Public-Coin Zero-Knowledge Arguments with (almost) Minimal Time and Space Overheads*

Alexander R. Block† Justin Holmgren‡ Alon Rosen§ Ron D. Rothblum¶
Pratik Soni‖

Abstract

Zero-knowledge protocols enable the truth of a mathematical statement to be certified by a verifier without revealing any other information. Such protocols are a cornerstone of modern cryptography and recently are becoming more and more practical. However, a major bottleneck in deployment is the efficiency of the prover and, in particular, the space-efficiency of the protocol.

For every NP relation that can be verified in time T and space S, we construct a public-coin zero-knowledge argument in which the prover runs in time $T \cdot \text{polylog}(T)$ and space $S \cdot \text{polylog}(T)$. Our proofs have length $\text{polylog}(T)$ and the verifier runs in time $T \cdot \text{polylog}(T)$ (and space $\text{polylog}(T)$). Our scheme is in the random oracle model and relies on the hardness of discrete log in prime-order groups.

Our main technical contribution is a new space efficient *polynomial commitment scheme* for multi-linear polynomials. Recall that in such a scheme, a sender commits to a given multi-linear polynomial $P : \mathbb{F}^n \to \mathbb{F}$ so that later on it can prove to a receiver statements of the form "$P(x) = y$". In our scheme, which builds on commitments schemes of Bootle et al. (Eurocrypt 2016) and Bünz et al. (S&P 2018), we assume that the sender is given multi-pass streaming access to the evaluations of P on the Boolean hypercube and we show how to implement both the sender and receiver in roughly time 2^n and space n and with communication complexity roughly n.

†Purdue University. Email: block9@purdue.edu.
‡NTT Research. Email: justin.holmgren@ntt-research.com.
§IDC Herzliya. Email: alon.rosen@idc.ac.il.
¶Technion. Email: rothblum@cs.technion.ac.il.
‖University of California, Santa Barbara. Email: pratik_soni@cs.ucsb.edu.

Contents

1 Introduction

Zero-knowledge protocols are a cornerstone of modern cryptography, enabling the truth of a mathematical statement to be certified by a prover to a verifier without revealing any other information. First conceived by Goldwasser, Micali, and Rackoff [GMR89], zero knowledge has myriad applications in both theory and practice and is a thriving research area today. Theoretical work primarily investigates the complexity tradeoffs inherent in zero-knowledge protocols:

- the number of rounds of interaction,
- the number of bits exchanged between the prover and verifier
- the computational complexity of the prover and verifier (e.g. running time, space usage)
- the degree of soundness—in particular, soundness can be statistical or computational, and the protocol may or may not be a proof of knowledge.

ZK-SNARKs (Zero-Knowledge Succinct Non-interactive ARguments of Knowledge) are protocols that achieve particularly appealing parameters: they are *non-interactive* protocols in which to certify an NP statement x with witness w, the prover sends a proof string π of length $|\pi| \ll |w|$. Such proof systems require setup (namely, a common reference string) and (under widely believed complexity-theoretic assumptions [GH98, GVW02]) are limited to achieving computational soundness.

One of the main bottlenecks limiting the scalability of ZK-SNARKs is the high computational complexity of generating proof strings. In particular, a major problem is that even for the lowest-overhead ZK-SNARKs (see e.g. [GGPR13, PHGR13, BBHR19] and follow-up works), the prover requires $\Omega(T)$ *space* to certify correctness of a time-T computation, even if that computation uses space $S \ll T$.

As typical computations require much less space than time, such space usage can easily become a hard bottleneck. While it is straight-forward to run a program for as long as one's patience allows, a computer's memory cannot be expanded without purchasing additional hardware. Moreover, the memory architecture of modern computer systems is hierarchical, consisting of different tiers (various cache levels, RAM, and nonvolatile storage), with latencies and capacities that increase by orders of magnitude at each successive level. In other words, high space usage can also incur a heavy penalty in running time.

In this work, we focus on uniform non-deterministic computations—that is, proving that a nondeterministic time-T space-S Turing machine accepts an input x. Our objective is to obtain "complexity-preserving" (ZK-)SNARKs [BC12a] for such computations, i.e., SNARKs in which the prover runs in time roughly T and space roughly S. Relatively efficient *privately verifiable* solutions are known [BC12b, HR18]. In such schemes the verifier holds some secret state that, if leaked, compromises soundness. However, many applications (such as cryptocurrencies or other massively decentralized protocols) require public verifiability, which is the emphasis of our work.

To date, *publicly verifiable* complexity-preserving SNARKs are known only via recursive composition [Val08, BCCT13]. This approach indeed yields SNARKs with prover running time $\tilde{O}(T)$ and space usage $S \cdot \mathrm{polylog}(T)$, but with significant concrete overheads. Recursively composed SNARKs require both the prover and verifier to make non-black-box usage of an "inner" verifier for a different SNARK, leading to enormous computational overhead in practice.

Several recent works [BGH19, BCMS20, COS20] attempt to solve the inefficiency problems with recursive composition, but the protocols in these works rely on heuristic and poorly understood

assumptions to justify their soundness. While any SNARK (with a black-box security reduction) inherently relies on non-falsifiable assumptions [GW11], these SNARKs possess additional troubling features. They rely on hash functions that are modeled as random oracles in the security proof, despite being used in a non-black-box way by the honest parties. Security thus cannot be reduced to a simple computational hardness assumption, even in the random oracle model. Moreover, the practicality of the schemes crucially requires usage of a novel hash function (e.g., Rescue [AAB+19]) with algebraic structure designed to maximize the *efficiency* of non-black-box operations. Such hash functions have endured far less scrutiny than standard SHA hash functions, and the algebraic structure could potentially lead to a security vulnerability.

In this work, we ask:

Can we devise a complexity-preserving ZK-SNARK in the random oracle model based on standard cryptographic assumptions?

1.1 Our Results

Our main result is an affirmative answer to this question.

Theorem 1. *Assume that the discrete-log problem is hard in obliviously sampleable*[1] *prime-order groups. Then, for every* NP *relation that can be verified by a random access machine in time T and space S, there exists a publicly verifiable ZK-SNARK, in the random oracle model, in which both the prover and verifier run in time $T \cdot \mathrm{polylog}(T)$, the prover uses space $S \cdot \mathrm{polylog}(T)$, and the verifier uses space $\mathrm{polylog}(T)$. The proof length is poly-logarithmic in T.*

We emphasize that the verifier in our protocol has similar running time to that of the prover, in contrast to other schemes in the literature that offer *poly-logarithmic* time verification. While this limits the usefulness of our scheme in delegating (deterministic) computations, our scheme is well-geared towards *zero-knowledge* applications in which the prover and verifier are likely to have similar computational resources.

At the heart of our ZK-SNARK for NP relations verifiable by time-T space-S random access machine (RAM) is a new public-coin *interactive* argument of knowledge, in the random oracle model, for the same relation where the prover runs in time $T \cdot \mathrm{polylog}(T)$ and requires space $S \cdot \mathrm{polylog}(T)$. We make this argument zero-knowledge by using standard techniques which incurs minimal asymptotic blow-up in the efficiency of the argument [BGG+90, CD98, WTs+18]. Finally, applying the Fiat-Shamir transformation [FS87] to our public-coin zero-knowledge argument yields Theorem 1.

1.1.1 Space-Efficient Polynomial Commitment for Multi-linear Polynomials

The key ingredient in our public-coin interactive argument of knowledge is a new space efficient *polynomial commitment scheme*, which we describe next.

Polynomial commitment schemes were introduced by Kate et al. [KZG10] and have since received much attention [BBHR18, BGKS19, BFS20, WBT+17, KPV19, ZXZS20], in particular due

[1] By *obliviously sampleable* we mean that there exist algorithms S and S^{-1} such that on input random coins r, the algorithm S samples a uniformly random group element g, whereas on input g, the algorithm S^{-1} samples random coins r that are consistent with the choice of g. In other words, if S uses ℓ random bits then the joint distributions $(U_\ell, S(U_\ell))$ and $(S^{-1}(S(U_\ell)), S(U_\ell))$ are identically distributed, where U_ℓ denotes the uniform distribution on ℓ bit strings.

to their usage in the construction of efficient zero-knowledge arguments. Informally, a polynomial commitment scheme is a cryptographic primitive that allows a committer to send to a receiver a commitment to an n-variate polynomial $Q : \mathbb{F}^n \to \mathbb{F}$, over some finite field $\mathbb{F}$, and later reveal evaluations y of Q on a point $\mathbf{x} \in \mathbb{F}^n$ of the receiver's choice along with a proof that indeed $y = Q(\mathbf{x})$.

In this work we construct polynomial commitment schemes where the space complexity is (roughly) *logarithmic* in the description size of the polynomial. In order to state this result more precisely, we must first determine the type of access that the committer has to the polynomial.

We first note that in this work we restrict our attention to *multi-linear* polynomials (i.e., polynomials which have individual degree 1). Note that such a polynomial $Q : \mathbb{F}^n \to \mathbb{F}$ is uniquely determined by its evaluations on the Boolean hybercube, that is, $(Q(0), \ldots, Q(2^n - 1))$, where the integers in $\mathbb{Z}_{2^n}$ are associated with vectors in $\{0, 1\}^n$ in the natural way.

Towards achieving our space efficient implementation, and motivated by our application to the construction of an efficient argument-scheme, we assume that the committer has *multi-pass streaming access* to the evaluations of the polynomial on the Boolean hypercube. Such an access pattern can be modeled by giving the committer access to a read-only tape that is pre-initialized with the values $(Q(0), \ldots, Q(2^n - 1))$. At every time-step the committer is allowed to either move the machine head to the right or to restart its position to 0.

Theorem 2 (Informal, see Theorem 5)**.** *Let $\mathbb{G}$ be an obliviously sampleable group of prime-order p and let $Q : \mathbb{F}^n \to \mathbb{F}$ be some n-variate multi-linear polynomial. Assuming the hardness of discrete-log over $\mathbb{G}$ and multi-pass streaming access to the sequence $(Q(0), \ldots, Q(2^n - 1))$, there exists a polynomial commitment scheme for Q in the random oracle model such that*

1. *The commitment consists of one group element, evaluation proofs consist of $O(n)$ group and field elements,*
2. *The committer and receiver perform $\tilde{O}(2^n)$ group and field operations, make $\tilde{O}(2^n)$ queries to the random oracle, and store only $O(n)$ group and field elements, and*
3. *The committer makes $O(n)$ passes over $(Q(0), \ldots, Q(2^n - 1))$.*

Following [KZG10], a number of works have focussed on achieving asymptotically optimal proof sizes (more generally, communication), and time complexity for both committer and receiver. However, the space complexity of the committer has been largely ignored; naively it is lower-bounded by the size of the committer's input (which is a description of the polynomial). As mentioned above, we believe that obtaining a space-efficient polynomial commitment scheme in the streaming model to be of independent interest and may even eventually lead to significantly improved performance of interactive oracle proofs, SNARKS, and related primitives in practice.

We also mention that the streaming model is especially well-suited to our application of building space-efficient SNARKs. The reason is that in such schemes, the prover typically uses a polynomial commitment scheme to commit to a low-degree extension of the transcript of a RAM program, which, naturally, can be generated as a stream in space that is proportional to the space complexity of the underlying RAM program.

At a high level, we use an algebraic homomorphic commitment (e.g., Pedersen commitment [Ped92]) to succinctly commit to the polynomial Q (by committing to the sequence $(Q(0), \ldots, Q(2^n - 1))$. Next, to provide evaluation proofs, our scheme leverages the fact that evaluating Q on point $\mathbf{x}$ reduces to computing an inner-product between $(Q(0), \ldots, Q(2^n - 1))$ and the sequence of Lagrange coefficients defined by the evaluation point $\mathbf{x}$. Relying on the homomorphic

properties of our commitment, the basic step of our evaluation protocol is a 2-move (randomized) reduction step which allows the committer to "fold" a statement of size 2^n into a statement of size $2^n/2$. Our scheme is inspired from the "inner-product argument" of Bootle et al. [BCC+16] (and its variants [BBB+18, WTs+18]) but differs in the 2-move reduction step. More specifically, their reduction step folds the left half of $(Q(0), \ldots, Q(2^n - 1))$ with its right half (referred to as msb-based folding as the index of the elements that are folded differ in the most significant bit). This, unfortunately, is not compatible with our streaming model (we explain this shortly). We instead perform the more natural *lsb-based folding* which, indeed, is compatible with the streaming model. We additionally exploit random access to the inner-product argument's setup parameters (defined by the random oracle) and the fact that any component of the coefficient sequence can be computed in polylogarithmic time, i.e. $\mathrm{poly}(n)$ time. We give a high level overview of our scheme in Section 2.1.

1.2 Prior Work

Complexity Preserving ZK-SNARKs. Bitansky and Chiesa [BC12b] proposed to construct complexity preserving ZK-SNARKS by first constructing complexity preserving multi-prover interactive proof (MIPs) and then compile them using cryptographic techniques. While our techniques share the same high-level approach, our compilation with a polynomial-commmitment scheme yields a publicly verifiable scheme whereas [BC12b] only obtain a designated verifier scheme.

Blumberg et al. [BTVW14] give a 2-prover complexity preserving MIP of knowledge, improving (concretely) on the complexity preserving MIP of [BC12b] (who obtain a 2-prover MIP via a reduction from their many-prover MIP). Both Bitansky and Chiesa and Blumberg et al. obtain their MIPs from reducing RAMs to circuits via the reduction of Ben-Sasson et al. [BCGT13], then appropriately arithmetize the circuit into an algebraic constraint satisfaction problem. Holmgren and Rothblum [HR18] obtain a non-interactive protocol based on standard (falsifiable assumptions) by also constructing a complexity preserving MIP for RAMs (achieving no-signaling soundness) and compiling it into an argument using fully-homomorphic encryption (á la [BMW98, KR09, KRR13]). We remark that [HR18] reduce a RAM directly to algebraic constraints via a different encoding of the RAM transcript, thereby avoiding the reduction to circuits entirely.

Another direction for obtaining complexity preserving ZK-SNARKS is via recursive composition [BCCT13, Val08], or "bootstrapping". Here, one begins with an "inefficient" SNARK and bootstraps it recursively to produce publicly verifiable complexity preserving SNARKs. While these constructions yield good asymptotics, these approaches require running the inefficient SNARK on many sub-computations. Recent works [BGH19, BCMS20, COS20] describe a novel approach to recursive composition which attempt to solve the inefficiencies of the aforementioned recursive compositions, though at a cost to the theoretical basis for the soundness of their scheme (as discussed above).

Interactive Oracle Proofs. Interactive oracle proofs (IOPs), introduced by Ben-Sasson et al. [BCS16] and independently by Reingold et al. [RRR16], are interactive protocols where a verifier has oracle access to all prover messages. IOPs capture (and generalize), both interactive proofs and PCPs.

A recent line of work [BCGT13, BTVW14, CMT12, GKR08, Set20, Tha13, WTs+18, RR19] follows the framework of Kilian [Kil92] and Micali [Mic94] to obtain efficient arguments by constructing efficient IOPs and compiling them into interactive arguments using collision resistant

hashing [BCS16, Kil92] or the random oracle model [BCS16, Mic94].

Polynomial Commitments. Polynomial commitment schemes were introduced by Kate et al. [KZG10] and have since been an active area of research. Lines of research for construction polynomial commitment schemes include privately verifiable schemes [KZG10, PST13], publicly-verifiable schemes with trusted setup [BFS20], and zero-knowledge schemes [WBT+17]. More recently, much focus has been on obtaining publicly-verifiable schemes without a trusted setup [BBHR18, BGKS19, BFS20, WBT+17, KPV19, ZXZS20]. We note that in all prior works on polynomial commitments, the space complexity of the sender is proportional to the description size of the polynomial, whereas we achieve *poly-logarithmic* space complexity.

2 Technical Overview

As mentioned above, the key component in our construction is that of a public-coin interactive argument for RAM computations. The latter construction itself consists of two key technical ingredients. First, we construct a *polynomial interactive oracle proof* (polynomial IOP) for time-T space-S RAM computations in which the prover runs in time $T \cdot \mathrm{polylog}(T)$ and space $S \cdot \mathrm{polylog}(T)$. We note that this ingredient is a conceptual contribution which formalizes prior work in the language of polynomial IOPs. Second, we compile this IOP with a space-efficient extractable *polynomial commitment scheme* where the prover has multi-pass streaming access to the polynomial to which it is committing—a property that plays nicely with the streaming nature of RAM computations. We emphasize that the construction of the space-efficient polynomial commitment scheme is our main technical contribution, and describe our scheme in more detail next.

2.1 Polynomial Commitment to Multi-linear Polynomials in the Streaming Model

Fix a finite field $\mathbb{F}$ of prime order p. Also fix an obliviously sampleable (see Footnote 1) group $\mathbb{G}$ of order p in which the discrete logarithm is hard. Let $H : \{0,1\}^* \to \mathbb{G}$ be the random oracle.

In order to describe our polynomial commitment scheme, we start with some notation. Let n be a positive integer and set $N = 2^n$. We will be considering N-dimensional vectors over $\mathbb{F}$ and will index such vectors using n dimensional binary vectors. For example, if $\mathbf{b} \in \mathbb{F}^{2^6}$ then $\mathbf{b}_{000101} = b_5$. For convenience, we will denote $\mathbf{b} \in \mathbb{F}^N$ by $(b_{\mathbf{c}} : \mathbf{c} \in \{0,1\}^n)$ where $b_{\mathbf{c}}$ is the $\mathbf{c}$-th element of $\mathbf{b}$. For $\mathbf{b} = (b_n, \ldots, b_1) \in \{0,1\}^n$ we refer to b_1 as the least-significant bit (lsb) of $\mathbf{b}$. Finally, for $\mathbf{b} \in \mathbb{F}^N$, we denote by $\mathbf{b}_e$ the restriction of $\mathbf{b}$ to the even indices, that is, $\mathbf{b}_e = (b_{\mathbf{c}0} : \mathbf{c} \in \{0,1\}^{n-1})$. Similarly, we denote by $\mathbf{b}_o = (b_{\mathbf{c}1} : \mathbf{c} \in \{0,1\}^{n-1})$ the restriction of $\mathbf{b}$ to odd indices.

Let $Q : \mathbb{F}^n \to \mathbb{F}$ be a multi-linear polynomial. Recall that such a polynomial can be fully described by the sequence of its evaluations over the Boolean hypercube. More specifically, for any $\mathbf{x} \in \mathbb{F}^n$, the evaluation of Q on $\mathbf{x}$ can be expressed as

$$Q(\mathbf{x}) = \sum_{\mathbf{b} \in \{0,1\}^n} Q(\mathbf{b}) \cdot z(\mathbf{x}, \mathbf{b}), \tag{1}$$

where $z(\mathbf{x}, \mathbf{b}) = \prod_{i \in [n]} (b_i \cdot x_i + (1 - b_i) \cdot (1 - x_i))$. We use $\mathbf{Q} \in \mathbb{F}^N$ to denote the restriction of Q to the Boolean hybercube (i.e., $\mathbf{Q} = (Q(\mathbf{b}) : \mathbf{b} \in \{0,1\}^n)$).

Next, we describe the our commitment scheme which has three phases: (a) Setup, (b) Commit and (c) Evaluation.

2.1.1 Setup and Commit Phase

During setup, the committer and receiver both consistently define a sequence of N generators for $\mathbb{G}$ using the random oracle, that is, $\mathbf{g} = (g_{\mathbf{b}} = H(\mathbf{b}) : \mathbf{b} \in \{0,1\}^n)$. Then, given streaming access to $\mathbf{Q}$, the committer computes the Pedersen multi-commitment [Ped92] C defined as

$$C = \prod_{\mathbf{b} \in \{0,1\}^n} (g_{\mathbf{b}})^{Q_{\mathbf{b}}} \ . \tag{2}$$

For $\mathbf{g} \in \mathbb{G}^{2^n}$ and $\mathbf{Q} \in \mathbb{F}^{2^n}$, we use $\mathbf{g}^{\mathbf{Q}}$ as a shorthand to denote the value $\prod_{\mathbf{b} \in \{0,1\}^n} (g_{\mathbf{b}})^{Q_{\mathbf{b}}}$. Assuming the hardness of discrete-log for $\mathbb{G}$, we note that C in Equation (2) is a binding commitment to $\mathbf{Q}$ under generators $\mathbf{g}$. Note that the committer only needs to perform a single-pass over $\mathbf{Q}$ and performs N exponentiations to compute C while storing only $O(1)$ number of group and field elements.[2]

2.1.2 Evaluation Phase

On input an evaluation point $\mathbf{x} \in \mathbb{F}^n$, the committer computes and sends $y = Q(\mathbf{x})$ and defines the auxiliary commitment $C_y \leftarrow C \cdot g^y$ for some receiver chosen generator g. Then, both engage in an argument (of knowledge) for the following NP statement which we refer to as the "inner-product" statement:

$$\exists \mathbf{Q} \in \mathbb{Z}_p^N \ : \ y = \langle \mathbf{Q}, \mathbf{z} \rangle \text{ and } \ C_y = g^y \cdot \mathbf{g}^{\mathbf{Q}} \ , \tag{3}$$

where $\mathbf{z} = (z(\mathbf{x}, \mathbf{b}) : \mathbf{b} \in \{0,1\}^n)$ as defined in Equation (1). This step can be viewed as proving knowledge of the decommitment $\mathbf{Q}$ of the commitment C_y, which furthermore is consistent with the inner-product claim that $y = \langle \mathbf{Q}, \mathbf{z} \rangle$.

Inner-product Argument. A basic step in the argument for the above inner-product statement is a 2-move randomized reduction step which allows the prover to decompose the N-sized statement $(C_y, \mathbf{z}, y)$ into two $N/2$-sized statements and then "fold" them into a single $N/2$-sized statement $(\bar{C}_{\bar{y}}, \bar{\mathbf{z}} = (\bar{z}_{\mathbf{c}} : \mathbf{c} \in \{0,1\}^{n-1}), \bar{y})$ using the verifier's random challenge. We explain the two steps below (as well as in Figure 1).

1. Committer computes the cross-product $y_e = \langle \mathbf{Q}_e, \mathbf{z}_o \rangle$ between the even-indexed elements $\mathbf{Q}_e$ with the odd-indexed vectors $\mathbf{z}_o$. Furthermore, it computes a binding commitment C_e that binds y_e (with g) and $\mathbf{Q}_e$ (with $\mathbf{g}_o$). That is,

$$C_e = g^{y_e} \cdot \mathbf{g}_o^{\mathbf{Q}_e} \ , \tag{4}$$

   where recall that for $\mathbf{g} = (g_1, \ldots, g_t)$ and $\mathbf{x} = (x_1, \ldots, x_t)$ the expression $\mathbf{g}^{\mathbf{x}} = \prod_{i \in [t]} g_i^{x_i}$. This results in an $N/2$-sized statement $(C_e, \mathbf{z}_o, y_e)$ with witness $\mathbf{Q}_e$. Similarly, as in Figure 1 it computes the second $N/2$-sized statement $(C_o, \mathbf{z}_e, y_o)$ with witness $\mathbf{Q}_o$. The committer sends (y_e, y_o, C_e, C_o) to the receiver.

[2] Here, we treat exponentiation as an atomic operation but note that computing g^α for $\alpha \in \mathbb{Z}_p$ can be emulated, via repeated squarings, by $O(\log p)$ group multiplications while storing only $O(1)$ number of group and field elements.

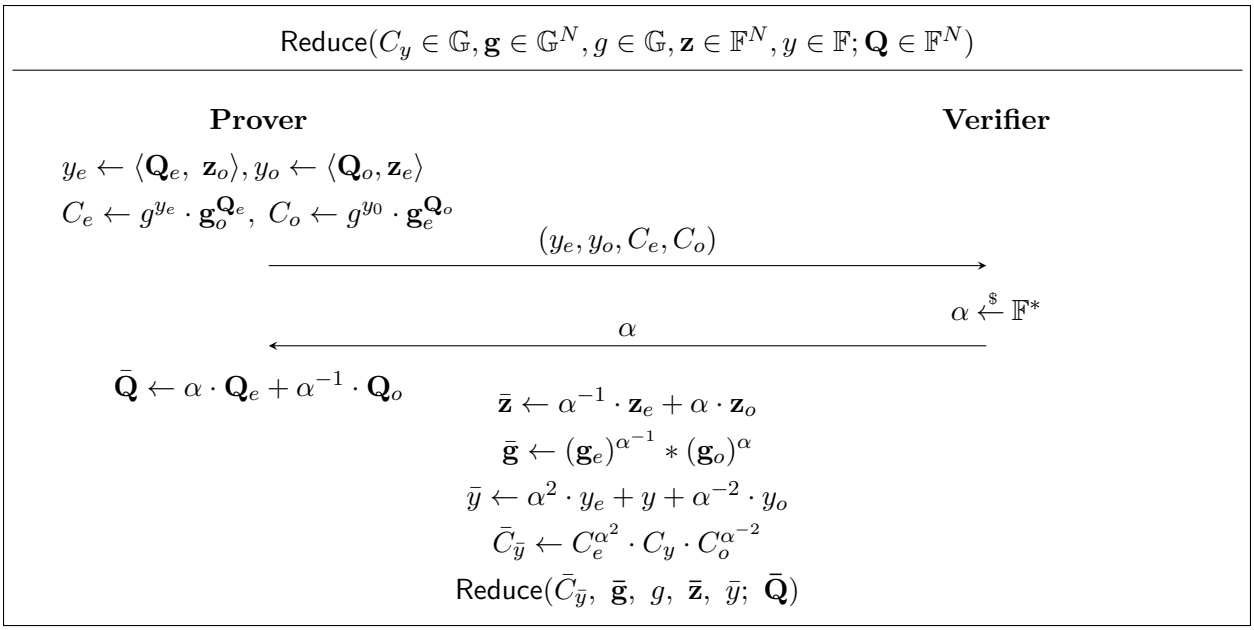

Figure 1: Our 2-move randomized reduction step for the inner-product protocol where recall that for any $\mathbf{Q} \in \mathbb{F}^N$, we denote by $\mathbf{Q}_e$ the elements of $\mathbf{Q}$ indexed by even numbers where $\mathbf{Q}_o$ denotes the elements with odd indices. On input a statement of size $N > 1$, Reduce results in a statement of size $N/2$.

2. After receiving a random challenge $\alpha \in \mathbb{F}^*$, committer folds its witness $\mathbf{Q}$ into an $N/2$-sized vector $\bar{\mathbf{Q}} = \alpha \cdot \mathbf{Q}_e + \alpha^{-1} \cdot \mathbf{Q}_o$. More specifically, for every $\mathbf{c} \in \{0,1\}^{n-1}$,

$$\bar{Q}_{\mathbf{c}} = \alpha \cdot Q_{\mathbf{c}0} + \alpha^{-1} \cdot Q_{\mathbf{c}1} \ . \tag{5}$$

   Similarly, the committer and receiver both compute the rest of the folded statement $(\bar{C}_{\bar{y}}, \bar{\mathbf{z}}, \bar{y})$ as shown in Figure 1.

Relying on the homomorphic properties of Pedersen commitments, it can be shown that if $\mathbf{Q}$ were a witness to $(C_y, \mathbf{z}, y)$ then $\bar{\mathbf{Q}}$ is a witness for $(\bar{C}_{\bar{y}}, \bar{\mathbf{z}}, \bar{y})$.[3] In the actual protocol, the parties then recurse on smaller statements $(\bar{C}_{\bar{y}}, \bar{\mathbf{z}}, \bar{y})$ forming a recursion tree. After $\log N$ steps, the statement is of size 1 in which case the committer sends its witness which is a single field element. This gives an overall communication of $O(\log N)$ field and group elements. Next we briefly discuss the efficiency of the scheme.

2.1.3 Efficiency

For the purpose of this overview, we focus only on the time and space efficiency of the committer in the inner-product argument (the analysis for the receiver is analogous). Recall that in a particular step of the recursion, suppose we are recursing on the $N/2$-sized statement $(\bar{C}_{\bar{y}}, \bar{\mathbf{z}}, \bar{y})$ with witness $\bar{\mathbf{Q}}$, the committer's computation includes computing (a) the cross-product $\langle \bar{\mathbf{Q}}_e, \bar{\mathbf{z}}_o \rangle$ between the

[3] albeit under different set of generators but we ignore this for now

Scheme	[BCC+16, BBB+18] (msb-based)	This work (lsb-based)
$\bar{Q}_{\mathbf{b}}$	$\alpha \cdot Q_{0\mathbf{b}} + \alpha^{-1} \cdot Q_{1\mathbf{b}}$	$\alpha \cdot Q_{\mathbf{b}0} + \alpha^{-1} \cdot Q_{\mathbf{b}1}$
$\bar{z}_{\mathbf{b}}$	$\alpha^{-1} \cdot z_{0\mathbf{b}} + \alpha \cdot z_{1\mathbf{b}}$	$\alpha^{-1} \cdot z_{\mathbf{b}0} + \alpha \cdot z_{\mathbf{b}1}$
$\bar{g}_{\mathbf{b}}$	$(g_{0\mathbf{b}})^{\alpha^{-1}} * (g_{1\mathbf{b}})^{\alpha}$	$(g_{\mathbf{b}0})^{\alpha^{-1}} * (g_{\mathbf{b}1})^{\alpha}$

Figure 2: Table highlights the differences between the 2-move randomized reduction steps of the inner-product argument of [BCC+16, BBB+18] (second column) and our scheme (third column). Specifically, given vectors $\mathbf{Q}, \mathbf{z}, \mathbf{g}$ of size 2^n, the rows describe the definition of the $2^n/2$ sized vectors $\bar{\mathbf{Q}}, \bar{\mathbf{z}}, \bar{\mathbf{g}}$ respectively where $\mathbf{b} \in \{0,1\}^{n-1}$.

even half of $\bar{\mathbf{Q}}$ and the odd half of $\bar{\mathbf{z}}$, and (b) the "cross-exponentiation" $\bar{\mathbf{g}}_o^{\bar{\mathbf{Q}}_e}$ of the even half of $\bar{\mathbf{Q}}$ with the odd half of the generators $\bar{\mathbf{g}}$.[4]

A straightforward approach to compute (a) is to have $\bar{\mathbf{Q}}$ (and $\bar{\mathbf{z}}$) in memory, but this requires the committer to have $\Omega(N)$ space which we want to avoid. Towards a space efficient implementation, first note every element of $\bar{\mathbf{Q}}$ depends on only two, more importantly, consecutive elements of $\mathbf{Q}$. This coupled with streaming access to $\mathbf{Q}$ is sufficient to simulate streaming access to $\bar{\mathbf{Q}}$ while making only **one** pass over $\mathbf{Q}$. Secondly, by definition, computing any element of $\mathbf{z}$ requires only $O(\log N)$ field operations while storing only $O(n)$ field elements This then allows to compute any element of $\bar{\mathbf{z}}$ on the fly with polylog(N) operations. Given the simulated streaming access to $\bar{\mathbf{Q}}$ along with the ability to compute any element of $\bar{\mathbf{z}}$ on the fly is sufficient to compute the $\langle \bar{\mathbf{Q}}_e, \bar{\mathbf{z}}_o \rangle$. Note this step, overall, requires performing only a single pass over $\mathbf{Q}$ and $N \cdot \text{polylog}\, N$ operations, and storing only the evaluation point $\mathbf{x}$ and verifier challenge α (along with some book-keeping). The computation of (b) is handled similarly, except that here we crucially leverage the fact that $\mathbf{g}$ is defined using the random oracle, and hence the committer has random access to all of the generators in $\mathbf{g}$. Relying on similar ideas as in (a), the committer can compute $\bar{\mathbf{g}}_o^{\bar{\mathbf{Q}}_e}$ while additionally making $O(N)$ queries to the random oracle. Overall, this gives the required prover efficiency. Please see Section 4.3 for a full discussion on the efficiency.

2.1.4 Comparison with the 2-move reduction step of [BCC+16, WTs+18]

In their protocol, a major difference is in how the folding is performed (Step 2, Figure 1). We list concrete differences in Figure 2. But at a high level, since they fold the first element $Q_{00^{n-1}}$ with the $N/2$-nd element $Q_{10^{n-1}}$, it takes at least a one pass over $\mathbf{Q}$ to even compute the first element of $\bar{\mathbf{Q}}$, thereby requiring $\Omega(N)$ passes over $\mathbf{Q}$ which is undesirable.[5] Although we differ in the 2-move reduction steps, the security of our scheme follows from ideas similar to [BCC+16, WTs+18].

2.2 Polynomial IOPs for RAM Programs

The second ingredient we use to obtain space-efficient interactive arguments for NP relations verifiable by time-T space-S RAMs is a space-efficient *polynomial interactive oracle proof system* [BCS16, RRR16, BFS20]. Informally, an interactive oracle proof (IOP) is an interactive protocol

[4] Efficiency for $\langle \bar{\mathbf{Q}}_o, \bar{\mathbf{z}}_e \rangle$ and $\bar{\mathbf{g}}_e^{\bar{\mathbf{Q}}_o}$ can be argued similarly.

[5] When a polynomial commitment is used in building arguments, it takes $O(N)$ time to stream $\mathbf{Q}$, and requiring $\Omega(N)$ passes results in a prover that runs in quadratic time.

such that in each round the verifier sends a message to the prover, and the prover responds with proof string that the verifier can query in only a few locations. A polynomial IOP is an IOP where the proof string sent by the prover is a polynomial (i.e, all evaluations of a polynomial on a domain), and if a cheating prover successfully convinces a verifier then the proof string is consistent with some polynomial.

We consider a variant of the polynomial IOP model in which the prover sends messages which are encoded by the channel; in particular, the time and space complexity of the encoding computed by the channel do not factor into the complexity of the prover. For our purposes, we use the polynomial IOP that is implicit in [BTVW14] and consider it with a channel which computes multi-linear extensions of the prover messages. We briefly describe the IOP construction for completeness (see Section 5 for more details). The polynomial IOP at its core first leverages the space-efficient RAM to arithemtic circuit satifsiability reduction of [BTVW14] (adapting techniques of [BCGT13]). This reduction transforms a time-T space-S RAM into a circuit of size $T \cdot \text{polylog}(T)$ and has the desirable property (for our purposes) that the circuit can be accessed by the prover in a streaming manner: the assignment of gate values in the circuit can be streamed "gate-by-gate" in time $T \cdot \text{polylog}(T)$ and space $S \cdot \text{polylog}(T)$, which, in particular, allows a prover to compute a correct transcript of the circuit in time $T \cdot \text{polylog}(T)$ and space $S \cdot \text{polylog}(T)$.

The prover sends the verifier an oracle that is the multi-linear extension of the gate values (i.e., the transcript), where we remark that this extension is computed by the channel. The correctness of the computation is reduced to an algebraic claim about a low degree polynomial which is identically 0 on the Boolean hypercube if and only if the circuit is satisfied by the given witness. Finally, the prover and verifier engage in the classical sum-check protocol [LFKN90, Sha90] to verify that the constructed polynomial indeed vanishes on the Boolean hypercube.

Theorem 3. *There exists a public-coin polynomial* IOP *over a channel which encodes prover messages as multi-linear extensions for* NP *relations verifiable by a time-T space-S random access machine M such that if $y = M(x; w)$ then*

1. *The* IOP *has perfect completeness and statistical soundness, and has $O(\log(T))$ rounds;*
2. *The prover runs in time $T \cdot \text{polylog}(T)$ and space $S \cdot \text{polylog}(T)$ (not including the space required for the oracle) when given input-witness pair $(x; w)$ for M, sends a single polynomial oracle in the first round, and has $\text{polylog}(T)$ communication in all subsequent rounds; and*
3. *The verifier runs in time $(|x| + |y|) \cdot \text{polylog}(T)$, space $\text{polylog}(T)$, and has query complexity 3.*

2.3 Obtaining Space-Efficient Interactive Arguments

We compile Theorem 3 and Theorem 2 into a space-efficient interactive argument scheme for NP relations verifiable by RAM computations.

Theorem 4 (Informal, see Theorem 6)**.** *There exists a public-coin interactive argument for* NP *relations verifiable by a time-T space-S random access machine M, in the random oracle model, under the hardness of discrete-log in obliviously sampleable prime-order groups such that:*

1. *The prover runs in time $T \cdot \text{polylog}(T)$ and space $S \cdot \text{polylog}(T)$;*
2. *The verifier runs in time $T \cdot \text{polylog}(T)$ and space $\text{polylog}(T)$; and*

3. The round complexity is $O(\log T)$ and the communication complexity is $\text{polylog}(T)$.

The interactive argument of Theorem 4 is obtained by modifying the polynomial IOP of Theorem 3 with the commitment scheme of Theorem 2 in the following manner. First, the prover uses the polynomial commitment scheme to send a commitment to the multi-linear extension of the gate values rather than an oracle. This is possible to do in a space-efficient manner because of the streaming nature of RAM computations and the streaming nature of the IOP. Second, the verifier oracle querie are replaced with the prover and verifier engaging in the evaluation protocol of the polynomial commitment scheme. The remainder of the IOP protocol remains unchanged. Thus we obtain Theorem 4. We obtain Theorem 1 by transforming the interactive argument to a zero-knowledge interactive argument using standard techniques, then apply the Fiat-Shamir transformation [FS87].

3 Preliminaries

We let λ denote the security parameter, let $n \in \mathbb{N}$ and $N = 2^n$. For a finite, non-empty set S, we let $x \xleftarrow{\$} S$ denote sampling element x from S uniformly at random. We let $\mathsf{Primes}(1^\lambda)$ denote the set of all λ-bit primes. We let $\mathbb{F}_p$ denote a finite field of prime cardinality p, often use lower-case Greek letters to denote elements of $\mathbb{F}$, e.g., $\alpha \in \mathbb{F}$. For a group $\mathbb{G}$, we denote elements of $\mathbb{G}$ with sans-serif font; e.g., $\mathsf{g} \in \mathbb{G}$. We use boldface lowercase letters to denote binary vectors, e.g. $\mathbf{b} \in \{0,1\}^n$. We assume for a bit string $(b_n, \ldots, b_1) = \mathbf{b} \in \{0,1\}^n$ that b_n is the most significant bit and b_1 is the least significant bit. For bit string $\mathbf{b} \in \{0,1\}^n$ and $b \in \{0,1\}$ we let $b\mathbf{b}$ (resp., $\mathbf{b}b$) denote the string $(b \circ \mathbf{b}) \in \{0,1\}^{n+1}$ (resp.,$(\mathbf{b} \circ b) \in \{0,1\}^{n+1}$), where "$\circ$" is the string concatenation operator. We use boldface lowercase Greek denotes $\mathbb{F}$ vectors, e.g., $\boldsymbol{\alpha} \in \mathbb{F}^n$, and let $\boldsymbol{\alpha} = (\alpha_n, \ldots, \alpha_1)$ for $\alpha_i \in \mathbb{F}$. We let uppercase letters denote sequences and let corresponding lowercase letters to denote its elements, e.g., $Y = (y_{\mathbf{b}} \in \mathbb{F} : \mathbf{b} \in \{0,1\}^n)$ is a sequence of 2^n elements in $\mathbb{F}$. We denote by $\mathbb{F}^N$ the set of all sequences over $\mathbb{F}$ of size N.

3.0.1 Random Oracle.

We let $\mathcal{U}(\lambda)$ denote the set of all functions that map $\{0,1\}^*$ to $\{0,1\}^\lambda$. A random oracle with security parameter λ is a function $H : \{0,1\}^* \to \{0,1\}^\lambda$ sampled uniformly at random from $\mathcal{U}(\lambda)$.

3.1 The Discrete-log Relation Assumption

Let GGen be an algorithm that on input $1^\lambda \in \mathbb{N}$ returns $(\mathbb{G}, p, \mathsf{g})$ such that $\mathbb{G}$ is the description of a finite cyclic group of prime order p, where p has length λ, and g is a generator of $\mathbb{G}$.

Assumption 1 (Discrete-log Assumption)**.** The Discrete-log Assumption holds for GGen if for all PPT adversaries A there exists a negligible function $\mu(\lambda)$ such that

$$\Pr\left[\alpha' = \alpha : (\mathbb{G}, \mathsf{g}, p) \xleftarrow{\$} \mathsf{GGen}(1^\lambda),\ \alpha \xleftarrow{\$} \mathbb{Z}_p, \alpha' \xleftarrow{\$} A(\mathbb{G}, \mathsf{g}, \mathsf{g}^\alpha)\right] \leq \mu(\lambda)\ .$$

For our purposes, we use the following variant of the discrete-log assumption which is equivalent to Assumption 1.

Assumption 2 (Discrete-log Relation Assumption [BCC+16])**.** The Discrete-log Relation Assumption holds for GGen if for all PPT adversaries A and for all $n \geq 2$ there exists a negligible function $\mu(\lambda)$ such that

$$\Pr\left[\exists \alpha_i \neq 0 \wedge \prod_{i=1}^{n} \mathsf{g}_i^{\alpha_i} = 1 : \begin{array}{l} (\mathbb{G}, \mathsf{g}, p) \xleftarrow{\$} \mathsf{GGen}(1^\lambda),\ \mathsf{g}_1, \ldots, \mathsf{g}_n \xleftarrow{\$} \mathbb{G}\ , \\ (\alpha_1, \ldots, \alpha_n) \in \mathbb{Z}_p^n \xleftarrow{\$} A(\mathbb{G}, \mathsf{g}_1, \ldots, \mathsf{g}_n)\ . \end{array}\right] \leq \mu(\lambda)\ .$$

We say $\prod_{i=1}^{n} \mathsf{g}_i^{\alpha_i} = 1$ is a non-trivial discrete log relation between $\mathsf{g}_1, \ldots, \mathsf{g}_n$. The Discrete Log Relation assumption states that an adversary can't find a non-trivial relation between randomly chosen group elements.

3.2 Interactive Arguments of Knowledge in ROM

Definition 1 (Witness Relation Ensemble)**.** A *witness relation ensemble* or relation ensemble is a ternary relation $\mathcal{R}_L$ that is polynomially bounded, polynomial time recognizable and defines a language $\mathcal{L} = \{(pp, x) : \exists w\ s.t.\ (pp, x, w) \in \mathcal{R}_\mathcal{L}\}$. We omit pp when considering languages recognized by binary relations.

Definition 2 (Interactive Arguments [GMR89])**.** Let $\mathcal{R}$ be some relation ensemble. Let (P, V) denote a pair of PPT interactive algorithms and Setup denote a non-interactive setup algorithm that outputs public parameters pp given security parameter 1^λ. Let $\langle P(pp, x, w), V(pp, x)\rangle$ denote the output of V's interaction with P on common inputs public parameter pp and statement x where additionally P has the witness w. The triple (Setup, P, V) is an *argument for* $\mathcal{R}$ in the random oracle model (ROM) if

1. Perfect Completeness. For any adversary A

$$\Pr\left[(x, w) \notin \mathcal{R} \text{ or } \langle P^H(pp, x, w), V^H(pp, x)\rangle = 1\right] = 1\ ,$$

   where probability is taken over $H \xleftarrow{\$} \mathcal{U}(\lambda), pp \xleftarrow{\$} \mathsf{Setup}^H(1^\lambda), (x, w) \xleftarrow{\$} A^H(pp)$.

2. Computational Soundness. For any non-uniform PPT adversary A

$$\Pr\left[\forall w\ (x, w) \notin \mathcal{R} \text{ and } \langle A^H(pp, x, st), V^H(pp, x)\rangle = 1\right] \leq \mathsf{negl}(\lambda)\ ,$$

   where probability is taken over $H \xleftarrow{\$} \mathcal{U}(\lambda), pp \xleftarrow{\$} \mathsf{Setup}^H(1^\lambda), (x, st) \xleftarrow{\$} A^H(pp)$.

Remark 1. Usually completeness is required to hold for all $(x, w) \in \mathcal{R}$. However, for the argument systems used in this work, statements x depends on pp output by Setup and the random oracle H. We model this by asking for completeness to hold for statements sampled by an adversary A, that is, for $(x, w) \xleftarrow{\$} A(pp)$.

For our applications, we will need (Setup, P, V) to be an argument of knowledge. Informally, in an argument of knowledge for $\mathcal{R}$, the prover convinces the verifier that it "knows" a witness w for x such that $(x, w) \in \mathcal{R}$. In this paper, knowledge means that the argument has *witness-extended emulation* [GI08, Lin03].

Definition 3 (Witness-extended Emulation)**.** Given a public-coin interactive argument tuple (Setup, P, V) and some arbitrary prover algorithm P^*, let $\mathsf{Record}(P^*, pp, x, st)$ denote the message transcript between P^* and V on shared input x, initial prover state st, and pp generated by Setup. Furthermore, let $\mathsf{E}^{\mathsf{Record}(P^*,pp,x,st)}$ denote a machine E with a transcript oracle for this interaction that can be rewound to any round and run again on fresh verifier randomness. The tuple (Setup, P, V) has witness-extended emulation if for every deterministic polynomial-time P^* there exists an expected polynomial-time emulator E such that for all non-uniform polynomial-time adversaries A the following holds:

$$\Pr\left[A^H(tr) = 1 : \begin{array}{c} H \xleftarrow{\$} \mathcal{U}(\lambda),\ pp \xleftarrow{\$} \mathsf{Setup}^H(1^\lambda), \\ (x, st) \xleftarrow{\$} A^H(pp),\ tr \xleftarrow{\$} \mathsf{Record}^H(P^*, pp, x, st) \end{array}\right] \approx$$

$$\Pr\left[\begin{array}{c} A^H(tr) = 1 \text{ and} \\ tr \text{ accepting} \implies (x, w) \in \mathcal{R} \end{array} : \begin{array}{c} H \xleftarrow{\$} \mathcal{U}(\lambda),\ pp \xleftarrow{\$} \mathsf{Setup}^H(1^\lambda), \\ (x, st) \xleftarrow{\$} A^H(pp), \\ (tr, w) \xleftarrow{\$} \mathsf{E}^{H,\mathsf{Record}^H(P^*,pp,x,st)}(pp, x) \end{array}\right]$$

It was shown in [BCC+16, BFS20] that witness-extended emulation is implied by an extractor that can extract the witness given a tree of accepting transcripts. For completeness we state this—dubbed Generalized Forking Lemma—more formally below but refer to [BFS20] for the proof.

Definition 4 (Tree of Accepting Transcripts)**.** An $(n_1, \ldots, n_r)$-tree of accepting transcripts for an interactive argument on input x is defined as follows: The root of the tree is labelled with the statement x. The tree has r depth. Each node at depth $i < r$ has n_i children, and each child is labeled with a distinct value for the i-th challenge. An edge from a parent node to a child node is labeled with a message from P to V. Every path from the root to a leaf corresponds to an accepting transcript, hence there are $\prod_{i=1}^{r} n_i$ distinct accepting transcripts overall.

Lemma 1 (Generalized Forking Lemma [BCC+16, BFS20])**.** *Let (Setup, P, V) be an r-round public-coin interactive argument system for a relation $\mathcal{R}$. Let T be a tree-finder algorithm that, given access to a $\mathsf{Record}(\cdot)$ oracle with rewinding capability, runs in polynomial time and outputs an $(n_1, \ldots, n_r)$-tree of accepting transcripts with overwhelming probability. Let Ext be a deterministic polynomial-time extractor algorithm that, given access to T's output, outputs a witness w for the statement x with overwhelming probability over the coins of T. Then, (P, V) has witness-extended emulation.*

Definition 5 (Public-coin)**.** An argument of knowledge is called public-coin if all messages sent from the verifier to the prover are chosen uniformly at random and independently of the prover's messages, i.e., the challenges correspond to the verifier's randomness H.

3.2.1 Zero-knowledge

We also need our argument of knowledge to be zero-knowledge, that is, to not leak partial information about w apart from what can be deduced from $(x, w) \in \mathcal{R}$.

Definition 6 (Zero-knowledge Arguments)**.** Let (Setup, P, V) be an public-coin interactive argument system for witness relation ensemble $\mathcal{R}$. Then, (Setup, P, V) has computational zero-knowledge with respect to an auxiliary input if for every PPT interactive machine V^*, there exists a PPT algorithm

S, called the simulator, running in time polynomial in the length of its first input, such that for every $(x, w) \in \mathcal{R}$ and any $z \in \{0,1\}^*$:

$$View(\langle P(w), V^*(z)\rangle(x)) \approx_c S(x, z) \ ,$$

where $View(\langle P(w), V^*(z)\rangle(x))$ denotes the distribution of the transcript of interaction between P and V^*, and $\approx_c$ denotes that the two quantities are computationally indistinguishable. If the statistical distance between the two distributions is negligible then the interactive argument is said to be statistical zero-knowledge. If the simulator is allowed to abort with probability at most $1/2$, but the distribution of its output conditioned on not aborting is identically distributed to $View(\langle P(w), V^*(z)\rangle(x))$, then the interactive argument is called perfect zero-knowledge.

3.3 Multi-linear Extensions

Definition 7 (Multi-linear Extensions)**.** Let $n \in \mathbb{N}$, $\mathbb{F}$ be some finite field and let $W : \{0,1\}^n \to \mathbb{F}$. Then, the multi-linear extension of W (denoted as $\mathsf{MLE}(W, \cdot) : \mathbb{F}^n \to \mathbb{F}$) is the (unique) multi-linear polynomial that agrees with W on $\{0,1\}^n$. Equivalently,

$$\mathsf{MLE}(W, \boldsymbol{\zeta} \in \mathbb{F}^n) = \sum_{\mathbf{b} \in \{0,1\}^n} W(\mathbf{b}) \cdot \prod_{i=1}^{n} \beta(b_i, \zeta_i) \ ,$$

where $\beta(b, \zeta) = b \cdot \zeta + (1 - b) \cdot (1 - \zeta)$.

For notational convenience, we denote $\prod_{i=1}^{k} \beta(b_i, \zeta_i)$ by $\overline{\beta}(\mathbf{b}, \boldsymbol{\zeta})$.

Remark 2. There is a bijective mapping between the set of all functions from $\{0,1\}^n \to \mathbb{F}$ to the set of all n-variate multi-linear polynomials over $\mathbb{F}$. More specifically, as seen above every function $W : \{0,1\}^n \to \mathbb{F}$ defines a (unique) multi-linear polynomial. Furthermore, every multi-linear polynomial $Q : \mathbb{F}^n \to \mathbb{F}$ is, in fact, the multi-linear extension of the function that maps $\mathbf{b} \in \{0,1\}^n \to Q(\mathbf{b})$.

3.3.1 Streaming access to multi-linear polynomials

For our commitment scheme, we assume that the committer will have *multi-pass streaming* access to the function table of W (which defines the multi-linear polynomial) in the lexicographic ordering. Specifically, the committer will be given access to a read-only tape that is pre-initialized with the sequence $W = (w_{\mathbf{b}} = W(\mathbf{b}) : \mathbf{b} \in \{0,1\}^n)$. At every time-step the committer is allowed to either move the machine head to the right or to restart its position to 0.

With the above notation, we can now view $\mathsf{MLE}(W, \boldsymbol{\zeta} \in \mathbb{F}^n)$ as an inner-product between W and $Z = (z_{\mathbf{b}} = \overline{\beta}(\mathbf{b}, \boldsymbol{\zeta}) : \mathbf{b} \in \{0,1\}^n)$ where computing $z_{\mathbf{b}}$ requires $O(n = \log N)$ field multiplications for fixed $\boldsymbol{\zeta}$ any $\mathbf{b} \in \{0,1\}^n$.

3.4 Polynomial Commitment Scheme to Multi-linear Extensions

Polynomial commitment schemes, introduced by Kate et al. [KZG10] and generalized in [BFS20, Set20, WTs+18], are a cryptographic primitive that allows one to commit to a multivariate polynomial of bounded degree and later provably reveal evaluations of the committed polynomial. Since we consider only multi-linear polynomials, we tailor our definition to them.

Convention. In defining the syntax of various protocols, we use the following convention for any list of arguments or returned tuple $(a, b, c; d, e)$ – variables listed before semicolon are known both to the prover and verifier whereas the ones after are only known to the prover. In this case, a, b, c are public whereas d, e are secret. In the absence of secret information the semicolon is omitted.

Definition 8 (Commitment to Multi-linear Extensions)**.** A polynomial commitment to multi-linear extensions is a tuple of protocols $(\mathsf{Setup}, \mathsf{Com}, \mathsf{Open}, \mathsf{Eval})$:

1. $pp \xleftarrow{\$} \mathsf{Setup}^H(1^\lambda, 1^N)$ takes as input the unary representations of security parameter $\lambda \in \mathbb{N}$ and size parameter $N = 2^n$ corresponding to $n \in \mathbb{N}$, and produces public parameter pp. We allow pp to contain the description of the field $\mathbb{F}$ over which the multi-linear polynomials will be defined.

2. $(\mathsf{C}; d) \xleftarrow{\$} \mathsf{Com}^H(pp, Y)$ takes as input public parameter pp and sequence $Y = (y_{\mathbf{b}} : \mathbf{b} \in \{0,1\}^n) \in \mathbb{F}^N$ that defines the multi-linear polynomial to be committed, and outputs public commitment C and secret decommitment d.

3. $b \leftarrow \mathsf{Open}^H(pp, \mathsf{C}, Y, d)$ takes as input pp, a commitment C, sequence committed Y and a decommitment d and returns a decision bit $b \in \{0, 1\}$.

4. $\mathsf{Eval}^H(pp, \mathsf{C}, \boldsymbol{\zeta}, \gamma; Y, d)$ is a public-coin interactive protocol between a prover P and a verifer V with common inputs—public parameter pp, commitment C, evaluation point $\boldsymbol{\zeta} \in \mathbb{F}^n$ and claimed evaluation $\gamma \in \mathbb{F}$, and prover has secret inputs Y and d. The prover then engages with the verifier in an interactive argument system for the relation

$$\mathcal{R}_{\mathsf{mle}}(pp) = \left\{ (\mathsf{C}, \boldsymbol{\zeta}, \gamma; Y, d) : \mathsf{Open}^H(pp, \mathsf{C}, Y, d) = 1 \wedge \gamma = \mathsf{MLE}(Y, \boldsymbol{\zeta}) \right\}. \tag{6}$$

   The output of V is the output of Eval protocol.

Furthermore, we require the following three properties.

1. Computational Binding. For all PPT adversaries A and $n \in \mathbb{N}$

$$\Pr\left[b_0 = b_1 \neq 0 \wedge Y_0 \neq Y_1 : \begin{array}{c} H \xleftarrow{\$} \mathcal{U}(\lambda),\ pp \xleftarrow{\$} \mathsf{Setup}^H(1^\lambda, 1^N) \\ (\mathsf{C}, Y_0, Y_1, d_0, d_1) \xleftarrow{\$} A^H(pp) \\ b_0 \leftarrow \mathsf{Open}^H(pp, \mathsf{C}, Y_0, d_0) \\ b_1 \leftarrow \mathsf{Open}^H(pp, \mathsf{C}, Y_1, d_1) \end{array} \right] \leq \mathsf{negl}(\lambda) \ .$$

2. Perfect Correctness. For all $n, \lambda \in \mathbb{N}$ and all $Y \in \mathbb{F}^N$ and $\boldsymbol{\zeta} \in \mathbb{F}^n$,

$$\Pr\left[1 = \mathsf{Eval}^H(pp, \mathsf{C}, Z, \gamma; Y, d) : \begin{array}{c} H \xleftarrow{\$} \mathcal{U}(\lambda),\ pp \xleftarrow{\$} \mathsf{Setup}^H(1^\lambda, 1^N), \\ (\mathsf{C}; d) \xleftarrow{\$} \mathsf{Com}^H(pp, Y),\ \gamma = \mathsf{MLE}(Y, \boldsymbol{\zeta}) \end{array} \right] = 1 \ .$$

3. Witness-extended Emulation. We say that the polynomial commitment scheme has witness-extended emulation if Eval has a witness-extended emulation as an interactive argument for the relation ensemble $\{\mathcal{R}_{\mathsf{mle}}(pp)\}_{pp}$ (Equation (6)) except with negligible probability over the choice of H and coins of $pp \xleftarrow{\$} \mathsf{Setup}^H(1^\lambda, 1^N)$.

4 Space-Efficient Commitment for Multi-linear Extensions

In this section we describe our polynomial commitment scheme for multilinear extensions, a high level overview of which was provided in Section 2.1. We dedicate the remainder of the section to proving our main theorem:

Theorem 5. *Let* GGen *be a generator of obliviously sampleable, prime-order groups. Assuming the hardness of discrete logarithm problem for* GGen*, the scheme* $(\mathsf{Setup}, \mathsf{Com}, \mathsf{Open}, \mathsf{Eval})$ *defined in Section 4.1 is a polynomial commitment scheme to multi-linear extensions with witness-extended emulation in the random oracle model. Furthermore, for every* $N \in \mathbb{N}$ *and sequence* $Y \in \mathbb{F}^N$*, the committer/prover has multi-pass streaming access to* Y *and*

1. Com *performs* $O(N \log p)$ *group operations, stores* $O(1)$ *field and group elements, requires one pass over* Y*, makes* N *queries to the random oracle, and outputs a single group element. Evaluating* $\mathsf{MLE}(Y, \cdot)$ *requires* $O(N)$ *field operations, storing* $O(1)$ *field elements and requires one pass over* Y*.*

2. Eval *is public-coin and has* $O(\log N)$ *rounds with* $O(1)$ *group elements sent in every round. Furthermore,*

   - *Prover performs* $O(N \cdot (\log^2 N) \cdot \log p)$ *field and group operations,* $O(N \log N)$ *queries to the random oracle, requires* $O(\log N)$ *passes over* Y *and stores* $O(\log N)$ *field and group elements.*
   - *Verifier performs* $O(N \cdot (\log N) \cdot \log p)$ *field and group operations,* $O(N)$ *queries to the random oracle, and stores* $O(\log N)$ *field and group elements.*

Section 4.1 describes our scheme, Section 4.2 and Section 4.3 establish its security and efficiency.

4.1 Commitment Scheme

We describe a commitment scheme $(\mathsf{Setup}, \mathsf{Com}, \mathsf{Open}, \mathsf{Eval})$ to multi-linear extensions below.

1. $\mathsf{Setup}^H(1^\lambda, 1^N)$: On inputs security parameter 1^λ and size parameter $N = 2^n$ and access to H, Setup samples $(\mathbb{G}, p, \mathsf{g}) \xleftarrow{\$} \mathsf{GGen}(1^\lambda)$, sets $\mathbb{F} = \mathbb{F}_p$ and returns $pp = (\mathbb{G}, \mathbb{F}, N, p)$. Furthermore, it implicitly defines a sequence of generators $\mathbf{g} = (\mathsf{g}_\mathbf{b} = H(\mathbf{b}) : \mathbf{b} \in \{0,1\}^n)$.

2. $\mathsf{Com}^H(pp, Y)$ returns $\mathsf{C} \in \mathbb{G}$ as the commitment and Y as the decommitment where
$$\mathsf{C} \leftarrow \prod_{\mathbf{b} \in \{0,1\}^n} (\mathsf{g}_\mathbf{b})^{y_\mathbf{b}} \ .$$

3. $\mathsf{Open}^H(pp, \mathsf{C}, Y)$ returns 1 iff $\mathsf{C} = \mathsf{Com}^H(pp, Y)$.

4. $\mathsf{Eval}^H(pp, \mathsf{C}, \boldsymbol{\zeta}, \gamma; Y)$ is an interactive protocol $\langle P, V \rangle$ that begins with V sending a random $\mathsf{g} \xleftarrow{\$} \mathbb{G}$. Then, both P and V compute the commitment $\mathsf{C}_\gamma \leftarrow \mathsf{C} \cdot \mathsf{g}^\gamma$ to additionally bind the claimed evaluation γ. Then, P and V engage in an interactive protocol $\mathsf{EvalReduce}$ on input $(\mathsf{C}_\gamma, Z, \mathbf{g}, \mathsf{g}, \gamma; Y)$ where the prover proves knowledge of Y such that
$$\mathsf{C}_\gamma = \mathsf{Com}(\mathbf{g}, Y) \cdot \mathsf{g}^\gamma \wedge \langle Y, Z \rangle = \gamma \ ,$$
where $Z = (z_\mathbf{b} = \bar{\beta}(\mathbf{b}, \boldsymbol{\zeta}) : \mathbf{b} \in \{0,1\}^n)$. We define the protocol in Figure 3.

$\mathsf{Eval}(pp, \mathsf{C}, \boldsymbol{\zeta}, \gamma; Y)$

V samples and sends $\mathsf{g} \xleftarrow{\$} \mathbb{G}$

P and V define $\mathsf{C}_\gamma \leftarrow \mathsf{C} \cdot \mathsf{g}^\gamma$

P and V define the sequence $Z = \left(z_{\mathbf{b}} = \prod_{i=1}^{n} \beta(b_i, \zeta_i) : \mathbf{b} \in \{0,1\}^n\right)$

P and V engage in $\mathsf{EvalReduce}(\mathsf{C}_\gamma, Z, \gamma, \mathbf{g}, \mathsf{g}; Y)$

$\mathsf{EvalReduce}(\mathsf{C}_\gamma \in \mathbb{G}, Z = (z_{\mathbf{b}}), \gamma \in \mathbb{F}, \mathbf{g} = (\mathsf{g}_{\mathbf{b}}), \mathsf{g}; Y = (y_{\mathbf{b}}))$

proves knowledge of Y such that: $\mathsf{C}_\gamma = \mathsf{Com}(\mathbf{g}, Y) \cdot \mathsf{g}^\gamma$ and $\langle Y, Z \rangle = \gamma$.

$N \leftarrow |Z|, n \leftarrow \log N$

if $N = 1$: **then**

Let $\mathbf{g} = (\mathsf{g}'),\ Z = (z),\ Y = (y)$

P sends y to V who accepts iff $\mathsf{C}_\gamma = \mathsf{g}'^{y} \cdot \mathsf{g}^{y \cdot z}$

else

P computes γ_{L} and γ_{R} where

$$\gamma_{\mathsf{L}} \leftarrow \sum_{\mathbf{b} \in \{0,1\}^{n-1}} y_{\mathbf{b}0} \cdot z_{\mathbf{b}1} \ ; \ \gamma_{\mathsf{R}} \leftarrow \sum_{\mathbf{b} \in \{0,1\}^{n-1}} y_{\mathbf{b}1} \cdot z_{\mathbf{b}0}.$$

P computes and sends C_{L} and C_{R} where

$$\mathsf{C}_{\mathsf{L}} \leftarrow \mathsf{g}^{\gamma_{\mathsf{L}}} \cdot \prod_{\mathbf{b} \in \{0,1\}^{n-1}} (\mathsf{g}_{\mathbf{b}1})^{y_{\mathbf{b}0}} \ ; \ \mathsf{C}_{\mathsf{R}} \leftarrow \mathsf{g}^{\gamma_{\mathsf{R}}} \cdot \prod_{\mathbf{b} \in \{0,1\}^{n-1}} (\mathsf{g}_{\mathbf{b}0})^{y_{\mathbf{b}1}}.$$

V samples $\alpha \xleftarrow{\$} \mathbb{F}$ and sends it to P.

P computes and sends $\gamma' = \alpha^2 \cdot \gamma_{\mathsf{L}} + \gamma + \alpha^{-2} \cdot \gamma_{\mathsf{R}}$.

P and V both compute

$$\mathsf{C}'_{\gamma'} \leftarrow (\mathsf{C}_{\mathsf{L}})^{\alpha^2} \cdot \mathsf{C}_\gamma \cdot (\mathsf{C}_{\mathsf{R}})^{\alpha^{-2}},$$

$$Z' = \left(z'_{\mathbf{b}} = \alpha^{-1} \cdot z_{\mathbf{b}0} + \alpha \cdot z_{\mathbf{b}1}\right)_{\mathbf{b} \in \{0,1\}^{n-1}},$$

$$\mathbf{g}' = \left(\mathsf{g}'_{\mathbf{b}} = (\mathsf{g}_{\mathbf{b}0})^{\alpha^{-1}} \cdot (\mathsf{g}_{\mathbf{b}1})^{\alpha}\right)_{\mathbf{b} \in \{0,1\}^{n-1}}.$$

P computes $Y' = \left(y'_{\mathbf{b}} = \alpha \cdot y_{\mathbf{b}0} + \alpha^{-1} \cdot y_{\mathbf{b}1}\right)_{\mathbf{b} \in \{0,1\}^{n-1}}$.

return $\mathsf{EvalReduce}(\mathsf{C}'_{\gamma'}, Z', \gamma', \mathbf{g}', \mathsf{g}; Y')$

Figure 3: Eval protocol for the commitment scheme from Section 4.1.

Remark 3. In fact, our scheme readily extends to proving any linear relation $\boldsymbol{\alpha}$ about a committed sequence Y (i.e., the value $\langle \boldsymbol{\alpha}, Y \rangle$), as long as each element of $\boldsymbol{\alpha}$ can be generated in poly-logarithmic time.

4.2 Correctness and Security

Lemma 2. *The scheme from Section 4.1 is perfectly correct, computationally binding and* Eval *has witness-extended emulation under the hardness of the discrete logarithm problem for groups sampled*

by GGen *in the random oracle model.*

The perfect correctness of the scheme follows from the correctness of EvalReduce protocol, which we prove in Lemma 3, computationally binding follows from that of Pedersen multi-commitments which follows from the hardness of discrete-log (in the random oracle model). The witness-extended emulation of Eval follows from the witness-extended emulation of the inner-product protocol in [BBB+18]. At a high level, we make two changes to their inner-product protocol: (1) sample the generators using the random oracle H, (2) perform the 2-move reduction step using the lsb-based folding approach (see Section 2.1 for a discussion). At a high level, given a witness Y for the inner-product statement $(\mathsf{C}_\gamma, \mathbf{g}, Z, \gamma)$, one can compute a witness for the permuted statement $(\mathsf{C}_\gamma, \pi(\mathbf{g}), \pi(Z), \gamma)$ for any efficiently computable/invertible public permutation π. Choosing π as the permutation that reverses its input allows us, in principle, to base the extractability of our scheme (lsb-based folding) to the original scheme of [BBB+18]. We provide a formal proof in the full version. Due to (1) our scheme enjoys security only in the random-oracle model.

Lemma 3. *Let* $(\mathsf{C}_\gamma, Z, \gamma, \mathbf{g}, \mathsf{g}; Y)$ *be inputs to* EvalReduce *and let* $(\mathsf{C}'_{\gamma'}, Z', \gamma', \mathbf{g}', \mathsf{g}; Y')$ *be generated as in Figure 3. Then,*

$$\begin{array}{c} \mathsf{C}_\gamma = \mathsf{Com}(\mathbf{g}, Y) \cdot \mathsf{g}^\gamma \\ \wedge \\ \langle Y, Z \rangle = \gamma \end{array} \implies \begin{array}{c} \mathsf{C}'_{\gamma'} = \mathsf{Com}(\mathbf{g}', Y') \cdot \mathsf{g}^{\gamma'} \\ \wedge \\ \langle Y', Z' \rangle = \gamma' \end{array} .$$

Proof. Let $N = |Z|$ and let $n = \log N$. Then,

1. To show $\gamma' = \langle Y', Z' \rangle$:

$$\begin{aligned} \langle Y', Z' \rangle &= \sum_{\mathbf{b} \in \{0,1\}^{n-1}} y'_{\mathbf{b}} \cdot z'_{\mathbf{b}}, \\ &= \sum_{\mathbf{b} \in \{0,1\}^{n-1}} (\alpha \cdot y_{\mathbf{b}0} + \alpha^{-1} \cdot y_{\mathbf{b}1}) \cdot (\alpha^{-1} \cdot z_{\mathbf{b}0} + \alpha \cdot z_{\mathbf{b}1}), \\ &= \sum_{\mathbf{b} \in \{0,1\}^{n-1}} y_{\mathbf{b}0} \cdot z_{\mathbf{b}0} + \alpha^2 \cdot y_{\mathbf{b}0} \cdot z_{\mathbf{b}1} + y_{\mathbf{b}1} \cdot z_{\mathbf{b}1} + \alpha^{-2} \cdot y_{\mathbf{b}1} \cdot z_{\mathbf{b}1}, \\ &= \gamma + \alpha^2 \cdot \gamma_{\mathsf{L}} + \alpha^{-2} \cdot \gamma_{\mathsf{R}} = \gamma'. \end{aligned}$$

2. $\mathsf{C}'_{\gamma'} = \mathsf{Com}(\mathbf{g}', Y') \cdot \mathsf{g}^{\gamma'}$:

$$\begin{aligned} \mathsf{Com}(\mathbf{g}', Y') &= \prod_{\mathbf{b} \in \{0,1\}^{n-1}} (\mathsf{g}'_{\mathbf{b}})^{y'_{\mathbf{b}}}, = \prod_{\mathbf{b} \in \{0,1\}^{n-1}} \left(\mathsf{g}_{\mathbf{b}0}^{\alpha^{-1}} \cdot g_{\mathbf{b}1}^{\alpha} \right)^{\alpha \cdot y_{\mathbf{b}0} + \alpha^{-1} \cdot y_{\mathbf{b}1}}, \\ &= \prod_{\mathbf{b} \in \{0,1\}^{n-1}} \left(\mathsf{g}_{\mathbf{b}0}^{y_{\mathbf{b}0}} \cdot g_{\mathbf{b}0}^{\alpha^{-2} \cdot y_{\mathbf{b}1}} \cdot \mathsf{g}_{\mathbf{b}1}^{\alpha^2 \cdot y_{\mathbf{b}0}} \cdot \mathsf{g}_{\mathbf{b}1}^{y_{\mathbf{b}1}} \right), \\ &= \prod_{\mathbf{b} \in \{0,1\}^{n-1}} (\mathsf{g}_{\mathbf{b}0}^{y_{\mathbf{b}1}})^{\alpha^{-2}} \cdot \mathsf{g}_{\mathbf{b}0}^{y_{\mathbf{b}0}} \cdot \mathsf{g}_{\mathbf{b}1}^{y_{\mathbf{b}1}} \cdot (\mathsf{g}_{\mathbf{b}1}^{y_{\mathbf{b}0}})^{\alpha^2}. \end{aligned}$$

   Then, above with the definition of γ' implies that $\mathsf{C}'_{\gamma'} = \mathsf{Com}(\mathbf{g}', Y') \cdot \mathsf{g}^{\gamma'}$.

□

4.3 Efficiency

In this section we discuss the efficiency aspects of each of the protocols defined in Section 4.1 with respect to four complexity measures: (1) queries to the random oracle H, (2) field/group operations performed, (3) field/group elements stored and (4) number of passes over the stream Y.

For the rest of this section, we fix $n, N = 2^n, H, \mathbb{G}, \mathbb{F}, \boldsymbol{\zeta} \in \mathbb{F}^n$ and furthermore fix $Y = (y_{\mathbf{b}} : \mathbf{b} \in \{0,1\}^n)$, $\mathsf{g} = (\mathsf{g}_{\mathbf{b}} = H(\mathbf{b}) : \mathbf{b} \in \{0,1\}^n)$ and $Z = (z_{\mathbf{b}} = \bar{\beta}(\mathbf{b}, \boldsymbol{\zeta}) : \mathbf{b} \in \{0,1\}^n)$. Note given $\boldsymbol{\zeta}$, any $z_{\mathbf{b}}$ can be computed by performing $O(n)$ field operations.

First, consider the prover P of Eval protocol (Figure 3). Given the inputs $(\mathsf{C}, Z, \gamma, \mathbf{g}, \mathsf{g}; Y)$, P and V call the recursive protocol $\mathsf{EvalReduce}$ on the N sized statement $(\mathsf{C}_\gamma, Z, \gamma, \mathbf{g}, \mathsf{g}; Y)$ where $\mathsf{C}_\gamma = \mathsf{C} \cdot \mathsf{g}^\gamma$. The prover's computation in this call to $\mathsf{EvalReduce}$ is dictated by computing (a) $\gamma_\mathsf{L}, \gamma_\mathsf{R}$ (line 6), (2) $\mathsf{C}_\mathsf{L}, \mathsf{C}_\mathsf{R}$ (line 7) and (c) inputs for the next recursive call on $\mathsf{EvalReduce}$ with $N/2$ sized statement $(\mathsf{C}'_{\gamma'}, Z', \gamma', \mathbf{g}', \mathsf{g}; Y')$ (line 9,11). The rest of its computation requires $O(1)$ number of operations. The recursion ends on the n-th call with statement of size 1. For $k \in \{0, \ldots, n\}$, the inputs at the k-th depth of the recursion be denoted with superscript k, that is, $\mathsf{C}^{(k)}, \gamma^{(k)}, Z^{(k)}, \mathbf{g}^{(k)}, Y^{(k)}$. For example, $Z^{(0)} = Z$, $Y^{(0)} = Y$ denote the initial inputs (at depth 0) where prover computes $\gamma^{(0)}_\mathsf{L}, \gamma^{(0)}_\mathsf{R}, \mathsf{C}^{(0)}_\mathsf{L}, \mathsf{C}^{(0)}_\mathsf{R}$ with verifier challenge $\alpha^{(0)}$. The sequences $Z^{(k)}, Y^{(k)}$ and $\mathsf{g}^{(k)}$ are of size 2^{n-k}.

At a high level, we ask prover to never explicitly compute the sequences $\mathsf{g}^{(k)}, Z^{(k)}, Y^{(k)}$ (item (c) above) but instead compute elements $\mathsf{g}^{(k)}_{\mathbf{b}}, z^{(k)}_{\mathbf{b}}, y^{(k)}_{\mathbf{b}}$, of the respective sequences, on demand, which then can be used to compute $\gamma^{(k)}_\mathsf{L}, \gamma^{(k)}_\mathsf{R}, \mathsf{C}^{(k)}_\mathsf{L}, \mathsf{C}^{(k)}_\mathsf{R}$ in required time and space. For this, first it will be useful to see how the elements of sequences $Z^{(k)}, Y^{(k)}, \mathbf{g}^{(k)}$ depend on the initial (i.e., depth-0) sequence $Z^{(0)}, Y^{(0)}, \mathbf{g}^{(0)}$.

Relating $Y^{(k)}$ with $Y^{(0)}$. First, lets consider $Y^{(k)} = (y^{(k)}_{\mathbf{b}} : \mathbf{b} \in \{0,1\}^{n-k})$ at depth $k \in \{0, \ldots, n\}$. Let $(\alpha^{(0)}, \ldots, \alpha^{(k-1)})$ be the verifier's challenges sent in all prior rounds.

Lemma 4 (Streaming of $Y^{(k)}$)**.** *For every* $\mathbf{b} \in \{0,1\}^{n-k}$,

$$y^{(k)}_{\mathbf{b}} = \sum_{\mathbf{c} \in \{0,1\}^k} \left(\prod_{j=1}^{k} \mathsf{coeff}(\alpha^{(j-1)}, c_j) \right) \cdot y_{\mathbf{b} \circ \mathbf{c}} \ , \tag{7}$$

where $\mathsf{coeff}(\alpha, c) = \alpha \cdot (1 - c) + \alpha^{-1} \cdot c$.

The proof follows by induction on depth k. Lemma 4 allows us to simulate the stream $Y^{(k)}$ with **one** pass over the initial sequence Y, additionally performing $O(N \cdot k)$ multiplications to compute appropriate coeff functions.

Relating $Z^{(k)}$ with $Z^{(0)}$. Next, consider $Z^{(k)} = (z^{(k)}_{\mathbf{b}} : \mathbf{b} \in \{0,1\}^{n-k})$ at depth $k \in \{0, \ldots, n\}$.

Lemma 5 (Computing $z^{(k)}_{\mathbf{b}}$)**.** *For every* $\mathbf{b} \in \{0,1\}^{n-k}$,

$$z^{(k)}_{\mathbf{b}} = \sum_{\mathbf{c} \in \{0,1\}^k} \left(\prod_{j=1}^{k} \mathsf{coeff}(\alpha^{(j-1)}, c_j) \right) \cdot z_{\mathbf{b} \circ \mathbf{c}} \ , \tag{8}$$

where $\mathsf{coeff}(\alpha, c) = \alpha \cdot c + \alpha^{-1} \cdot (1 - c)$. *Furthermore, computing* $z^{(k)}_{\mathbf{b}}$ *requires* $O(2^k \cdot n)$ *field multiplications and storing* $O(n)$ *elements (see algorithm* $\mathsf{Computez}$ *in Figure 4).*

$\mathsf{Computez}(k, \mathbf{c}, \boldsymbol{\zeta}, \boldsymbol{\alpha})$	$\mathsf{Computeg}^H(k, \mathbf{c}, \boldsymbol{\alpha})$
$z_{\mathbf{c}}^{(k)} \leftarrow 0$	$\mathsf{g}_{\mathbf{c}}^{(k)} \leftarrow 0$
foreach $\mathbf{a} \in \{0,1\}^k$ **do**	**foreach** $\mathbf{a} \in \{0,1\}^k$ **do**
$\quad \mathsf{temp} \leftarrow 1_{\mathbb{F}}$	$\quad \mathsf{temp} \leftarrow 1_{\mathbb{F}}$
$\quad$ **foreach** $j \in \{1, \ldots, k\}$ **do**	$\quad$ **foreach** $j \in \{1, \ldots, k\}$ **do**
$\quad\quad \mathsf{temp} \leftarrow \mathsf{temp} \cdot \mathsf{coeff}(\alpha^{(j-1)}, a_j)$	$\quad\quad \mathsf{temp} \leftarrow \mathsf{temp} \cdot \mathsf{coeff}(\alpha^{(j-1)}, a_j)$
$\quad z_{\mathbf{c}}^{(k)} \leftarrow \mathsf{temp} \cdot \beta(\mathbf{c} \circ \mathbf{a}, \boldsymbol{\zeta})$	$\quad \mathsf{g}_{\mathbf{c}}^{(k)} \leftarrow H(\mathbf{c} \circ \mathbf{a})^{\mathsf{temp}}$
return $z_{\mathbf{c}}^{(k)}$	**return** $\mathsf{g}_{\mathbf{c}}^{(k)}$

Figure 4: Algorithms for computing $z_{\mathbf{b}}^{(k)}$ and $\mathsf{g}_{\mathbf{b}}^{(k)}$. In both algorithms $\mathbf{c} \in \{0,1\}^{n-k}$ and $\boldsymbol{\alpha} = (\alpha^{(0)}, \ldots, \alpha^{(k-1)})$, where $\beta(\mathbf{b}, \boldsymbol{\zeta}) = \prod_{i=1}^{n} \beta(b_i, \zeta_i)$ for $\mathbf{b} = \mathbf{c} \circ \mathbf{a}$ and $\mathsf{coeff}(\alpha, c) = \alpha \cdot c + \alpha^{-1} \cdot (1 - c)$.

Relating $\mathbf{g}^{(k)}$ with $\mathbf{g}^{(0)}$. Finally, consider $\mathbf{g}^{(k)} = (\mathsf{g}_{\mathbf{b}}^{(k)} : \mathbf{b} \in \{0,1\}^{n-k})$ at depth $k \in \{0, \ldots, n\}$.

Lemma 6 (Computing $\mathsf{g}_{\mathbf{b}}^{(k)}$)**.** *For every* $\mathbf{b} \in \{0,1\}^{n-k}$,

$$\mathsf{g}_{\mathbf{b}}^{(k)} = \prod_{\mathbf{c} \in \{0,1\}^k} \mathsf{g}_{\mathbf{b} \circ \mathbf{c}}^{\mathsf{coeff}(\boldsymbol{\alpha}, \mathbf{c})} \;;\; \mathsf{coeff}(\boldsymbol{\alpha}, \mathbf{c}) = \prod_{i=1}^{k} \alpha^{(j-1)} \cdot c_j + (\alpha^{(j-1)})^{-1} \cdot (1 - c_j) \;. \tag{9}$$

Furthermore, computing $\mathsf{g}_{\mathbf{b}}^{(k)}$ *requires* $2^k \cdot k$ *field multiplications,* 2^k *queries to* H, 2^k *group multiplications and exponentiations, and storing* $O(k)$ *elements (see algorithm* $\mathsf{Computeg}$ *in Figure 4).*

We now discuss the efficiency of the commitment scheme.

4.3.1 Commitment Phase

We first note that Com^H on input pp and given streaming access to Y can compute the commitment $\mathsf{C} = \prod_{\mathbf{b}} (H(\mathbf{b}))^{y_{\mathbf{b}}}$ for $\mathbf{b} \in \{0,1\}^n$ making N queries to H, performing N group exponentiations and a single pass over Y. Furthermore, requires storing only a single group element.

Note that a single group exponentiation g^{α} can be emulated while performing $O(\log p)$ group multiplications while storing $O(1)$ group and field elements. Since, $\mathbb{G}, \mathbb{F}$ are of order p, field and group operations can, furthermore, be performed in $\mathrm{polylog}(p(\lambda))$ time.

4.3.2 Evaluating $\mathsf{MLE}(Y, \boldsymbol{\zeta})$

The honest prover (when used in higher level protocols) needs to evaluate $\mathsf{MLE}(Y, \boldsymbol{\zeta})$ which requires performing $O(N \log N)$ field operations overall and a single pass over stream Y.

4.3.3 Prover Efficiency

For every depth-k of the recursion, it is sufficient to discuss the efficiency of computing $\gamma_{\mathsf{L}}^{(k)}, \gamma_{\mathsf{R}}^{(k)}, \mathsf{C}_{\mathsf{L}}^{(k)}$, and $\mathsf{C}_{\mathsf{R}}^{(k)}$. We argue the complexity of computing $\gamma_{\mathsf{L}}^{(k)}$ and $\mathsf{C}_{\mathsf{L}}^{(k)}$ and the analysis for the remaining is similar. We give a formal algorithm Prover in Figure 5.

$\mathsf{Prover}^H(pp, k, Y, \boldsymbol{\zeta}, \mathsf{g}, \alpha^{(0)}, \ldots, \alpha^{(k-1)})$

1 : $\gamma_{\mathsf{L}}, \gamma_{\mathsf{R}}, y^{(k)} \leftarrow 0_{\mathbb{F}}, \mathsf{g}^{(k)}, \mathsf{C}_{\mathsf{L}}, \mathsf{C}_{\mathsf{R}} \leftarrow 1_{\mathbb{G}}, \mathit{count} \leftarrow 0$
2 : **foreach** $\mathbf{b} = (b_n, \ldots, b_1) \in \{0,1\}^n$ **do**
3 : $\quad \mathsf{temp} \leftarrow 1_{\mathbb{F}}$
4 : $\quad$ **foreach** $j \in \{1, \ldots, k\}$ **do**
5 : $\quad\quad \mathsf{temp} \leftarrow \mathsf{temp} \cdot \mathsf{coeff}(\alpha^{(j-1)}, b_j)$
6 : $\quad y^{(k)} \leftarrow y^{(k)} + \mathsf{temp} \cdot y_{\mathbf{b}}$
7 : $\quad \mathit{count} \leftarrow \mathit{count} + 1$
8 : $\quad$ **if** $\mathit{count} == 2^k$ **then**
9 : $\quad\quad z^{(k)} \leftarrow \mathsf{Computez}(k, (b_n, \ldots, b_{n-k+1}, 1 - b_{n-k}), \boldsymbol{\zeta}, \alpha^{(0)}, \ldots, \alpha^{(k-1)})$
10 : $\quad\quad \mathsf{g}^{(k)} \leftarrow \mathsf{Computeg}^H(k, (b_n, \ldots, b_{n-k+1}, 1 - b_{n-k}), \alpha^{(0)}, \ldots, \alpha^{(k-1)})$
11 : $\quad\quad$ **if** $b_{n-k} == 0$ **then**
12 : $\quad\quad\quad \gamma_{\mathsf{L}} \leftarrow \gamma_{\mathsf{L}} + z^{(k)} \cdot y^{(k)}$; $\mathsf{C}_{\mathsf{L}} \leftarrow \mathsf{C}_{\mathsf{L}} \cdot (\mathsf{g}^{(k)})^{y^{(k)}}$
13 : $\quad\quad$ **else**
14 : $\quad\quad\quad \gamma_{\mathsf{R}} \leftarrow \gamma_{\mathsf{R}} + z^{(k)} \cdot y^{(k)}$; $\mathsf{C}_{\mathsf{R}} \leftarrow \mathsf{C}_{\mathsf{R}} \cdot (\mathsf{g}^{(k)})^{y^{(k)}}$
15 : $\quad\quad y^{(k)} \leftarrow 0_{\mathbb{F}}$; $\mathsf{g}^{(k)} \leftarrow 1_{\mathbb{G}}$; $\mathit{count} \leftarrow 0$
16 : $\mathsf{C}_{\mathsf{L}} \leftarrow \mathsf{C}_{\mathsf{L}} \cdot \mathsf{g}^{\gamma_{\mathsf{L}}}$; $\mathsf{C}_{\mathsf{R}} \leftarrow \mathsf{C}_{\mathsf{R}} \cdot \mathsf{g}^{\gamma_{\mathsf{R}}}$
17 : **return** $(\gamma_{\mathsf{L}}, \mathsf{C}_{\mathsf{L}}, \gamma_{\mathsf{R}}, \mathsf{C}_{\mathsf{R}})$

Figure 5: Space-efficient Prover

Computing $\gamma_{\mathsf{L}}^{(k)}$. Recall that $\gamma_{\mathsf{L}}^{(k)} = \sum_{\mathbf{b}} y_{\mathbf{b}0}^{(k)} \cdot z_{\mathbf{b}1}^{(k)}$ for $\mathbf{b} \in \{0,1\}^{n-k-1}$. To compute $\gamma_{\mathsf{L}}^{(k)}$ we stream the initial N-sized sequence Y and generate elements of the sequence $(y_{\mathbf{b}0}^{(k)} : \mathbf{b} \in \{0,1\}^{n-k-1})$ in a streaming manner. Since each $y_{\mathbf{b}0}^{(k)}$ depends on a contiguous block of 2^k elements in the initial stream Y, we can compute $y_{\mathbf{b}0}^{(k)}$ by performing $2^k \cdot k$ field operations (lines 2-7 in Figure 5). For every $\mathbf{b} \in \{0,1\}^{n-k-1}$, after computing $y_{\mathbf{b}0}^{(k)}$, we leverage "random access" to Z and compute $z_{\mathbf{b}1}^{(k)}$ (Lemma 5) which requires $O(2^k \cdot k)$ field operations. Overall, $\gamma_{\mathsf{L}}^{(k)}$ can be computed in $O(N \cdot k)$ field operations and a single pass over Y.

Computing $\mathsf{C}_{\mathsf{L}}^{(k)}$. The two differences in computing $\mathsf{C}_{\mathsf{L}}^{(k)}$ (see Figure 3 for the definition) is that (a) we need to compute $\mathsf{g}_{\mathbf{b}1}^{(k)}$ instead of computing $z_{\mathbf{b}1}^{(k)}$ and (b) perform group exponentiations, that is, $\mathsf{g}_{\mathbf{b}1}^{(k)^{y_{\mathbf{b}0}^{(k)}}}$ as opposed to group multiplications as in the computation of $\gamma_{\mathsf{L}}^{(k)}$. Both steps overall can be implemented in $O(N \cdot k \cdot \log p)$ field and group operations and N queries to H (Lemma 6). Overall, at depth k the prover (1) makes $O(N)$ queries to H, (2) performs $O(N \cdot k \cdot \log(p))$ field and group operations and (3) requires a single pass over Y.

Therefore, the entire prover computation (over all calls to $\mathsf{EvalReduce}$) requires $O(\log N)$ passes over Y, makes $O(N \log N)$ queries to H and performs $O(N \cdot \log^2 N \cdot \log p)$ field/group operations. Furthermore, this requires storing only $O(\log N)$ field and group elements.

4.3.4 Verifier Efficiency

V only needs to compute folded sequence $Z^{(n)}$ and folded generators $\mathsf{g}^{(n)}$ at depth-n of the recursion. These can computed by invoking Computez and Computeg (Figure 4) with $k = n$ and require $O(N \cdot \log(N, p))$ field and group operations, $O(N)$ queries to H and storing $O(\log N)$ field and group elements.

Lemma 7. *The time and space efficiency of each of the phases of the protocols are listed below:*

Computation	*H queries*	*Y passes*	$\mathbb{F}/\mathbb{G}$ *ops*[6]	$\mathbb{G}/\mathbb{F}$ *elements*
Com	N	*1*	$O(N)$	*O(1)*
$\mathsf{MLE}(Y, \zeta)$	*0*	*1*	$O(N \log N)$	*O(1)*
P *(in* Eval*)*	$O(N \log N)$	$O(\log N)$	$O(N \log^2 N)$	$O(\log N)$
V *(in* Eval*)*	$O(N)$	0	$O(N \log N)$	$O(\log N)$

Finally, Theorem 5 follows directly from Lemma 2 and Lemma 7.

5 A Polynomial IOP for Random Access Machines

We obtain space efficient arguments for any NP relation verifiable by time-T space-S RAM computations by compiling our polynomial commitment scheme with a suitable space-efficient *polynomial interactive oracle proof* (IOP) [BCS16, BFS20, RRR16]. Informally, a polynomial IOP is a multi-round interactive PCP such that in each round the verifier sends a message to the prover and the prover responds with a proof oracle that the verifier can query via random access, with the additional property that the proof oracle is a polynomial.

We give a detailed overview the construction of our polynomial IOP and prove Theorem 3. We first recall Theorem 3.

Theorem 3. *There exists a public-coin polynomial* IOP *over a channel which encodes prover messages as multi-linear extensions for* NP *relations verifiable by a time-T space-S random access machine M such that if $y = M(x; w)$ then*

1. *The* IOP *has perfect completeness and statistical soundness, and has $O(\log(T))$ rounds;*
2. *The prover runs in time $T \cdot \mathrm{polylog}(T)$ and space $S \cdot \mathrm{polylog}(T)$ (not including the space required for the oracle) when given input-witness pair $(x; w)$ for M, sends a single polynomial oracle in the first round, and has $\mathrm{polylog}(T)$ communication in all subsequent rounds; and*
3. *The verifier runs in time $(|x| + |y|) \cdot \mathrm{polylog}(T)$, space $\mathrm{polylog}(T)$, and has query complexity 3.*

To begin, formally define Random Access Machines.

Definition 9 (Random Access Machine)**.** A *(non-deterministic) Random Access Machine (RAM)* is a tuple $M = \langle k, r, \mathbb{A}, \mathbb{L} \rangle$ where $\ell \in \mathbb{N}$ is the register size, $r \in \mathbb{N}$ is the number of registers, $\mathbb{A} \subseteq \{f \colon \{0,1\}^\ell \times \{0,1\}^\ell \to \{0,1\}^\ell\}$ is the arithmetic unit, and $\mathbb{L} = (I_0, \ldots, I_m)$, where $m \in [2^\ell]$ and each I_j is an instruction, is the code. For any input $x \in \{0,1\}^n$ and witness $w \in \{0,1\}^*$, a RAM

[6] $\log(p)$ factors are omitted.

M runs in time $T(n)$ and space $S(n)$ if $M(x;w)$ halts after executing at most $T(n)$ instructions and uses at most $S(n)$ space.

We let $\mathcal{R}_{\mathsf{RAM}}$ denote the set of all tuples $(M,x,y,T,S;w)$ such that M is a RAM and $M(x;w)$ outputs y in time T and space S. We define the language $\mathcal{L}_{\mathsf{RAM}}$ as

$$\mathcal{L}_{\mathsf{RAM}} = \{(M,x,y,T,S) \mid \exists w\colon (M,x,y,T,S;w) \in \mathcal{R}_{\mathsf{RAM}}\}.$$

Our construction is in fact a polynomial IOP for the NP relation $\mathcal{R}_{\mathsf{RAM}}$. At a high-level, our IOP construction reduces checking membership of a $\mathcal{R}_{\mathsf{RAM}}$ instance to performing the classical sum-check protocol [LFKN90, Sha90] of some appropriately constructed polynomial. We proceed in two parts. First we reduce an $\mathcal{R}_{\mathsf{RAM}}$ instance into a circuit satisfiability instance for some appropriate circuit. We then reduce the circuit satisfiability instance to a polynomial statement compatible with the sum-check protocol. Our construction is inspired from previous approaches [BCGT13, BTVW14, CMT12, GKR08, Set20, Tha13, WTs+18] re-imagined in the language of IOPs.

5.1 RAMs to Circuits.

Let $(M,x,y,T,S;w) \in \mathcal{R}_{\mathsf{RAM}}$. The first step in the IOP is for the prover to compile the RAM M into an appropriate (non-deterministic) arithmetic circuit over some appropriate finite field $\mathbb{F}$.

For finite field $\mathbb{F}$ and arithemtic circuit $C\colon \mathbb{F}^n \to \mathbb{F}^k$, we let $|C|$ denote the size of the circuit.. We assume a canonical ordering on the gates of C (known by both the prover and verifier), and label every gate in C with unique $a \in \{0,1\}^s$ for $s = \lceil \log |C| \rceil$. Without loss of generality we assume the first n input gates of C correspond to the s-bit representation of the integers $\{1,\ldots,n\}$.

Definition 10 (Circuit Transcript)**.** A *transcript* for arithmetic circuit C with input x and output y is an assignment $W\colon \{0,1\}^s \to \mathbb{F}$ of values to the circuit gates, where $s = \lceil \log |C| \rceil$. We say that a transcript W is *correct* for (C,x,y) if $W(a) = x_a$ for all input gates a, $W(a) = y_i$ if a is the i-th output gate, and for every $a,b,c \in \{0,1\}^s$ such that b and c are parents of a, $W(a) = W(b) + W(c)$ if a is an add gate and $W(a) = W(b) \cdot W(c)$ if a is a multiplication gate. Given a tuple (C,x,y), the problem of determining whether there exists a correct transcript W for (C,x,y) is referred to as the *non-deterministic circuit evaluation problem.*

We use the RAM to circuit transformation of Blumberg et al. [BTVW14].

Lemma 8 ((Non-Deterministic) RAM to Circuit [BTVW14, Lemma 4.2])**.** *For $(M,x,y,T,S;w) \in \mathcal{R}_{\mathsf{RAM}}$, M can be transformed into an equivalent (non-deterministic) arithmetic circuit C_M over a finite field $\mathbb{F}$ of size* $\mathrm{polylog}(T)$ *with the following properties:*

1. *C_M has size $T \cdot \mathrm{polylog}(T)$.*
2. *An (input, witness) pair $(x;w)$ such that $(M,x,y,T,S;w) \in \mathcal{R}_{\mathsf{RAM}}$ can be mapped to a correct transcript W for C_M in time $T \cdot \mathrm{polylog}(T)$ and space $S \cdot \mathrm{polylog}(T)$ such that $|W| = 2^s$ for some $s = O(\log |C_M|)$. Furthermore, w is a substring of the transcript W, and any correct W' for C_M possesses a witness w' certifying $(M,x,y,T,S) \in \mathcal{L}_{\mathsf{RAM}}$ as a substring.*
3. *C_M can be evaluated "gate-by-gate" in time $T \cdot \mathrm{polylog}(T)$ and space $S \cdot \mathrm{polylog}(T)$.*

5.2 Circuits to Polynomials

Consider $(M, x, y, T, S; w) \in \mathcal{R}_{\mathsf{RAM}}$ and let C_M be the arithmetic circuit given by Lemma 8, and let $s = \lceil \log |C_M| \rceil$. Let W be a correct transcript for (C_M, x, y). We reduce showing W is a correct transcript for (C_M, x, y) to showing that a polynomial we construct over $\mathbb{F}$ is the 0-polynomial. In particular, the constructed polynomial is the 0-polynomial if and only if W is a correct transcript. Let $\overline{W}$ be the multi-linear extension of a transcript W.

We associate three *wiring predicates* $\mathsf{add}, \mathsf{mult}, \mathsf{io} \colon \{0,1\}^{3s} \to \{0,1\}$ with C_M such that the following hold. For all $a, b, c \in \{0,1\}^s$, $\mathsf{add}(a, b, c) = 1$ if and only if a is an addition gate with parent gates b and c (with $\mathsf{mult}(a, b, c)$ being defined analogously), and $\mathsf{io}(a, b, c) = 1$ if and only b and c are parents of a and a is an output gate or a (non-auxiliary) input gate. We also define $I_{x,y} \colon \{0,1\}^s \to \mathbb{F}$ for $a \in \{0,1\}^s$ as $I_{x,y}(a) = x_a$ if a is an input gate, $I_{x,y}(a) = y_i$ if a is the i-th output gate, and $I_{x,y}(a) = 0$ otherwise.

Let $\overline{I}_{x,y}, \overline{\mathsf{io}}$ be the multi-linear extensions of $I_{x,y}, \mathsf{io}$, and let $\overline{\mathsf{add}}, \overline{\mathsf{mult}}$ be some degree-3 extensions of $\mathsf{add}, \mathsf{mult}$. Define polynomials $\overline{G}_{x,y} \colon \mathbb{F}^{3s} \to \mathbb{F}$ and $F_{x,y} \colon \mathbb{F}^{3s} \to \mathbb{F}$ as

$$\begin{aligned} \overline{G}_{x,y}(\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}, \boldsymbol{\zeta}^{(3)}) &= \overline{\mathsf{io}}(\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}, \boldsymbol{\zeta}^{(3)}) \cdot (\overline{I}_{x,y}(\boldsymbol{\zeta}^{(1)}) - \overline{W}(\boldsymbol{\zeta}^{(1)})) \\ &+ \overline{\mathsf{add}}(\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}, \boldsymbol{\zeta}^{(3)}) \cdot (\overline{W}(\boldsymbol{\zeta}^{(1)}) - \overline{W}(\boldsymbol{\zeta}^{(2)}) - \overline{W}(\boldsymbol{\zeta}^{(3)})) \\ &+ \overline{\mathsf{mult}}(\boldsymbol{\zeta}^{(1)}, \boldsymbol{\zeta}^{(2)}, \boldsymbol{\zeta}^{(3)}) \cdot (\overline{W}(\boldsymbol{\zeta}^{(1)}) - \overline{W}(\boldsymbol{\zeta}^{(2)}) \cdot \overline{W}(\boldsymbol{\zeta}^{(3)})) \,, \end{aligned} \tag{10}$$

and

$$F_{x,y}(\mathbf{X}) = \sum_{\mathbf{c} \in \{0,1\}^{3s}} \overline{G}_{x,y}(\mathbf{c}) \cdot \prod_{i=1}^{3s} \beta(c_i, X_i) = \sum_{\mathbf{c} \in \{0,1\}^{3s}} \overline{G}_{x,y}(\mathbf{c}) \cdot g(\mathbf{c}, \mathbf{X}) \,. \tag{11}$$

Lemma 9. *The polynomial $F_{x,y}$ is the 0-polynomial if and only if $\overline{G}_{x,y}(\mathbf{c}) = 0$ for all $\mathbf{c} \in \{0,1\}^{3s}$ if and only if $\overline{W}$ is a multi-linear extension of a correct transcript W of C_M. Further, there exist degree-3 extensions of $\overline{\mathsf{add}}$ and $\overline{\mathsf{mult}}$ such that $\overline{\mathsf{add}}$ and $\overline{\mathsf{mult}}$ can be evaluated at any point in $O(\mathrm{polylog}(T))$ time without explicit access to C_M.*

Proof. We first note that the definition of $\overline{G}_{x,y}$ immediately gives us that $\overline{G}_{x,y}|_{\{0,1\}^{3s}} \equiv 0$ if and only if $\overline{W}$ is a multi-linear extension of a correct transcript W of C_M [BTVW14, GKR08, Set20, WTs+18]. Next, we note that the polynomial $F_{x,y}$ is defined as the multi-linear extension of the sequence $(\overline{G}_{x,y}(\mathbf{c}) \colon \mathbf{c} \in \{0,1\}^{3s})$. In particular, $F_{x,y}(\mathbf{c}) = \overline{G}_{x,y}(\mathbf{c})$ for all $\mathbf{c} \in \{0,1\}^{3s}$. This directly implies that $F_{x,y}$ is the 0-polynomial if and only if $\overline{G}_{x,y}(\mathbf{c}) = 0$ for all $\mathbf{c} \in \{0,1\}^{3s}$. To finish the proof, we note that existence of degree-3 $\overline{\mathsf{add}}$ and $\overline{\mathsf{mult}}$ with the desired properties follows directly from [BTVW14, Theorem 4.1 and Lemma 4.2]. □

Finally, for any $\boldsymbol{\tau} \in \mathbb{F}^{3s}$, we define the polynomial $h_{\boldsymbol{\tau}}(\boldsymbol{\zeta}) := \overline{G}_{x,y}(\boldsymbol{\zeta}) \cdot g(\boldsymbol{\zeta}, \boldsymbol{\tau})$. By Lemma 9 the satisfiability of the circuit C_M is reduced to checking if $F_{x,y}$ is the 0-polynomial. In particular by Schwartz-Zippel a verifier is convinced that $F_{x,y}$ is the 0-polynomial if $F_{x,y}(\boldsymbol{\tau}) = 0$ for $\boldsymbol{\tau} \xleftarrow{\$} \mathbb{F}^{3s}$. However, a verifier would perform $O(|C_M|^3) = T^3 \cdot \mathrm{polylog}(T)$ operations to compute $F_{x,y}(\boldsymbol{\tau})$, which gives a non-succinct verifier. Instead, checking $F_{x,y}(\boldsymbol{\tau}) = 0$ is offloaded to a prover via a sum-check protocol for the statement

$$0 = \sum_{\mathbf{c} \in \{0,1\}^{3s}} h_{\boldsymbol{\tau}}(\mathbf{c}) = F_{x,y}(\boldsymbol{\tau}).$$

$\mathsf{PIOP}(M, x, T, S; w)$

P compiles circuit C_M and transcript W via the reduction of [BTVW14].

P provides V with an oracle for $\overline{W}$.

V samples $\boldsymbol{\tau} \xleftarrow{\$} \mathbb{F}^{3s}$ and sends $\boldsymbol{\tau}$ to P.

P computes polynomial $h_{\boldsymbol{\tau}}$ and sets $\gamma \leftarrow 0$. P sends γ to V.

V sets $\gamma' \leftarrow \gamma$.

foreach $j \in \{1, \ldots, 3s\}$ **do** // sum-check

  P sends $h^{(j)}_{\boldsymbol{\tau}}(X_j)$ to V, where $h^{(j)}_{\boldsymbol{\tau}}(X_j) \leftarrow \sum_{\mathbf{c}' \in \{0,1\}^{3s-j}} h_{\boldsymbol{\tau}}(\alpha_1, \ldots, \alpha_{j-1}, X_j, \mathbf{c}')$.

  V checks $\gamma' \stackrel{?}{=} h^{(j)}_{\boldsymbol{\tau}}(0) + h^{(j)}_{\boldsymbol{\tau}}(1)$, rejecting if equality doesn't hold.

  V samples $\alpha_j \xleftarrow{\$} \mathbb{F}$ and sets $\gamma' \leftarrow h^{(j)}_{\boldsymbol{\tau}}(\alpha_j)$.

  if $j < 3s$ **then** V sends α_j to P **endif**

V queries the oracle $\overline{W}$ and obtains $\gamma_i \leftarrow \overline{W}(\boldsymbol{\alpha}^{(i)})$ for $i \in \{1, 2, 3\}$, where $\boldsymbol{\alpha}^{(i)} \leftarrow (\alpha_{i\cdot 1}, \ldots, \alpha_{i\cdot s})$.

V computes $h_{\boldsymbol{\tau}}(\boldsymbol{\alpha})$ using oracle queries γ_i and accepts if and only if $\gamma' = h_{\boldsymbol{\tau}}(\boldsymbol{\alpha})$.

Figure 6: Our Polynomial IOP for time-T space-S RAM computations.

5.3 Polynomial IOP Construction

We present our polynomial IOP construction in Figure 6, which we call PIOP. The protocol PIOP takes as input $(M, x, y, T, S; w)$ and the prover compiles it into circuit instance (C_M, x, y) via the reduction guaranteed by Lemma 8. The prover next sends an oracle to the multi-linear extension of the transcript of (C_M, x, y). The prover compiles C_M into a suitable polynomial $F_{x,y}$ given by Lemma 9, receives $\boldsymbol{\tau} \xleftarrow{\$} \mathbb{F}^{3s}$ from the verifier, and engages in a sum-check with the verifier for the statement $\sum_{\mathbf{c}} h_{\boldsymbol{\tau}}(\mathbf{c}) = F_{x,y}(\boldsymbol{\tau}) = 0$. Finally the verifier uses the challenges $\boldsymbol{\alpha} \in \mathbb{F}^{3s}$ given by the sum-check to query the oracle at 3 points and evaluates polynomial $h_{\boldsymbol{\tau}}(\boldsymbol{\alpha})$ locally and compares it to the value output by the sum-check.

We now argue that PIOP satisfies Theorem 3. Perfect completeness follows directly from the protocol description, and the round complexity follows since $s = O(\log |C_M|) = \text{polylog}(T)$. To finish proving the theorem, we show the efficiency and soundness of the protocol.

5.3.1 Verifier Efficiency

The verifier samples $O(\log(T))$ random field elements $\boldsymbol{\tau}$ and α_j for $j \in \{1, \ldots, 3s\}$. During each round of the sum-check the verifier evaluates the polynomial $h^{(j)}_{\boldsymbol{\tau}}(X_j)$ at 3 points $\{0, 1, \alpha_j\} \subset \mathbb{F}$. Since $h_{\boldsymbol{\tau}}$ has individual degree at most 6, each univariate $h^{(j)}_{\boldsymbol{\tau}}$ has degree at most 6, giving $O(1)$ multiplications to evaluate $h^{(j)}_{\boldsymbol{\tau}}$. This gives that the verifier has complexity $\text{polylog}(T)$ multiplications and $\text{polylog}(T)$ communication during the sum-check phase. Next the verifier queries the oracle $\overline{W}$ at 3 points.As a last step, the verifier uses the 3 oracle queries to compute $h_{\boldsymbol{\tau}}(\boldsymbol{\alpha})$, which by Lemma 9 takes time $(|x| + |y|) \cdot \text{polylog}(T)$ and is computable without explicit access to C_M.

During any round of the sum-check, the verifier stores $O(1)$ field elements for the description of

$h_{\boldsymbol{\tau}}^{(j)}$ and stores all challenges $\boldsymbol{\tau} \in \mathbb{F}^{3s}$ and $\boldsymbol{\alpha} \in \mathbb{F}^{3s}$, which is polylog$(T)$ field elements of storage. Computing extensions $\overline{\mathsf{add}}$ and $\overline{\mathsf{mult}}$ can be done in polylog(T) time, so they can be computed using at most polylog(T) space. This gives a verifier with space complexity of polylog(T).

5.3.2 Prover Efficiency

We examine the complexity of the prover in PIOP after sending the oracle $\overline{W}$, an oracle to the multi-linear extension of a correct transcript W for $(C_M, x, y, T, S; w)$. The prover receives verifier challenge $\boldsymbol{\tau}$ and now must run the sum-check with polynomial $h_{\boldsymbol{\tau}}(\boldsymbol{\zeta})$ of Lemma 9. We leverage the following lemma to allow the prover to perform this computation efficiently in each round.

Lemma 10 ([BTVW14, Theorem 4.1, Lemma 4.2])**.** *Let $(M, x, y, T, S; w) \in \mathcal{R}_{\mathsf{RAM}}$ and let C_M be the equivalent arithmetic circuit given by Lemma 8. Then given (M, x, y, T, S, w), one can run in time $T \cdot \text{polylog}(T)$ and space $S \cdot \text{polylog}(T)$ to compute sum-check messages for the polynomial $h_{\boldsymbol{\tau}}(\mathbf{Y})$.*

In particular, Lemma 10 implies that the prover's computation of the polynomial $h_{\boldsymbol{\tau}}^{(j)}(X_j)$ in each round of the sum-check can be done in $T \cdot \text{polylog}(T)$ time and $S \cdot \text{polylog}(T)$ space. Note also that each polynomial $h_{\boldsymbol{\tau}}^{(j)}(X_j)$ is a degree at most 6 polynomial, and therefore uses $O(1)$ space. Finally, the prover also stores verifier challenges α_j for each $j \in \{1, \ldots, 3s-1\}$, which requires $O(\text{polylog}(T))$ space. Since $s = O(\log |C_M|)$ and $|C_M| = T \cdot \text{polylog}(T)$, we have that the total prover time is $T \cdot \text{polylog}(T)$ and total space is $S \cdot \text{polylog}(T)$.

5.3.3 Soundness

The soundness of PIOP follows from the soundness of the sum-check protocol.

Lemma 11 (Sum-check Soundness [LFKN90, Sha90])**.** *For $\gamma \in \mathbb{F}$, $v, d \in \mathbb{N}$, let $\mathcal{L}_{\gamma,v,d}$ be the language of all v-variate polynomials f of individual degree at most d such that $\gamma = \sum_{\mathbf{c} \in \{0,1\}^v} f(\mathbf{c})$. Then the sum-check protocol is an interactive proof system for $\mathcal{L}_{\gamma,v,d}$ with perfect completeness and soundness error $\varepsilon_{\mathsf{sc}} \leq dv/|\mathbb{F}|$, where the verifier is given oracle access to the function f.*

We now show the soundness of PIOP.

Proposition 1. *Let V be the verifier of PIOP. For every $\mathbb{x} = (M, x, y, T, S) \notin \mathcal{L}_{\mathsf{RAM}}$ and every P^*, over the randomness of V we have that $\langle P^*, V(\mathbb{x}) \rangle = 1$ with probability at most $\frac{36s}{|\mathbb{F}|}$.*

Proof. Let P^* be an arbitrary prover. Let $(M, x, y, T, S) \notin \mathcal{L}_{\mathsf{RAM}}$ be the input to both verifier V and P^*. By assumption, there does not exist a correct transcript W for the circuit C_M. Let W^* be the polynomial oracle sent by P^*.

By Lemma 9, the polynomial $F_{x,y}$ is the 0-polynomial if and only if W^* is a multi-linear extension of a correct transcript, so by assumption $F_{x,y}$ is not the 0-polynomial and has individual degree at most 6. For a verifier to be convinced that W^* is a correct transcript, it suffices for the verifier to sample $\boldsymbol{\tau} \xleftarrow{\$} \mathbb{F}^{3s}$ and check $F_{x,y}(\boldsymbol{\tau}) \stackrel{?}{=} 0$. Since $F_{x,y}$ is a polynomial of individual degree at most 6, by Schwartz-Zippel we have

$$\Pr_{\boldsymbol{\tau} \xleftarrow{\$} \mathbb{F}^{3s}} [F_{x,y}(\boldsymbol{\tau}) = 0 | F_{x,y} \not\equiv 0] \leq \frac{6 \cdot 3s}{|\mathbb{F}|}.$$

So after receiving oracle W^* the verifier samples and sends $\boldsymbol{\tau} \xleftarrow{\$} \mathbb{F}^{3s}$ to P^*.

Now P^* and V engage in a sum-check protocol for the statement

$$0 = \sum_{\mathbf{c} \in \{0,1\}^{3s}} h_{\boldsymbol{\tau}}(\mathbf{c}) = F_{x,y}(\boldsymbol{\tau}). \tag{12}$$

Conditioning on $F_{x,y}(\boldsymbol{\tau}) \neq 0$, the soundness now reduces to the soundness of the sum-check. In particular, P^* must convinced V that Equation (12) holds, but $F_{x,y}(\boldsymbol{\tau}) \neq 0$. In this case the probability P^* succeeds is at most $(3s \cdot 6)/|\mathbb{F}|$ by Lemma 11. Since $F_{x,y}$ is not the 0-polynomial we have that

$$\Pr[\langle P^*, V(\mathbb{x}) \rangle = 1] \leq \Pr_{\boldsymbol{\tau}}[F_{x,y}(\boldsymbol{\tau}) = 0] + \Pr[P^* \text{ breaks sum-check}] \leq \frac{36s}{|\mathbb{F}|}.$$

□

6 Time- and Space-Efficient Arguments for RAM

We obtain space-efficient arguments $\langle P_{\mathsf{arg}}, V_{\mathsf{arg}} \rangle$ for NP relations that can be verified by time-T space-S RAMs by composing the polynomial commitment scheme of Theorem 5 and the polynomial IOP of Figure 6. Specifically, the prover P_{arg} and V_{arg} runs the prover and the verifier of the underlying PIOP except two changes: (1) P_{arg} (line 2, Figure 6) instead provides V_{arg} with a commitment to the multi-linear extension of the circuit transcript W. Here P_{arg} crucially relies on streaming access to W to compute the commitment in small-space using Com. (2) P_{arg} and V_{arg} run the protocol Eval in place of all verifier queries to the oracle $\overline{W}$ (line 11, Figure 6). We state the formal theorem.

Theorem 6 (Small-space Arguments for RAMs)**.** *There exists a public-coin interactive argument for* NP *relations verifiable by time-T space-S random access machines M, in the random oracle model, under the hardness of discrete-log in obliviously sampleable prime-order groups with the following complexity.*

1. *The protocol has perfect completeness, has* $O(\log(T))$ *rounds and* $\mathrm{polylog}(T)$ *communication, and has witness-extended emulation.*
2. *The prover runs in time* $T \cdot \mathrm{polylog}(T)$ *and space* $S \cdot \mathrm{polylog}(T)$ *given input-witness pair* $(x; w)$ *for M; and*
3. *The verifier runs in time* $T \cdot \mathrm{polylog}(T)$ *and space* $\mathrm{polylog}(T)$.

Proof. We compose our commitment scheme of Theorem 5 with the Polynomial IOP of Theorem 3. The algorithm Setup is identical to that of the commitment scheme. The protocol is identical to the protocol PIOP except for the following changes. First, instead of providing the verifier with an oracle $\overline{W}$, the prover instead sends $\mathsf{Com}^H(pp, W)$, noting that W uniquely defines the multi-linear extension $\overline{W}$. Second, the 3 verifier queries to the oracle $\overline{W}$ are replaced with 3 invocations of the protocol Eval^H.

By Lemma 8, the transcript W is computable in a streaming manner in time $T \cdot \mathrm{polylog}(T)$ and space $S \cdot \mathrm{polylog}(T)$, so computing $\mathsf{Com}^H(pp, W)$ requires $S \cdot \mathrm{polylog}(T)$ space. Replacing the oracle queries to $\overline{W}$ with 3 invocations of the Eval protocol requires $T \cdot \mathrm{polylog}(T)$ time from the verifier, since $|W| = T \cdot \mathrm{polylog}(T)$.

Let P^* be a cheating prover for our argument system. The we construct an emulator $\mathcal{E}$ for our scheme which does the following.

1. Runs P^* until the first Eval query is generated, yielding partial transcript t_1.
2. Run the emulator $\mathcal{E}^H_{\mathsf{pc}}$ of the polynomial commitment scheme, yielding extracted witness $W\colon \{0,1\}^s \to \mathbb{F}$ and some transcript t_2. We note that if $\mathcal{E}^H_{\mathsf{pc}}$ aborts, then $\mathcal{E}$ aborts.
3. Finish the interaction with P^*, yielding transcript t_3. Let $\mathsf{tr} = t_1 \circ t_2 \circ t_3$.
4. If W is not consistent with (C_M, x, y) or t is rejecting, output $(\mathsf{tr}, 0)$. If W is consistent with (C_M, x, y), extract witness w from W (since w is a substring of W by Lemma 8). If tr is accepting and $M(x; w) = y$, output (tr, w). Else output $\perp$.

Note that $\mathcal{E}$ as run above runs in expected polynomial time if $\mathcal{E}^H_{\mathsf{pc}}$ runs in expected polynomial time. Fix polynomial time adversary A. We need to show that

$$\Pr[\mathsf{tr} \leftarrow \langle P^*, V\rangle\colon A(\mathsf{tr}) = 1] \approx_{\mathsf{negl}(\lambda)} \Pr\left[(\mathsf{tr}, w) \leftarrow \mathcal{E}\colon \begin{array}{c} A(\mathsf{tr}) = 1 \wedge \\ \mathsf{tr} \text{ is accepting} \implies M(x; w) = y \end{array}\right], \tag{13}$$

where λ is the security parameter.

If $\mathcal{E}$ outputs a pair (tr, w), the transcript tr is a transcript produced by an honest verifier interacting with P^*. In this case, the probability that $A(\mathsf{tr}) = 1$ when $\mathcal{E}$ does not abort differs from the probability that $A(\mathsf{tr}) = 1$ for tr output by $\langle P^*, V\rangle$ by at most $\mathsf{negl}(\lambda)$. This is due to the negligible error of the emulator $\mathcal{E}^H_{\mathsf{pc}}$. We now turn to show that the probability $\mathcal{E}$ aborts is negligible.

Let E be the event that $\mathcal{E}$ aborts. In particular, $\mathcal{E}$ aborts if $\mathcal{E}^H_{\mathsf{pc}}$ fails to extract a witness, or if $\mathcal{E}^H_{\mathsf{pc}}$ extracts a witness (circuit transcript) W and W is a correct transcript for (C_M, x, y) and tr is an accepting transcript but $M(x; w) \neq y$ for witness w that is a substring of W. By the witness-extended emulation property of our polynomial commitment scheme, it holds that $\mathcal{E}^H_{\mathsf{pc}}$ aborts with negligible probability. Now suppose $\mathcal{E}^H_{\mathsf{pc}}$ does not abort and that it successfully extracts a witness W. Let E' be the event that W is a correct transcript for (C_M, x, y) and tr is accepting but $M(x; w) \neq y$, conditioned on $\mathcal{E}^H_{\mathsf{pc}}$ not aborting and successfully extracting witness W. Suppose $\Pr[E'] = \varepsilon$ for some $\varepsilon \in [0, 1]$. We now construct an adversary $\hat{P}$ which breaks soundness of our IOP with probability at least $\varepsilon - \mathsf{negl}(\lambda)$.

We first note the differences between the our argument verifier V and our IOP verifier V_{IOP}. The IOP verifier V_{IOP} receives a polynomial oracle W^* from the prover while V does not receive an oracle to the multi-linear extension of a transcript and instead receives a commitment to the the oracle W^*. Further, V_{IOP} simply queries said oracle to check the final statement of the sum-chekc, while V interacts with P^* to obtain evaluations of this oracle.

We now describe how $\hat{P}$ breaks the soundness of the IOP. $\hat{P}$ simulates the interaction between P^* and V until V makes an eval query. $\hat{P}$ then runs the emulator $\mathcal{E}^H_{\mathsf{pc}}$ to extract out witness W, rewinding P^* and V as necessary. Once a witness W is extracted, $\hat{P}$ rewinds P^* to the point just after the commitment to W was sent. Now $\hat{P}$ computes the multi-linear extension of W as $\overline{W}$ and sends this to V_{IOP}. $\hat{P}$ then forwards all verifier messages to P^* and forwards all messages from P^* to V_{IOP}. Finally V_{IOP} outputs accept or reject after querying the oracle $\overline{W}$ and computing a final check.

We note that $\mathcal{E}^H_{\mathsf{pc}}$ has error probability $\mathsf{negl}(\lambda)$; that is, the emulator may output an incorrect W with negligible probability. Suppose this is not the case and that W is a correct transcript for

(C_M, x, y) but the witness w extracted from W is such that $M(x; w) \neq y$. By assumption, we have that P^* convinces V to accept this scenario with probability ε. By our construction of $\hat{P}$, we also have that in this scenario V_{IOP} accepts with probability ε. So we have that

$$\Pr\left[\langle \hat{P}, V_{\mathsf{IOP}} \rangle = 1\right] \geq \varepsilon(1 - \mathsf{negl}(\lambda)) \geq \varepsilon - \mathsf{negl}(\lambda).$$

Note that by soundness of the IOP we have that

$$\Pr\left[\langle \hat{P}, V_{\mathsf{IOP}} \rangle = 1\right] \leq \frac{36s}{|\mathbb{F}|},$$

where $|\mathbb{F}|$ is exponential in the security parameter and $s = O(\log T)$. Thus we have that

$$\varepsilon \leq \frac{36s}{|\mathbb{F}|} + \mathsf{negl}(\lambda) = \mathsf{negl}(\lambda).$$

Therefore $\Pr[E] = \varepsilon \leq \mathsf{negl}(\lambda)$ as desired, and the total probability that $\mathcal{E}$ aborts is at most some negligible function of λ. This along with the negligible error of $\mathcal{E}^H_{\mathsf{pc}}$ gives that Equation (13) holds for some function $\mathsf{negl}(\lambda)$, showing witness-extended emulation. □

6.1 Obtaining Theorem 1

We discuss how to modify our interactive argument of knowledge from Theorem 6 to satisfy zero-knowledge and then make the resulting argument non-interactive, thus obtaining Theorem 1.

6.1.1 Zero-knowledge

We use commit-and-prove techniques introduced in [BGG+90, CD98] and later implemented in [WTs+18]. At a high level, this requires making two changes in our base protocols: (1) modify polynomial commitment from Section 4 to satisfy zero-knowledge—we modify all commitments sent in both Com and Eval protocols (Figure 3) to additionally include blinding factors. For example, commitment to $x \in \mathbb{F}$ under generator $g \in \mathbb{G}$ is changed from g^x to $g^x \cdot h^r$ for some randomly sampled $h \xleftarrow{\$} \mathbb{G}$ and $r \xleftarrow{\$} \mathbb{F}$. Further, at the end of the $\mathsf{EvalReduce}$ protocol when $N = 1$, prover instead of sending the witness in the clear instead engages with the verifier in Schnorr's zero-knowledge proof of dot-product protocol [Sch91]. This along with hiding of the commitments now ensure that the resulting polynomial commitment is zero-knowledge. (2) We replace all messages sent in the argument Theorem 6 in the clear with Pedersen hiding commitments and use techniques developed in [WTs+18] to ensure verifier checks go through. We emphasize that these changes do not asymptotically blow up the complexity of the protocol and, in particular, keep the space-complexity low. Furthermore, this transformation preserves the knowledge-soundness and public-coin features of the underlying argument [WTs+18].

6.1.2 Non-interactivity

We apply the Fiat-Shamir (FS) transform [FS87] to our zero-knowledge argument of knowledge, thereby obtaining a non-interactive, zero-knowledge argument of knowledge. However, note that it is folklore that applying FS to a t-round public-coin argument of knowledge yields a non-interactive argument of knowledge where the extractor runs in time exponential in t. Since our protocol has $O(\log T)$ rounds our extractor runs in poly(T)-time.

7 Acknowledgments

This work was done in part while Alexander R. Block and Pratik Soni were visiting the FACT Research Center at IDC Herzliya, Israel. Ron Rothblum was supported in part by a Milgrom family grant, by the Israeli Science Foundation (Grants No. 1262/18 and 2137/19), and the Technion Hiroshi Fujiwara cyber security research center and Israel cyber directorate. Alon Rosen is supported in part by ISF grant No. 1399/17 and Project PROMETHEUS (Grant 780701). Pratik Soni was supported in part by NSF grants CNS-1528178, CNS-1929901 and CNS-1936825 (CAREER), Glen and Susanne Culler Chair, ISF grant 1861/16 and AFOSR Award FA9550-17-1-0069. Alexander R. Block was supported in part by NSF grant CCF-1910659.

References

[AAB+19] Abdelrahaman Aly, Tomer Ashur, Eli Ben-Sasson, Siemen Dhooghe, and Alan Szepieniec. Design of symmetric-key primitives for advanced cryptographic protocols. Cryptology ePrint Archive, Report 2019/426, 2019. https://eprint.iacr.org/2019/426.

[BBB+18] Benedikt Bünz, Jonathan Bootle, Dan Boneh, Andrew Poelstra, Pieter Wuille, and Greg Maxwell. Bulletproofs: Short proofs for confidential transactions and more. In *2018 IEEE Symposium on Security and Privacy*, pages 315–334. IEEE Computer Society Press, May 2018. doi:10.1109/SP.2018.00020.

[BBHR18] Eli Ben-Sasson, Iddo Bentov, Yinon Horesh, and Michael Riabzev. Fast reed-solomon interactive oracle proofs of proximity. In Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella, editors, *ICALP 2018*, volume 107 of *LIPIcs*, pages 14:1–14:17. Schloss Dagstuhl, July 2018. doi:10.4230/LIPIcs.ICALP.2018.14.

[BBHR19] Eli Ben-Sasson, Iddo Bentov, Yinon Horesh, and Michael Riabzev. Scalable zero knowledge with no trusted setup. In Alexandra Boldyreva and Daniele Micciancio, editors, *CRYPTO 2019, Part III*, volume 11694 of *LNCS*, pages 701–732. Springer, Heidelberg, August 2019. doi:10.1007/978-3-030-26954-8_23.

[BC12a] Nir Bitansky and Alessandro Chiesa. Succinct arguments from multi-prover interactive proofs and their efficiency benefits. Cryptology ePrint Archive, Report 2012/461, 2012. http://eprint.iacr.org/2012/461.

[BC12b] Nir Bitansky and Alessandro Chiesa. Succinct arguments from multi-prover interactive proofs and their efficiency benefits. In Reihaneh Safavi-Naini and Ran Canetti, editors, *CRYPTO 2012*, volume 7417 of *LNCS*, pages 255–272. Springer, Heidelberg, August 2012. doi:10.1007/978-3-642-32009-5_16.

[BCC+16] Jonathan Bootle, Andrea Cerulli, Pyrros Chaidos, Jens Groth, and Christophe Petit. Efficient zero-knowledge arguments for arithmetic circuits in the discrete log setting. In Marc Fischlin and Jean-Sébastien Coron, editors, *EUROCRYPT 2016, Part II*, volume 9666 of *LNCS*, pages 327–357. Springer, Heidelberg, May 2016. doi:10.1007/978-3-662-49896-5_12.

[BCCT13] Nir Bitansky, Ran Canetti, Alessandro Chiesa, and Eran Tromer. Recursive composition and bootstrapping for SNARKS and proof-carrying data. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *45th ACM STOC*, pages 111–120. ACM Press, June 2013. doi:10.1145/2488608.2488623.

[BCGT13] Eli Ben-Sasson, Alessandro Chiesa, Daniel Genkin, and Eran Tromer. Fast reductions from RAMs to delegatable succinct constraint satisfaction problems: extended abstract. In Robert D. Kleinberg, editor, *ITCS 2013*, pages 401–414. ACM, January 2013. doi:10.1145/2422436.2422481.

[BCMS20] Benedikt Bünz, Alessandro Chiesa, Pratyush Mishra, and Nicholas Spooner. Proof-carrying data from accumulation schemes. Cryptology ePrint Archive, Report 2020/499, 2020. https://eprint.iacr.org/2020/499.

[BCS16] Eli Ben-Sasson, Alessandro Chiesa, and Nicholas Spooner. Interactive oracle proofs. In Martin Hirt and Adam D. Smith, editors, *TCC 2016-B, Part II*, volume 9986 of *LNCS*, pages 31–60. Springer, Heidelberg, October / November 2016. doi:10.1007/978-3-662-53644-5_2.

[BFS20] Benedikt Bünz, Ben Fisch, and Alan Szepieniec. Transparent SNARKs from DARK compilers. In Anne Canteaut and Yuval Ishai, editors, *EUROCRYPT 2020, Part I*, volume 12105 of *LNCS*, pages 677–706. Springer, Heidelberg, May 2020. doi:10.1007/978-3-030-45721-1_24.

[BGG+90] Michael Ben-Or, Oded Goldreich, Shafi Goldwasser, Johan Håstad, Joe Kilian, Silvio Micali, and Phillip Rogaway. Everything provable is provable in zero-knowledge. In Shafi Goldwasser, editor, *CRYPTO'88*, volume 403 of *LNCS*, pages 37–56. Springer, Heidelberg, August 1990. doi:10.1007/0-387-34799-2_4.

[BGH19] Sean Bowe, Jack Grigg, and Daira Hopwood. Halo: Recursive proof composition without a trusted setup. Cryptology ePrint Archive, Report 2019/1021, 2019. https://eprint.iacr.org/2019/1021.

[BGKS19] Eli Ben-Sasson, Lior Goldberg, Swastik Kopparty, and Shubhangi Saraf. DEEP-FRI: Sampling outside the box improves soundness. Cryptology ePrint Archive, Report 2019/336, 2019. https://eprint.iacr.org/2019/336.

[BMW98] Ingrid Biehl, Bernd Meyer, and Susanne Wetzel. Ensuring the integrity of agent-based computations by short proofs. In Kurt Rothermel and Fritz Hohl, editors, *Mobile Agents*, pages 183–194, Berlin, Heidelberg, 1998. Springer Berlin Heidelberg.

[BTVW14] Andrew J. Blumberg, Justin Thaler, Victor Vu, and Michael Walfish. Verifiable computation using multiple provers. Cryptology ePrint Archive, Report 2014/846, 2014. http://eprint.iacr.org/2014/846.

[CD98] Ronald Cramer and Ivan Damgård. Zero-knowledge proofs for finite field arithmetic; or: Can zero-knowledge be for free? In Hugo Krawczyk, editor, *CRYPTO'98*, volume 1462 of *LNCS*, pages 424–441. Springer, Heidelberg, August 1998. doi:10.1007/BFb0055745.

[CMT12] Graham Cormode, Michael Mitzenmacher, and Justin Thaler. Practical verified computation with streaming interactive proofs. In Shafi Goldwasser, editor, *ITCS 2012*, pages 90–112. ACM, January 2012. doi:10.1145/2090236.2090245.

[COS20] Alessandro Chiesa, Dev Ojha, and Nicholas Spooner. Fractal: Post-quantum and transparent recursive proofs from holography. In Anne Canteaut and Yuval Ishai, editors, *EUROCRYPT 2020, Part I*, volume 12105 of *LNCS*, pages 769–793. Springer, Heidelberg, May 2020. doi:10.1007/978-3-030-45721-1_27.

[FS87] Amos Fiat and Adi Shamir. How to prove yourself: Practical solutions to identification and signature problems. In Andrew M. Odlyzko, editor, *CRYPTO'86*, volume 263 of *LNCS*, pages 186–194. Springer, Heidelberg, August 1987. doi:10.1007/3-540-47721-7_12.

[GGPR13] Rosario Gennaro, Craig Gentry, Bryan Parno, and Mariana Raykova. Quadratic span programs and succinct NIZKs without PCPs. In Thomas Johansson and Phong Q. Nguyen, editors, *EUROCRYPT 2013*, volume 7881 of *LNCS*, pages 626–645. Springer, Heidelberg, May 2013. doi:10.1007/978-3-642-38348-9_37.

[GH98] Oded Goldreich and Johan Håstad. On the complexity of interactive proofs with bounded communication. *Inf. Process. Lett.*, 67(4):205–214, 1998.

[GI08] Jens Groth and Yuval Ishai. Sub-linear zero-knowledge argument for correctness of a shuffle. In Nigel P. Smart, editor, *EUROCRYPT 2008*, volume 4965 of *LNCS*, pages 379–396. Springer, Heidelberg, April 2008. doi:10.1007/978-3-540-78967-3_22.

[GKR08] Shafi Goldwasser, Yael Tauman Kalai, and Guy N. Rothblum. Delegating computation: interactive proofs for muggles. In Richard E. Ladner and Cynthia Dwork, editors, *40th ACM STOC*, pages 113–122. ACM Press, May 2008. doi:10.1145/1374376.1374396.

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof systems. *SIAM J. Comput.*, 18(1):186–208, 1989. doi:10.1137/0218012.

[GVW02] Oded Goldreich, Salil P. Vadhan, and Avi Wigderson. On interactive proofs with a laconic prover. *Comput. Complex.*, 11(1-2):1–53, 2002.

[GW11] Craig Gentry and Daniel Wichs. Separating succinct non-interactive arguments from all falsifiable assumptions. In Lance Fortnow and Salil P. Vadhan, editors, *43rd ACM STOC*, pages 99–108. ACM Press, June 2011. doi:10.1145/1993636.1993651.

[HR18] Justin Holmgren and Ron Rothblum. Delegating computations with (almost) minimal time and space overhead. In Mikkel Thorup, editor, *59th FOCS*, pages 124–135. IEEE Computer Society Press, October 2018. doi:10.1109/FOCS.2018.00021.

[Kil92] Joe Kilian. A note on efficient zero-knowledge proofs and arguments (extended abstract). In *24th ACM STOC*, pages 723–732. ACM Press, May 1992. doi:10.1145/129712.129782.

[KPV19] Assimakis Kattis, Konstantin Panarin, and Alexander Vlasov. RedShift: Transparent SNARKs from list polynomial commitment IOPs. Cryptology ePrint Archive, Report 2019/1400, 2019. https://eprint.iacr.org/2019/1400.

[KR09] Yael Tauman Kalai and Ran Raz. Probabilistically checkable arguments. In Shai Halevi, editor, *CRYPTO 2009*, volume 5677 of *LNCS*, pages 143–159. Springer, Heidelberg, August 2009. doi:10.1007/978-3-642-03356-8_9.

[KRR13] Yael Tauman Kalai, Ran Raz, and Ron D. Rothblum. Delegation for bounded space. In Dan Boneh, Tim Roughgarden, and Joan Feigenbaum, editors, *45th ACM STOC*, pages 565–574. ACM Press, June 2013. doi:10.1145/2488608.2488679.

[KZG10] Aniket Kate, Gregory M. Zaverucha, and Ian Goldberg. Constant-size commitments to polynomials and their applications. In Masayuki Abe, editor, *ASIACRYPT 2010*, volume 6477 of *LNCS*, pages 177–194. Springer, Heidelberg, December 2010. doi:10.1007/978-3-642-17373-8_11.

[LFKN90] Carsten Lund, Lance Fortnow, Howard J. Karloff, and Noam Nisan. Algebraic methods for interactive proof systems. In *31st FOCS*, pages 2–10. IEEE Computer Society Press, October 1990. doi:10.1109/FSCS.1990.89518.

[Lin03] Yehuda Lindell. Parallel coin-tossing and constant-round secure two-party computation. *Journal of Cryptology*, 16(3):143–184, June 2003. doi:10.1007/s00145-002-0143-7.

[Mic94] Silvio Micali. CS proofs (extended abstracts). In *35th FOCS*, pages 436–453. IEEE Computer Society Press, November 1994. doi:10.1109/SFCS.1994.365746.

[Ped92] Torben P. Pedersen. Non-interactive and information-theoretic secure verifiable secret sharing. In Joan Feigenbaum, editor, *CRYPTO'91*, volume 576 of *LNCS*, pages 129–140. Springer, Heidelberg, August 1992. doi:10.1007/3-540-46766-1_9.

[PHGR13] Bryan Parno, Jon Howell, Craig Gentry, and Mariana Raykova. Pinocchio: Nearly practical verifiable computation. In *2013 IEEE Symposium on Security and Privacy*, pages 238–252. IEEE Computer Society Press, May 2013. doi:10.1109/SP.2013.47.

[PST13] Charalampos Papamanthou, Elaine Shi, and Roberto Tamassia. Signatures of correct computation. In Amit Sahai, editor, *TCC 2013*, volume 7785 of *LNCS*, pages 222–242. Springer, Heidelberg, March 2013. doi:10.1007/978-3-642-36594-2_13.

[RR19] Noga Ron-Zewi and Ron Rothblum. Local proofs approaching the witness length. *Electron. Colloquium Comput. Complex.*, 26:127, 2019. URL: https://eccc.weizmann.ac.il/report/2019/127.

[RRR16] Omer Reingold, Guy N. Rothblum, and Ron D. Rothblum. Constant-round interactive proofs for delegating computation. In Daniel Wichs and Yishay Mansour, editors, *48th ACM STOC*, pages 49–62. ACM Press, June 2016. doi:10.1145/2897518.2897652.

[Sch91] Claus-Peter Schnorr. Efficient signature generation by smart cards. *Journal of Cryptology*, 4(3):161–174, January 1991. doi:10.1007/BF00196725.

[Set20] Srinath Setty. Spartan: Efficient and general-purpose zkSNARKs without trusted setup. In Daniele Micciancio and Thomas Ristenpart, editors, *CRYPTO 2020, Part III*, volume 12172 of *LNCS*, pages 704–737. Springer, Heidelberg, August 2020. doi:10.1007/978-3-030-56877-1_25.

[Sha90] Adi Shamir. IP=PSPACE. In *31st FOCS*, pages 11–15. IEEE Computer Society Press, October 1990. doi:10.1109/FSCS.1990.89519.

[Tha13] Justin Thaler. Time-optimal interactive proofs for circuit evaluation. In Ran Canetti and Juan A. Garay, editors, *CRYPTO 2013, Part II*, volume 8043 of *LNCS*, pages 71–89. Springer, Heidelberg, August 2013. doi:10.1007/978-3-642-40084-1_5.

[Val08] Paul Valiant. Incrementally verifiable computation or proofs of knowledge imply time/space efficiency. In Ran Canetti, editor, *TCC 2008*, volume 4948 of *LNCS*, pages 1–18. Springer, Heidelberg, March 2008. doi:10.1007/978-3-540-78524-8_1.

[WBT+17] Primal Wijesekera, Arjun Baokar, Lynn Tsai, Joel Reardon, Serge Egelman, David A. Wagner, and Konstantin Beznosov. The feasibility of dynamically granted permissions: Aligning mobile privacy with user preferences. In *2017 IEEE Symposium on Security and Privacy*, pages 1077–1093. IEEE Computer Society Press, May 2017. doi:10.1109/SP.2017.51.

[WTs+18] Riad S. Wahby, Ioanna Tzialla, abhi shelat, Justin Thaler, and Michael Walfish. Doubly-efficient zkSNARKs without trusted setup. In *2018 IEEE Symposium on Security and Privacy*, pages 926–943. IEEE Computer Society Press, May 2018. doi:10.1109/SP.2018.00060.

[ZXZS20] Jiaheng Zhang, Tiancheng Xie, Yupeng Zhang, and Dawn Song. Transparent polynomial delegation and its applications to zero knowledge proof. In *2020 IEEE Symposium on Security and Privacy*, pages 859–876. IEEE Computer Society Press, May 2020. doi:10.1109/SP40000.2020.00052.