Pr}\left[\mathrm{Sim}^{x,w_0}(\hat{Z}^0_{<i}, R^0_{\mathrm{Sim},i}) \neq \mathrm{Sim}^{x,w_1}(\hat{Z}^0_{<i}, R^0_{\mathrm{Sim},i})\right]\\
&\leq \sum_{i=1}^{k} \left(\mathbf{Pr}\left[\mathrm{Sim}^{x,w_0}(Z^0_{<i}, R^0_{\mathrm{Sim},i}) \neq \mathrm{Sim}^{x,w_1}(Z^0_{<i}, R^0_{\mathrm{Sim},i})\right] + 2\mu \cdot (i-1) \cdot \ell_P\right).
\end{aligned}$$

Finally by Lemma 3.11 it holds that for every $i \in [k]$,

$$\mathbf{Pr}\left[\mathrm{Sim}^{x,w_0}(Z^0_{<i}, R^0_{\mathrm{Sim},i}) \neq \mathrm{Sim}^{x,w_1}(Z^0_{<i}, R^0_{\mathrm{Sim},i})\right] \leq \ell_P \cdot (3\varepsilon + 4\mu) + {\ell_P}^2 \cdot \mu \qquad (12)$$

Putting Eqs. (11) and (12) together we get,

$$\mathbf{Pr}\left[\hat{T}^0 \neq \hat{T}^1 \mid E\right] \leq k \cdot \ell_P \cdot 3\varepsilon + \mu \cdot (k{\ell_P}^2 + (k^2+3k)\ell_P) \leq k \cdot \ell_P \cdot 3\varepsilon + 1/q \tag{13}$$

where the last equality follows by choosing μ to be sufficiently small.

□

This concludes the proof of Theorem 3.1. Keeping the individual prover-message lengths $\ell_{P,i}$ in the same analysis gives WI error at most $3\varepsilon L_P + 1/q$, where $L_P = \sum_{i=1}^{k} \ell_{P,i}$ is the total prover-to-verifier communication. The terms multiplied by μ remain polynomially bounded and are absorbed into $1/q$. □

3.5 Canonical Simulation: Proving Lemma 3.11

In this section we restate and prove Lemma 3.11. We make use of the following claims to prove our canonical simulation lemma stated below.

Claim 3.14. *Let Z^0 and Z^1 be two distributions over $\{0,1\}^n$ such that $\mathrm{SD}(Z^0, Z^1) \leq \varepsilon$. For every $b \in \{0,1\}$ and $i \in [n]$, let $p^b(z_{<i}) = \mathbf{Pr}\left[Z_i^b = 1 \mid Z_{<i}^b = z_{<i}\right]$. Then,*

$$\mathop{\mathbf{E}}_{z_{<i} \leftarrow Z^0_{<i}} \left[|p^0(z_{<i}) - p^1(z_{<i})|\right] \leq 2\varepsilon.$$

Claim 3.14 is proven in Section B.4.

Claim 3.15. *Let Z^0 and Z^1 be two distribution over $\{0,1\}^n$, and let Pred^0 and Pred^1 be a next-bit μ-predictor such that for every $b \in \{0,1\}$, Pred^b is a next-bit predictor for Z^b and let Samp^b denote $\mathrm{Samp}^{\mathsf{Pred}^b}$. The following holds: If $\mathrm{SD}(Z^0, Z^1) \leq \varepsilon$ then for every $i \in [n]$*

$$\mathop{\mathbf{Pr}}_{\substack{z_{<i} \leftarrow Z^0_{<i} \\ r \leftarrow \{0,1\}^{\ell_{\mathsf{Pred}} + \ell_S}}} \left[\mathrm{Samp}^0(z_{<i}; r) \neq \mathrm{Samp}^1(z_{<i}; r)\right] \leq 3\varepsilon + 4\mu$$

Proof of Claim 3.15. For every $b \in \{0,1\}$, let $p^b(z_{<i}) = \mathbf{Pr}\left[Z_i^b = 1 | Z_{<i}^b = z_{<i}\right]$. By Claim 3.14

$$\mathop{\mathbf{E}}_{z_{<i} \leftarrow Z^0_{<i}} \left[|p^0(z_{<i}) - p^1(z_{<i})|\right] \leq 2\varepsilon. \tag{14}$$

By definition of next-bit predictors for every $b \in \{0,1\}$,

$$\mathop{\mathbf{Pr}}_{\substack{z_{<i} \leftarrow Z^b_{<i} \\ r_{\mathsf{Pred}} \leftarrow \{0,1\}^{\ell_{\mathsf{Pred}}}}} \left[\left|\mathsf{Pred}^b(z_{<i}; r_{\mathsf{Pred}}) - p^b(z_{<i})\right| > \mu\right] \leq \mu,$$

Which implies that

$$\mathop{\mathbf{E}}_{\substack{z_{<i} \leftarrow Z^b_{<i} \\ r_{\mathsf{Pred}} \leftarrow \{0,1\}^{\ell_{\mathsf{Pred}}}}} \left[\left|\mathsf{Pred}^b(z_{<i}; r_{\mathsf{Pred}}) - p^b(z_{<i})\right|\right] \leq 2\mu,$$

and since $\mathrm{SD}(Z^0, Z^1) \leq \varepsilon$, it follow that:

$$\mathop{\mathbf{E}}_{\substack{z_{<i} \leftarrow Z^0_{<i} \\ r_{\mathsf{Pred}} \leftarrow \{0,1\}^{\ell_{\mathsf{Pred}}}}} \left[|\mathsf{Pred}^1(z_{<i}; r_{\mathsf{Pred}}) - p^1(z_{<i})|\right] \leq 2\mu + \varepsilon.$$

Applying union bound and triangle inequality of the above equation and Eq. (14), it follows that

$$\mathop{\mathbf{E}}_{\substack{z_{<i} \leftarrow Z^0_{<i} \\ r_{\mathsf{Pred}} \leftarrow \{0,1\}^{\ell_{\mathsf{Pred}}}}} \left[|\mathsf{Pred}^0(z_{<i}; r_{\mathsf{Pred}}) - \mathsf{Pred}^1(z_{<i}; r_{\mathsf{Pred}})|\right] \leq 3\varepsilon + 4\mu. \tag{15}$$

Finally by definition, for every fixed $z_{<i}$ and r_{Pred} such that $|\mathsf{Pred}^0(z_{<i}; r_{\mathsf{Pred}}) - \mathsf{Pred}^1(z_{<i}; r_{\mathsf{Pred}})| = q \in [0,1]$,

$$\mathop{\mathbf{Pr}}_{r_S \leftarrow \{0,1\}^{\ell_S}} \left[\mathrm{Samp}^0(z_{<i}; r_{\mathsf{Pred}}, r_s) \neq \mathrm{Samp}^1(z_{<i}; r_{\mathsf{Pred}}, r_s)\right] = q \tag{16}$$

Putting Eqs. (15) and (16) together, we get that

$$\begin{aligned}\mathop{\mathbf{Pr}}_{\substack{z_{<i}\leftarrow Z^0_{<i}\\ r\leftarrow\{0,1\}^{\ell_{\mathsf{Pred}}+\ell_S}}}\left[\mathrm{Samp}^0(z_{<i};r)\neq\mathrm{Samp}^1(z_{<i};r)\right] &= \mathop{\mathbf{E}}_{\substack{z_{<i}\leftarrow Z^0_{<i}\\ r_{\mathsf{Pred}}\leftarrow\{0,1\}^{\ell_{\mathsf{Pred}}}}}\left[\left|\mathsf{Pred}^0(z_{<i};r_{\mathsf{Pred}})-\mathsf{Pred}^1(z_{<i};r_{\mathsf{Pred}})\right|\right]\\ &\leq 3\varepsilon+4\mu.\end{aligned}$$

This concludes the proof. □

Proof of Lemma 3.11. We are now ready to prove Lemma 3.11, restated below.

Lemma 3.16 (Canonical simulation, restating Lemma 3.11)**.** *Let $X^0=(Z,Y^0)$ and $X^1=(Z',Y^1)$ be two distribution over $\{0,1\}^{n_0}\times\{0,1\}^n$, and let Pred^0 and Pred^1 be a next-bit μ-predictor such that for $b\in\{0,1\}$, Pred^b is a predictor for X^b, and let Sim^b denote $\mathrm{Sim}^{\mathsf{Pred}^b}$. If $\mathrm{SD}(X^0,X^1)\leq\varepsilon$ then it holds that,*

$$\mathop{\mathbf{Pr}}_{\substack{z\leftarrow Z\\ r\leftarrow(\{0,1\}^\ell)^n}}\left[\mathrm{Sim}^0(z,r)\neq\mathrm{Sim}^1(z,r)\right]\leq n\cdot\beta+n\cdot(n-1)\cdot\mu$$

for $\beta=3\varepsilon+4\mu$.

Proof. It holds that,

$$\begin{aligned}&\mathop{\mathbf{Pr}}_{\substack{z\leftarrow Z\\ r=(r_1,\cdots,r_n)\leftarrow(\{0,1\}^\ell)^n}}\left[\mathrm{Sim}^0(z,r)\neq\mathrm{Sim}^1(z,r)\right] \qquad (17)\\ &\leq \mathop{\mathbf{Pr}}_{\substack{z\leftarrow Z\\ r=(r_1,\cdots,r_n)\leftarrow(\{0,1\}^\ell)^n}}\left[\exists i\in[n]\text{ s.t. }\mathrm{Sim}^0(z,r)_i\neq\mathrm{Sim}^1(z,r)_i\ \cap\ \mathrm{Sim}^0(z,r)_{<i}=\mathrm{Sim}^1(z,r)_{<i}\right]\\ &\leq\sum_{i=1}^n\mathop{\mathbf{Pr}}_{\substack{z\leftarrow Z\\ r=(r_1,\cdots,r_n)\leftarrow(\{0,1\}^\ell)^n}}\left[\mathrm{Sim}^0(z,r)_i\neq\mathrm{Sim}^1(z,r)_i\ \cap\ \mathrm{Sim}^0(z,r)_{<i}=\mathrm{Sim}^1(z,r)_{<i}\right]\end{aligned}$$

For every $b\in\{0,1\}$, let $\mathrm{Samp}^b:=\mathrm{Samp}^{\mathsf{Pred}^b}$. Note that for every fixed $i\in[n]$ the first $i-1$ bits outputted by Sim^0 and Sim^1 are fully determent by the value of the initial state Z and first $i-1$ batches of the random strings $(r_1,\cdots,r_{i-1})$. Thus by definition for every $i\in[n]$ it holds that

$$\begin{aligned}&\mathop{\mathbf{Pr}}_{\substack{z\leftarrow Z\\ r=(r_1,\cdots,r_n)\leftarrow(\{0,1\}^\ell)^n}}\left[\mathrm{Sim}^0(z,r)_i\neq\mathrm{Sim}^1(z,r)_i\ \cap\ \mathrm{Sim}^0(z,r)_{<i}=\mathrm{Sim}^1(z,r)_{<i}\right] \qquad (18)\\ &=\mathop{\mathbf{Pr}}_{\substack{z\leftarrow Z\\ r=(r_1,\cdots,r_n)\leftarrow(\{0,1\}^\ell)^n\\ w^0=\mathrm{Sim}^0(z,r)_{<i},w^1=\mathrm{Sim}^1(z,r)_{<i}}}\left[\mathrm{Samp}^0((z,w^0);r_i)\neq\mathrm{Samp}^1((z,w^1);r_i)\ \cap\ w^0=w^1\right]\\ &\leq\mathop{\mathbf{Pr}}_{\substack{z\leftarrow Z\\ r=(r_1,\cdots,r_n)\leftarrow(\{0,1\}^\ell)^n\\ w^0=\mathrm{Sim}^0(z,r)_{<i}}}\left[\mathrm{Samp}^0((z,w^0);r_i)\neq\mathrm{Samp}^1((z,w^0);r_i)\right]\end{aligned}$$

that is, for any fixing of Z and any fixing of the first $i-1$ random coin batches used to generate the first $i-1$ bits in the simulated distribution, the random variable r_i used to generate the ith bit in the simulated distribution is an independent random variable.

By Claim 3.9 for every $i\in[n]$, it holds that $\mathrm{SD}((Z,\mathrm{Sim}^0(Z,R)_{<i}),(Z,Y^0_{<i}))\leq 2(i-1)\mu$, which implies that

$$\begin{aligned}&\mathop{\mathbf{Pr}}_{\substack{z\leftarrow Z\\ r=(r_1,\cdots,r_n)\leftarrow(\{0,1\}^\ell)^n\\ w^0=\mathrm{Sim}^0(z,r)_{<i}}}\left[\mathrm{Samp}^0((z,w^0);r_i)\neq\mathrm{Samp}^1((z,w^0);r_i)\right] \qquad (19)\\ &\leq\mathop{\mathbf{Pr}}_{\substack{(z,y_{<i})\leftarrow(Z,Y^0_{<i})\\ r\leftarrow\{0,1\}^\ell}}\left[\mathrm{Samp}^0((z,y_{<i});r)\neq\mathrm{Samp}^1((z,y_{<i});r)\right]+2(i-1)\mu\\ &\leq\beta+2(i-1)\mu.\end{aligned}$$

where the finally inequality follows by Claim 3.15 for $\beta = 3\varepsilon + 4\mu$. Putting Eqs. (17) to (19), together we get:

$$\begin{aligned}
&\Pr_{\substack{z\leftarrow Z\\ r\leftarrow(\{0,1\}^\ell)^n}}\left[\mathrm{Sim}^0(z,r)\neq \mathrm{Sim}^1(z,r)\right]\\
&\leq \sum_{i=1}^{n}\Pr_{\substack{z\leftarrow Z\\ r=(r_1,\cdots,r_n)\leftarrow(\{0,1\}^\ell)^n}}\left[\mathrm{Sim}^0(z,r)_i\neq \mathrm{Sim}^1(z,r)_i \ \cap\ \mathrm{Sim}^0(z,r)_{<i}=\mathrm{Sim}^1(z,r)_{<i}\right]\\
&\leq \sum_{i=1}^{n}(\beta+2(i-1)\mu) = n\cdot\beta + n\cdot(n-1)\cdot\mu.
\end{aligned}$$

□

3.6 Round Collapse for Constant-Round Deterministic-Prover SWI

We show that constant-round honest-verifier SWI can be transformed to having only one round (two messages, where the verifier goes first):

Lemma 3.17 (Lemma 3.2, restated)**.** *Assume there is a k-round honest-verifier deterministic-prover δ-SWI public-coin interactive argument for $\mathcal{R}$ with $(1-\alpha)$-completeness and ϵ-soundness errors, prover-to-verifier communication $\ell_P \in \mathrm{poly}$ and total communication complexity ℓ where k is a constant and $\epsilon < \frac{1}{64}$. Then either infinitely-often auxiliary-input OWFs exist, or there exists a two-message (AM) honest-verifier deterministic-prover $(\delta(n)\cdot(\mathrm{poly}(\ell_P,\log(n)))^k)$-SWI public-coin interactive argument for $\mathcal{R}$ with negligible soundness error, $(1-\alpha\cdot\mathrm{poly}(\ell_P,\log(n))^k)$-completeness and (total) communication complexity $\ell\cdot(\mathrm{poly}(\ell_P,\log(n)))^k$.*

Moreover, if the argument is ϵ statistically *sound, then the resulting protocol is statistically sound with negligible error.*

In order to prove the lemma, we will have to open up the proof of the following theorem due to Pass and Venkitasubramaniam, which we will not use directly but write here nonetheless for the reader's convenience.

Theorem 3.18 ([PV20], Theorem 7.4)**.** *Assume there exists a k-round public-coin (efficient-prover) argument system for $\mathcal{L}$ with prover-to-verifier communication complexity ℓ_P and total communication complexity ℓ, where k is a constant. Then, either non-uniform ioOWFs exist, or there exists a two-message public-coin (efficient-prover) argument for $\mathcal{L}$ with communication complexity $O(\ell\cdot\mathrm{poly}(\ell_P, \mathit{polylog}(n))^k)$.*

Proof of Lemma 3.17. The proof follows almost immediately from the proof of Theorem 3.18 (we note that Pass and Venkitasubramaniam do remark that the proof holds for resettable WI). We sketch their construction, which is identical to that of Babai and Moran [BM88].

We describe how this is done for a 3-message protocol (with the prover going first). For k rounds this is done by a similar transformation, where we successively turn the last 3 messages into 2 messages using the same techniques (we further describe later). The strategy is:

1. First amplify soundness of the protocol by parallel repetition of the last two rounds of the protocol $m = \ell_{P,k-1}^2\cdot\mathrm{polylog}(n)$ times, where $\ell_{P,k-1}$ is the prover-to-verifier communication complexity in round $k-1$. That is, the protocol is identical up to the final round. In the final round, the verifier sends m different challenges and the prover replies to all of these challenges. The verifier accepts if all of the replies to challenges are accepting.

2. Now, we switch the order of the last two messages of the protocol. That is, the verifier begins by sending its random message ρ, and the prover replies with both of its messages m_1 and m_2 together.

3. Amplify soundness of the protocol using standard parallel repetition $\mathrm{polylog}(n)$ times.

It is immediate that, following the transformation described above, the final 3 messages of the protocol have collapsed to 2, so that the protocol now has $k-1$ rounds, and that the prover remains efficient

and deterministic. Moreover, the communication complexity goes from ℓ to $\ell \cdot m \cdot \text{polylog}(n) = O(\ell \cdot \ell_P^2 \cdot \text{polylog}(n))$. The prover-to-verifier complexity of the protocol on all but the final round remains identical (this will be important later when we will repeat this transformation multiple times).

We have completeness $1 - \alpha \cdot (\text{poly}(\ell_P, \log(n)))$ by taking a simple union-bound over all of the repetitions taken. Negligible soundness was shown to hold assuming infinitely-often auxiliary-input one-way functions do not exist by [PV20] in proving Theorem 3.18 (note that while in the theorem statement, [PV20] assume that the protocol starts with negligible soundness error, in their proof they only require that the soundness error be bounded by $\frac{1}{64}$). If the protocol has statistical soundness rather than only computational soundness then negligible statistical soundness of the resulting protocol follows by the original analysis of [BM88] as the construction is identical.

All that remains is to show that WI is preserved, and is where we will crucially use the assumption that the prover is deterministic. Fix $x \in \mathcal{L}$ and valid witnesses w_0 and w_1. We consider each step in the transformation separately:

- Item 1: For $b \in \{0,1\}$ and verifier randomness r let $\mathsf{tr}_b(r) = \langle P(x, w_b), V(x; r) \rangle$ (i.e, the prover uses w_b and the verifier uses randomness r). Observe that for fixed b and r this function is deterministic since the prover is deterministic and so has no private internal state. Moreover, observe that both $\mathsf{tr}_0(r)$ and $\mathsf{tr}_1(r)$ are computable given r, w_0 and w_1.

  Let $I(r)$ be the indicator function that is 1 if and only if $\mathsf{tr}_0(r) = \mathsf{tr}_1(r)$. Then by δ honest-verifier SWI it holds that $\mathbf{Pr}_r[I(r) = 0] \leq \delta$. Indeed, given the view of V (which includes r), a distinguisher can compute $\mathsf{tr}_0(r)$ and $\mathsf{tr}_1(r)$ and, since they differ, use them to decide which witness the prover used.

  Finally, let r' be the randomness of the verifier up to round $k-1$ and $r''_1, \ldots, r''_m$ be its randomness in the k-th round. Then the verifier's views interacting with the prover with w_0 and w_1 are indistinguishable (at least) as long as $I(r'||r''_i) = 1$ for every $i \in [m]$. By a union-bound over every value of i we have that they differ with probability at most $\delta \cdot m$. It follows that the statistical distance between the views in the new protocol are bounded by $\delta \cdot m$.

- Item 2: This step preserves honest-verifier WI as the views of the verifier in both cases are identical up to reordering of the messages. Thus, following this step we still have $\delta \cdot m$ honest-verifier WI.

- Item 3: This step is conventional parallel repetition which by a standard hybrid argument adds a multiplicative factor of $\text{poly}(n)$ to the WI distinguishing advantage. Thus, we end up with $\delta \cdot m \cdot \text{polylog}(n) = \delta \cdot \ell_{P,k-1}^2 \cdot \text{polylog}(n)$ honest-verifier WI.

Finally, the parameters claimed are due to repeating the above analysis $k-1$ times, where each time the protocol's round complexity is reduced by 1. This yields factors that depend on $(\text{polylog}(n))^{k-1} \cdot \prod_{i=1}^{k-1} \ell_{P,i}^2$ (where these are parameters of the original protocol), which, since $\ell_{P,i} \leq \ell_P$, is bounded by $\text{poly}(\ell_P, \log n)^{k-1}$. The final protocol has 2 messages with the verifier going first. □

3.7 Error Elimination via Derandomization

In this section, we show that under the derandomization assumption of Liu and Pass [LP25], the completeness and soundness errors of a statistically sound witness canonicalization scheme can be eliminated while preserving inverse-polynomial uniqueness error.

3.7.1 Derandomization: Definitions and Assumptions

Definition 3.19 (Nondeterministic time with non-uniform advice)**.** *Let $t, a \colon \mathbb{N} \to \mathbb{N}$. We say that a language L belongs to $\mathsf{NTIME}[t(n)]/a(n)$ if there exists a nondeterministic Turing machine M and a sequence of advice strings $\{z_n\}_{n \in \mathbb{N}}$, where $|z_n| \leq a(n)$, such that M runs in time at most $t(n)$ and, for every $x \in \{0,1\}^n$,*

$$x \in L \iff M(x, z_n) \text{ has an accepting computation path.}$$

We say that $L \in \mathsf{ioNTIME}[t(n)]/a(n)$ if there exist such M and $\{z_n\}_{n \in \mathbb{N}}$ for which the above equivalence holds for every input of length n for infinitely many $n \in \mathbb{N}$.

We write $\mathsf{E} = \mathsf{DTIME}[2^{O(n)}]$. We will rely on the following assumption of Liu and Pass.

Assumption 3.20 (Liu–Pass Derandomization Assumption)**.** *There exists a constant $\gamma > 0$ such that, for every $k \in \mathbb{N}$,*

$$\mathsf{E} \not\subseteq \mathsf{ioNTIME}[2^{kn}]/2^{\gamma n}.$$

Pseudorandom generators for derandomization. We use the following standard, complexity-theoretic notion of a pseudorandom generator.

Definition 3.21 (Pseudorandom generator for derandomization)**.** *Let $\mathcal{C}$ be a class of Boolean algorithms on N-bit inputs. A function $G\colon \{0,1\}^d \to \{0,1\}^N$ is an ε-PRG for $\mathcal{C}$ if, for every $C \in \mathcal{C}$,*

$$\left| \Pr_{s \leftarrow \{0,1\}^d} [C(G(s)) = 1] - \Pr_{r \leftarrow \{0,1\}^N} [C(r) = 1] \right| \leq \varepsilon.$$

The classical hardness-versus-randomness framework of Impagliazzo and Wigderson [IW97], together with its nondeterministic extension due to Klivans and van Melkebeek [KM02], gives the following consequence.

Theorem 3.22 (Standard nondeterministic PRG, [IW97; KM02])**.** *Assume Assumption 3.20. For every polynomial t and every constant $c > 0$, there exists a polynomial-time computable family*

$$G_N\colon \{0,1\}^{d(N)} \to \{0,1\}^N,$$

for $d(N) = O_{t,c}(\log N)$, that is an N^{-c}-PRG for $\mathsf{NTIME}[t(N)]/t(N)$.

We note that the Liu and Pass derandomization assumption defined above, is stronger than the hardness assumption needed for the above theorem. However, for our derandomization construction we require an additional PRG that has the crucial property that the constant multiplying $\log N$ in the seed length does not depend on the polynomial running time being fooled. We will need the following result of Liu and Pass (which require the stronger assumption of Assumption 3.20).

Theorem 3.23 (Liu–Pass, [LP25])**.** *Assume Assumption 3.20. There exists a universal constant $\hat{a} > 0$ such that, for every polynomial t, there exists a polynomial-time computable family*

$$G_N\colon \{0,1\}^{\lceil \hat{a} \log N \rceil} \to \{0,1\}^N$$

that is a $1/6$-PRG for $\mathsf{NTIME}[t(N)]/N$. The family G and its running time may depend on t, but the constant $\hat{a}$ does not.

The independence of $\hat{a}$ from t will be important below. The running time of the tests to which we apply Theorem 3.23 may depend arbitrarily on the uniqueness error of the starting WC. Nevertheless, the number of seeds that we enumerate will be $2^{\hat{a} \log N} = N^{\hat{a}}$, where N is a fixed polynomial in the instance length whose degree is independent of this error. This prevents the size of the enumerated set from growing with the accuracy of the starting scheme. Our use of a complexity-theoretic PRG to enumerate a polynomial-size collection of verifier random strings is also similar in spirit to the derandomization of ZAPs into NIWIs by Barak, Ong, and Vadhan [BOV07].

3.7.2 Witness Canonicalization with Perfect Completeness and Soundness

Recall that a WC has *perfect completeness* if $V(r, x, \mathsf{map}(r, x, w)) = 1$ for every $(x, w) \in \mathcal{R}$ and every public random string r. Similarly, it has *perfect soundness* if $V(r, x, \pi) = 0$ for every $x \notin \mathcal{L}(\mathcal{R})$, every r, and every purported proof π.

Lemma 3.24 (Error elimination for witness canonicalization)**.** *Assume Assumption 3.20. Let $\mathcal{R}$ be an NP relation. Suppose that, for every $\eta \in 1/\mathrm{poly}$, there exists a statistically sound η-WC for $\mathcal{R}$. Then, for every $\varepsilon \in 1/\mathrm{poly}$, there exists a statistically sound ε-WC for $\mathcal{R}$ with perfect completeness and perfect soundness. The same implication holds with both the starting schemes and the resulting scheme non-uniform.*

Proof. It suffices to prove the lemma for $\varepsilon(n) = n^{-b}$, where $b > 0$ is an arbitrary constant. Let $c > 0$ be a constant that will be fixed at the end of the proof, and let (map, V) be a statistically sound n^{-c}-WC for $\mathcal{R}$. Let $\alpha, \beta \in \mathsf{negl}$ denote its completeness and soundness errors, respectively. Let $\ell(n)$ denote the length of its public randomness and let $p(n)$ be the length of its output. Without loss of generality, V rejects every string whose length is not $p(n)$. We also pad the public randomness with ignored bits so that $\ell(n) \geq n$.

For $(x, w) \in \mathcal{R}$, define the circuit

$$C_{x,w}(r) := V\big(r, x, \mathsf{map}(r, x, w)\big).$$

For $x \notin \mathcal{L}(\mathcal{R})$, define the nondeterministic circuit

$$S_x(r) = 1 \quad \iff \quad \exists \pi \in \{0,1\}^{p(n)} \colon V(r, x, \pi) = 1.$$

Finally, for $x \in \mathcal{L}(\mathcal{R})$ and $(x, w_0), (x, w_1) \in \mathcal{R}$, define

$$D_{x,w_0,w_1}(r) = 1 \quad \iff \quad \mathsf{map}(r, x, w_0) \neq \mathsf{map}(r, x, w_1).$$

By completeness,

$$\mathop{\mathbf{Pr}}_{r \leftarrow \{0,1\}^{\ell(n)}} [C_{x,w}(r) = 1] \geq 1 - \alpha(n). \tag{20}$$

Since (map, V) is statistically sound, for every $x \notin \mathcal{L}(\mathcal{R})$,

$$\mathop{\mathbf{Pr}}_{r \leftarrow \{0,1\}^{\ell(n)}} [S_x(r) = 1] \leq \beta(n). \tag{21}$$

Indeed, an inefficient prover can find an accepting π whenever one exists. Moreover, by the uniqueness property and a union bound, it follows that for every two valid witnesses w_0, w_1,

$$\mathop{\mathbf{Pr}}_{r \leftarrow \{0,1\}^{\ell(n)}} [D_{x,w_0,w_1}(r) = 1] \leq 2n^{-c}. \tag{22}$$

Apply Theorem 3.22, with sufficiently large parameters, to obtain a polynomial-time computable PRG

$$G_1 \colon \{0,1\}^{d(n)} \to \{0,1\}^{\ell(n)},$$

for $d(n) = O(\log n)$, that fools all the circuits $C_{x,w}$, S_x, and D_{x,w_0,w_1} with error $\delta(n) \leq n^{-c}$. The constant hidden in the $O(\cdot)$ notation may depend on c. In particular,

$$\mathop{\mathbf{Pr}}_{z \leftarrow \{0,1\}^{d(n)}} [C_{x,w}(G_1(z)) = 1] \geq 1 - \alpha(n) - \delta(n), \tag{23}$$

$$\mathop{\mathbf{Pr}}_{z \leftarrow \{0,1\}^{d(n)}} [S_x(G_1(z)) = 1] \leq \beta(n) + \delta(n), \tag{24}$$

$$\mathop{\mathbf{Pr}}_{z \leftarrow \{0,1\}^{d(n)}} [D_{x,w_0,w_1}(G_1(z)) = 1] \leq 3n^{-c}. \tag{25}$$

Let $p_{\mathcal{R}}$ be a polynomial upper bound on the witness length of $\mathcal{R}$. Fix a constant v that depends only on $\mathcal{R}$ such that, for $N := n^v$, the tuple (x, w, u), where $|w| \leq p_{\mathcal{R}}(n)$ and $u \in \{0,1\}^{d(n)}$, can be encoded using at most N bits for all sufficiently large n. Such a constant v, independent of c, exists because $d(n) = O(\log n)$.

For fixed x, w, u, define the following Boolean algorithm on $y \in \{0,1\}^N$:

$$\widehat{C}_{x,w,u}(y) := C_{x,w}\big(G_1(u \oplus y_{\leq d(n)})\big).$$

Similarly, for fixed x, u, define the nondeterministic algorithm

$$\widehat{S}_{x,u}(y) := S_x\big(G_1(u \oplus y_{\leq d(n)})\big).$$

There exists a polynomial t such that these algorithms can be evaluated nondeterministically in time $t(N)$ with at most N bits of non-uniform advice. The degree of t may depend arbitrarily on c. Importantly, the proof π in the definition of $\widehat{S}_{x,u}$ is guessed nondeterministically and is not included in the advice. Apply Theorem 3.23 to the polynomial t, obtaining

$$G_2 \colon \{0,1\}^{d_2(N)} \to \{0,1\}^N, \qquad d_2(N) := \lceil \hat{a} \log N \rceil,$$

that 1/6-fools both $\widehat{C}_{x,w,u}$ and $\widehat{S}_{x,u}$. Let $\mathcal{S} := \{0,1\}^{d_2(N)}$ and $K := |\mathcal{S}|$ and note that this means that

$$K \le 2N^{\hat{a}} = 2n^{v\hat{a}}. \tag{26}$$

For every $s \in \mathcal{S}$ and $u \in \{0,1\}^{d(n)}$, define

$$r_s(u) := G_1\big(u \oplus G_2(s)_{\le d(n)}\big).$$

We define the new witness canonicalization scheme (map^*, V^*). Its public randomness is $u \leftarrow \{0,1\}^{d(n)}$, and

$$\mathsf{map}^*(u,x,w) := \big(\mathsf{map}(r_s(u),x,w)\big)_{s\in\mathcal{S}}.$$

Given a vector $(\pi_s)_{s\in\mathcal{S}}$, the algorithm V^* accepts if and only if

$$|\{s \in \mathcal{S} \colon V(r_s(u),x,\pi_s) = 1\}| > \frac{K}{2}.$$

By Eq. (26), both map^* and V^* run in polynomial time.

Perfect completeness. Fix $(x,w) \in \mathcal{R}$ and $u \in \{0,1\}^{d(n)}$. For $Y \leftarrow \{0,1\}^N$, the random variable $u \oplus Y_{\le d(n)}$ is uniform over $\{0,1\}^{d(n)}$. Thus, by Eq. (23),

$$\Pr_{Y\leftarrow\{0,1\}^N}\left[\widehat{C}_{x,w,u}(Y) = 1\right] \ge 1 - \alpha(n) - \delta(n).$$

The pseudorandomness of G_2 now gives

$$\Pr_{s\leftarrow\mathcal{S}}\left[C_{x,w}(r_s(u)) = 1\right] \ge 1 - \alpha(n) - \delta(n) - \frac{1}{6} > \frac{1}{2}$$

for all sufficiently large n. Hence, for every u, a strict majority of the coordinates produced by map^* are accepted, and the new scheme has perfect completeness.

Perfect soundness. Fix $x \notin \mathcal{L}(\mathcal{R})$ and $u \in \{0,1\}^{d(n)}$. By Eq. (24),

$$\Pr_{Y\leftarrow\{0,1\}^N}\left[\widehat{S}_{x,u}(Y) = 1\right] \le \beta(n) + \delta(n).$$

The pseudorandomness of G_2 implies that

$$\Pr_{s\leftarrow\mathcal{S}}\left[S_x(r_s(u)) = 1\right] \le \beta(n) + \delta(n) + \frac{1}{6} < \frac{1}{2}$$

for all sufficiently large n. Consequently, fewer than half of the coordinates admit any accepting proof. Therefore, every purported vector is rejected, and the new scheme has perfect soundness.

Uniqueness. Fix $x \in \mathcal{L}(\mathcal{R})$, and fix an arbitrary witness w_x such that $(x,w_x) \in \mathcal{R}$. For every $u \in \{0,1\}^{d(n)}$, define

$$\pi^*_{x,u} := \mathsf{map}^*(u,x,w_x).$$

Fix any other witness w such that $(x,w) \in \mathcal{R}$. For every fixed $s \in \mathcal{S}$, a uniform $u \leftarrow \{0,1\}^{d(n)}$ makes $u \oplus G_2(s)_{\le d(n)}$ uniform over $\{0,1\}^{d(n)}$. Therefore, by Eq. (25),

$$\Pr_{u\leftarrow\{0,1\}^{d(n)}}\left[\mathsf{map}(r_s(u),x,w) \ne \mathsf{map}(r_s(u),x,w_x)\right] \le 3n^{-c}.$$

Taking a union bound over all $s \in \mathcal{S}$ and using Eq. (26),

$$\begin{aligned}\Pr_{u\leftarrow\{0,1\}^{d(n)}}\left[\mathsf{map}^*(u,x,w) \ne \pi^*_{x,u}\right] &\le 3Kn^{-c}\\ &\le 6n^{v\hat{a}-c}.\end{aligned}$$

Finally, we can choose

$$c \geq b + v\hat{a} + 3.$$

This implies that the above probability is at most $n^{-b} = \varepsilon(n)$ for all sufficiently large n.

For non-uniform schemes, use the same construction, retaining the original advice a_n in both algorithms. The construction of G_1 still applies, since Theorem 3.22 allows polynomial-size advice. For G_2, replace the application of Theorem 3.23 by the following non-uniform choice. Once a_n and G_1 are fixed, there are at most 2^{N+1} tests $\widehat{C}_{x,w,u}$ and $\widehat{S}_{x,u}$, indexed by the encodings above. Taking $\hat{a} \geq 2$ without loss of generality, a table of $K = 2^{\lceil \hat{a} \log N \rceil}$ independent uniform N-bit outputs 1/6-fools all these tests with positive probability, by a Chernoff bound and a union bound. Fix such a table as G_2 and include its KN bits as additional polynomial-size advice. Thus Eq. (26) and the completeness, soundness, and uniqueness analyses remain unchanged, with v and $\hat{a}$ independent of c. Finitely many exceptional input lengths can be handled by hardwiring. □

We can now strengthen Theorem 3.3 in the statistically sound setting.

Theorem 3.25 (WI with no perfectly complete and sound WC implies ioAI-OWF under derandomization)**.** *Assume Assumption 3.20. Let $\mathcal{R}$ be an NP relation for which there is no ε-WC with perfect completeness and perfect statistical soundness for some $\varepsilon \in 1/\mathrm{poly}$. If there exists a constant-round honest-verifier witness-indistinguishable interactive proof for $\mathcal{R}$, then infinitely-often auxiliary-input one-way functions exist.*

Proof. Assume towards contradiction that infinitely-often auxiliary-input one-way functions do not exist. By Theorem 3.3, for every $\eta \in 1/\mathrm{poly}$ there exists a statistically sound η-WC for $\mathcal{R}$. Applying Lemma 3.24 with the value of ε from the theorem statement gives a statistically sound ε-WC with perfect completeness and perfect soundness, contradicting the assumption. □

The same argument applies to the batch and succinct proof consequences from Corollary 3.5.

Corollary 3.26 (Batch and succinct proofs under derandomization)**.** *Assume Assumption 3.20. There exists a universal constant $c > 1$ such that the following holds. Let $\mathcal{R}$ be an NP relation for which, there is no non-uniform ε-WC with perfect completeness and perfect statistical soundness for some $\varepsilon \in 1/\mathrm{poly}$. If one of the following exists:*

- *A k-round batch proof for $\mathcal{R}$, for a constant k, such that when batch proving a sufficiently large polynomial number of instances $t = t(n)$, the prover-to-verifier communication satisfies $\ell_P \leq t^{\gamma}$ for some constant $\gamma < 1/(ck)$.*
- *A k-round γ-succinct interactive proof for $\mathcal{R}^{\otimes t}$, for a constant k, some $\gamma < 1/(ck)$, and a sufficiently large $t \in \mathrm{poly}(n)$.*

Then there exists an infinitely-often auxiliary-input one-way function.

Proof. Assume that infinitely-often auxiliary-input one-way functions do not exist. Repeat the proof of Corollary 3.5, using the statistical soundness guarantees throughout. Since the starting batch protocol is a proof, the batch-to-SWI transformation and all subsequent transformations preserve statistical soundness. Moreover, the internal parameters can be chosen to obtain a statistically sound non-uniform η-WC for every $\eta \in 1/\mathrm{poly}$, with fixed polynomial-size advice for both map and V at each instance length.

Applying the non-uniform version of Lemma 3.24 gives a non-uniform ε-WC with perfect completeness and perfect statistical soundness, contradicting the assumption. □

Acknowledgments

The authors are thankful to Idan Baril, Iftach Haitner, and Willy Quach for very fruitful discussions. This work was initiated and much of it was conducted while Gal Arnon, Noam Mazor and Jad Silbak were research fellows at the Simons Institute for the Theory of Computing.

Gal Arnon is supported by the European Research Council (ERC) under the EU's Horizon 2020 research and innovation programme (Grant agreement No. 101019547) and by a Cariplo CRYPTONOMEX grant. The majority of this work was done while Gal Arnon was a postdoc at Bocconi University. Noam

Mazor is supported by the Simons Foundation. Rafael Pass is supported in part by AFOSR Award FA9550-23-1-0387, AFOSR Award FA9550-23-1-0312, AFOSR Award FA9550-24-1-0267, ISF Award 2338/23 and ERC Advanced Grant KolmoCrypt - 101142322.

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Government, the AFOSR, the European Union or the European Research Council Executive Agency.

AI Use Statement

The results and proofs in this work were found by humans. ChatGPT was used solely for writing assistance. The authors take full responsibility for all claims and proofs made in this paper.

References

[AMPS26] Gal Arnon, Noam Mazor, Rafael Pass, and Jad Silbak. *Witness-Indistinguishable Arguments of Knowledge and One-Way Functions*. Cryptology ePrint Archive, Paper 2026/682. 2026. URL: https://eprint.iacr.org/2026/682.

[BC20] Nir Bitansky and Arka Rai Choudhuri. "Characterizing deterministic-prover zero knowledge". In: *Theory of Cryptography Conference*. Springer. 2020, pp. 535–566.

[BH26] Idan Baril and Iftach Haitner. *On the Complexity of Interactive Arguments*. Cryptology ePrint Archive, Paper 2026/272. 2026. URL: https://eprint.iacr.org/2026/272.

[BKPRV24] Nir Bitansky, Chethan Kamath, Omer Paneth, Ron D Rothblum, and Prashant Nalini Vasudevan. "Batch proofs are statistically hiding". In: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. 2024, pp. 435–443.

[Blu83] Manuel Blum. "Coin flipping by telephone a protocol for solving impossible problems". In: *ACM SIGACT News* 15.1 (1983), pp. 23–27.

[BM84] Manuel Blum and Silvio Micali. "How to Generate Cryptographically Strong Sequences of Pseudo-Random Bits". In: *SIAM J. Comput.* 13.4 (1984), pp. 850–864. DOI: 10.1137/0213053. URL: https://doi.org/10.1137/0213053.

[BM88] László Babai and Shlomo Moran. "Arthur-Merlin Games: A Randomized Proof System, and a Hierarchy of Complexity Classes". In: *J. Comput. Syst. Sci.* 36.2 (1988), pp. 254–276.

[BOV07] Boaz Barak, Shien Jin Ong, and Salil P. Vadhan. "Derandomization in Cryptography". In: *SIAM J. Comput.* 37.2 (2007), pp. 380–400.

[CHKT26] Suvradip Chakraborty, James Hulett, Dakshita Khurana, and Kabir Tomer. "Non-Trivial Zero-Knowledge Implies One-Way Functions". In: *arXiv preprint arXiv:2602.17651* (2026).

[CHS05] Ran Canetti, Shai Halevi, and Michael Steiner. "Hardness amplification of weakly verifiable puzzles". In: *Theory of Cryptography Conference*. Springer. 2005, pp. 17–33.

[CLMP12] Kai-Min Chung, Edward Lui, Mohammad Mahmoody, and Rafael Pass. "Unprovable security of 2-message zero knowledge". In: *Cryptology ePrint Archive* (2012).

[CPW20] Suvradip Chakraborty, Manoj Prabhakaran, and Daniel Wichs. "Witness Maps and Applications". In: *Public-Key Cryptography (PKC 2020), Part I*. Ed. by Aggelos Kiayias, Markulf Kohlweiss, Petros Wallden, and Vassilis Zikas. Vol. 12110. Lecture Notes in Computer Science. Springer, 2020, pp. 220–246.

[CPW23] Suvradip Chakraborty, Manoj Prabhakaran, and Daniel Wichs. "A Map of Witness Maps: New Definitions and Connections". In: *Public-Key Cryptography - PKC 2023 - 26th IACR International Conference on Practice and Theory of Public-Key Cryptography, Atlanta, GA, USA, May 7-10, 2023, Proceedings, Part II*. Ed. by Alexandra Boldyreva and Vladimir Kolesnikov. Vol. 13941. Lecture Notes in Computer Science. Springer, 2023, pp. 635–662.

[CT21] Lijie Chen and Roei Tell. "Simple and fast derandomization from very hard functions: eliminating randomness at almost no cost". In: *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*. 2021, pp. 283–291.

[DH76] Whitfield Diffie and Martin E. Hellman. "New directions in cryptography". In: *IEEE Trans. Inf. Theory* 22.6 (1976), pp. 644–654. DOI: 10.1109/TIT.1976.1055638. URL: https://doi.org/10.1109/TIT.1976.1055638.

[DN07] Cynthia Dwork and Moni Naor. "Zaps and Their Applications". In: *SIAM J. Comput.* 36.6 (2007), pp. 1513–1543.

[FLS99] Uriel Feige, Dror Lapidot, and Adi Shamir. "Multiple NonInteractive Zero Knowledge Proofs Under General Assumptions". In: *SIAM J. Comput.* 29.1 (1999), pp. 1–28.

[FNV17] Antonio Faonio, Jesper Buus Nielsen, and Daniele Venturi. "Predictable arguments of knowledge". In: *IACR International Workshop on Public Key Cryptography*. Springer. 2017, pp. 121–150.

[FS86] Amos Fiat and Adi Shamir. "How to prove yourself: Practical solutions to identification and signature problems". In: *Conference on the theory and application of cryptographic techniques*. Springer. 1986, pp. 186–194.

[FS90] Uriel Feige and Adi Shamir. "Witness Indistinguishable and Witness Hiding Protocols". In: *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing, May 13-17, 1990, Baltimore, Maryland, USA*. Ed. by Harriet Ortiz. ACM, 1990, pp. 416–426.

[GGM86] Oded Goldreich, Shafi Goldwasser, and Silvio Micali. "How to construct random functions". In: *J. ACM* 33.4 (1986), pp. 792–807. DOI: 10.1145/6490.6503. URL: https://doi.org/10.1145/6490.6503.

[GM84] Shafi Goldwasser and Silvio Micali. "Probabilistic Encryption". In: *J. Comput. Syst. Sci.* 28.2 (1984), pp. 270–299. DOI: 10.1016/0022-0000(84)90070-9. URL: https://doi.org/10.1016/0022-0000(84)90070-9.

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. "The Knowledge Complexity of Interactive Proof Systems". In: *SIAM J. Comput.* 18.1 (1989), pp. 186–208.

[GMW91] Oded Goldreich, Silvio Micali, and Avi Wigderson. "Proofs that yield nothing but their validity or all languages in NP have zero-knowledge proof systems". In: *Journal of the ACM (JACM)* 38.3 (1991), pp. 690–728.

[GO94a] Oded Goldreich and Yair Oren. "Definitions and Properties of Zero-Knowledge Proof Systems". In: *J. Cryptol.* 7.1 (1994), pp. 1–32.

[GO94b] Oded Goldreich and Yair Oren. "Definitions and properties of zero-knowledge proof systems". In: *Journal of Cryptology* 7.1 (1994), pp. 1–32.

[Gol90] Oded Goldreich. "A note on computational indistinguishability". In: *Information Processing Letters* 34.6 (1990), pp. 277–281.

[GOS06] Jens Groth, Rafail Ostrovsky, and Amit Sahai. "Non-interactive Zaps and New Techniques for NIZK". In: *Advances in Cryptology - CRYPTO 2006, 26th Annual International Cryptology Conference, Santa Barbara, California, USA, August 20-24, 2006, Proceedings*. Ed. by Cynthia Dwork. Vol. 4117. Lecture Notes in Computer Science. Springer, 2006, pp. 97–111.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. "A Pseudorandom Generator from any One-way Function". In: *SIAM J. Comput.* 28.4 (1999), pp. 1364–1396. DOI: 10.1137/S0097539793244708. URL: https://doi.org/10.1137/S0097539793244708.

[HN24] Shuichi Hirahara and Mikito Nanashima. "One-Way Functions and Zero Knowledge". In: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing, STOC 2024, Vancouver, BC, Canada, June 24-28, 2024*. Ed. by Bojan Mohar, Igor Shinkar, and Ryan O'Donnell. ACM, 2024, pp. 1731–1738.

[HNOS96] Lane A. Hemaspaandra, Ashish V. Naik, Mitsunori Ogihara, and Alan L. Selman. "Computing Solutions Uniquely Collapses the Polynomial Hierarchy". In: *SIAM J. Comput.* 25.4 (1996), pp. 697–708. DOI: 10.1137/S0097539794268315. URL: https://doi.org/10.1137/S0097539794268315.

[HR07] Iftach Haitner and Omer Reingold. "Statistically-hiding commitment from any one-way function". In: *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*. 2007, pp. 1–10.

[IL89] Russell Impagliazzo and Michael Luby. "One-way functions are essential for complexity based cryptography". In: *30th Annual Symposium on Foundations of Computer Science*. IEEE Computer Society. 1989, pp. 230–235.

[IW97] R. Impagliazzo and A. Wigderson. "$P = BPP$ if E Requires Exponential Circuits: Derandomizing the XOR Lemma". In: *STOC*. 1997, pp. 220–229.

[Kiy24] Susumu Kiyoshima. "Resettable Statistical Zero-Knowledge for NP". In: *Annual International Cryptology Conference*. Springer. 2024, pp. 288–320.

[KM02] A. Klivans and D. van Melkebeek. "Graph Nonisomorphism Has Subexponential Size Proofs Unless the Polynomial-Time Hierarchy Collapses". In: *SIAM J. Comput.* 31.5 (2002), pp. 1501–1526.

[KT24] Dakshita Khurana and Kabir Tomer. "Commitments from quantum one-wayness". In: *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*. 2024, pp. 968–978.

[LMP24] Yanyi Liu, Noam Mazor, and Rafael Pass. "A note on zero-knowledge for NP and one-way functions". In: *Cryptology ePrint Archive* (2024).

[LMP25] Yanyi Liu, Noam Mazor, and Rafael Pass. "On witness encryption and laconic zero-knowledge arguments". In: *Annual International Cryptology Conference*. Springer. 2025, pp. 429–461.

[LP25] Yanyi Liu and Rafael Pass. "Hardness Along the Boundary: Towards One-Way Functions from the Worst-Case Hardness of Time-Bounded Kolmogorov Complexity". In: *Advances in Cryptology – CRYPTO 2025*. Ed. by Yael Tauman Kalai and Seny F. Kamara. Vol. 16000. Lecture Notes in Computer Science. Springer, 2025, pp. 617–650. DOI: 10.1007/978-3-032-01855-7_20.

[MPS25] Noam Mazor, Rafael Pass, and Tomer Solomon. "A Meta-Complexity Theoretic Approach to Indistinguishability Obfuscation and Witness Pseudo-Canonicalization". In: *Theory of Cryptography Conference*. Springer. 2025.

[Nao91] Moni Naor. "Bit commitment using pseudorandomness". In: *Journal of cryptology* 4.2 (1991), pp. 151–158.

[NW94] Noam Nisan and Avi Wigderson. "Hardness vs randomness". In: *Journal of computer and System Sciences* 49.2 (1994), pp. 149–167.

[NY89] Moni Naor and Moti Yung. "Universal One-Way Hash Functions and their Cryptographic Applications". In: *Proceedings of the 21st Annual ACM Symposium on Theory of Computing, May 14-17, 1989, Seattle, Washington, USA*. Ed. by David S. Johnson. ACM, 1989, pp. 33–43. DOI: 10.1145/73007.73011. URL: https://doi.org/10.1145/73007.73011.

[Ost91] Rafail Ostrovsky. "One-Way Functions, Hard on Average Problems, and Statistical Zero-Knowledge Proofs." In: *SCT*. 1991, pp. 133–138.

[OW93] Rafail Ostrovsky and Avi Wigderson. "One-Way Functions 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*. IEEE Computer Society, 1993, pp. 3–17.

[Pas03] Rafael Pass. "Simulation in quasi-polynomial time, and its application to protocol composition". In: *International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 2003, pp. 160–176.

[Pas04] Rafael Pass. "Alternative variants of zero-knowledge proofs". PhD thesis. 2004.

[PV20] Rafael Pass and Muthuramakrishnan Venkitasubramaniam. "Is it Easier to Prove Theorems that are Guaranteed to be True?" In: *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*. Ed. by Sandy Irani. IEEE, 2020, pp. 1255–1267.

[Rom90] John Rompel. "One-Way Functions are Necessary and Sufficient for Secure Signatures". In: *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing, May 13-17, 1990, Baltimore, Maryland, USA*. Ed. by Harriet Ortiz. ACM, 1990, pp. 387–394. DOI: 10.1145/100216.100269. URL: https://doi.org/10.1145/100216.100269.

[Wee06] Hoeteck Wee. "Finding pessiland". In: *Theory of Cryptography Conference*. Springer. 2006, pp. 429–442.

A Application to Hard Oracles

In this section, we study deterministic-prover witness indistinguishability in the presence of hard oracles and show how it can be leveraged toward witness encryption. Informally, we consider an oracle that samples yes-instances of an NP relation together with corresponding witnesses, while the induced distribution over instances remains computationally hard. We then show that deterministic-prover witness indistinguishability, when combined with such oracle access, yields witness encryption relative to the oracle. More broadly, in light of Theorem 3.1, this perspective suggests that witness indistinguishability either implies one-way functions on its own, or, when combined with hard oracles that by themselves need not imply cryptography, yields stronger cryptographic primitives in settings where hardness is provided through oracle access rather than efficient samplability.

This section is organized as follows:

- In Section A.1 we formally define hard oracles, witness encryption and predictable arguments.
- In Section A.2 we state our formal theorem and corollary.
- In Section A.3 we prove the theorem stated in Section A.2.

A.1 Definitions and Notions for Application to Hard Oracles

A.1.1 Hard Oracles

To state our next result, we start by defining the notion of hard oracles for a relation. We then show why such an oracle that is hard for NP could be useful when combined with a WI protocol.

Observe first that assuming that NP is not in $\mathsf{ioP/poly}$ then there are distributions that are hard for any non-uniform PPT algorithm in the following sense: There exists two distributions $\mathcal{D}^Y$ and $\mathcal{D}^N$ such that $\mathcal{D}^Y$ is distributed over instances in SAT and $\mathcal{D}^N$ is distributed over instances that don't belong to SAT, however, given only an instance drawn from either $\mathcal{D}^Y$ or $\mathcal{D}^N$ no non-uniform PPT algorithm could identify which distribution the instance was drawn from beyond negligible advantage. Stated more formally:

Corollary A.1 (Hard distribution for NP)**.** *Assuming that $\mathsf{NP} \not\subseteq \mathsf{ioP/poly}$, then there exists a pair of distribution ensembles $\mathcal{D}^Y = \{\mathcal{D}^Y_n\}_{n\in\mathbb{N}}, \mathcal{D}^N = \{\mathcal{D}^N_n\}_{n\in\mathbb{N}}$ with $\mathcal{D}^Y_n \subseteq \mathsf{SAT}_n$ and $\mathcal{D}^N_n \subseteq \{0,1\}^n \setminus \mathsf{SAT}_n$, such that for any non-uniform PPT algorithm A and every sufficiently large $n \in \mathbb{N}$:*

$$\left| \Pr_{x\leftarrow\mathcal{D}^Y_n}[A(x)=1] - \Pr_{x\leftarrow\mathcal{D}^N_n}[A(x)=1] \right| \leq \mathsf{negl}(n).$$

The above corollary is immediate from Lemma 2.4. We stress that these two distributions $\mathcal{D}^Y$ and $\mathcal{D}^N$ are not necessarily efficiently samplable. Indeed, if they were, then the sampler implies the existence of OWF. Still, it could be informative to consider a world in which we are granted access only to an oracle that when queried provided random sampled from $\mathcal{D}^Y$ with a satisfying witness (we refer to these oracles as hard-oracles). Stated formally:

Definition A.2 (Hard Oracle)**.** *For a relation $\mathcal{R}_{\mathcal{L}}$ to a language $\mathcal{L}$, we say that an oracle $\mathcal{O}$ is* hard for *$\mathcal{R}_{\mathcal{L}}$ if there exists a pair of distribution ensembles $\mathcal{D}^Y = \{\mathcal{D}^Y_n\}_{n\in\mathbb{N}}, \mathcal{D}^N = \{\mathcal{D}^N_n\}_{n\in\mathbb{N}}$ with $\mathcal{D}^Y_n \subseteq \mathcal{R}_{\mathcal{L}_n}$ and $\mathcal{D}^N_n \subseteq \{0,1\}^n \setminus \mathcal{L}_n$, such that on input 1^n, the oracle $\mathcal{O}$ returns a sample $(x,w) \leftarrow \mathcal{D}^Y_n$. Moreover, for every non-uniform* PPT *algorithm A with oracle access to $\mathcal{O}$,*

$$\left| \Pr_{(x,\cdot)\leftarrow\mathcal{D}^Y}\left[A^{\mathcal{O}}(x)=1\right] - \Pr_{x\leftarrow\mathcal{D}^N}\left[A^{\mathcal{O}}(x)=1\right] \right| \leq \mathsf{negl}(1^n).$$

A.1.2 Witness Encryption

Definition A.3. *A* witness encryption (WE) *scheme for an NP language L consists of two* PPT *algorithms,* Enc *and* Dec*, such that the following are satisfied:*

- **Correctness:** *For all $x \in L$, $n = |x|$, all witnesses w of x, any $b \in \{0,1\}$:*

$$\Pr[\mathsf{Dec}(\mathsf{Enc}(x,b),w) = b] \geq 1 - 2^{-n}.$$

- **Soundness Security:** *For every non-uniform polynomial-time machine D, there exists a negligible function ν such that for all $x \notin L$, $n = |x|$,*

$$|\Pr[D(\mathsf{Enc}(x,0)) = 1] - \Pr[D(\mathsf{Enc}(x,1)) = 1]| \leq \nu(n)$$

We say that (Enc, Dec) *is* statistically secure *if the two distributions in the soundness security guarantee are statistically indistinguishable. We say that* (Enc, Dec) *is a witness encryption with* $(1-\alpha)$-correctness *and* β-soundness *if the probability in the correctness and soundness condition is at least $(1-\alpha(n))$ and $\beta(n)$ respectively, on input lengths n.*

For an oracle $\mathcal{O}$, we say that a scheme is WE with respect to $\mathcal{O}$*, if the above holds when* (Enc, Dec) *and any* PPT *adversary A have oracle access to $\mathcal{O}$.*

When (Enc, Dec) *are* non-uniform PPT *algorithm that satisfy correctness and soundness security as stated above, we say that the scheme is a* non-uniform witness encryption.

We use the following amplification result for witness encryption [LMP25; CHS05], that turns weak WE schemes with into fully-fledged WE scheme. A careful inspection of the proof of [LMP25; CHS05] shows that this amplification holds even with respect to a no-input oracles.

Definition A.4 (No-input Oracles)**.** *We say that $\mathcal{O}$ is a no-input randomized oracles, if there exists a family of distribution $\mathcal{D} = \{\mathcal{D}_n\}_{n\in\mathbb{N}}$ such that on input 1^n, $\mathcal{O}$ outputs an independent sample of $y \leftarrow \mathcal{D}_n$.*

In Section 3 (Definition A.2) we specify the kind of no-input oracles that are needed for our results which we refer to as hard oracles.

Proposition A.5 (WE amplification [LMP25; CHS05])**.** *Assume that there exists a weak witness encryption scheme with $(1-\alpha)$-correctness and β-security soundness for two polynomials α, β such that $\alpha(n) \leq (1-\beta(n)) \cdot \frac{1}{8n}$. Then, there exists a (fully-fledged) witness encryption scheme. Moreover, the above statement preserves the statistical security between the schemes, and holds with respect to any no-input oracle $\mathcal{O}$.*

A.1.3 Predictable Arguments

Definition A.6 (Predictable Arguments, [FNV17])**.** *Let $\mathcal{R}$ be an NP relation and let $\Pi = (P,V)$ be an interactive argument with k rounds. We say that Π is a* predictable argument (PA) *for $\mathcal{R}$ if there exist a pair of* PPT *algorithms* (Chall, Resp) *such that, on input $(x,w) \in \mathcal{R}$:*

1. *The verifier samples a challenge and a prediction pair $(c,b) \leftarrow \mathsf{Chall}(x)$ where $c = (c_1, \cdots, c_k)$ and $b = (b_1, \cdots, b_k)$.*
2. *For all $i \in [k]$ in increasing sequence:*
   (a) *V forewords c_i to P.*
   (b) *P computes $(a_1, \cdots, a_i) := \mathsf{Resp}(x, w, c_1, \cdots, c_i)$ and forwards a_i to V.*

(c) V checks that $a_i = b_i$ and returns 0 (reject) if this is not the case.

3. *If all challenges are answered correctly, V returns 1 (accept).*

Π *has completeness and soundness as defined in Definition 2.8. For an oracle $\mathcal{O}$, we say that a scheme is* PA with respect to $\mathcal{O}$*, if the above holds when (P, V),* Chall *and any* PPT *adversary A have oracle access to samples from $\mathcal{O}$.*

We use the following result by [FNV17], that shows that predictable arguments are equivalent to witness encryption. The result was stated without addressing the presence of hard oracles however, a careful inspection of the proofs reveals that it holds even when the adversary has access to an oracle $\mathcal{O}$. This holds since the equivalence is shown throw a black-box reduction that still holds with respect to the oracle.

Theorem A.7 (PA and WE equivalence, [FNV17])**.** *There exists a predictable argument for $\mathcal{R}$ with respect to a no-input oracle $\mathcal{O}$ if and only if there exists a witness encryption (*WE*) scheme for $\mathcal{R}$ with respect to a no-input oracle $\mathcal{O}$.*

Since we are aiming for a more fine grain quantification on the security of the PA and WE we derive the following parametrized version that is immediate from the result of [FNV17] upon inspection.

Theorem A.8 (PA implies WE parametrized, [FNV17])**.** *There exists an efficient black-box reduction $\mathcal{B}$ that terns any predictable argument* (PA) *for a $\mathcal{R}$ with respect to a no-input oracle $\mathcal{O}$ with $(1-\alpha)$-completeness and non-uniform computational β-soundness and transcript length ℓ into a witness encryption* (WE) *scheme for $\mathcal{R}$ with respect to $\mathcal{O}$ with $(1-\alpha)$-correctness and non-uniform $(1-\frac{1-\beta}{\ell})$-soundness security.*

A.2 Application to Hard Functions: Theorem Statement

Intuitively, given only oracle access to a hard-oracle (even for the NP complete relation $\mathcal{R}_{\mathsf{SAT}}$) seem unlikely to imply any form of cryptographic hardness on its own. In particular, these hard oracles could conceivably still exists even if AI-OWF don't.

This raises the following natural question. Could the "hardness" of hard-oracles still be leveraged for meaningful cryptography? Our next theorem shows that either WI implies cryptographic hardness (specifically, AI-OWF) on its own, or these WI protocols when granted access to a hard-oracle imply the existence of witness encryption schemes that are secure against adversaries that also have access to the oracle. This is stated formally in the following theorem.

Witness Encryption From Hard Oracles and Deterministic-Prover WI. Recall that the relation $\mathcal{R}_{\mathsf{OR\text{-}SAT}}$ consists of instances that are the "OR" $\varphi^{OR} = \varphi \vee \varphi'$ of two equally sized formulas such that φ^{OR} is satisfiable.

Theorem A.9. *Let $\mathcal{O}$ be an oracle that is hard for $\mathcal{R}_{\mathsf{SAT}}$, and let (P, V) be an $\varepsilon(n)$-WI argument for $\mathcal{R}_{\mathsf{OR\text{-}SAT}}$, such that $\varepsilon(n) < \frac{1}{150 \cdot k(n) \cdot \ell(n)^2 \cdot n}$ where ℓ denotes the polynomial upper bound on the length of the transcript and k denotes the polynomial upper bound on the number of rounds. Then at least one of the following holds:*

1. *There exists an infinitely often auxiliary-input one-way function (AI-OWF).*
2. *There exists a witness encryption scheme with respect to $\mathcal{O}$.*

Moreover, the same implication holds when the initial WI provided has a non-uniform prover and verifier, where the resulting witness encryption has non-uniform encoding and decoding schemes.[11]

The proof for Theorem A.9 is given in Section 3. To prove the above theorem we make use of Theorem 3.1. Loosely speaking, if no AI-OWF exists then by Theorem 3.1 any WI protocol could be turned into a deterministic prover SWI protocol. We then show that deterministic-prover SWI when equipped with a hard Oracle, could be turned into a predictive argument PA (see Definition A.6) that by the result of [FNV17] are known to imply witness encryption. The main observation, is that it is

[11] An alternative way to view this, is that there is a public parameter that is uniformly fixed once and given to any encoder and decoding, and anyone including any potential adversary also has access to this public information.

possible to preform the OR trick [FLS99] by using only samples from this hard distribution (without having access to a function that generates these distributions).

The following corollary extends the above implication of Theorem A.9 to other primitives that where shown to imply WI protocols.

Corollary A.10. *Let $\mathcal{O}$ be an oracle that is hard for $\mathcal{R}_{\mathsf{SAT}}$. Assume that at least one of the following exists:*

- *A batch proof for $\mathcal{R}_{\mathsf{SAT}}$ with ρ-compression rate for $\rho(n,t) < 1/(c \cdot t^3 \cdot n)^{\frac{2}{7}}$, for a significantly large constant c or,*
- *A γ-succinct protocol for $\mathcal{R}_{\mathsf{SAT}}$ for $\gamma < 1/7$.*

Then at least one of the following holds:

1. *There exists an infinitely often auxiliary-input one-way function (AI-OWF).*
2. *There exists a non-uniform witness encryption scheme with respect $\mathcal{O}$.*

Remark A.11. *To prove the above corollary we use the recent result by Bitansky, Kamath, Paneth, Rothblum, and Vasudevan [BKPRV24] that showed that batch proofs (and succinct arguments for NP as they trivially imply batch proofs) imply SWI. However, since the result of [BKPRV24] only yields non-uniform provers our construction for predictive arguments also requires this non-uniformity which ultimately translates into a non-uniform encryption scheme.*

This implication also extends to succinct arguments for NP (as they trivially imply batch proofs), and for well studied non-interactive protocols such as SNARGs and non-interactive batch proofs.

Perspective. To see why the above notion of WE with respect to an oracle might still be interesting and useful, consider the following two informal examples. First, suppose there are no one-way functions, but there exists a hard distribution that can be sampled using a quantum circuit (i.e., one-way puzzles [KT24]). In such a setting, our framework suggests the possibility of a WE scheme (and hence public-key encryption) in which the encryption algorithm is quantum while the communication remains classical.

As a second example, consider a key-agreement protocol with an offline phase: in the offline phase, the parties (perhaps inefficiently) independently generate samples from the hard distribution $\mathcal{D}$, and in the online phase they communicate securely using the WE scheme. Our result shows that, assuming that WI exists but that AI-OWF does not, there exists a secure key-agreement protocol with an offline phase.

We interpret this result as providing evidence that witness indistinguishability inherently relies on some form of "cryptographic hardness". Although it remains open whether WI fundamentally require cryptographic assumptions, our findings suggest that the absence of such assumptions would be quite surprising: either WI implies the existence of auxiliary-input one-way functions, or it strengthens the power of cryptographic hardness assumptions in a nontrivial way.

Proof of Corollary A.10. We handle the two items starting with the first.

Batch proof with ρ-compression. By Theorem 2.15, a batch proof/argument Π for $(\mathcal{R}_{\mathsf{SAT}})^{\otimes t}$ with compression rate ρ yields an ε-SWI protocol $(\widehat{P}, \widehat{V})$ with non-uniform prover and verifier satisfying

$$\varepsilon \;\le\; c_1\sqrt{\rho} \qquad \text{and} \qquad \ell \;=\; \ell_P \;=\; \rho \cdot t.$$

If $\rho < (c\,t^3\,n)^{-2/7}$, then $\varepsilon \le c_1(c\,t^3\,n)^{-1/7}$ and $\ell \le t\cdot(c\,t^3\,n)^{-2/7} = \frac{t^{1/7}}{c^{2/7}n^{2/7}}$. Choosing $c \ge 150\,c_1$ gives $\varepsilon \le \frac{1}{150\,\ell^3\,n} \le \frac{1}{150\cdot k\cdot \ell^2\cdot n}$, hence by Theorem A.9 there exists a non-uniform witness encryption scheme with respect to $\mathcal{O}$.

γ-succinct protocol with $\gamma < 1/7$. A γ-succinct protocol for $\mathcal{R}_{\mathsf{SAT}}$ has prover communication at most $(|x|+|w|)^\gamma$. For a batch of t instances $(x_1,\ldots,x_t)$ with witnesses $(w_1,\ldots,w_t)$, run this protocol once on the concatenated instance; putting $t = (|x_i|+|w_i|)^{c_t}$ for a sufficiently large constant $c_t > 1$,

$$\ell_P \;\le\; \big(t(|x_i|+|w_i|)\big)^\gamma = t^\gamma \cdot (|x_i|+|w_i|)^\gamma = t^{\gamma+\gamma/c_t}.$$

Thus the induced batch compression rate is

$$\rho_{\text{succinct}} = \frac{\ell_P}{t} \leq t^{\gamma+\gamma/c_t-1} = \frac{1}{t^{1-\gamma-\gamma/c_t}}.$$

Since $n < t^{1/c_t}$, we have

$$(c\,t^3\,n)^{2/7} \leq c^{2/7} \cdot t^{6/7+2/(7c_t)}.$$

Thus, as $\gamma < 1/7$, for sufficiently large c_t it holds that

$$1 - \gamma - \frac{\gamma}{c_t} > \frac{6}{7} + \frac{2}{7c_t},$$

which implies that

$$\rho_{\text{succinct}} \leq \frac{1}{t^{1-\gamma-\gamma/c_t}} < \frac{1}{(c\,t^3\,n)^{2/7}}$$

(for a suitable constant c). Hence we can reduces to first item of batch compression, yielding $\varepsilon \leq \frac{1}{150\,\ell^3\,n} \leq \frac{1}{150\,k\,\ell^2\,n}$ and, by Theorem A.9, a non-uniform witness encryption scheme with respect to $\mathcal{O}$. □

A.3 Proof of Theorem A.9

To prove Theorem A.9 we will make use of the following lemma that shows that if there exits deterministic provers SWI for the $\mathcal{R}_{\mathsf{OR\text{-}SAT}}$ relation then, given a hard oracle $\mathcal{O}$ there exists a predictive arguments for $\mathcal{R}_{\mathsf{SAT}}$ with respect to $\mathcal{O}$.

Lemma A.12. *Let $\mathcal{O}$ be a hard oracle for the relation $\mathcal{R}_{\mathsf{SAT}}$, and let (P, V) be a deterministic prover honest-verifier δ-SWI interactive argument for $\mathcal{R}_{\mathsf{OR\text{-}SAT}}$ with $(1-\alpha)$-completeness and non-uniform computational β-soundness. Then there is a predictive argument $(\widehat{P}, \widehat{V}^{\mathcal{O}})$ for $\mathcal{R}_{\mathsf{SAT}}$ with respect to $\mathcal{O}$, with $(1-\alpha-\delta)$-completeness and non-uniform $(\beta+\alpha+\mu)$-soundness where μ is a negligible function. Moreover, $\widehat{V}$ makes a single call to $\mathcal{O}$.*

We prove Lemma A.12 below, but first we make use of it to prove Theorem A.9.

Proof of Theorem A.9. Let ℓ be the polynomial that upper bound on the length of the transcript of the protocol. Note that by Theorem 3.1, either there exists an AI-OWF, or for every $q \in \text{poly}$ there exists a γ-SWI protocol (P, V) for $\mathcal{R}$ with a deterministic (non-uniform) prover and an honest verifier, with $(1-1/q)$-completeness and

$$\gamma \leq 3\varepsilon \cdot \ell \cdot k + 1/q.$$

By Lemma A.12, this yields a predictive argument for $\mathcal{R}$ with respect to the oracle $\mathcal{O}$, with $(1-\gamma-1/q)$-completeness and non-uniform computational soundness, up to a negligible loss.

Since by assumption

$$\varepsilon \leq \frac{1}{150 \cdot \ell^2 \cdot k \cdot n},$$

we get that

$$3\varepsilon \cdot \ell \cdot k \leq \frac{1}{50 \cdot \ell \cdot n}.$$

By taking q sufficiently large, we may ensure that

$$\frac{1}{q} \leq \frac{1}{50 \cdot \ell \cdot n}.$$

Therefore,

$$\gamma \leq 3\varepsilon \cdot \ell \cdot k + \frac{1}{q} \leq \frac{1}{25 \cdot \ell \cdot n}. \tag{27}$$

We may now apply Theorem A.8 to obtain a WE scheme with $(1-\gamma-1/q)$-correctness and $(\beta' := (1-\frac{1}{2\ell}))$-soundness. Finally, by Proposition A.5, we can amplify this scheme to obtain a fully-fledged WE scheme provided that $\gamma + 1/q < \frac{1}{8} \cdot (1-\beta') \cdot \frac{1}{n} = \frac{1}{8} \cdot \frac{1}{2\ell} \cdot \frac{1}{n}$, which is satisfied by Eq. (27). This concludes the proof.

□

Proof of Lemma A.12. The proof uses the OR trick [FLS99]. Recall that $(\varphi, v) \in \mathcal{R}_{\mathsf{SAT}}$ iff $\varphi[v] = 1$ and that $(\widehat{P}, \widehat{V})$ is a PA for $\mathcal{R}_{\mathsf{SAT}}$ if there exist a pair of PPT algorithms $(\mathsf{Chall}, \mathsf{Resp})$ that satisfies Definition A.6. Recall also that $(c, b) \leftarrow \mathsf{Chall}^{\mathcal{O}}(\varphi)$, $(a_1, \cdots, a_i) = \mathsf{Resp}(\varphi, v, c_1, \cdots, c_i)$ and that (P, V) is a δ-SWI with a deterministic prover for $\mathcal{R}_{\mathsf{OR\text{-}SAT}}$. We now define $(\mathsf{Chall}, \mathsf{Resp})$ (using (P, V)) which implicitly defines the predictive argument $(\widehat{P}, \widehat{V})$.

$\mathsf{Chall}^{\mathcal{O}}(\varphi)$**:**

- Call the oracle $\mathcal{O}$ once, and get $(\varphi', v') \leftarrow \mathcal{O}(1^{|\varphi|})$ (recall that by definition $\varphi'[v'] = 1$)
- Computes the "or" instance $\varphi^{OR} = \varphi \vee \varphi'$.
- Compute the interaction between P and V where the prover gets the instance φ^{OR} and the witness v' and the verifier has the instance φ^{OR}. That is, $(\rho'_1, m'_1, \cdots, \rho'_k, m'_k) = \langle P(\varphi^{OR}, v'), V(\varphi^{OR}) \rangle$.
- Output $c = (c_1 := (\varphi', \rho'_1), c_2 := \rho'_2, \cdots, c_k := \rho'_k)$ and $b = (m'_1, \cdots, m'_k)$, if the verifier V accepts (that is $\mathsf{Out}(\langle P(\varphi^{OR}, v'), V(\varphi^{OR}) \rangle) = 1$) otherwise set b to be a random string of size $|\varphi|$.

$\mathsf{Resp}(\varphi, v, c_1, \cdots, c_i)$:

- Parse $c_1 = (\varphi', \rho'_1)$ and $c_i = \rho'_i$ and compute the deterministic provers responses till round i on given instance $\varphi^{OR} = \varphi \vee \varphi'$ witness v and answers verifier messages $\rho'_1, \cdots, \rho'_i$ (run the interaction between P and V where in each round treat ρ'_i as the ith message of V and produce the prover message a_i). That is, compute $m_i = P_i(\varphi^{OR}, v, c_1, \cdots, c_i)$ and output $a_{\leq i} = (m_1, \cdots, m_i)$.

We now show that $(\widehat{P}, \widehat{V}^{\mathcal{O}})$ satisfy completeness and soundness.

Completeness: Recall that by construction for any $(\varphi, v) \in \mathcal{R}_{\mathsf{SAT}}$, Chall creates $\varphi^{OR} = \varphi \vee \varphi'$ where (φ', v') is a sample received from $\mathcal{O}$. Since by assumption $(\varphi', v') \in \mathcal{R}_{\mathsf{SAT}}$ it follows that $\varphi^{OR} \in \mathcal{R}_{\mathsf{OR\text{-}SAT}}$.

Note that for any fixed (φ^{OR}, v') the values of $(c_1, \ldots, c_k)$ are fully determined by the random coins of V. Moreover, by definition $(\widehat{P}, \widehat{V}^{\mathcal{O}})$ for every $(\varphi, v) \in \mathcal{R}_{\mathsf{SAT}}$,

$$\mathbf{Pr}\left[\mathsf{Out}(\langle \widehat{P}(\varphi), \widehat{V}^{\mathcal{O}}(\varphi) \rangle) = 1\right] = \mathop{\mathbf{Pr}}_{\substack{(c,b) \leftarrow \mathsf{Chall}^{\mathcal{O}}(\varphi) \\ a = \mathsf{Resp}(\varphi, v, c)}} [a = b] .$$

Observe that if the interaction of (P, V) accepts on instance φ^{OR} with witness v' that is $\mathsf{Out}(\langle P(\varphi^{OR}, v'), V(\varphi^{OR}) \rangle) = 1$ then the tuple $b = (m'_1, \ldots, m'_k)$ corresponds to the sequential (deterministic) prover responses to the verifier messages $(c_1, \ldots, c_k)$ produced by $P(\varphi^{OR}, v')$, which uses the witness v'. Similarly, $a = (m_1, \ldots, m_k)$ denotes the (deterministic) prover responses produced by $P(\varphi^{OR}, v)$, this time using the witness v. Since by definition $\varphi^{OR}[v] = 1$ and $\varphi^{OR}[v'] = 1$, that is, both v and v' are valid witness for φ^{OR}, and since (P, V) is a δ-SWI with $(1 - \alpha)$-completeness it follows by union bound that

$$\mathop{\mathbf{Pr}}_{\substack{(c,b) \leftarrow \mathsf{Chall}^{\mathcal{O}}(\varphi) \\ a = \mathsf{Resp}(\varphi, v, c)}} [a = b] \geq 1 - \alpha - \delta$$

which implies that $(\widehat{P}, \widehat{V}^{\mathcal{O}})$ has $(1 - \alpha - \delta)$-completeness.

Soundness: Assume toward contradiction that soundness does not hold. That is, there exists a $\bar{\varphi} \notin \mathsf{SAT}$ and PPT prover $\bar{P}$ with oracle access to $\mathcal{O}$ and $q \in \text{poly}$, such that

$$\mathbf{Pr}\left[\mathsf{Out}(\langle \bar{P}^{\mathcal{O}}(\bar{\varphi}), \widehat{V}^{\mathcal{O}}(\bar{\varphi}) \rangle) = 1\right] > \beta + \alpha + 1/q(|\varphi|)$$

Define $\tilde{P}^{\mathcal{O}}$ a cheating prover for the relation $\mathcal{R}_{\mathsf{OR\text{-}SAT}}$, using $\bar{P}^{\mathcal{O}}$ as follows: On instance $\varphi^{OR} := \bar{\varphi} \vee \varphi'$, for $i = 1$, $\tilde{P}_1^{\mathcal{O}}$ receives ρ_1 and outputs $\bar{P}_1^{\mathcal{O}}(\bar{\varphi}, c = (\varphi', \rho_1))$, and for $1 < i \leq k$ define $\tilde{P}_i^{\mathcal{O}}(\cdot) := \bar{P}_i^{\mathcal{O}}(\cdot)$. Note that by construction, this implies that

$$\mathbf{Pr}\left[\mathsf{Out}(\langle \bar{P}^{\mathcal{O}}(\bar{\varphi}), \widehat{V}^{\mathcal{O}}(\bar{\varphi}) \rangle) = 1\right] \leq \mathop{\mathbf{Pr}}_{\substack{(\varphi', v') \leftarrow \mathcal{O}(|\bar{\varphi}|) \\ \varphi^{OR} := \bar{\varphi} \vee \varphi'}} \left[\mathsf{Out}(\langle \tilde{P}^{\mathcal{O}}(\varphi^{OR}), V^{\mathcal{O}}(\varphi^{OR}) \rangle) = 1\right] + \alpha$$

If the predicted honest transcript accepts and all predicted answers are matched, the corresponding execution of the underlying verifier accepts as well. The predicted honest transcript rejects with probability at most α. This proves the preceding inequality. This implies that

$$\mathop{\mathbf{Pr}}_{\substack{(\varphi',v')\leftarrow\mathcal{O}(|\bar{\varphi}|)\\ \varphi^{OR}:=\bar{\varphi}\vee\varphi'}}\left[\mathsf{Out}(\langle\tilde{P}^{\mathcal{O}}(\varphi^{OR}),V^{\mathcal{O}}(\varphi^{OR})\rangle)=1\right]\geq\beta+1/q>\beta+1/2q. \tag{28}$$

Moreover, by the assumption on the hardness of $\mathcal{O}$ it holds that there exists a distribution $\mathcal{D}^N\subseteq\{0,1\}^{|\bar{\varphi}|}\setminus\mathcal{R}_{\mathsf{SAT}}$ such that for any non-uniform adversary A

$$|\mathop{\mathbf{Pr}}_{(x,\cdot)\leftarrow\mathcal{O}(1^{|\bar{\varphi}|})}\left[A^{\mathcal{O}}(x)=1\right]-\mathop{\mathbf{Pr}}_{x\leftarrow\mathcal{D}^N}\left[A^{\mathcal{O}}(x)=1\right]|\leq\mu. \tag{29}$$

Where $\mu:=\mu(1^{|\bar{\varphi}|})$ is some negligible function. Thus by Eqs. (28) and (29) it follows that:

$$\begin{aligned}\mathop{\mathbf{Pr}}_{\substack{\varphi^N\leftarrow\mathcal{D}^N\\ \varphi'':=\bar{\varphi}\vee\varphi^N}}\left[\mathsf{Out}(\langle\tilde{P}^{\mathcal{O}}(\varphi''),V^{\mathcal{O}}(\varphi'')\rangle)=1\right]&\geq\mathop{\mathbf{Pr}}_{\substack{(\varphi',v')\leftarrow\mathcal{O}(|\bar{\varphi}|)\\ \varphi^{OR}:=\bar{\varphi}\vee\varphi'}}\left[\mathsf{Out}(\langle\tilde{P}^{\mathcal{O}}(\varphi^{OR}),V^{\mathcal{O}}(\varphi^{OR})\rangle)=1\right]-\mu\\ &\geq\beta+1/2q-\mu\\ &>\beta\end{aligned} \tag{30}$$

Since φ'' is a no-instance, fixing the no-instance φ^N and the independent answer tapes of $\mathcal{O}$ gives a non-uniform cheating prover contradicting the β-soundness of the underlying protocol.

□

B Deferred Proofs

B.1 Proving Lemma 2.4

Lemma B.1 (Lemma 2.4, restated)**.** *Assume that $\mathcal{L} \notin \mathsf{ioSIZE}[T(n)]$. Then, there exists a pair of distributions ensemble $\mathcal{D}^Y = \{\mathcal{D}^Y_n\}_{n\in\mathbb{N}}, \mathcal{D}^N = \{\mathcal{D}^N_n\}_{n\in\mathbb{N}}$ with $\mathcal{D}^Y_n \subseteq \mathcal{L}_n$ and $\mathcal{D}^N_n \subseteq \{0,1\}^n \setminus \mathcal{L}_n$, such that*

$$|\Pr_{x\leftarrow\mathcal{D}^Y}[A(x)=1] - \Pr_{x\leftarrow\mathcal{D}^N}[A(x)=1]| \leq \epsilon(n)$$

for any $A \in \mathsf{SIZE}[\frac{\epsilon^2(n)(T(n)-\mathrm{poly}(n,1/\epsilon(n)))}{160n}]$ and any large enough n.

Proof of Lemma 2.4. Let $s(n) = \frac{\epsilon^2(n)(T(n)-\mathrm{poly}(n,1/\epsilon(n)))}{160n}$. We start by showing that there exists a distribution ensemble $\mathcal{D} = \{\mathcal{D}^c_n\}_{n\in\mathbb{N}}$ with $\mathcal{D}^c_n \in \{0,1\}^n$, such that

$$\Pr_{x\leftarrow\mathcal{D}_n}[A(x) = L(x)] \leq 1/2 + \epsilon(n)/4$$

for any $A \in \mathsf{SIZE}[s(n)]$ and any large enough n. Later we will show how to construct the ensembles $\mathcal{D}^Y, \mathcal{D}^N$ from $\mathcal{D}$.

We want to show that there is a hard distribution against any algorithm in $\mathsf{SIZE}[s(n)]$. Towards this, for $n \in \mathbb{N}$, consider a two-party game, where Alice chooses an algorithm A, and Bob chooses an input $x \in \{0,1\}^n$. Alice wins if $A(x) = L(x)$, and Bob wins otherwise. We want to show that for a large enough $n \in \mathbb{N}$ Bob have a mixed strategy, such that when he plays first Alice wins the game with probability at most $1/2+\epsilon(n)/4$. By the min-max theorem it is enough to show that when Alice plays first, Bob has such a strategy. That is, we want to show that for every distribution $\mathcal{P}_n$ over circuits of size $s(n)$, there is an input $x \in \{0,1\}^n$, for which

$$\Pr_{A\leftarrow\mathcal{P}_n}[A(x) = L(x)] \leq 1/2 + \epsilon(n)/4.$$

Assume towards a contradiction this is not the case. That is, for every $x \in \{0,1\}^n$,

$$\Pr_{A\leftarrow\mathcal{P}_n}[A(x) = L(x)] > 1/2 + \epsilon(n)/4.$$

We will construct a circuit C of size $T(n)$ that decides L_n, which is a contradiction to the assumption. To construct C, let $k = 160n/\epsilon^2(n)$, and let $A_1, \ldots, A_k$ be k random samples from $\mathcal{P}_n$. Let $C_{A_1,\ldots,A_k}$ be the circuit that given an input x computes $A_1(x), \ldots, A_k(x)$, and outputs the majority of the answers. By assumption and Chernoff bound (Lemma 2.2), letting $B_i = \mathbf{1}[A_i(x) = L(x)]$,:

$$\mathbf{Pr}[C_{A_1,\ldots,A_k}(x) \neq L(x)] \leq \mathbf{Pr}\left[\sum_{i=1}^{k} B_i \leq k/2\right] \leq \exp(-k\epsilon(n)^2/8) < 2^{-n}.$$

and thus by the union bound, there is a choice of $A_1, \ldots, A_k$ such that $C_{A_1,\ldots,A_k}$ will answer correctly for any $x \in \{0,1\}^n$. Since the size of $C_{A_1,\ldots,A_k}$ is at most $\sum_i |A_i| + k^2 \leq T$, we get a contradiction to the assumption that $L \notin \mathsf{SIZE}[T]$.

Observe that by the fact that $\mathcal{D}_n$ is hard for small circuits, it must hold that $\mathbf{Pr}_{x\leftarrow\mathcal{D}_n}[x \in L] \in [1/2-\epsilon/4, 1/2+\epsilon/4]$. It follows tat we can choose a distribution $\mathcal{D}'$ such that $\mathrm{SD}(\mathcal{D}_n, \mathcal{D}'_n) \leq \epsilon/4$, and

$$\Pr_{x\leftarrow\mathcal{D}'_n}[x \in L] = 1/2.$$

Since $\mathrm{SD}(\mathcal{D}_n, \mathcal{D}'_n) \leq \epsilon$, we get that for every $A \in \mathsf{SIZE}[s(n)]$, it holds that

$$\Pr_{x\leftarrow\mathcal{D}'_n}[A(x) = L(x)] \leq 1/2 + \epsilon(n)/2.$$

We next define the two ensembles $\mathcal{D}^Y, \mathcal{D}^N$. For $n \in \mathbb{N}$, let $\mathcal{D}^Y_n$ be the distribution of $(x \leftarrow \mathcal{D}'_n)|_{x\in L}$, and let $\mathcal{D}^N_n$ e the distribution of $(x \leftarrow \mathcal{D}'_n)|_{x\notin L}$. Assume that there is an algorithm $A \in |[|s(n)]$ such that

$$|\Pr_{x\leftarrow\mathcal{D}^Y}[A(x)=1] - \Pr_{x\leftarrow\mathcal{D}^N}[A(x)=1]| > \epsilon(n)$$

and assume without loss of generality that

$$\Pr_{x\leftarrow\mathcal{D}^Y}[A(x)=1] > \Pr_{x\leftarrow\mathcal{D}^N}[A(x)=1]+\epsilon(n).$$

We get that

$$\Pr_{x\leftarrow\mathcal{D}'_n}[A(x)=L(x)] = 1/2\Pr_{x\leftarrow\mathcal{D}^Y_n}[A(x)=1]+1/2(1-\Pr_{x\leftarrow\mathcal{D}^N_n}[A(x)=1]) > 1/2+\epsilon(n)/2,$$

with contradiction to the choice of $\mathcal{D}'$. □

B.2 Proving Claim 3.9

Claim B.2 (Claim 3.9, restated)**.** *Let $X=(Z,Y)$ be a distribution over $\{0,1\}^{n_0}\times\{0,1\}^n$, and let* Pred *be a next-bit μ-predictor for X. Let $\tilde{Y}\equiv \mathrm{Sim}^{\mathsf{Pred}}(Z,R)$ be the random variable generated by the simulated distribution of* Pred *on initial state Z and independent uniform stings $R=(R_1,\ldots,R_n)$. Then for every $i\in[n+1]$,*

$$\mathrm{SD}((Z,Y_{<i}),(Z,\tilde{Y}_{<i})) \;\le\; 2\cdot(i-1)\cdot\mu.$$

Proof. We prove by induction on i. For $i=1$, both sides are equal to Z, so the bound holds. For some $i\ge 1$ assume $\mathrm{SD}((Z,Y_{<i}),(Z,\tilde{Y}_{<i}))\le 2(i-1)\mu$; we show $\mathrm{SD}((Z,Y_{\le i}),(Z,\tilde{Y}_{\le i}))\le 2i\mu$. For a fixed prefix $w=(z,y_{<i})\in\{0,1\}^{n_0}\times\{0,1\}^{i-1}$ let

$$p(w)=\mathbf{Pr}\left[Y_i=1\,\middle|\,(Z,Y_{<i})=w\right].$$

By the definition of a next-bit μ-predictor,

$$\Pr_{\substack{w=(z,y_{<i})\leftarrow(Z,Y_{<i})\\ r_{\mathsf{Pred}}\leftarrow\{0,1\}^{\ell_{\mathsf{Pred}}}}}[|\mathsf{Pred}(w;r_{\mathsf{Pred}})-p(w)|\le\mu] \;\ge\; 1-\mu.$$

Recall that by definition for any fixed $w=(z,y_{<i})$:

- $Y_i|_{(Z,Y_{<i})=w}\equiv U_{p(w)}$
- $\tilde{Y}_i|_{(Z,\tilde{Y}_{<i})=w}\equiv \mathrm{Samp}^{\mathsf{Pred}}(w;R_{\mathsf{Pred}},R_S)\equiv U_{\mathsf{Pred}(w;R_{\mathsf{Pred}})}$

where U_p denotes the distribution over $\{0,1\}$ that takes 1 with probability p. Let

$$G=\{(r_{\mathsf{Pred}},w)\in\{0,1\}^{\ell_{\mathsf{Pred}}}\times\{0,1\}^{n_0+i-1}\colon |\mathsf{Pred}(w;r_{\mathsf{Pred}})-p(w)|\le\mu\}$$

and note that for every $(r_{\mathsf{Pred}},w)\in G$ the statistical distance between $U_{p(w)}$ and $U_{\mathsf{Pred}(w;r_{\mathsf{Pred}})}$ is at most μ. When (r_{Pred},w) are "bad", that is $(r_{\mathsf{Pred}},w)\notin G$ (which happens with probability at most μ), this distance is trivially at most 1. Averaging over r_{Pred} and w gives

$$\mathop{\mathbf{E}}_{\substack{w=(z,y_{<i})\leftarrow(Z,Y_{<i})\\ r\leftarrow\{0,1\}^{\ell_{\mathsf{Pred}}}}}\left[\mathrm{SD}\left(Y_i\,|_{(Z,Y_{<i})=w}\;,\;\tilde{Y}_i\,|_{(Z,\tilde{Y}_{<i})=w}\right)\right] \;\le\; (1-\mu)\cdot\mu \;+\; \mu\cdot 1 \;\le\; 2\mu.$$

We now apply the standard (hybrid argument) for statistical distance:

$$\mathrm{SD}((Z,Y_{\le i}),(Z,\tilde{Y}_{\le i}))\le\mathrm{SD}((Z,Y_{<i}),(Z,\tilde{Y}_{<i}))+\mathop{\mathbf{E}}_{\substack{w=(z,y_{<i})\leftarrow(Z,Y_{<i})\\ r\leftarrow\{0,1\}^{\ell_{\mathsf{Pred}}}}}\left[\mathrm{SD}\left(Y_i\,|_{(Z,Y_{<i})=w}\;,\;\tilde{Y}_i\,|_{(Z,\tilde{Y}_{<i})=w}\right)\right].$$

Using the inductive hypothesis for the first term and the bound 2μ for the expectation,

$$\mathrm{SD}((Z,Y_{\le i}),(Z,\tilde{Y}_{\le i}))\le 2\cdot(i-1)\cdot\mu+2\mu=2\cdot i\cdot\mu.$$

This completes the induction and the proof. □

B.3 Proving Lemma 3.10

Lemma B.3 (Lemma 3.10, restated)**.** *Let (P, V) be a computational honest-verifier ϵ-WI for a relation $\mathcal{R}$ with an efficient prover (given a witness), and let $\ell(n)$ denote the polynomial that bounds the length of the transcript of the protocol. Consider the function family $\{f_a\}_{a\in\{0,1\}^{m(n)}}$ defined for $a = (x, w_0, w_1) \subseteq \{0,1\}^{m(n)}$ with $(x, w_0) \in \mathcal{R}_n$ and $(x, w_1) \in \mathcal{R}_n$, and for every bit $b \in \{0,1\}$, index $i \in [\ell(n)]$ and prover and verifier randomness r_P and r_V:*

$$f_a(i, b, r_P, r_V) \stackrel{\text{def}}{=} \langle P(x, w_b; r_P), V(x; r_V)\rangle_{\le i}.$$

If $\{f_a\}_a$ is not an infinitely often distributional auxiliary-input one-way function, then the protocol (P, V) is a honest-verifier ϵ'–SWI, for $\epsilon' = \epsilon + \mathsf{negl}$.

Proof of Lemma 3.10. Fix an instance $a = (x, w_0, w_1)$ with $(x, w_0), (x, w_1) \in \mathcal{R}_n$, and let

$$T_b := \langle P(x, w_b; R_P), V(x; R_V)\rangle$$

denote the full transcript of the protocol when the prover uses witness w_b, where R_P, R_V are sampled uniformly and independently.

Assume that the function family $\{f_a\}_a$ is not an infinitely often distributional auxiliary-input one-way function. Let $\mu \in 1/\mathrm{poly}$. By the definition of distributional $ioAI-OWF$, applied with parameter μ/ℓ, there exists an efficient randomized algorithm Inv such that

$$\mathrm{SD}\Big(\big(\mathsf{Inv}(a, f_a(I, B, R_P, R_V); R_{\mathsf{Inv}}),\ f_a(I, B, R_P, R_V)\big),\ \big((I, B, R_P, R_V),\ f_a(I, B, R_P, R_V)\big)\Big) \le \frac{\mu}{\ell},$$

where I is uniform in $[\ell]$, B is uniform in $\{0,1\}$, and all randomness is independent.

Conditioning on the event $I = \ell$, which occurs with probability $1/\ell$, increases the statistical distance by at most a factor of ℓ. Hence

$$\mathrm{SD}\Big(\big(\mathsf{Inv}(a, T_B; R_{\mathsf{Inv}}),\ T_B\big),\ \big((\ell, B, R_P, R_V),\ T_B\big)\Big) \le \mu. \tag{31}$$

For every transcript y, define

$$p(y) := \mathbf{Pr}[B = 1 \mid T_B = y].$$

Also, writing

$$\mathsf{Inv}(a, y; r) = (\hat{\ell}, \hat{B}, \hat{R}_P, \hat{R}_V),$$

define

$$p^{\mathsf{Inv}}(y) := \mathop{\mathbf{Pr}}_{r}[\hat{B} = 1].$$

Since the two distributions in (31) have the same marginal on $y = T_B$, we may write their statistical distance as the average, over $y \leftarrow T_B$, of the statistical distance between the corresponding conditional distributions. Projecting to the B-coordinate gives

$$\mathop{\mathbf{E}}_{y\leftarrow T_B}\big[\,|p(y) - p^{\mathsf{Inv}}(y)|\,\big] \le \mu. \tag{32}$$

We now turn $p^{\mathsf{Inv}}(y)$ into an efficient estimator. Let $N = \Theta(1/\mu^2)$ be a sufficiently large polynomial, and define the efficient randomized algorithm A_a as follows: on input y, sample independent random strings $r_1, \ldots, r_N$, compute

$$\mathsf{Inv}(a, y; r_j) = (\hat{\ell}_j, \hat{B}_j, \hat{R}_{P,j}, \hat{R}_{V,j}),$$

and output

$$A_a(y) := \frac{1}{N}\sum_{j=1}^{N} \mathbf{1}[\hat{B}_j = 1].$$

For every fixed y, the random variable $A_a(y)$ is the average of N random variables of mean $p^{\mathsf{Inv}}(y)$, and therefore by a standard Chernoff bound,

$$\mathbf{E}\big[\,|A_a(y) - p^{\mathsf{Inv}}(y)|\,\big] \le 2\mu,$$

provided N is chosen large enough. Averaging also over $y \leftarrow T_B$ and combining with (32), we obtain

$$\mathop{\mathbf{E}}_{y \leftarrow T_B,\, r}\big[\,|A_a(y) - p(y)|\,\big] \le 4\mu. \tag{33}$$

For $b \in \{0,1\}$, let $\mathcal{D}_b$ denote the distribution of T_b. Since B is uniform,

$$p(y) = \mathbf{Pr}[B = 1 \mid T_B = y] = \frac{\mathcal{D}_1(y)}{\mathcal{D}_0(y) + \mathcal{D}_1(y)}.$$

Let

$$D^*(y) := \mathbf{1}\left[p(y) \ge \frac{1}{2}\right]$$

be the optimal (possibly inefficient) likelihood-ratio test, and let

$$D'(y) := \mathbf{1}\left[A_a(y) \ge \frac{1}{2}\right].$$

The distinguisher D' is efficient.

Recall that

$$\mathrm{SD}(\mathcal{D}_0, \mathcal{D}_1) = \mathop{\mathbf{Pr}}_{y \leftarrow \mathcal{D}_1}[D^*(y) = 1] - \mathop{\mathbf{Pr}}_{y \leftarrow \mathcal{D}_0}[D^*(y) = 1].$$

Moreover, if $D'(y) \neq D^*(y)$, then necessarily

$$|2p(y) - 1| \le 2|A_a(y) - p(y)|.$$

Indeed, if $D'(y) \neq D^*(y)$, then $p(y)$ and $A_a(y)$ lie on opposite sides of $1/2$, which implies that $|p(y) - 1/2| \le |A_a(y) - p(y)|$, and hence $|2p(y) - 1| \le 2|A_a(y) - p(y)|$. Therefore,

$$\begin{aligned}
&\mathrm{SD}(\mathcal{D}_0, \mathcal{D}_1) - \Big(\mathop{\mathbf{Pr}}_{y \leftarrow \mathcal{D}_1}[D'(y) = 1] - \mathop{\mathbf{Pr}}_{y \leftarrow \mathcal{D}_0}[D'(y) = 1]\Big) \\
&\quad \le \sum_y |\mathcal{D}_1(y) - \mathcal{D}_0(y)| \cdot \mathbf{Pr}_r[D'(y;r) \neq D^*(y)] \\
&\quad = \sum_y (\mathcal{D}_0(y) + \mathcal{D}_1(y))\,|2p(y) - 1|\,\mathbf{Pr}_r[D'(y;r) \neq D^*(y)] \\
&\quad \le 2\sum_y (\mathcal{D}_0(y) + \mathcal{D}_1(y))\ \mathbf{E}\,|A_a(y) - p(y)| \\
&\quad = 4 \mathop{\mathbf{E}}_{y \leftarrow T_B,\, r}\big[\,|A_a(y) - p(y)|\,\big] \ \le\ 8\mu,
\end{aligned}$$

where the last inequality uses (33).

Hence

$$\mathrm{SD}(\mathcal{D}_0, \mathcal{D}_1) \le \Big(\mathop{\mathbf{Pr}}_{y \leftarrow \mathcal{D}_1}[D'(y) = 1] - \mathop{\mathbf{Pr}}_{y \leftarrow \mathcal{D}_0}[D'(y) = 1]\Big) + 8\mu.$$

Since D' is efficient and (P, V) is an ϵ-WI protocol,

$$\mathop{\mathbf{Pr}}_{y \leftarrow \mathcal{D}_1}[D'(y) = 1] - \mathop{\mathbf{Pr}}_{y \leftarrow \mathcal{D}_0}[D'(y) = 1] \le \epsilon.$$

Therefore,

$$\mathrm{SD}(\mathcal{D}_0, \mathcal{D}_1) \le \epsilon + 8\mu.$$

Since the above holds for any arbitrary $\mu \in 1/\mathrm{poly}$, it follows that

$$\mathrm{SD}(\mathcal{D}_0, \mathcal{D}_1) \le \epsilon + \mathsf{negl}(n).$$

Finally, recall that by definition $\mathcal{D}_b$ denotes the distribution of transcripts generated using witness w_b. Since the verifier's randomness is independent of the witness used by the prover conditioned on a fixed transcript, the verifier's view can be post-sampled from the transcript alone. We conclude that the protocol is ε'-SWI for $\varepsilon' = \varepsilon + \mathsf{negl}$ as desired. □

B.4 Proving Claim 3.14

Claim B.4 (Claim 3.14, restated)**.** *Let Z^0 and Z^1 be two distributions over $\{0,1\}^n$ such that $\mathrm{SD}(Z^0, Z^1) \leq \varepsilon$. For every $b \in \{0,1\}$ and $i \in [n]$, let $p^b(z_{<i}) = \mathbf{Pr}\left[Z_i^b = 1 \mid Z_{<i}^b = z_{<i}\right]$. Then,*

$$\mathop{\mathbf{E}}_{z_{<i} \leftarrow Z_{<i}^0} \left[|p^0(z_{<i}) - p^1(z_{<i})|\right] \leq 2\varepsilon.$$

Proof of Claim 3.14. Let $\mathcal{D}^b(z) = \mathbf{Pr}\left[Z_{<i}^b = z\right]$ denote the probability of generating prefix $z_{<i}$. By the definition of statistical distance, the distance between the two distributions truncated at bit i is:

$$\begin{aligned} 2\mathrm{SD}(Z_{\leq i}^0, Z_{\leq i}^1) = \sum_z \Big(& |\mathcal{D}^0(z)p^0(z) - \mathcal{D}^1(z)p^1(z)| \\ & + |\mathcal{D}^0(z)(1-p^0(z)) - \mathcal{D}^1(z)(1-p^1(z))| \Big). \end{aligned}$$

For any real numbers $a, c \geq 0$ and $b, d \in [0,1]$, the triangle inequality yields $|ab - cd| \geq a|b-d| - d|a-c|$ and $|a(1-b) - c(1-d)| \geq a|b-d| - (1-d)|a-c|$. Summing these gives:

$$|ab - cd| + |a(1-b) - c(1-d)| \geq 2a|b-d| - |a-c|.$$

Applying this inequality to the terms in our sum with $a = \mathcal{D}^0(z)$, $b = p^0(z)$, $c = \mathcal{D}^1(z)$, and $d = p^1(z)$, we obtain:

$$\begin{aligned} 2\mathrm{SD}(Z_{\leq i}^0, Z_{\leq i}^1) &\geq \sum_z 2\mathcal{D}^0(z)|p^0(z) - p^1(z)| - \sum_z |\mathcal{D}^0(z) - \mathcal{D}^1(z)| \\ &= 2 \mathop{\mathbf{E}}_{z \leftarrow Z_{<i}^0} \left[|p^0(z) - p^1(z)|\right] - 2\mathrm{SD}(Z_{<i}^0, Z_{<i}^1). \end{aligned}$$

Rearranging the terms, we get:

$$\mathop{\mathbf{E}}_{z \leftarrow Z_{<i}^0} \left[|p^0(z) - p^1(z)|\right] \leq \mathrm{SD}(Z_{\leq i}^0, Z_{\leq i}^1) + \mathrm{SD}(Z_{<i}^0, Z_{<i}^1).$$

Since truncating a distribution cannot increase its statistical distance, both $\mathrm{SD}(Z_{\leq i}^0, Z_{\leq i}^1)$ and $\mathrm{SD}(Z_{<i}^0, Z_{<i}^1)$ are bounded by the total statistical distance $\mathrm{SD}(Z^0, Z^1) \leq \varepsilon$. Therefore, the expectation is bounded by 2ε. □