RoK, Paper, SISsors – Toolkit for Lattice-based Succinct Arguments

(Full Version)

Michael Klooß[1*], Russell W. F. Lai[2], Ngoc Khanh Nguyen[3], and Michał Osadnik[2]

[1] ETH Zurich, Zurich, Switzerland
[2] Aalto University, Espoo, Finland
[3] King's College London, London, UK

Abstract. Lattice-based succinct arguments allow to prove bounded-norm satisfiability of relations, such as $f(\mathbf{s}) = \mathbf{t} \bmod q$ and $\|\mathbf{s}\| \leq \beta$, over specific cyclotomic rings $\mathcal{O}_\mathcal{K}$, with proof size polylogarithmic in the witness size. However, state-of-the-art protocols require either 1) a super-polynomial size modulus q due to a soundness gap in the security argument, or 2) a verifier which runs in time linear in the witness size. Furthermore, construction techniques often rely on specific choices of $\mathcal{K}$ which are not mutually compatible. In this work, we exhibit a diverse toolkit for constructing efficient lattice-based succinct arguments:

(i) We identify new subtractive sets for general cyclotomic fields $\mathcal{K}$ and their maximal real subfields $\mathcal{K}^+$, which are useful as challenge sets, e.g. in arguments for exact norm bounds.
(ii) We construct modular, verifier-succinct reductions of knowledge for the bounded-norm satisfiability of structured-linear/inner-product relations, without any soundness gap, under the vanishing SIS assumption, over any $\mathcal{K}$ which admits polynomial-size subtractive sets.
(iii) We propose a framework to use twisted trace maps, i.e. maps of the form $\tau(z) = \frac{1}{N} \cdot \mathsf{Trace}_{\mathcal{K}/\mathbb{Q}}(\alpha \cdot z)$, to embed $\mathbb{Z}$-inner-products as $\mathcal{R}$-inner-products for some structured subrings $\mathcal{R} \subseteq \mathcal{O}_\mathcal{K}$ whenever the conductor has a square-free odd part.
(iv) We present a simple extension of our reductions of knowledge for proving the consistency between the coefficient embedding and the Chinese Remainder Transform (CRT) encoding of $\mathbf{s}$ over any cyclotomic field $\mathcal{K}$ with a smooth conductor, based on a succinct decomposition of the CRT map into automorphisms, and a new, simple succinct argument for proving automorphism relations.

Combining all techniques, we obtain, for example, verifier-succinct arguments for proving that $\mathbf{s}$ satisfying $f(\mathbf{s}) = \mathbf{t} \bmod q$ has binary coefficients, without soundness gap and with polynomial-size modulus q.

1 Introduction

A fundamental and recurring task in constructing lattice-based succinct arguments is to prove knowledge of a committed vector $\mathbf{s} \in \mathcal{R}^m$ over a ring $\mathcal{R}$ which satisfies norm-bound constraints, such as $\|\mathbf{s}\| \leq \beta$. For instance, such protocols could be extended directly into a succinct argument for structured languages [CLM23], combined with quadratic functional commitments to yield succinct arguments for NP [ACL+22, CLM23][4], or transformed into polynomial commitment schemes [FMN23, AFLN24, CMNW24] which allow compiling polynomial interactive oracle proofs [BCS16] into succinct arguments.

As evidenced in prior works [Lyu12, LNP22, BS23], the currently most efficient lattice-based (non-)succinct arguments operate over rings of integers $\mathcal{R} := \mathbb{Z}[\zeta]$ of cyclotomic number fields $\mathcal{K} := \mathbb{Q}(\zeta)$, where ζ is a primitive $\mathfrak{f}$-th root of unity for $\mathfrak{f} = \mathsf{poly}(\lambda)$. Indeed, the ability to construct exponential-sized low-norm challenge sets over $\mathcal{R}$ allows the aforementioned protocols to achieve negligible soundness in one-shot while maintaining relatively small lattice parameters. However, this comes at a cost of the following two complications.

*Work done at Aalto University. The author's affiliation changed before publication.

[4] [ACL+22, CLM23] relied on the knowledge-kRISIS assumption for the knowledge soundness of well-formedness of commitments. However, the assumption has subsequently been cryptanalysed [WW23, DFS24], rendering the security proofs vacuous.

Correctness Gap. The first one can be described as the *correctness gap*. Namely, most of the recursion-based protocols start with the initial witness $\mathbf{s}_0 := \mathbf{s}$, and in the i-th iteration, an honest prover somehow folds the "current" witness $\mathbf{s}_{i-1}$ into a new one $\mathbf{s}_i$; thus shrinking the dimension of the witness, but simultaneously, increasing its norm. At the end, say after μ iterations, the prover outputs the final witness $\mathbf{s}_\mu$ of small (potentially constant) dimension. Suppose there exists some γ such that for all $i = 1, \ldots, \mu$ we have $\|\mathbf{s}_i\| \leq \gamma \cdot \|\mathbf{s}_{i-1}\|$. Then, in order to maintain correctness, one must inherently choose $q > \gamma^\mu \cdot \beta \geq \|\mathbf{s}_\mu\|$. We call this phenomenon the correctness gap, since if our only task were to commit to $\mathbf{s}$ using a standard lattice-based commitment scheme, setting $q = O(\beta)$ would suffice[5].

Soundness Gap. A more concerning issue is the *soundness gap*. A vast majority of prior works based on cyclotomic rings encounter the problem that the extracted witness $\bar{\mathbf{s}}$ is not necessarily short, but it is of the fractional form $\bar{\mathbf{s}} := \bar{\mathbf{z}}/\bar{c} \bmod q$, where q is the proof system modulus and both $\bar{\mathbf{z}} \in \mathcal{R}^m$ and $\bar{c} \in \mathcal{R}$ are somewhat short (but $\|\bar{\mathbf{z}}\|$ is larger than β). Even though this *relaxed* soundness suffices to construct basic primitives, such as signature schemes [Lyu12, DKL+18], verifiable encryption [LN17], or few-time verifiable random functions [EKS+21], it is not enough when the required functionality naturally involves proving exact norm bounds (e.g. in set membership and range proofs). But especially in the context of succinct arguments built in a recursive manner, dealing with the slack and other norm-growth related issues have shown to have enormous impact on setting up the parameters [BLNS20, BCS23, AL21, AFLN24] , such as picking super-polynomial modulus q, which makes the aforementioned schemes seem barely practical.

Prior works. Since the soundness gap seemed to be the main efficiency bottleneck of lattice-based succinct arguments, several works naturally tried to address this issue first. To begin with, Albrecht and Lai [AL21] designed a lattice-based argument of polylogarithmic size, where the extracted witness $\bar{\mathbf{s}}$ is somewhat short. The key ingredient of [AL21] was the notion of *subtractive sets*. Namely, a set $S \subseteq \mathcal{R}$ is called subtractive if for any two distinct elements $c, c' \in S$, $c - c'$ is invertible over the ring $\mathcal{R}$. Since the invertibility is independent of the proof system modulus q, the latter can be picked freely so that the inverse $(c - c')^{-1}$ is short relative to q. Further, it was shown how to construct such subtractive sets of cardinality p in cyclotomic rings of prime power conductors $\mathfrak{f} := p^k$. Thus, using subtractive sets as a challenge space for the verifier, one can argue that the extracted witness $\bar{\mathbf{s}} := \bar{\mathbf{z}}/\bar{c}$ has low norm, because $1/\bar{c}$ itself is short. However, this approach comes at a cost of non-negligible soundness error (due to the size of subtractive sets), and therefore some sort of soundness amplification is necessary. Furthermore, the protocol itself still does not manage to prove the exact norm bound, i.e. $\|\mathbf{s}\| \leq \beta$. In fact, in the context of recursive succinct arguments, the norm of the extracted witness can only be upper bounded by $\gamma^\mu \cdot \theta^{O(\mu)} \cdot \beta$ for some $\theta \approx \mathfrak{f}$.

In the setting of power-of-two cyclotomic rings, the strategy above falls apart completely since there exists no subtractive set of size larger than two [Len76, AL21]. Hence, a different methodology has recently been developed. Notably, Beullens and Seiler [BS23] proposed a succinct argument, LaBRADOR, for proving $\|\mathbf{s}\|^2 \leq \beta^2$ (among other relations), inspired by the following two-fold approach from [LNP22]:

(i) *Approximate shortness proof.* Prove that $\mathbf{s}$ is somewhat short.
(ii) $\mathbb{Z}_q$*-Inner product proof.* Prove that $(\langle \psi(\mathbf{s}), \psi(\mathbf{s}) \rangle \pmod q) \leq \beta^2$, where $\psi(\mathbf{s})$ is the coefficient vector of $\mathbf{s}$.

Combining (i) and (ii), one can argue that for a large enough modulus q no modulo wrap-around occurs, and therefore $\langle \psi(\mathbf{s}), \psi(\mathbf{s}) \rangle \leq \beta^2$ holds over $\mathbb{Z}$.

In order to prove (i) without relying on subtractive sets, LaBRADOR uses the Johnson-Lindenstrauss random projection technique [BL17, LNS21, GHL22]. The idea is that the verifier will first generate an integer matrix $\mathbf{B}$ with short (binary or ternary) values as a challenge, and the prover then outputs $\psi(\mathbf{v}) := \mathbf{B}\psi(\mathbf{s}) \pmod q$. Afterwards, the verifier checks whether $\psi(\mathbf{v})$ is of low norm (which is true in the honest executions, since both $\mathbf{B}$ and $\psi(\mathbf{s})$ are). Finally, the prover needs to prove wellformedness of $\psi(\mathbf{v})$, i.e. the linear equation $\mathbf{B}\psi(\mathbf{s}) = \psi(\mathbf{v})$ over $\mathbb{Z}_q$. The crucial soundness argument is that if the extracted $\mathbf{s}$ was not short, then with high probability (dictated by the number of rows of $\mathbf{B}$), $\psi(\mathbf{v}) = \mathbf{B}\psi(\mathbf{s})$ would not have low norm, which leads to a contradiction. Unfortunately, the random projection strategy inherently requires the verifier to generate the matrix $\mathbf{B}$, which itself has length $O(m)$. As a consequence, the verifier

[5]For presentation, we omitted the factors related to the security parameter λ.

runtime becomes essentially linear in the witness size, which may not be satisfying in certain real-world use cases.

We highlight that both (i) and (ii) require some kind of inner product proof over $\mathbb{Z}_q$; either between two committed vectors, or between one public and one committed vector. Since the underlying protocol natively operates over cyclotomic rings $\mathcal{R} = \mathbb{Z}[\zeta]$, it is essential to transform $\mathbb{Z}$-relations into equivalent ones over the ring $\mathcal{R}$. To this end, it was shown in [LNP22] that for any two elements $a, b \in \mathcal{R}$ of a power-of-two cyclotomic ring, the constant term[6] of $a \cdot \bar{b} \in \mathcal{R}$ is exactly equal to the inner product $\langle \psi(\mathbf{a}), \psi(\mathbf{b}) \rangle \in \mathbb{Z}$, where $\psi(\mathbf{a}), \psi(\mathbf{b})$ are the coefficient vectors of a, b respectively and $\bar{\cdot}$ here denotes the complex conjugation. This observation allows us to translate proving inner products and linear relations over integers into proving statements about constant terms over the ring $\mathcal{R}$. Finally, LaBRADOR makes use of the fact that inner product relations over $\mathcal{R}$ are "folding-friendly" and can be efficiently proven in a recursive manner.

Interestingly, LaBRADOR also managed to circumvent the correctness gap by taking inspiration from the "decompose-then-hash" paradigm used in lattice-based Merkle trees [PSTY13]. Intuitively, using the notation above for describing recursive-based protocols, instead of folding the intermediate witness $\mathbf{s}_{i-1}$ directly into a new one $\mathbf{s}_i$, an honest prover would first decompose $\mathbf{s}_{i-1}$ (w.r.t. some decomposition base b) into multiple vectors $(\mathbf{s}_{i-1,j})_{j\in[\ell]}$ of much smaller norm and then fold all of them together into a new witness $\mathbf{s}_i$[7]. By carefully picking various parameters, such as b, one can ensure that, in an honest execution, if $\|\mathbf{s}_{i-1}\| \leq \beta$, then we must have $\|\mathbf{s}_i\| \leq \beta$. This technique was also adopted in a recent folding scheme called LatticeFold [BC24].

Bridging the gap. At a high level, the aforementioned approaches to prove shortness seem somewhat orthogonal. For $\mathfrak{f} = p^k$, where $p = \mathsf{poly}(\lambda)$ is a large enough prime, one can rely on subtractive sets to efficiently prove approximate shortness (i) with succinct verification [CLM23]. However, it is unknown how to translate proving $\mathbb{Z}_q$-relations, as in (ii), into equivalent relations over odd prime-power cyclotomic rings. On the other hand, for $\mathfrak{f} = 2^k$, one can apply the Johnson-Lindenstrauss projection strategy to prove both (i) and (ii), but at the cost of slow verification time.

Hence, it is an important research question whether there exist cyclotomic (or other) rings $\mathcal{R}$, which contain subtractive sets of fairly large size, and at the same time, expose efficient packing and batching techniques for turning relations over $\mathbb{Z}$ (or more generally, other base rings) to relations over $\mathcal{R}$. An affirmative answer, together with existing optimisations, would then yield a practical lattice-based succinct argument for proving exact norm bounds with fast verification.

1.1 Our Contributions

In this work, we present a versatile toolkit for constructing lattice-based succinct arguments that eliminate correctness and soundness gaps while maintaining succinct verification. Our contributions are outlined as follows:

Succinct Arguments for Bounded-Norm Satisfiability. We design a lattice-based succinct argument system for bounded-norm satisfiability of structured linear and inner-product relations. Our system retains features of previous protocols, such as transparent setup, quasi-linear-time prover, polylogarithmic-time verifier, and negligible soundness in one-shot, while simultaneously eliminating any correctness and soundness gaps. Consequently, our argument system achieves asymptotically the most attractive proof sizes, which are smaller by at least a factor of $\Omega(\log^2 \lambda)$ than the prior state-of-the-art constructions (see Figure 1 for more details). Furthermore, our protocol's modular design allows for straightforward analysis and customisation, making it adaptable to various applications.

Subtractive Sets. Our protocol uses subtractive sets as challenge sets. While subtractive sets for prime-power cyclotomic rings are well-known, the non-prime-power case seems less studied. Motivated by the need of non-prime-power rings (e.g. for the twisted trace technique, see below) in some applications, we identify a subtractive set for cyclotomic rings $\mathbb{Z}[\zeta_\mathfrak{f}]$ of non-prime-power conductor $\mathfrak{f}$ with a cardinality of $\mathfrak{f}/\mathfrak{f}_{\max}$, where $\mathfrak{f}_{\max}$ is the largest prime-power divisor of $\mathfrak{f}$. Additionally, we identify subtractive sets

[6] We say that $a_0 \in \mathbb{Z}$ is the constant term of the ring element $a = \sum_{i=0}^{\varphi(\mathfrak{f})} a_i \zeta^i \in \mathbb{Z}[\zeta]$.

[7] For soundness, the prover needs to prove additional relations involving $(\mathbf{s}_{i-1,j})_{j\in[\ell]}$.

scheme	assumptions	transparent setup	proof size
[CLM23]	vSIS	✓	$O\left(\log^5 m \cdot \frac{\lambda^2}{\log^2 \lambda}\right)$
[BCS23]	M-SIS	✓	$O\left(\log^6 m \cdot \frac{\lambda^2}{\log \lambda}\right)$
[FMN23]	PowerBASIS	×	$O\left(\log^5 m \cdot \frac{\lambda^2}{\log^2 \lambda}\right)$
[AFLN24]	M-SIS	×	$O\left(\log^5 m \cdot \frac{\lambda^2}{\log^2 \lambda}\right)$
[CMNW24]	SIS	✓	$O(\log^3 m \cdot \lambda^2)$
This work	vSIS	✓	$O\left(\log^3 m \cdot \frac{\lambda^2}{\log^2 \lambda}\right)$

Fig. 1. Asymptotic efficiency of our commitment opening proof (in bits) and comparison with prior works which support succinct $\mathsf{poly}(\log m, \lambda)$ verification time. Here, λ is the security parameter and m is the length of the committed vector. For each construction, the proof size corresponds to the soundness error $\mathsf{poly}(\lambda, \log m) \cdot 2^{-\lambda}$. The SIS-related parameters were chosen with respect to the methodology from [MR09] for running BKZ on block size $b = O(\lambda)$. For [BCS23, CLM23, FMN23], which only achieve inverse-polynomial soundness in one-shot, we applied a standard soundness amplification by parallel-repeating the protocol by a factor of $O(\lambda/\log \lambda)$. We note that for [AFLN24], [CMNW24], and this work, super-polynomial knowledge extraction runtime $O(m^{\log \lambda})$ is obtained.

over the real subrings $\mathbb{Z}[\zeta_{\mathfrak{f}} + \zeta_{\mathfrak{f}}^{-1}]$, with a cardinality of $(p+1)/2$ for prime-power conductors $\mathfrak{f} = p^k$ and $\lfloor \mathfrak{f}/(2\mathfrak{f}_{\max}) \rfloor$ for non-prime-power $\mathfrak{f}$.

Embedded $\mathbb{Z}$-Inner-Products via Twisted Trace. While our protocol supports proving inner products over rings such as $\mathbb{Z}[\zeta_{\mathfrak{f}}]$, higher-layer applications may require proving inner products over $\mathbb{Z}$, e.g. for proving that a committed $\mathbb{Z}$-vector is binary. Unfortunately, efficient methods for embedding $\mathbb{Z}$-inner products to $\mathbb{Z}[\zeta_{\mathfrak{f}}]$-inner products were only known for $\mathfrak{f} = 2^d$ being a power of 2, which is problematic because subtractive sets over $\mathbb{Z}[\zeta_{2^d}]$ are of cardinality at most 2. We extend the existing embedding method to any ring of the form $\mathbb{Z}[\zeta_{2^d}] \otimes \mathbb{Z}[\zeta_{p_0} + \zeta_{p_0}^{-1}] \otimes \ldots \otimes \mathbb{Z}[\zeta_{p_{k-1}} + \zeta_{p_{k-1}}^{-1}]$, where $p_0, \ldots, p_{k-1}$ are distinct odd primes. This is achieved by replacing the "constant term map" with a "twisted trace map" of the form $\tau(z) = \frac{1}{N}\mathsf{Trace}(\alpha \cdot z)$.

Succinct Consistency Proof for CRT. Another typical way of embedding $\mathbb{Z}$-relations into $\mathbb{Z}[\zeta_{\mathfrak{f}}]$-relations is via the Chinese Remainder Transform (CRT). However, this requires proving that the witness vector is committed in both the coefficient embedding and its CRT coefficients consistently, and known consistency proofs are not succinct. Using the fact that the CRT over cyclotomic fields with smooth conductors can be succinctly represented through a few automorphism evaluations, we derive a succinct argument for the consistency between the commitment of the coefficient embedding and that of the CRT coefficients. At the core of our succinct consistency proof is a new succinct argument that verifies whether two committed vectors are related by an entry-wise automorphism.

2 Technical Overview

Throughout this work, we will assume that $\mathcal{K} = \mathbb{Q}(\zeta)$ is a cyclotomic field with conductor $\mathfrak{f}$ and degree $\varphi = \varphi(\mathfrak{f}) = \mathsf{poly}(\lambda)$, and $\mathcal{O}_{\mathcal{K}} = \mathbb{Z}[\zeta]$ is its ring of integers. For some of our results, we will further require $\mathcal{K}^+ = \mathbb{Q}(\zeta + \zeta^{-1})$, the maximal real subfield of $\mathcal{K}$, and its ring of integers $\mathcal{O}_{\mathcal{K}^+} = \mathbb{Z}[\zeta + \zeta^{-1}]$. Depending on the context of a specific section, we will use $\mathcal{R} \subset \mathcal{O}_{\mathcal{K}}$ to denote a ring of interest to that section. Unless specified, we measure the norm of elements and vectors by their ℓ_2-norm over the canonical embedding over $\mathcal{K}$. Our results can be divided into three parts, which we overview in Section 2.1, 2.2, and 2.3 respectively.

2.1 Subtractive Sets

In Section 4, we expose subtractive sets over $\mathcal{O}_{\mathcal{K}}$ with non-prime-power conductor $\mathfrak{f}$, and over $\mathcal{O}_{\mathcal{K}^+}$ with both prime-power and non-prime-power conductors, with favourable properties, i.e. they have $\mathsf{poly}(\lambda)$

cardinality and small expansion factors. These subtractive sets can be used in any lattice-based arguments, and in particular those developed in this work.

A set $S \subset \mathcal{R}$ is said to be subtractive over $\mathcal{R}$ if for any two distinct elements $c, c' \in S$, it holds that $c - c' \in \mathcal{R}^\times$, i.e. $c - c'$ is a unit. This concept is prevalently linked with the examination of Euclidean number fields [Len76] and has also found relevance in lattice-based cryptography, specifically in argument systems and secret sharing [AL21]. An explicit creation of an upper-bound-matching cardinality p is evident in a cyclotomic ring $\mathcal{R} = \mathcal{O}_\mathcal{K}$ with a prime-power conductor $\mathfrak{f} = p^k$. On the other hand, we are not aware of explicit studies of subtractive sets regarding other cyclotomic rings and their subrings.

For applications in lattice-based cryptography, the most relevant measures of the quality of a subtractive set S are its

(i) cardinality $|S|$, which inversely affects the knowledge error of argument systems using S as a challenge set,
(ii) "expansion factor" $\gamma = \gamma_S$, i.e. how much the norm of an element grows when multiplied with an element in S, which affects the "correctness gap" of lattice-based argument systems,
(iii) "inverse-expansion factor" $\theta = \theta_S$, i.e. how much the norm of an element grow when multiplied with $(c - c')^{-1}$ for distinct $c, c' \in S$, which affects the "soundness gap" of lattice-based argument systems.

For $\mathcal{R} = \mathcal{O}_\mathcal{K}$ with prime-power conductor $\mathfrak{f} = p^k$, it is known [Len76, AL21] that there exists a subtractive set S of cardinality p and expansion factors $\gamma, \theta \approx p$.

Our main result in this part is the exposition of the subtractive set $S := \{\zeta^i\}_{i \in [\mathfrak{f}/\mathfrak{f}_{\max}]}$ of cardinality $|S| = \mathfrak{f}/\mathfrak{f}_{\max}$ for any conductor $\mathfrak{f}$ with at least two distinct prime factors, where $\mathfrak{f}_{\max}$ is the largest prime-power factor of $\mathfrak{f}$. Notably, the expansion factor (conserning the canonical 2-norm) is $\gamma = 1$, i.e. the norm of an element does not grow when multiplied with an element from S, while the inverse-expansion factor $\theta \approx \mathfrak{f}$ is similar to the existing result for prime-power rings.

For completeness, we also expose related subtractive sets over $\mathcal{O}_{\mathcal{K}^+}$ for both prime-power and non-prime-power conductors.

2.2 Tight Succinct Argument for Bounded Norm Satisfiability

In Section 5, we work with $\mathcal{R} = \mathcal{O}_\mathcal{K}$ or $\mathcal{O}_{\mathcal{K}^+}$. We present a new lattice-based succinct argument for proving the bounded norm satisfiability of structured linear and/or inner-product relations, denoted by Ξ^{lin} and Ξ^{ip} respectively. More concretely, the argument system allows to prove knowledge of a short vector $\mathbf{w} \in \mathcal{R}^m$, with $m = d^\mu$, satisfying

- a linear relation $\mathbf{F}\mathbf{w} = \mathbf{y} \bmod q$, where $\mathbf{F} = \mathbf{F}_{\mu-1} \bullet \ldots \bullet \mathbf{F}_0 \in \mathcal{R}_q^{n \times m}$ can be expressed as a row-wise tensor product of μ matrices $\mathbf{F}_i \in \mathcal{R}_q^{n \times d}$, and
- (optionally) an inner-product relation $\langle \mathbf{w}, \alpha(\mathbf{w}) \rangle \bmod q$, where α is either the identity function or the complex conjugate (specified publicly).

Our argument system consists of $O(\mu) = O(\log_d m)$ rounds and is public-coin, and can thus be made non-interactive via the Fiat-Shamir transform. The prover time is quasi-linear in the size of the statement, and both the proof size and the verifier time are polylogarithmic in the statement size. It can be instantiated with a transparent setup. For example, the rows of $\mathbf{F}$ could contain a random commitment key of the vSIS commitment scheme [CLM23] and evaluations of monomials at different evaluation points. This turns the vSIS commitment scheme into a polynomial commitment scheme, which can then be used to compile a PIOP into a SNARK.

Correctness and Soundness Gaps. A distinguishing feature of our argument system is that it is free of the so-called "correctness gap" and "soundness gap".

The correctness gap refers to the phenomenon that although the prover's witness $\mathbf{w}$ is of norm at most β, the norm check performed by the verifier in the protocol is against a bound $\beta' \gg \beta$. Typically, e.g. in lattice-based Bulletproofs, we have $\beta' \approx (1 + \gamma)^\mu \beta$. Using the subtractive set suggested in [AL21] and picking $\mu \approx \log \lambda$, the gap $\beta'/\beta \approx (1 + \gamma)^\mu$ is super-polynomial in λ. Note that if the subtractive set suggested in Section 4 with $\gamma = 1$ is used, then the correctness gap is immediately reduced to $\mathsf{poly}(\lambda)$ but still greater than 1 (i.e. no gap).

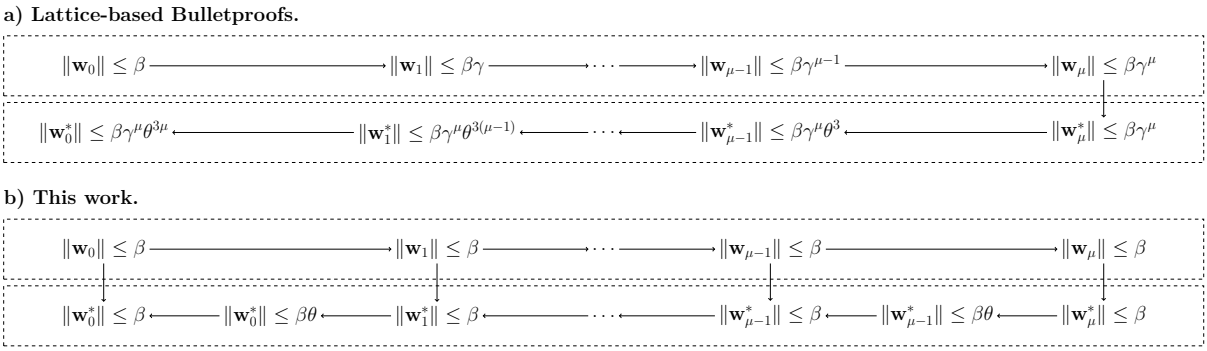

Fig. 2. Overview of the evolution of a prover witness $\mathbf{w}_0$ to an extracted witness $\mathbf{w}_0^*$ in lattice-based Bulletproofs and in this work.

The more challenging issue is that of the soundness gap, which refers[8] to the limitation that, in addition to the correctness gap β'/β, the witness produced by a knowledge extractor is of even larger norm $\beta^* \ggg \beta'$. Using the example of lattice-based Bulletproofs again, we have $\beta^* \approx (2\theta)^{3\mu}\beta' \approx (1+\gamma)^\mu(2\theta)^{3\mu}\beta$. Since no currently known subtractive set (including those suggested in Section 4) achieves $\theta = O(1)$, the soundness gap problem cannot be solved by simply using a different subtractive set, at least until more favourable sets are found.[9]

Figure 2 overviews the evolution of a prover witness $\mathbf{w}_0$ to an extracted witness $\mathbf{w}_0^*$ in lattice-based Bulletproofs and in this work.

Lattice-based Bulletproofs. In Fig. 2 part a) for Bulletproofs, each arrow in the top row represents one Bulletproofs folding step, where $\mathbf{w}_i$ denotes the intermediate witness after the i-th folding step. The norm of the i-th round prover witness $\mathbf{w}_i$ grows by a multiplicative factor of (around) γ compared to the previous round prover witness $\mathbf{w}_{i-1}$. The last round witness $\mathbf{w}_\mu$ is then of norm around $\beta\gamma^\mu$, i.e. with correctness gap γ^μ. The vertical arrow is trivial since the last-round prover witness is sent in plain, i.e. $\mathbf{w}_\mu^* = \mathbf{w}_\mu$. Each arrow in the bottom row represents a "traditional witness extraction step", i.e. moving one layer up in the tree-special soundness witness extraction, where $\mathbf{w}_i^*$ denotes the extracted witness at depth i. The norm of the i-th round extracted witness $\mathbf{w}_i^*$ grows by (roughly) a multiplicative factor of θ^3 compared to the previous round extracted witness $\mathbf{w}_{i-1}^*$. The final extracted witness $\mathbf{w}_0^*$ is then of norm around $\beta\gamma^\mu\theta^{3\mu}$, i.e. the soundness gap is $\gamma^\mu\theta^{3\mu}$.

Split-and-Fold and Norm-Check. We propose a modular approach to building a protocol which has no correctness and soundness gaps. The basis are atomic reductions of knowledge for handling different tasks. Before explaining our protocols, we need to look ahead and introduce our principal relation Ξ^{lin}.

Bird-eye view of principal relation. Recall that our the principal relation Ξ^{lin} consists of statements $(\mathbf{H}, \mathbf{F}, \mathbf{Y})$ and witnesses $\mathbf{W}$, all matrices over $\mathcal{R}$, which satisfy the relation

$$\mathbf{HFW} = \mathbf{Y} \pmod q \qquad \text{and} \qquad \|\mathbf{W}\| \le \beta$$

For simplicity, we first ignore the matrix $\mathbf{H}$ and treat it as the identity matrix. As noted above, the matrix $\mathbf{F}$ has a tensor structure, $\mathbf{F} = \mathbf{F}_{\mu-1} \bullet \ldots \bullet \mathbf{F}_0 \in \mathcal{R}_q^{n\times m}$ where $\mathbf{F}_i \in \mathcal{R}_q^{n\times d}$. The dimension $m \times r$ of the witness $\mathbf{W} \in \mathcal{R}_q^{m\times r}$ and its norm bound β are the pivotal measures our atomic protocols operate on. Note that the claim $\mathbf{FW} = \mathbf{Y}$ is equivalent to r claims $\mathbf{Fw}_i = \mathbf{y}_i$, for $i \in [r]$, where $\mathbf{W} = (\mathbf{w}_0, \ldots, \mathbf{w}_{r-1})$.

[8] In general, the soundness gap consists of a "stretch", i.e. increase in witness norm, and a "slack", i.e. a multiplicative approximation factor. Using a subtractive set, the slack can be eliminated.

[9] We believe that a slightly better but still super-polynomial soundness gap of $\beta^*/\beta' \approx (1+\gamma)^\mu(2\theta)^\mu$ can be achieved using a technique called "short-circuit extraction" [HKR19].

The atomic protocols. Next, we give a high-level overview of our atomic protocols. These are all reductions of knowledge, which reduce a claim $(\mathbf{H}, \mathbf{F}, \mathbf{Y})$ to another claim $(\widetilde{\mathbf{H}}, \widetilde{\mathbf{F}}, \widetilde{\mathbf{Y}})$ and witness $\mathbf{W}$ to $\widetilde{\mathbf{W}}$. Each protocol affects different parameters of the statement or witness.

Split. The purpose of the split protocol Π^{split} is to reduce the witness height m to m/d in exchange for growing the width r to rd. In other words, we reduce the dimension of the columns $\mathbf{w}_i$ by increasing the number of instances/columns. To achieve this, we use the row-wise tensor structure of $\mathbf{F}$ to factor it into $\mathbf{F} = \mathbf{R} \bullet \widetilde{\mathbf{F}}$, where $\bullet$ denotes row-wise tensoring. Decomposing $\mathbf{W}$ into $\sum_{i \in [d]} \mathbf{e}_i \otimes \mathbf{W}_i$, i.e. viewing $\mathbf{W}$ as a *vertical* stack of matrices $\mathbf{W}_i$ compatible with the tensor decomposition, we let $\widetilde{\mathbf{Y}}_j = \widetilde{\mathbf{F}}\mathbf{W}_j$, and $\widetilde{\mathbf{Y}} = (\widetilde{\mathbf{Y}}_0, \ldots, \widetilde{\mathbf{Y}}_{d-1})$ and the prover sends these cross terms. The reduced statement is then $(\widetilde{\mathbf{H}}, \widetilde{\mathbf{F}}, \widetilde{\mathbf{Y}})$ with witness $\widetilde{\mathbf{W}} = (\mathbf{W}_0, \ldots, \mathbf{W}_{d-1})$ the *horizontal* concatenation of the matrices.

Note that Π^{split} reshapes the dimensions of $\mathbf{W}$ as required. Moreover, the witness norm is left unchanged. Lastly, we note that handling the case where $\mathbf{H}$ (and thus $\widetilde{\mathbf{H}}$) is not the identity matrix slightly is more involved and explained in detail in Section 5.3.

Fold. The fold protocol Π^{fold} reduces the witness width r to r' (by random linear combining the columns). The protocol simply multiplies $\mathbf{W}$ and $\mathbf{Y}$ with a random short challenge matrix $\mathbf{C} \in \mathcal{R}_q^{r' \times r}$ from the left to get $\widetilde{\mathbf{W}} = \mathbf{W} \cdot \mathbf{C}$ and $\widetilde{\mathbf{Y}} = \mathbf{Y} \cdot \mathbf{C}$ and the new instance is $(\mathbf{H}, \mathbf{F}, \widetilde{\mathbf{Y}})$. Observe that the norm of the witness grows. Also note that the soundness of this step depends on the dimension r': The larger r' the shorter $\mathbf{C}$ can be. Hence, we can pick binary $\mathbf{C}$ (resp. roots of unity) to reduce norm growth at the expense of a wider $\widetilde{\mathbf{W}}$.

b-ary Decomposition. The b-ary decomposition protocol $\Pi^{b\text{-}\mathsf{decomp}}$ reduces the witness norm β by b-ary decomposing the matrix $\mathbf{W}$ as $\sum_{i=0}^{\ell-1} b^i \mathbf{V}_i$, at the expense of increased width $\widetilde{r}$ in the resulting $\widetilde{\mathbf{W}}$. The prover needs to communicate $\mathbf{Y}_i = \mathbf{H}\mathbf{F}\mathbf{V}_i$. The new witness is $\widetilde{\mathbf{W}} = (\mathbf{V}_0, \ldots, \mathbf{V}_{\ell-1})$ and the resulting $\widetilde{\mathbf{Y}}$ is $(\mathbf{Y}_0, \ldots, \mathbf{Y}_{\ell-1})$. This protocol is used to counteract the norm growth in Π^{fold} and eliminate the correctness gap.

Norm-Check and Inner Product. The norm-check protocol Π^{norm} ensures that the norm bound $\|\sigma(\mathbf{W})\|_2 \leq \beta$ holds, at the expense of slightly extending the witness by adding columns and constraints (i.e. increasing the width r and height n^{out} of $\mathbf{Y}$). All above protocols $(\Pi^{\mathsf{split}}, \Pi^{\mathsf{fold}}, \Pi^{b\text{-}\mathsf{decomp}})$ negatively affect the norm of the *extracted* witness. The norm check counteracts this, and ensures that the norm of the extracted witness is at most β. This eliminates the soundness gap.

The norm-check is implemented through the *inner product* protocol Π^{ip}, which proves that $t = \langle \mathbf{w}, \overline{\mathbf{w}} \rangle$, where $\overline{\mathbf{w}}$ denotes the complex conjugate. Given the inner product t, the canonical ℓ_2-norm $\|\sigma(\mathbf{w})\|_2$ of $\mathbf{w}$ satisfies $\|\sigma(\mathbf{w})\|_2^2 = \mathsf{Trace}(t)$, and thus, the norm-check can be implemented on top of the inner product by checking $\mathsf{Trace}(t) \leq \beta^2$. (This check is expanded to a matrix $\mathbf{W}$ column-wise; we leave details to Section 5.6.) To implement Π^{ip}, the prover encodes $\mathbf{w}$ as the coefficients of a polynomial $g(X)$, and commits to the coefficients of the Laurent polynomial $L(X) = g(X) \cdot \bar{g}(X^{-1})$, whose constant term is $\langle \mathbf{w}, \overline{\mathbf{w}} \rangle$. This reduces the problem to checking that L is computed correctly and has constant term t, both of which can be expressed as relations captured by Ξ^{lin}.

Two issues remain: First, the norm of the coefficients of $L(X)$ is around β^2 instead of β. To tackle this, we shrink the coefficients of $L(X)$ by immediately b-ary decomposing (for suitable b); we note that for technical reasons, we do not apply $\Pi^{b\text{-}\mathsf{decomp}}$ modularly here. We add this decomposition to $\mathbf{W}$, as well as additional rows to $\mathbf{F}$ for the new evaluation constraints of $L(X)$. Second, checking that L is computed correctly and has constant term t by introducing more constraints translates to higher communication costs when handled naively, namely, when $\mathbf{H}$ is always the identity. To tackle this, the parties run the batch protocol Π^{batch} to compress the newly added constraints with the existing ones. We explain this now.

Batch. As noted above, during the Π^{norm} protocol (and also the complete version of Π^{split}), new constraints (i.e. rows) are added to $\mathbf{F}$ and $\mathbf{Y}$, which increases the size of $\mathbf{Y}$ and thus the size of the cross terms communicated in our atomic protocols. To counteract this, our principal relation includes the matrix $\mathbf{H}$, which is will be of the form

$$\mathbf{H} = \begin{pmatrix} \mathbf{I}_{\overline{n}} & \mathbf{0}_{\overline{n} \times \underline{n}} \\ \underline{\mathbf{H}}_0 & \underline{\mathbf{H}}_{\underline{n^{\mathsf{out}}} \times \underline{n}} \end{pmatrix}$$

and which captures batch verification of the rows of $\mathbf{F}$: The identity block $\mathbf{I}_k$ ensures the vSIS instance in $\mathbf{F}$ is never compressed during batch verification (as this leads to technical problems), while the bottom rows $\underline{\mathbf{H}} = (\underline{\mathbf{H}}_0, \underline{\mathbf{H}}_1)$, where $\underline{\mathbf{H}}_0 = \mathbf{0}_{\underline{n^{\mathsf{out}}} \times \underline{n}}$, capture the current state of batch verification of the remaining rows of $\mathbf{F}$.

The batch protocol Π^{batch} reduces the height of $\mathbf{Y}$ by randomly linearly combining its bottom rows by left-multiplying with $\mathbf{C} = \begin{pmatrix} \mathbf{I}_{\overline{n}} & \mathbf{0}_{\overline{n} \times \underline{n}} \\ \underline{\mathbf{c}}_0^\intercal & \underline{\mathbf{c}}_1^\intercal \end{pmatrix}$ for a challenge vector $\mathbf{c} = (\underline{\mathbf{c}}_0, \underline{\mathbf{c}}_1)$. This yields $\widetilde{\mathbf{H}} = \mathbf{C} \cdot \mathbf{H}$ and $\widetilde{\mathbf{Y}} = \mathbf{C} \cdot \mathbf{Y}$ for the new instance, with $\mathbf{F}$ and $\mathbf{W}$ left unchanged. The protocol needs no prover communication, and has (almost) no effect on correctness and soundness gaps. Hence, it is applied whenever the height of $\mathbf{Y}$ is not minimal.

Composing the atomic protocols. In Section 6 we propose ways of composing the protocols with respect to asymptotic and concrete efficiency. The goal is to compose these atomic protocols to obtain succinct arguments for Ξ^{lin} without correctness and soundness gaps. We discuss the composition strategies, keeping track of parameter changes and communication costs to ensure that the security budget and norms remain within limits. Finally, an asymptotic complexity analysis shows how our proposed composition yields communication-efficient protocols while ensuring the hardness of the underlying cryptographic assumptions.

One suggested composition, which yields an easy-to-analyse (asymptotically) composition, is:

$$(\Pi^{\mathsf{norm}} \to \Pi^{\mathsf{batch}} \to \Pi^{b\text{-}\mathsf{decomp}} \to \Pi^{\mathsf{split}} \to \Pi^{\mathsf{fold}})_{i \in [\mu]} \to \Pi^{\mathsf{finish}},$$

where Π^{finish} introduces the trivial step of sending the witness in plain.

2.3 Embedding $\mathbb{Z}$-Inner Products

Lattice-based succinct arguments such as those constructed in Section 5 typically support proving relations over a ring $\mathcal{R}$ natively. However, in many applications, we would like to prove algebraic statements given over $\mathbb{Z}$, which motivates the question of how to reduce a statement over $\mathbb{Z}$ to statements over $\mathcal{R}$, so that a proof of the latter implies a proof of the former. Specifically, we consider the task of proving that some (committed) vectors $\mathbf{x}, \mathbf{y} \in \mathbb{Z}^{m\delta}$ satisfies $\langle \mathbf{x}, \mathbf{y} \rangle = z$ for some given $z \in \mathbb{Z}$. This task is of particular interest since, for some applications (e.g. constructing verifiable delay function [LM23]) it is necessary for the prover to prove that the witness is not only short but in fact binary. More generally, the application might require the prover to show a proof for $\mathbf{x} \in [a, b]^{m\delta}$ for some $a, b \in \mathbb{Z}$, which is not immediately implied by a bounded-norm guarantee.

To prove binariness, the basic idea is, for a witness $\mathbf{w} \in \mathbb{Z}^{m\delta}$, to use the equivalence $\mathbf{w} \in \{0,1\}^{m\delta} \iff \langle \mathbf{1}^{m\delta} - \mathbf{w}, \mathbf{w} \rangle_{\mathbb{Z}} = 0$ to reduce checking the binariness of $\mathbf{w}$ to checking that some transformed witness vector over $\mathcal{R}$ is short and satisfies some linear and inner-product relations, where $\mathcal{R} \subset \mathcal{O}_{\mathcal{K}}$ is of dimension $\delta \mid \varphi$ when viewed as $\mathbb{Z}$-modules.

Existing Embedding Methods. We are aware of three ways to embed $\mathbb{Z}$-inner products into $\mathcal{R}$-inner products in the literature, each with a significant drawback:

(i) Naive embedding: Interpret each $\mathbb{Z}$ element as an $\mathcal{R}$ element via the inclusion $\mathbb{Z} \subset \mathcal{R}$, and interpret the $\mathbb{Z}$-inner product as an $\mathcal{R}$-inner product. This incurs a multiplicative overhead of δ in terms of statement and witness sizes, which translate into overheads in prover and verifier computation, proof size, etc.

(ii) Coefficient embedding: Divide the witness into blocks containing δ $\mathbb{Z}$-elements, and encode each block as an $\mathcal{R}$ element via the (inverse-)coefficient embedding[10] $\psi^{-1} : \mathbb{Z}^{m\delta} \to \mathcal{R}^m$. For certain $\mathcal{R}$, we have

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{Z}} = \mathsf{ct}(\langle \psi^{-1}(\mathbf{x}), \overline{\psi^{-1}(\mathbf{y})} \rangle_{\mathcal{R}})$$

where $\mathsf{ct}(\cdot)$ denotes the constant term of the coefficient embedding.
This embedding has a convenient property that it is (somewhat) norm-preserving, i.e. $\mathbf{x}$ is short if and only if $\psi^{-1}(\mathbf{x})$ is also short (in both coefficient and canonical embedding). However, this approach only works for $\mathbb{Z}[\zeta_{2^d}]$. This is problematic since the largest subtractive set over $\mathbb{Z}[\zeta_{2^d}]$ is $\{0, 1\}$.

[10] For example, with respect to the power basis $\{1, \zeta, \ldots, \zeta^{\varphi-1}\}$ of a cyclotomic field, the coefficient embedding of an element $x = \sum_{i \in [\varphi]} x_i \zeta^i$ is denoted as $\psi(x) = (x_i)_{i \in [\varphi]}$.

(iii) CRT embedding: Let the witness vectors be such that $\mathbf{x}, \mathbf{y} \in \mathbb{Z}_p^{m\delta}$ for some (typically small) prime p which splits completely in $\mathcal{R}$. Divide the witness into blocks of δ $\mathbb{Z}$ elements, and encode each block as an $\mathcal{R}$ element via the (inverse-)CRT embedding $\mathsf{CRT}_p^{-1} : \mathbb{Z}_p^{m\delta} \to \mathcal{R}_p^m$. It holds that

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{Z}} = \left\langle \mathbf{1}^{\delta}, \mathsf{CRT}_p\left(\langle \mathsf{CRT}_p^{-1}(\mathbf{x}), \mathsf{CRT}_p^{-1}(\mathbf{y}) \rangle_{\mathcal{R}}\right) \right\rangle_{\mathbb{Z}} \bmod p.$$

This approach is powerful in that it not only supports proving about $\mathbb{Z}_p$-inner products, but in fact about $\mathbb{Z}_p$-Hadamard products $\mathbf{x} \odot \mathbf{y} \bmod p$, which is more fine-grained. However, to turn a claim about $\mathbb{Z}_p$-inner products into a claim about $\mathbb{Z}$-inner products (without reduction modulo p), we would additionally need to prove that $\|\langle \mathbf{x}, \mathbf{y} \rangle\|_{\infty} < p/2$, so that the reduction modulo p has no effect. Since CRT_p does not respect the geometry of $\mathbb{Z}$ and $\mathcal{R}$, this approach usually requires the prover to commit to the witness vectors in both the $\psi^{-1}(\cdot)$ and $\mathsf{CRT}_p^{-1}(\cdot)$ encodings, prove that the former is short, and prove that the two commitments are consistent. An issue here is that existing proofs of consistency between the two encodings (e.g. [BS23, LNS20]) do not have a succinct verifier, i.e. they run in time linear in the witness size.

In the following, we highlight how the aforementioned issues regarding the coefficient and CRT embeddings can be solved over certain (wide) range of rings.

Twisted Trace Maps. In Section 7, we generalise the coefficient embedding technique over power-of-2 rings to a wide range of other rings. Recall from the above that, over $\mathcal{O}_{\mathcal{K}}$ with a power-of-2 conductor, it holds that $\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{Z}} = \mathsf{ct}(\langle \psi^{-1}(\mathbf{x}), \overline{\psi^{-1}(\mathbf{y})} \rangle_{\mathcal{R}})$. In fact, the constant term function can be expressed as $\mathsf{ct}(\cdot) = \frac{1}{\varphi} \cdot \mathsf{Trace}_{\mathcal{K}/\mathbb{Q}}(\cdot)$ where $\mathsf{Trace}_{\mathcal{K}/\mathbb{Q}}$ denotes the field trace, and the power basis $\{1, \zeta, \ldots, \zeta^{\varphi-1}\}$ satisfies i.e. the power basis is orthogonal with respect to the field trace.

The above point of view motivates the search for ideal lattices with $\mathbb{Z}$-bases orthogonal with respect to the field trace. This leads us to the literature of lattice constellations. In particular, we extract the following embedding method from [BFOV04]: Over $\mathcal{O}_{\mathcal{K}^+}$ with prime conductor $\mathfrak{f}$, there exists an (efficiently computable) basis $\mathbf{b}^+ \in \mathcal{O}_{\mathcal{K}^+}^{\varphi/2}$ and a twist element $\alpha \in \mathcal{O}_{\mathcal{K}^+}$ such that

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbb{Z}} = \frac{1}{2\mathfrak{f}} \mathsf{Trace}_{\mathcal{K}/\mathbb{Q}}(\alpha \cdot \langle \psi_{\mathbf{b}^+}^{-1}(\mathbf{x}), \overline{\psi_{\mathbf{b}^+}^{-1}(\mathbf{y})} \rangle_{\mathcal{R}})$$

where $\psi_{\mathbf{b}^+} : \mathcal{O}_{\mathcal{K}^+} \to \mathbb{Z}^{\varphi/2}$ denotes the coefficient embedding with respect to the basis $\mathbf{b}^+$. Furthermore, adapting a result from the same work [BFOV04] regarding tensor products of rings, we extract similar embedding methods based on twisted trace maps for rings $\mathcal{R}$ of the form $\mathcal{R} = \mathcal{O}_{\mathcal{K}_{2^d}} \otimes \mathcal{O}_{\mathcal{K}_{p_0}^+} \otimes \ldots \otimes \mathcal{O}_{\mathcal{K}_{p_{k-1}}^+}$, where the subscripts of $\mathcal{K}$ denote the conductors the respective factor rings and $p_0, \ldots, p_{k-1}$ are distinct odd primes. This captures power-of-2 rings as a special case. Notably, since such $\mathcal{R}$ generally have non-prime-power conductors, they are compatible with the subtractive set for non-prime-power rings exposed in Section 4.

Succinct Proof for Consistency of CRT. As highlighted earlier, the missing piece, required to harness the power of the CRT embedding for Hadamard and inner products, is a verifier-succinct argument for proving the consistency between the coefficient embedding and the CRT embedding. More precisely, we need a succinct argument for proving that two ring vectors $\mathbf{w}, \mathbf{w}' \in \mathcal{R}^m$ satisfy

$$\psi(\mathbf{w}) = \mathsf{CRT}_p(\mathbf{w}') \bmod p. \tag{1}$$

In Section 8, we present a protocol for performing this task over $\mathcal{R} = \mathcal{O}_{\mathcal{K}}$ where the conductor $\mathfrak{f}$ is w-smooth, i.e. all its prime factors are at most some small integer w, with proof size and verifier time scaling linearly in $w^2 \log_w \mathfrak{f}$. In other words, if $w = O(1)$, then the complexity is logarithmic in $\mathfrak{f}$.

Underlying our protocol is the observation that, if the conductor $\mathfrak{f}$ is w-smooth, then the map $\psi^{-1} \circ \mathsf{CRT}_p$ can be expressed as the composition of $t \leq O(\log \mathfrak{f})$ maps, each being a linear combination of $h \leq w^2$ automorphisms from $\mathsf{Gal}(\mathcal{K}/\mathbb{Q})$ with coefficients lying in $\mathcal{R}$. This means that, to succinctly prove that $\mathbf{w} = \psi^{-1}(\mathsf{CRT}_p(\mathbf{w}'))$ mod p, it suffices to design a succinct argument for proving automorphism relations.

Motivated by the above, we present a succinct reduction of knowledge from checking $\alpha(\mathbf{w}) = \mathbf{w}'$ to checking that $(\mathbf{w}, \mathbf{w}')$ satisfies some linear relations. We obtain a succinct argument for proving Eq. (1).

3 Preliminaries

Let $\mathbb{N} = \{1, 2, \dots\}$ denotes natural numbers and $\lambda \in \mathbb{N}$ be the security parameter. For $n \in \mathbb{N}$, we write $[n] \coloneqq \{0, \dots, n-1\}$ counting from 0. For multidimensional ranges, we use the shorthand $(i, j, k) \in [n, m, \ell]$ for $i \in [n]$, $j \in [m]$, and $k \in [\ell]$.

Throughout this work, we let $\mathcal{K} = \mathbb{Q}(\zeta)$ be a cyclotomic field with conductor $\mathfrak{f}$ of degree $\varphi = \varphi(\mathfrak{f})$, where ζ is a root of unity of order $\mathfrak{f}$ and φ is Euler's totient function, and $\mathcal{O}_\mathcal{K} = \mathbb{Z}[\zeta]$ be its ring of integers. We will also consider the maximal real subfield $\mathcal{K}^+ = \mathbb{Q}(\zeta + \zeta^{-1})$ of $\mathcal{K}$ and its ring of integers $\mathcal{O}_{\mathcal{K}^+} = \mathbb{Z}[\zeta + \zeta^{-1}]$. In contexts where we refer to multiple cyclotomic fields with different conductors $(\mathfrak{f}_i)_{i\in[k]}$, we write $\mathcal{K}_{\mathfrak{f}_i}$ for $i \in [k]$ to emphasise the conductors. We will usually use $\mathcal{R} \subseteq \mathcal{O}_\mathcal{K}$ to denote a subring which has dimension δ when viewed as a $\mathbb{Z}$-module.

For a modulus $q \in \mathbb{N}$, we write $\mathcal{R}_q \coloneqq \mathcal{R}/q\mathcal{R}$. We denote by $\mathcal{R}^\times$ and $\mathcal{R}_q^\times$ the sets of units in $\mathcal{R}$ and $\mathcal{R}_q$ respectively. We endow $\mathcal{R}$ with two geometries via the coefficient embedding $\psi_\mathbf{b} : \mathcal{R} \to \mathbb{Z}^\delta$ (for a given basis $\mathbf{b}$) and the canonical embedding $\sigma : \mathcal{K} \to \mathbb{C}^\varphi$ (of $\mathcal{K}$). Specifically, for a given $\mathbb{Z}$-basis $\mathbf{b} = (b_i)_{i\in[\delta]}$ of $\mathcal{R}$ and an element $x = \sum_{i\in[\delta]} x_i b_i \in \mathcal{R}$, we write

$$\psi_\mathbf{b}(x) \coloneqq (x_i)_{i\in[\delta]} \qquad \text{and} \qquad \sigma(x) \coloneqq (\sigma_j(x))_{j\in[\varphi]}$$

where $\sigma_j \in \mathsf{Gal}(\mathcal{K}/\mathbb{Q})$. Note that we define $\sigma(x)$ by treating $x \in \mathcal{K}$ in order to avoid discussing the canonical embedding of subfields of $\mathcal{K}$. If $\mathcal{R} = \mathcal{O}_\mathcal{K}$ and is the standard powerful basis, we may omit $\mathbf{b}$ from the subscript of $\psi_\mathbf{b}$. We define powerful basis as

$$\mathbf{b} = (1, \zeta, \dots, \zeta^{\varphi-1})$$

for prime-power conductor $\mathfrak{f}$. The basis generalises to the composite conductor $\mathfrak{f} = \prod_{i\in[k]} \mathfrak{f}_i^{e_i}$ for prime $\mathfrak{f}_i$ via tensor product,

$$\mathbf{b} = \bigotimes_{i\in[k]} \left(1, \zeta_{\mathfrak{f}_i}, \dots, \zeta_{\mathfrak{f}_i}^{\varphi(\mathfrak{f}_i)-1}\right).$$

We extend the notation of $\psi_\mathbf{b}$ and σ naturally to vectors, i.e. if $\mathbf{x} = (x_i)_{i\in[m]} \in \mathcal{R}^m$, then

$$\psi_\mathbf{b}(\mathbf{x}) \coloneqq (\psi_\mathbf{b}(x_i))_{i\in[\delta]} \qquad \text{and} \qquad \sigma(\mathbf{x}) \coloneqq (\sigma_j(x_i))_{j\in[\varphi]}$$

are defined as concatenations.

For any $p \in \mathbb{N}$, we consider the balanced representation of $\mathbb{Z}_p$, i.e. elements are represented by $[-p/2, p/2) \cap \mathbb{Z}$. When considering the quotient ring $\mathcal{R}_p \coloneqq \mathcal{R}/p\mathcal{R}$ where $\mathcal{R}$ has $\mathbb{Z}$-basis $\mathbf{b}$, we assume that an element $x \in \mathcal{R}_p$ is represented by $\psi_\mathbf{b}(x) \in ([-p/2, p/2) \cap \mathbb{Z})^\varphi$. As such, for any $x \in \mathcal{R}$, we abuse the notation $x \in \mathcal{R}_p$ to mean that $\psi(x) \in ([-p/2, p/2) \cap \mathbb{Z})^\varphi$. The above extends naturally to vectors over $\mathcal{R}$.

To distinguish between $\mathbb{Z}$-inner products and $\mathcal{R}$-inner products, we write $\langle \mathbf{x}, \mathbf{y} \rangle_\mathbb{Z} = \sum_{i\in[m]} x_i y_i$ or $\langle \mathbf{x}, \mathbf{y} \rangle_\mathcal{R} = \sum_{i\in[m]} x_i y_i$ depending on whether $\mathbf{x}, \mathbf{y} \in \mathbb{Z}^m$ or $\mathbf{x}, \mathbf{y} \in \mathcal{R}^m$. Note that $\langle \mathbf{x}, \mathbf{y} \rangle_\mathcal{R}$ is defined without complex conjugation.

For any Galois extension $\mathcal{M}/\mathcal{L}$, the field trace can be computed as $\mathsf{Trace}_{\mathcal{M}/\mathcal{L}} : \mathcal{K} \to \mathcal{L}$, $\mathsf{Trace}_{\mathcal{M}/\mathcal{L}}(x) \coloneqq \sum_{\sigma_j \in \mathsf{Gal}(\mathcal{K}/\mathcal{L})} \sigma_j(x)$. When $\mathcal{L} = \mathbb{Q}$, we drop the subscript and write $\mathsf{Trace} = \mathsf{Trace}_{\mathcal{M}/\mathbb{Q}}$.

The coefficient ℓ_p-norm and canonical ℓ_p-norm of a vector $\mathbf{x} \in \mathcal{R}^m$ is denoted by $\|\psi(\mathbf{x})\|_p$ and $\|\sigma(\mathbf{x})\|_p$ respectively. We will mostly use $\|\psi(\cdot)\|_\infty$ and $\|\sigma(\cdot)\|_2$. For matrices, the norm is defined as $\|\mathbf{M}\| = \|\mathsf{vec}(\mathbf{M})\|$ for all norms, where $\mathsf{vec}(\cdot)$ denotes vectorisation, i.e. rearranging the elements of the matrix into a vector. In the context of 2-norms, such norm is called "Frobenius norm". The ring expansion factor of $\mathcal{R}$ w.r.t. the coefficient ℓ_∞-norm is defined as $\gamma_\mathcal{R} \coloneqq \max_{a,b\in\mathcal{R}} \|\psi(a \cdot b)\|_\infty / (\|\psi(a)\|_\infty \cdot \|\psi(b)\|_\infty)$. Assuming balanced representation, for any $x \in \mathcal{R}_p$, we have $\|\psi(x)\|_\infty \leq p/2$. Note that $\|\sigma(\mathbf{x})\|_2^2 = \mathsf{Trace}(\mathbf{x}^\mathsf{T}\overline{\mathbf{x}})$, where $\bar{\cdot}$ denotes the complex conjugate.

For horizontal and vertical concatenation of matrices, we write respectively:

$$(\mathbf{M}_i)_{i\in[\ell]} \qquad \text{and} \qquad \overbrace{\mathbf{M}_i}^{i\in[\ell]} \quad \left(\text{or } \sum_{i\in[\ell]} \mathbf{e}_i \otimes \mathbf{M}_i\right).$$

3.1 Cryptographic Assumption

We state an equivalent formulation of the vanishing short integer solution (vSIS) assumption [CLM23], which has a simpler description and better aligns with the notation adopted in this work. For more discussion on vSIS, we refer to Section A.1.

Definition 1 (vSIS Assumption (adapted from [CLM23])). *Let* $\mathsf{params} = (\mathcal{R}, q, \beta, \chi)$ *be parametrised by* λ*, where* $\mathcal{R}$ *is a ring,* $q \in \mathbb{N}$ *a modulus,* $\beta > 0$ *a norm bound, and* χ *a distribution over* $\mathcal{R}_q^{n \times \otimes_{i \in [\mu]} d_i}$ *for some dimensions* $n, d_0, \ldots, d_{\mu-1}, \mu \in \mathbb{N}$*. The* $\mathsf{vSIS}_{\mathsf{params}}$ *assumption states that, for any PPT adversary* $\mathcal{A}$*, the advantage function satisfies*

$$\mathsf{Adv}^{\mathsf{vsis}}_{\mathsf{params},\mathcal{A}}(\lambda) \coloneqq \Pr\left[\begin{array}{l|l} \mathbf{Fw} = \mathbf{0} \bmod q & \mathbf{F} \leftarrow\!\$\ \chi \\ \|\sigma(\mathbf{w})\|_2 \leq \beta & \mathbf{w} \leftarrow \mathcal{A}(\mathbf{F}) \end{array}\right] \leq \mathsf{negl}(\lambda).$$

For simplicity, in this work, we will consider the setting where the block sizes $d_0, \ldots, d_{\mu-1}$ are identically set to some $d \in \mathbb{N}$, so that $\mathbf{F}$ can be factored into $\mathbf{F} = \mathbf{F}_{\mu-1} \bullet \ldots \bullet \mathbf{F}_0$ with $\mathbf{F}_i \in \mathcal{R}_q^{n \times d}$, where $\bullet$ denotes the row-wise tensor product.

3.2 Reduction of Knowledge

In this paper we consider ternary relations $\Xi \subseteq \{0,1\}^* \times \{0,1\}^* \times \{0,1\}^*$, where a tuple $(\mathsf{pp}, \mathsf{stmt}, \mathsf{wit}) \in \Xi$ consists of public parameters pp, statement stmt and witness wit. For presentation, we omit including pp when it is known from the context. We consider a modified and simplified definition of a reduction of knowledge [KP23] for the following reasons: All of our protocols are *public coin* and *(coordinate-wise) special sound* [FMN23] or similar.[11] Thus, public reducibility is automatic and we have (super-constant) sequential composition results due to known (tree) black-box extractors, whereas composition in [KP23] is limited a constant number of protocols. Lastly, we define a *relaxed* knowledge soundness notion which is not present in [KP23]. For lack of space, we provide a condensed overview of reductions of knowledge. See Section A.2 for details.

Definition 2 (Reduction of Knowledge (modified)). *Let* Ξ_0*,* Ξ_1 *be ternary relations. A* reduction of knowledge (RoK) Π *from* Ξ_0 *to* Ξ_1*, short* $\Pi \colon \Xi_0 \to \Xi_1$*, is defined by two PPT algorithms* $\Pi = (\mathcal{P}, \mathcal{V})$*, the prover* $\mathcal{P}$ *, and the verifier* $\mathcal{V}$*, with the following interface:*

- $\mathcal{P}(\mathsf{pp}, \mathsf{stmt}_1, \mathsf{wit}_1) \to (\mathsf{stmt}_2, \mathsf{wit}_2)$*: Interactively reduce the input statement* $(\mathsf{pp}, \mathsf{stmt}, \mathsf{wit}) \in \Xi_0$ *to a new statement* $(\mathsf{pp}, \widetilde{\mathsf{stmt}}, \widetilde{\mathsf{wit}}) \in \Xi_1$ *or* $\perp$*.*
- $\mathcal{V}(\mathsf{pp}, \mathsf{stmt}) \to \widetilde{\mathsf{stmt}}$*: Interactively reduce the task of checking the input statement* $(\mathsf{pp}, \mathsf{stmt})$ *w.r.t* Ξ_0 *to checking a new statement* $(\mathsf{pp}, \widetilde{\mathsf{stmt}})$ *w.r.t.* Ξ_1*.*

A RoK Π is *correct*, if for any honest protocol run (with correct inputs), the prover outputs a witness for the reduced statement (which the verifier outputs). A RoK Π is *relaxed knowledge sound* from Ξ_0^{KS} to Ξ_1^{KS} with knowledge error $\kappa(\mathsf{pp}, \mathsf{stmt})$ if there is a *black-box* expected polynomial-time extractor $\mathcal{E}$, which succeeds with probability $\epsilon - \kappa(\mathsf{pp}, \mathsf{stmt})$ if the malicious prover outputs a valid witness for the reduced statement with probability ϵ (on verifier's input $(\mathsf{pp}, \mathsf{stmt})$).

Lemma 1 (Relations between norms (derived from [LPR13] and [DPSZ12, DPSZ12])). *Let* $x \in \mathcal{K} = \mathbb{Q}(\zeta_{\mathfrak{f}})$ *and* $\varphi = \varphi(\mathfrak{f})$*. Let* $\hat{\mathfrak{f}}$ *be* $\mathfrak{f}$ *if* $\mathfrak{f}$ *is odd and* $\mathfrak{f}/2$ *if it is even; let* $\mathsf{rad}(\mathfrak{f})$ *be the radical (i.e. the product of all primes dividing* $\mathfrak{f}$*). Let* $\sigma \colon \mathcal{K} \to \mathbb{R}^{\varphi}$ *be the canonical embedding and let* $\psi \colon \mathcal{K} \to \mathbb{R}^{\varphi}$ *be the coefficient embedding w.r.t. the powerful basis. Then we have*

(i) $\|\psi(x)\|_2 \leq \sqrt{\frac{\mathsf{rad}(\mathfrak{f})}{\mathfrak{f}}} \|\sigma(x)\|_2$

(ii) $\|\sigma(x)\|_2 \leq \sqrt{\hat{\mathfrak{f}}} \cdot \|\psi(x)\|_2$

(iii) $\|\sigma(x)\|_\infty \leq \varphi \cdot \|\psi(x)\|_\infty$

[11] To turn soundness errors of probabilistic tests (such as Schwartz–Zippel) into knowledge errors, we merely need uniformly random transcripts. These are produced by (CW)SS extractors for example. We call such extractors k-transcript extractors.

(iv) $\|\psi(x)\|_\infty \le \mathfrak{c}_{\mathsf{rad}(\mathfrak{f})} \cdot \|\sigma(x)\|_\infty$ *where* $\mathfrak{c}_j$ *is a constant such that* $\mathfrak{c}_{\mathsf{rad}(\mathfrak{f})} \le (4/\pi)^\ell$*, where* ℓ *is the number of different odd prime factors in* $\mathsf{rad}(\mathfrak{f})$*. Moreover,* $\mathfrak{c}_{f_1 \cdot f_2} = \mathfrak{c}_{f_1} \cdot \mathfrak{c}_{f_2}$ *for coprime* f_1 *and* f_2 *and* $\mathfrak{c}_{2^e} = 1$*.*

We note that the constants $\mathfrak{c}_\mathfrak{f}$ are quite small in practice: For all $\mathfrak{f} \le 255254$ we have $\mathfrak{c}_\mathfrak{f} \le (4/\pi)^5 \le 3.35$. Because ℓ is the number of odd prime factors in $\mathfrak{f}$, we find that, up to $\mathfrak{f} \le 1154 = 3 \cdot 5 \cdot 7 \cdot 11 - 1$ we have $\mathfrak{c}_\mathfrak{f} \le (4/\pi)^3 \le 2.065$; and up to $\mathfrak{f} \le 15014 = 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 - 1$ we have $\mathfrak{c}_\mathfrak{f} \le (4/\pi)^4 \le 2.63$; and so on.

Proof. The relations follow essentially from bounds in [LPR13] and [DPSZ12]. The crucial piece is the powerful basis $\mathbf{b}$ and the canonica embedding matrix CRT, which is defined as

$$\sigma(\mathbf{b}) \coloneqq (\sigma_1(\mathbf{b}), \dots, \sigma_\varphi(\mathbf{b}))$$

i.e. the columns are different canonical embeddings. In this basis, we have

$$\forall x \in \mathcal{K}: \quad \sigma(x) = \sigma(\mathbf{b}) \cdot \psi(x)$$

Moreover, by [LPR13, Lemma 4.3], we have for the singular values of $\sigma(\mathbf{b})$

$$s_{min}(\sigma(\mathbf{b})) = \sqrt{\mathfrak{f}/\mathsf{rad}(\mathfrak{f})} \quad \text{and} \quad s_{max}(\sigma(\mathbf{b})) = \sqrt{\hat{\mathfrak{f}}}$$

where $\hat{\mathfrak{f}} = \mathfrak{f}$ if odd, else $\mathfrak{f}/2$. With this, we can prove the claims.

The first point follows from

$$\|\sigma(x)\|_2^2 = \|\sigma(\mathbf{b})\psi(x)\|_2^2 \ge s_{min}(\sigma(\mathbf{b}))^2 \cdot \|\psi(x)\|_2^2$$

where taking the square root yields the claim.[12]

The second point is immediate from

$$\|\sigma(x)\|_2 = \|\sigma(\mathbf{b})\psi_2(x)\|_2 \le s_{\max}(\sigma(\mathbf{b}))\|\psi(x)\|_2$$

by bounds for the operator norm of $\sigma(\mathbf{b})$.

The third point follows from

$$\|\sigma(x)\|_\infty = \|\sigma(\mathbf{b})\psi(x)\|_\infty \le \|\sigma(\mathbf{b})\|_{\mathsf{op},\infty}\|\psi(x)\|_\infty$$

The operator norm $\|\sigma(\mathbf{b})\|_{\mathsf{op},\infty}$ w.r.t. ∞-norm is row-wise 1-norm of $\sigma(\mathbf{b})$, which is yields exactly φ.

The last point is a consequence of [DPSZ12, Lemma 4 and 5], which applied to

$$\|\psi(x)\|_\infty = \|\sigma(\mathbf{b})^{-1}\sigma(\mathbf{b})\psi(x)\|_\infty = \|\sigma(\mathbf{b})^{-1}\|_{\mathsf{op},\infty}\|\sigma(x)\|_\infty$$

shows that $\|\sigma(\mathbf{b})^{-1}\|_{\mathsf{op},\infty} \le \mathfrak{c}_\mathfrak{f}$ for a family of constants which satisfies $\mathfrak{c}_{p^e} = \mathfrak{c}_p$ for prime powers $p \ne 2$ (and $\mathfrak{c}_{2^e} = 1$), and $\mathfrak{c}_{mn} \le \mathfrak{c}_m\mathfrak{c}_n$ for coprime m, n, and[13]

$$\mathfrak{c}_p = \frac{1}{p}\sum_{r=1}^{p-1} 2\sin(r\pi/p) = \frac{2 \cdot \sin(\pi/p)}{p \cdot (1 - \cos(\pi/p))} \le 4/\pi.$$

From this we deduce that given ℓ odd prime factors in $\mathfrak{f}$, we have $\mathfrak{c}_\mathfrak{f} \le (4/\pi)^\ell$. (The claim $\mathfrak{c}_{mn} \le \mathfrak{c}_m\mathfrak{c}_n$ for coprime m, n is not shown explicitly in [DPSZ12], but is a direct consequence of the tensor decomposition of CRT and $\|\mathbf{A} \otimes \mathbf{B}\|_\infty = \|\mathbf{A}\|_\infty \cdot \|\mathbf{B}\|_\infty$ for any complex matrices $\mathbf{A}, \mathbf{B}$.) □

The following corollary is immediate from Lemma 1.

Corollary 1. *Let* $x \in \mathcal{K}$*. It holds that* $\|\sigma(x)\|_2 \le \sqrt{\hat{\mathfrak{f}}\varphi}\,\|\psi(x)\|_\infty$ *and* $\|\psi(\mathbf{x})\|_\infty \le \|\sigma(\mathbf{x})\|_2$*.*

[12] We note that the inequality follows by expressing terms as inner products, using SVD decomposition to cancel U in $U\Sigma V^*$, and then obvious inequality for a diagonal $D \ge 0$ and $\langle Dz, Dz\rangle$ with $z = V^*\psi(x)$, and finally using that V^* is unitary, so can be removed in the norm.

[13] There is a sign error in [DPSZ12, Lemma 5], which (incorrectly) states $\frac{1}{p}\sum_{r=1}^{p-1} 2\sin(r\pi/p) = \frac{2\cdot\sin(\pi/p)}{p\cdot(\cos(\pi/p)-1)}$, but which does not affect the final conclusion. This is fixed here.

4 Subtractive Sets

A subtractive set S over a ring $\mathcal{R}$ is such that $c - c'$ is a unit for any distinct $c, c' \in S$. While the notion is connected to the study of Euclidean number fields [Len76], it also found applications in lattice-based cryptography in the contexts of argument systems and secret sharing [AL21]. For a cyclotomic ring $\mathcal{R}$ with prime-power conductor $\mathfrak{f} = p^k$, an explicit construction of upper-bound-matching cardinality p is known. For other cyclotomic rings and their subrings, however, not much seem to be explicitly studied. In this section, we construct subtractive sets over non-prime-power cyclotomic rings, as well as *real* cyclotomic rings.

Definition 3 (Subtractive Set). *We say that a set $S \subseteq \mathcal{R}$ is subtractive over $\mathcal{R}$ if $c - c' \in \mathcal{R}^\times$ for any distinct $c, c' \in S$.*

While [AL21] measured the quality of a subtractive set over cyclotomic rings in terms of the ℓ_∞-norm over the coefficient embedding, in this work, we will instead work with the ℓ_∞-norm over the canonical embedding for compatibility with Section 5 via the inequality $\forall\, c, x \in \mathcal{R},\ \|\sigma(c \cdot x)\|_2 \leq \|\sigma(c)\|_\infty \cdot \|\sigma(x)\|_2$. We measure the quality of a subtractive set by its cardinality, expansion factor $\gamma_{\|\sigma(\cdot)\|_2, S}$, and inverse-expansion factor $\theta_{\|\sigma(\cdot)\|_2, S}$, with the latter two defined below.

Definition 4 ((Inverse-)Expansion Factor of Subtractive Set). *Let $S \subseteq \mathcal{R}$ be subtractive over $\mathcal{R}$. The expansion and inverse-expansion factors of S are $\gamma_S \coloneqq \max_{c \in S, t \in \mathcal{R}, t \neq 0} \|t \cdot c\| / \|t\|$ and $\theta_S \coloneqq \max_{c, c' \in S, c \neq c', t \in \mathcal{R}, , t \neq 0} \|t \frac{1}{c - c'}\| / \|t\|$ respectively.*

To distiguish between canonical 2-norm and coefficient ∞-norm, we use $\gamma_{\|\sigma(\cdot)\|_2, S}$, $\gamma_{\|\psi(\cdot)\|_\infty, S}$, $\theta_{\|\sigma(\cdot)\|_2, S}$ and $\theta_{\|\psi(\cdot)\|_\infty, S}$. Recall that $\|\sigma(cy)\|_2 \leq \|\sigma(c)\|_\infty \|\sigma(y)\|_2$ for $x, y \in \mathcal{R}$, and thus $\|\sigma(c)\|_\infty$ is (a bound on) the expansion factor of x w.r.t. canonical (2-)norm. The following lemma often is handy for analysing inverse-expansion factors.

Lemma 2. *Let $K = \mathbb{Q}(\zeta)$ with ζ a primitive $\mathfrak{f}$-th root of unity such that $\mathfrak{f} \geq 4$. It holds that $\|\sigma\left(\frac{1}{1-\zeta}\right)\|_\infty \leq \frac{\mathfrak{f}}{4\sqrt{2}}$. Furthermore, if ζ is a (not necessarily primitive) k-th root of unity, i.e. $\zeta^k = 1$ and $k \in \mathbb{N}$ is minimal, then $\|\sigma\left(\frac{1}{1-\zeta^i}\right)\|_\infty \leq \frac{k}{4\sqrt{2}}$*

Proof. By the definition of $\|\sigma(\cdot)\|_\infty$, we need to upper bound $\max_{\sigma_j} \left|\sigma_j\left(\frac{1}{1-\zeta}\right)\right| = \max_{\sigma_j} \left|\frac{1}{1-\sigma_j(\zeta)}\right|$, where σ_j ranges from $\mathsf{Gal}(\mathcal{K}/\mathbb{Q})$. Since $\sigma_j(\zeta)$ ranges over all primitive $\mathfrak{f}$-th root of unity, this is the same as $\max_{j \in \mathbb{Z}_\mathfrak{f}^\times} \left|\frac{1}{1-\zeta^j}\right| = \max_{j \in \mathbb{Z}_\mathfrak{f}^\times} \left|\frac{1}{1-e^{j \cdot 2\pi i/\mathfrak{f}}}\right|$. Thus, it suffices to lower-bound $\left|1 - e^{j \cdot 2\pi i/\mathfrak{f}}\right|$ over $j \in \mathbb{Z}_\mathfrak{f}^\times$. Geometrically, $e^{j \cdot 2\pi i/\mathfrak{f}}$ are points on the unit circle in the complex plane with angles incremented by $2\pi j/\mathfrak{f}$. Thus, the value is approximately $\left|1 - e^{j \cdot 2\pi 1/\mathfrak{f}}\right| \approx 2\pi j/\mathfrak{f}$ for small $2\pi j/\mathfrak{f}$. For an explicit bound, observe that for $\alpha \leq \frac{1}{4}$ we have $\left|1 - e^{\alpha 2\pi i}\right| = 2 \cdot \sin(\frac{2\pi\alpha}{2}) \geq \alpha \cdot 4\sqrt{2}$. Setting $\alpha = \frac{1}{\mathfrak{f}}$ proves the claim. Observe that the above argument only depends on the multiplicative order of ζ, thus, the claim $\|\sigma\left(\frac{1}{1-\zeta}\right)\|_\infty \leq \frac{k}{2}$ follows for any (not necessarily primitive) k-th root of unity ζ. □

Corollary 2 (Field expansion factor $\gamma_{\|\psi(\cdot)\|_\infty, \mathcal{K}}$). *Let $\mathcal{K} = \mathbb{Q}(\zeta)$ with ζ a primitive $\mathfrak{f}$-th root of unity, Let $S \subseteq K$ be the powerful basis w.r.t. ζ of K. Then for all $x, y \in \mathcal{K}$, we have $\gamma_{\|\psi(\cdot)\|_\infty, \mathcal{K}} \leq \mathfrak{c}_{\mathsf{rad}(\mathfrak{f})} \varphi \|\sigma(x)\|_\infty$ because*

$$\|\psi(xy)\|_\infty \leq \mathfrak{c}_{\mathsf{rad}(\mathfrak{f})} \varphi \|\sigma(x)\|_\infty \|\psi(y)\|_\infty$$

Proof. The claim follows immediately from Lemma 1, because

$$\|\psi(xy)\|_\infty \leq \mathfrak{c}_{\mathsf{rad}(\mathfrak{f})} \|\sigma(xy)\|_\infty \leq \mathfrak{c}_{\mathsf{rad}(\mathfrak{f})} \|\sigma(x)\|_\infty \|\sigma(y)\|_\infty \leq \mathfrak{c}_{\mathsf{rad}(\mathfrak{f})} \varphi \|\sigma(x)\|_\infty \|\psi(y)\|_\infty$$

holds for $x, y \in \mathcal{K}$.

4.1 Prime-Power Cyclotomics

We recall the subtractive set for prime-power cyclotomics [Len76, AL21] with conductor $\mathfrak{f} = p^k$ and analyse its (inverse-)expansion factor in canonical ℓ_2-norm. Although we are interested mostly in $p \gg 2$, the result also holds for $p = 2$.

Theorem 1. *Let $\mathfrak{f} = p^k > 4$ for some prime p. The set $S := \{\mu_0, \ldots, \mu_{p-1}\} \subseteq_p \mathcal{O}_\mathcal{K}$ is subtractive, where $\mu_i = (\zeta^i - 1)/(\zeta - 1)$. Further, $\gamma_{\|\sigma(\cdot)\|_2,S} \leq p$, $\theta_{\|\sigma(\cdot)\|_2,S} \leq \frac{\mathfrak{f}}{2\sqrt{2}}$, and $\gamma_{\|\psi(\cdot)\|_\infty,S} \leq \varphi$ and $\theta_{\|\psi(\cdot)\|_\infty,S} \leq \varphi$, where $\varphi = \varphi(\mathfrak{f})$.*

Proof. Let $i < j \in [p]$. Observe that $\mu_j - \mu_i = \zeta^i + \zeta^{i+1} + \ldots + \zeta^{j-1} = \zeta^i \cdot \frac{\zeta^{j-i}-1}{\zeta-1}$ which is clearly a unit in $\mathcal{R}$, hence S is subtractive.

For the canonical 2-norm expansion factor, note that μ_i is a sum of i roots of unity and $i < p$. Therefore $\gamma_{\|\sigma(\cdot)\|_2,S} = \max_{i\in[p]} \|\sigma(\mu_i)\|_\infty < p$. For the inverse-expansion factor, observe that

$$\|\sigma\left(\frac{1}{\mu_j - \mu_i}\right)\|_\infty = \|\sigma\left(\zeta^{-i} \cdot \frac{\zeta - 1}{\zeta^{j-i} - 1}\right)\|_\infty \leq \|\sigma(\zeta - 1)\|_\infty \cdot \|\sigma\left(\frac{1}{\zeta^{j-i} - 1}\right)\|_\infty \leq 2\|\sigma\left(\frac{1}{\zeta^{j-i} - 1}\right)\|_\infty \leq \frac{\mathfrak{f}}{2\sqrt{2}},$$

where the last inequality follows from Lemma 2 and the rest are elementary.

For the coefficient ∞-norm expansion factor, we recall results of [AL21] that $\|\psi(1/\mu_i - \mu_j)\|_\infty \leq 1$. Therefore, $\gamma_{\|\psi(\cdot)\|_\infty,S} \leq \varphi$ and $\theta_{\|\psi(\cdot)\|_\infty,S} \leq \varphi$. □

4.2 Non-Prime-Power Cyclotomics

A drawback of the subtractive set recalled above is its rather large expansion factor $\gamma_{\|\sigma(\cdot)\|_2,S} \leq p$. In some applications, e.g. Section 5, we would like $\gamma_{\|\sigma(\cdot)\|_2,S}$ to be constant. Below, we expose a subtractive set over non-prime-power cyclotomic rings with very small expansion factor.

Theorem 2. *Let $\mathfrak{f}$ factor into $k \geq 2$ coprime prime-power factors $(\hat{\mathfrak{f}}_i)_{i\in[k]}$, i.e. $\mathfrak{f} = \prod_{i\in[k]} \hat{\mathfrak{f}}_i$. Write $\hat{\mathfrak{f}}_{\max} := \max_{i\in[k]} \hat{\mathfrak{f}}_i$. The set $S := \left\{1, \zeta, \zeta^2, \ldots, \zeta^{\mathfrak{f}/\hat{\mathfrak{f}}_{\max}-1}\right\} \subseteq_{\mathfrak{f}/\mathfrak{f}_{\max}} \mathcal{O}_\mathcal{K}$, is subtractive. Furthermore, $\gamma_{\|\sigma(\cdot)\|_2,S} = 1$ and $\theta_{\|\sigma(\cdot)\|_2,S} \leq \frac{\mathfrak{f}}{4\sqrt{2}}$, and $\gamma_{\|\psi(\cdot)\|_\infty,S} \leq \mathfrak{c}_{\mathsf{rad}(f)}\varphi$ and $\theta_{\|\psi(\cdot)\|_\infty,S} \leq \mathfrak{c}_{\mathsf{rad}(f)}\varphi$.*

To prove Theorem 2, we begin with the following lemma which we believe should be well-established together with a suppostive proposition. Since we could not find an explicit reference to the lemma, we provide a proof.

Lemma 3. *Let $\mathcal{R} = \mathbb{Z}[\zeta_\mathfrak{f}]$ with a conductor $\mathfrak{f}$ having $k \geq 2$ coprime prime-power factors[14] $(\hat{\mathfrak{f}}_i)_{i\in[k]}$, i.e. $\mathfrak{f} = \prod_{i\in[k]} \hat{\mathfrak{f}}_i$. Write $\hat{\mathfrak{f}}_{\max} := \max_{i\in[k]} \hat{\mathfrak{f}}_i$. For $j \in \left\{1, 2, \ldots, \frac{\mathfrak{f}}{\mathfrak{f}_{\max}} - 1\right\}$, it holds that $1 - \zeta^j \in \mathcal{R}^\times$.*

Proof. Write $\zeta = \zeta_\mathfrak{f}$. First, consider the case when ζ^j is a primitive $\mathfrak{f}$-th root of unity. Then, by Proposition 1, $1 - \zeta^j$ is a unit in $\mathbb{Z}[\zeta_\mathfrak{f}]$. If ζ^j is not a primitive $\mathfrak{f}$-th root of unity, then it is a primitime $\mathfrak{h}$-th root of unity for some $\mathfrak{h} \mid \mathfrak{f}$ and $\zeta^j \in \mathbb{Z}[\zeta_\mathfrak{h}]$. Observe that $\frac{\mathfrak{f}}{\mathfrak{h}} \mid j$. Assume that $\mathfrak{h}$ is a prime-power, i.e. $\mathfrak{h} = \hat{\mathfrak{f}}_i^n$ for some $i \in [k]$ and $n \geq 2$. Hence, as $j \in \left\{1, 2, \ldots, \frac{\mathfrak{f}}{\mathfrak{f}_{\max}} - 1\right\}$,

$$\frac{\mathfrak{f}}{\hat{\mathfrak{f}}_i^n} \leq j < \frac{\mathfrak{f}}{\mathfrak{f}_{\max}},$$

which implies $\mathfrak{f}_{\max} < \hat{\mathfrak{f}}_i^n$, a contradiction. Therefore, $\mathfrak{h}$ is not a prime power, i.e. it has more than one distinct prime factors. By Proposition 1, $1 - \zeta^j$ is invertible in $\mathcal{R}_\mathfrak{h}$, thus in $\mathcal{R}_\mathfrak{f}$. □

Next, we recall an elementary result.

Proposition 1 ([Was97, Proposition 2.8]). *Suppose $\mathfrak{f}$ has at least two distinct prime factors. Then, $1 - \zeta$ is a unit in $\mathcal{R} = \mathbb{Z}[\zeta_\mathfrak{f}]$ for any $\mathfrak{f}$-th primitve root of unity ζ.*

[14] For example, $(2^3, 3^2)$ are coprime prime-power factors of $72 = 2^3 3^2$, but $(2, 2^2, 3^2)$ are not.

Finally, we state our proof of Theorem 2.

Proof (Proof of Theorem 2). For $i, j \in [\mathfrak{f}/\mathfrak{f}_{\mathsf{max}}]$, where $i < j$, $\zeta^i - \zeta^j = \zeta^i \cdot (1 - \zeta^{j-i})$ is invertible due to Lemma 3. The expansion factor satisfying $\gamma_{\|\sigma(\cdot)\|_2,S} = 1$ is immediate. For the inverse-expansion factor, we have

$$\theta_{\|\sigma(\cdot)\|_2,S} = \max_{i \neq j} \left\| \frac{1}{\zeta^i - \zeta_j} \right\|_\infty = \max_{i \neq j} \left\| \frac{1}{1 - \zeta_{i-j}} \right\|_\infty \leq \frac{\mathfrak{f}}{4\sqrt{2}}.$$

where the inequality is due to Lemma 2.

For coefficient ∞-norm, we observe that both bounds follow from Corollary 2. □

4.3 Real Cyclotomics

We identify subtractive sets for real cyclotomic rings, i.e. the rings of integers of maximal real subfields of cyclotomic fields. The results over these rings mirror those for cyclotomic fields presented in Theorems 1 and 2.

Theorem 3. *Let $\mathcal{R}^+ = \mathbb{Z}[\zeta_\mathfrak{f} + \zeta_\mathfrak{f}^{-1}]$ with $\mathfrak{f} = p^k$, $\mathfrak{f} > 4$, p prime. The set*

$$S \coloneqq \{ \mu_1^+, \ldots, \mu_{(p+1)/2}^+ \} \subseteq_{(p+1)/2} \mathcal{R}^+$$

is subtractive, where $\mu_i^+ = \mu_i + \bar{\mu}_i$ and $\mu_i = (\zeta^i - 1)/(\zeta - 1)$ for $i \in [(p+1)/2]$, where $\bar{\cdot}$ denotes the complex conjugate. Furthermore, $\gamma_{\|\sigma(\cdot)\|_2,S} \leq p$ and $\theta_{\|\sigma(\cdot)\|_2,S} \leq \frac{\mathfrak{f}^2}{8}$

Proof. Observe that, for $i > j$,

$$\mu_i^+ - \mu_j^+ = \left(1 + \zeta^{-j-i+1}\right) \cdot \frac{\zeta^i - \zeta^j}{\zeta - 1}.$$

The first factor is invertible if $j + i - 1 \nmid \mathfrak{f}$, which holds for distinct $i, j \in [(p+1)/2]$. The second factor is invertible due to Theorem 1. Hence, S is subtractive.

Since any $c \in S$ is a sum of at most p roots of unity, we have $\gamma_{\|\sigma(\cdot)\|_2,S} \leq p$. For $\theta_{\|\sigma(\cdot)\|_2,S}$, we observe that

$$\frac{1}{\mu_i^+ - \mu_j^+} = \frac{\zeta - 1}{(1 + \zeta^{-j-i+1}) \cdot (\zeta^i - \zeta^j)}.$$

Write $\|\cdot\| = \|\sigma(\cdot)\|_\infty$. By Lemma 2,

$$\theta_{\|\sigma(\cdot)\|_2,S} \leq \left\| \frac{\zeta - 1}{(1 + \zeta^{-j-i+1}) \cdot (\zeta^i - \zeta^j)} \right\| \leq \|\zeta^i \cdot (\zeta - 1)\| \cdot \left\| \frac{1}{1 - \zeta^{j-i}} \right\| \cdot \left\| \frac{1}{1 + \zeta^{-j-i+1}} \right\| \leq \frac{\mathfrak{f}^2}{8}.$$ □

where we use that $-\zeta^{-j-i+1}$ is at most a root of unity of order $2\mathfrak{f}$.

Theorem 4. *Let $\mathcal{R}^+ = \mathbb{Z}[\zeta_\mathfrak{f} + \zeta_\mathfrak{f}^{-1}]$ with a non-prime-power conductor $\mathfrak{f}$ having $k \geq 2$ coprime prime-power factors $(\hat{\mathfrak{f}}_i)_{i \in [k]}$, i.e. $\mathfrak{f} = \prod_{i \in [k]} \hat{\mathfrak{f}}_i$. Write $\hat{\mathfrak{f}}_{\max} \coloneqq \max_{i \in [k]} \hat{\mathfrak{f}}_i$. The set*

$$S \coloneqq \left\{ \zeta^i + \zeta^{-i} \right\}_{\left[\left\lfloor \frac{\mathfrak{f}/\mathfrak{f}_{\max}}{2} \right\rfloor\right]} \subseteq_{\left\lfloor \frac{\mathfrak{f}/\mathfrak{f}_{\max}}{2} \right\rfloor} \mathcal{R}^+,$$

is subtractive. Furthermore, $\gamma_{\|\sigma(\cdot)\|_2,S} \leq 2$ and $\theta_{\|\sigma(\cdot)\|_2,S} \leq \frac{\mathfrak{f}^2}{32}$.

Proof. Consider $c_i = \zeta^i + \zeta^{-i} \in S$ and $c_j = \zeta^j + \zeta^{-j} \in S$ with $i > j$. Note that $c_i - c_j = (\zeta^i + \zeta^{-i}) - (\zeta^j + \zeta^{-j}) = \zeta^{-i} \cdot (\zeta^{i+j} - 1) \cdot (\zeta^{i-j} - 1)$. As, $i + j, i - j \in [\mathfrak{f}/\hat{\mathfrak{f}}_{\max}]$, $c_i - c_j$ is invertible in $\mathcal{R}$ by Theorem 2.

The expansion factor satisfying $\gamma_{\|\sigma(\cdot)\|_2,S} \leq 2$ is immediate. Write $\|\cdot\| = \|\sigma(\cdot)\|_\infty$. For $\theta_{\|\sigma(\cdot)\|_2,S}$, we observe that

$$\left\| \frac{1}{c_i - c_j} \right\| \leq \left\| \frac{1}{\zeta^{i+j} - 1} \right\| \cdot \left\| \frac{1}{\zeta^{i-j} - 1} \right\| \leq \left(\frac{\mathfrak{f}}{4\sqrt{2}} \right)^2 = \frac{\mathfrak{f}^2}{32},$$

where the inequality follows from Theorem 2. □

5 Atomic RoK Protocols for Bounded-Norm Satisfiability

In this section, we assume that $\mathcal{R}$ is either $\mathcal{O}_\mathcal{K}$ or $\mathcal{O}_{\mathcal{K}^+}$ which admit large enough subtractive sets, e.g. those constructed in Section 4. Let $\mathcal{C}_\mathcal{R} \subset \mathcal{R}$ denote a fixed subtractive set with expansion factor γ and inverse-expansion factor θ. Throughout, we use both canonical 2-norm and coefficient ∞-norm, and simply write $\|\cdot\|$, when this is not relevant. Both norms might be useful in various application-specific context. In theorems, we track both norms (if needed) and use visual distinctions. The norm of the matrix is defined via vectorisation, i.e. for $\mathbf{A} \in \mathcal{R}^{m\times n}$, $\|\mathbf{A}\| = \|\mathrm{vec}(\mathbf{A})\|$. We use the shorthand notation $\mathcal{R}_q^{n\times d^{\otimes\mu}} := ((\mathcal{R}_q^{1\times d})^{\otimes\mu})^n$ for a matrix whose rows are elementary tensors. We also write $\overline{\mathbf{Z}}$ (resp. $\underline{\mathbf{Z}}$) to indicate the top (resp. bottom) half of a block matrix; the block dimension will be clear from the context. Lastly, we let $\mathcal{C}_{\mathcal{R}_q} \subseteq \mathcal{R}_q^\times$ be obtained by taking a subfield of $\mathcal{R}_q$ and removing 0. Note that $\mathcal{C}_\mathcal{R}$ and $\mathcal{C}_{\mathcal{R}_q}$ have the *invertible differences property* with respect to $\mathcal{R}$ and $\mathcal{R}_q$ respectively, i.e. $\forall x \neq y \in \mathcal{C}_\mathcal{R}$ (resp. $\mathcal{C}_{\mathcal{R}_q}$): $x - y \in \mathcal{R}^\times$ (resp. $\mathcal{R}_q^\times$).

The goal of this section is to construct atomic RoK protocols

$$\Pi^{b\text{-decomp}}, \Pi^{\mathsf{split}}, \Pi^{\mathsf{fold}}, \Pi^{\mathsf{batch}}, \Pi^{\mathsf{norm}} \text{ and } \Pi^{\mathsf{ip}}$$

for proving that a short vector $\mathbf{w}$ satisfies:

- $\mathcal{R}_q$-linear elementary tensor relations, i.e. $(\mathbf{g}_{\mu-1} \otimes \ldots \otimes \mathbf{g}_0) \cdot \mathbf{w} = y \bmod q$;
- a self-inner-product relation, i.e. $t = \langle \mathbf{w}, \alpha(\mathbf{w})\rangle_\mathcal{R} = \sum_{i=0}^{m-1} w_i \cdot \alpha(w)_i \in \mathcal{R}_q$; where $\alpha \in \{\mathsf{id}, \overline{\mathsf{id}}\}$ is either the identity map or the complex conjugate; and
- a norm bound $\|\mathbf{w}\| \leq \beta$.

More specifically, each atomic protocol is a reduction of knowledge which maps between families of relations of the above form with different parameters. In Table 1 on page 27, we provide an overview of parameters for correctness and relaxed knowledge soundness.

The section will be structured as follows. In Section 5.1, we establish some notational convention for this section and formally define the principal relation Ξ^{lin} for which (self-)reductions of knowledge will be constructed. In Sections 5.2 to 5.5, we present the self-reductions of knowledge $\Pi^{b\text{-decomp}}, \Pi^{\mathsf{split}}, \Pi^{\mathsf{fold}}$ and Π^{batch} for Ξ^{lin}. Finally, in Section 5.6, we define two extended relations Ξ^{norm} and Ξ^{ip} and two reductions of knowledge, Π^{norm} and Π^{ip} respectively, which reduce the extended relations to the principal Ξ^{lin} relation.

5.1 The (principal) relation Ξ^{lin}

We begin by defining the relation Ξ^{lin} and outline how protocols reduce instances in this relation to other instances. This relation serves as the principal building block for further protocols.

Basic (single-block) relation. We define our central relation(s) over the ring $\mathcal{R}$, modulo q, for witness dimension $m = d^\mu$. In fact, there are two central relations: Ξ^{lin} for correctness; and $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$ for relaxed knowledge soundness. We define both at once, so that $\Xi^{\mathsf{lin}\vee\mathsf{sis}} \supseteq \Xi^{\mathsf{lin}}$ contains all highlighted parts *additionally*. Let

$$\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{\mathcal{R},q,m,n^{\mathsf{out}},r,\mu,\beta,\beta^{\mathsf{sis}}} := \left\{ \begin{array}{l} ((\mathbf{H},\mathbf{F},\mathbf{Y}),\mathbf{W} \text{ or } \mathbf{w}): \\ \quad \mathbf{H} \in \mathcal{R}_q^{n^{\mathsf{out}}\times n};\ \mathbf{F} \in \mathcal{R}_q^{n\times d^{\otimes\mu}} \subseteq \mathcal{R}_q^{n\times m};\ \mathbf{Y} \in \mathcal{R}_q^{n^{\mathsf{out}}\times r};\ \mathbf{W} \in \mathcal{R}^{m\times r};\ \mathbf{w} \in \mathcal{R}^m \\ \left\{ \begin{array}{c} \|\mathbf{W}\| \leq \beta \\ \mathbf{HFW} = \mathbf{Y} \bmod q \end{array} \right\} \text{ or } \left\{ \begin{array}{l} \mathbf{w} \neq \mathbf{0} \ \wedge\ \|\mathbf{w}\| \leq \beta^{\mathsf{sis}} \\ \overline{\mathbf{H}}\mathbf{F}\mathbf{w} = \mathbf{0}_{\overline{n}} \bmod q \end{array} \right\} \end{array} \right\}$$

where we *always assume* that $\mathbf{H}$ has the block structure[15]

$$\mathbf{H} = \begin{pmatrix} \overline{\mathbf{H}} \\ \underline{\mathbf{H}} \end{pmatrix} \in \mathcal{R}_q^{n^{\mathsf{out}}\times n} \quad \text{where} \quad \overline{\mathbf{H}} = (\mathbf{I}_{\overline{n}}\ \mathbf{0}) \in \mathcal{R}_q^{\overline{n}\times n} \quad \text{and} \quad \underline{\mathbf{H}} \in \mathcal{R}_q^{\underline{n}\times n} \tag{2}$$

Similarly, we write $\overline{\mathbf{Y}} \in \mathcal{R}_q^{\overline{n}}$ and $\underline{\mathbf{Y}} \in \mathcal{R}_q^{\underline{n}}$ for the $\overline{n}$ top (resp. $\underline{n}$ bottom) rows of $\mathbf{Y}$.

[15] This can be marginally relaxed: As long as there is an invertible $\mathbf{X} \in \mathcal{R}^{n^{\mathsf{out}}\times n^{\mathsf{out}}}$ such that $\mathbf{XH}$ has this block structure, we can replace the claim $(\mathbf{H},\mathbf{F},\mathbf{y})$ with the equivalent claim $(\mathbf{XH},\mathbf{F},\mathbf{Xy})$ which has the block structure our protocols require.

Remark 1 (Notational conventions). We often omit irrelevant parameters in Ξ^{lin} and similar relations. Especially all fixed parameters in our protocols, which are $\mathcal{R}$, q, $\overline{n}$, β^{sis}. For example, for parameterised relation like $\Xi_{\mathcal{R},q,x,y}$, we write $\Xi_{x=f(\xi)}$ for $\Xi_{\mathcal{R},q,f(\xi),z}$ or even just $\Xi_{f(\xi)}$ if $x = f(\xi)$ is clear from the context. Also, we fix d and always set $m = d^{\mu}$. As such, we often omit d and μ.

Remark 2 (Matrix witness). For generality and efficiency, we present a relation which deals with a *matrix* $\mathbf{W}$ instead of a vector $\mathbf{w}$ for the witness, and likewise a matrix $\mathbf{Y}$ instead of a vector $\mathbf{y}$. However, it is convenient to think of the columns of $\mathbf{W}$ as a tuple of witnesses $(\mathbf{w}_1, \ldots, \mathbf{w}_r)$ and claims $\mathbf{HFw}_i = \mathbf{y}_i$. Indeed, the linear constraint in Ξ^{lin} is equivalent to r linear constraints (column-wise). However, for efficiency reasons we consider the norm constraint over $\mathbf{W}$ (instead of column-wise norm constraints).

Clearly, relation Ξ^{lin} asserts that the witness $\mathbf{W}$ has norm $\|\mathbf{W}\| \leq \beta$. For the linear relation, let us first assume that $\mathbf{H} = \mathbf{I}_n$ is an identity matrix. In this case, the relation asserts that $\mathbf{FW} = \mathbf{Y}$ holds over $\mathcal{R}_q$. The matrix $\mathbf{F}$ is structured, namely each row $\mathbf{f}$ is an elementary tensor in $\mathcal{R}_q^{1 \times d^{\otimes \mu}}$, i.e. $\mathbf{f} = \mathbf{g}_{\mu-1} \otimes \ldots \otimes \mathbf{g}_0$ for $\mathbf{g}_i = (g_{i,0}, \ldots, g_{i,d-1}) \in \mathcal{R}_q^{1 \times d}$.

For $\Xi^{\mathsf{lin} \vee \mathsf{sis}}$, we relax these assertions by introducing the highlighted OR-part, which captures a break of some underlying cryptographic assumption, e.g. a break of the vSIS assumption [CLM23] (Section A.1). For this, $\overline{\mathbf{F}} = \overline{\mathbf{H}}\mathbf{F}$ will be the commitment key in a protocol. If the assumption is broken, then Ξ^{lin} may not be satisfied, hence the relaxed soundness relation $\Xi^{\mathsf{lin} \vee \mathsf{sis}}$ is necessary.

Now, we further explain $\mathbf{H}$. The primary use of $\mathbf{H}$ is to capture *random linear combination* of rows of $\mathbf{F}$. The block structure asserts that the top $\overline{n}$ rows of $\mathbf{F}$ are simply copied — $\overline{\mathbf{F}} = \overline{\mathbf{H}}\mathbf{F}$ will correspond to the commitment key. Naively, our protocols would have communication costs linear in the number of rows of $\mathbf{F}$, but by using $\mathbf{H}$, we can compress this from n down to $n^{\mathsf{out}} = \overline{n} + \underline{n}$. In prior works, one would simply (re)define $\mathbf{F}$ as $\mathbf{HF}$. However, to keep *(verifier-)succinctness*, we cannot do this: A (random) linear combination of elementary tensors is in general not an elementary tensor. However, our protocol crucially relies on the rows of $\mathbf{F}$ being elementary tensors in order to apply FRI-style (verifier-succinct) folding of the statement. Therefore, we remember the (random) linear combinations of rows in $\mathbf{H}$, instead of carrying out the multiplication. Importantly, the *communication* of the protocol can indeed by compressed by applying $\mathbf{H}$. (Note there that the dimensions of $\mathbf{H}$ and $\mathbf{Y}$ are in general much smaller than that of $\mathbf{W}$.)

Reductions between Ξ^{lin}. Our protocols reduce instances of $\Xi^{\mathsf{lin}}_{m,\beta}$ with different parameters, and we chain them to obtain our final split-and-fold protocol with intermediate norm checks. Primary protocols and parameters of interest are:

(i) $\Pi^{b\text{-}\mathsf{decomp}}$: Reduce an instance with norm bound β to an instance with more columns ($r' > r$) but smaller norm bound.
(ii) Π^{split}: Reduce an instance with witness of shape $m \times r$ to shape $\frac{m}{d} \times (r \cdot d)$.
(iii) Π^{fold}: Reduce an instance with witness of shape $m \times r$ and norm β to an instance of shape $m \times r'$ and norm $\beta' = \gamma\beta$. to shape $m' \times r'$ with $m' = m/d$ and $r' = r \cdot d$. Usually, $r' \in \mathcal{O}(1)$ or $r' \in \mathcal{O}(\lambda)$ is fixed and independent of r.
(iv) Π^{batch}: Reduce one instance to another instance by randomly combining the last $\underline{n}$ rows of $\mathbf{H}$ and $\mathbf{Y}$ into a single one, so that $n^{\mathsf{out}} = \overline{n} + 1$.

Handling vSIS breaks. Knowledge reductions can simply pass a $\Xi^{\mathsf{lin} \vee \mathsf{sis}}$-witness on as their extracted witness. Thus, we sometimes omit that discussion entirely.

5.2 $\Pi^{b\text{-}\mathsf{decomp}}$: b-ary Decomposition Knowledge Reduction

Let $b \geq 1$ be an integer. The protocol $\Pi^{b\text{-}\mathsf{decomp}}$ (Fig. 3) is very simple: It takes a claim $((\mathbf{H}, \mathbf{F}, \mathbf{Y}), \mathbf{W}) \in \Xi^{\mathsf{lin}}_{m,\beta}$ and does a balanced b-ary decomposition of the witness $\mathbf{W}$ with $\|\mathbf{W}\| \leq \beta$ into $\mathbf{W} = \sum_{i=0}^{\ell-1} b^i \mathbf{V}_i$, where $\mathbf{V}_i \in \mathcal{R}_b^r$ (hence $\|\mathbf{V}_i\|_\infty \leq b/2$) and $\ell = \lceil \log_b(2\beta + 1) \rceil$. Then, appropriate claims $\mathbf{Z}_i = \mathbf{HFV}_i$ for the decomposed witness are computed, and the verifier makes sure the new claims imply the original one. Thus, the original statement is reduced to $((\mathbf{H}, \mathbf{F}, \mathbf{Z}_i), \mathbf{V}_i)_{i \in [\ell]}$, which is further combined to $((\mathbf{H}, \mathbf{F}, \widetilde{\mathbf{Z}}), \widetilde{\mathbf{V}})$.

Remark 3. In protocol $\Pi^{b\text{-}\mathsf{decomp}}$, we could apply the optimisation of not sending $\mathbf{Z}_0$, and instead let the verifier compute the unique accepting $\mathbf{Z}_0$, i.e. such that $\mathbf{Y} = \sum_{i \in [\ell]} b^i \mathbf{Z}_i$.

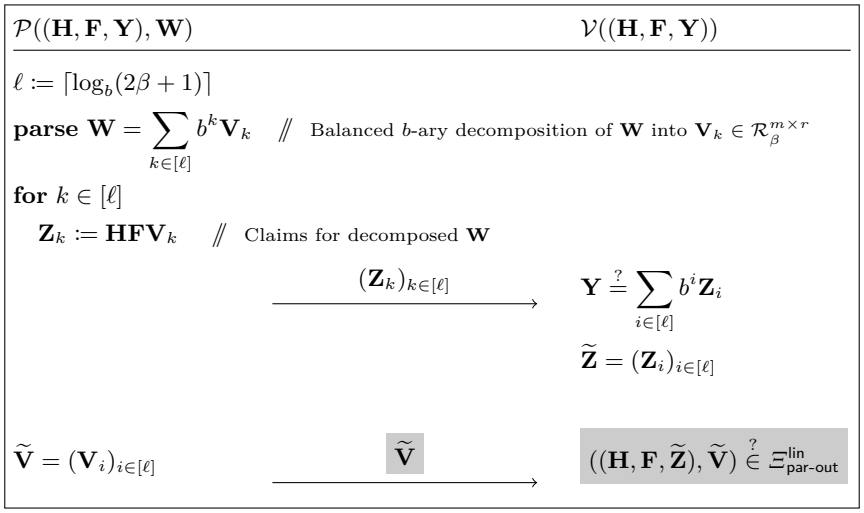

Fig. 3. Protocol $\Pi^{b\text{-decomp}}$, a reduction of $\Xi^{\mathsf{lin}}_{\mathsf{par\text{-}in}}$ to $\Xi^{\mathsf{lin}}_{\mathsf{par\text{-}out}}$ with par-in, par-out specified in Lemma 4. As a *proof* (but not *reduction*) of knowledge, $\Pi^{b\text{-decomp}}$ sends the marked parts.

Lemma 4 (Decomposition). *Let $m, b, r \in \mathbb{N}$ where $0 \le \beta \le \beta^{\mathsf{sis}} \le q$. Protocol $\Pi^{b\text{-decomp}}$ is a perfectly correct self-reduction of knowledge for Ξ^{lin} with parameters*

$$(r,\beta) \mapsto \left(r\cdot\ell, \boxed{\frac{1}{2}\sqrt{\ell r m}\sqrt{\hat{\mathsf{f}}\varphi}\, b}_{\|\sigma(\cdot)\|_2}, \left(\mathsf{resp.}, \boxed{\frac{b}{2}}_{\|\psi(\cdot)\|_\infty}\right)\right).$$

It is a perfectly relaxed knowledge sound self-reduction for $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$ with parameters

$$(r,\beta'_0,\beta^{\mathsf{sis}}) \leftarrow\!\shortmid (r\cdot\ell,\beta'_1,\beta^{\mathsf{sis}}).$$

where $\beta'_0 = \frac{b^\ell-1}{b-1}\cdot\beta'_1$ and $\ell = \lceil \log_b(2\beta+1)\rceil$.

Proof. Perfect correctness of $\Pi^{b\text{-decomp}}$ from $\Xi^{\mathsf{lin}}_{m,r,\beta}$ to $\Xi^{\mathsf{lin}}_{m,r\cdot\ell,b}$ is easy to see: By construction, each $\mathbf{V}_i$ has $\|\psi(\mathbf{V}_i)\|_\infty \le b/2$, and by Corollary 1 it follows that $\|\sigma(\mathbf{V}_i)\|_2 \le \frac{1}{2}\sqrt{rm}\sqrt{\hat{\mathsf{f}}\varphi}\, b$, and therefore $\|\sigma(\mathbf{V}_0,\dots,\mathbf{V}_{\ell-1})\|_2 \le \frac{1}{2}\sqrt{\ell rm}\sqrt{\hat{\mathsf{f}}\varphi}\, b$ as claimed. The linear equations $\mathbf{H}\mathbf{F}\mathbf{V}_i = \mathbf{Z}_i$ hold by definition.

For relaxed knowledge soundness, observe that again by linearity, the original linear equation holds for $\mathbf{W} = \sum_{i=0}^{\ell-1} b^i\mathbf{V}_i$. For the norm, we have

$$\|\mathbf{W}\| \le \sum_{i=0}^{\ell-1} b^i\|\mathbf{V}_i\| \le \frac{b^\ell-1}{b-1}\cdot\beta'_1$$

by the geometric series. Now, we derive the second, simplified bound (which has more slack). For that, observe that $\frac{b^\ell-1}{b-1} \le \frac{b^\ell}{b-1} = \frac{b}{b-1}b^{\ell-1} \le 2b^{\ell-1}$ Moreover, for $x = 2\beta+1$, observe that $b \le \lceil (2\beta+1)^{1/\ell}\rceil \le x^{1/\ell}+1$, and therefore $b^{\ell-1} \le (x^{1/\ell}+1)^{\ell-1} = x^{1-1/\ell}(1+x^{-1/\ell})^{\ell-1}$ and $(1+x^{-1/\ell})^{\ell-1} = \exp((\ell-1)\ln(1+x^{-1/\ell})) \le \exp((\ell-1)x^{-1/\ell})$.

Finally the OR-branch in $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$ is handled by letting the $\mathbf{w}$ with $\overline{\mathbf{H}}\mathbf{F}\mathbf{w} = \mathbf{0}$ be the extracted witness. If $\beta'_0 \le \beta^{\mathsf{sis}}$, this is a witness for $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$. □

Remark 4 (Choice of b). To balance between correctness, soundness, and efficiency, it is convenient to choose ℓ instead of b, and then consider $b = \lceil (2\beta+1)^{1/\ell}\rceil$. In other words, it might be possible that for various values b, the corresponding values ℓ will be equivalent. For the efficiency perspective, there is no point is selecting other b except the smallest one for specified ℓ.

Remark 5. Protocol Π^{split} can be optimised. Intead of sending $(\mathbf{Z}_k)_{k\in[\ell]}$ and verifying $\mathbf{Y} \stackrel{?}{=} \sum_{i\in[\ell]} b^i\mathbf{Z}_i$, is it enough to send $(\mathbf{Z}_k)_{k\in[\ell-1]}$ and recompute the remaining part.

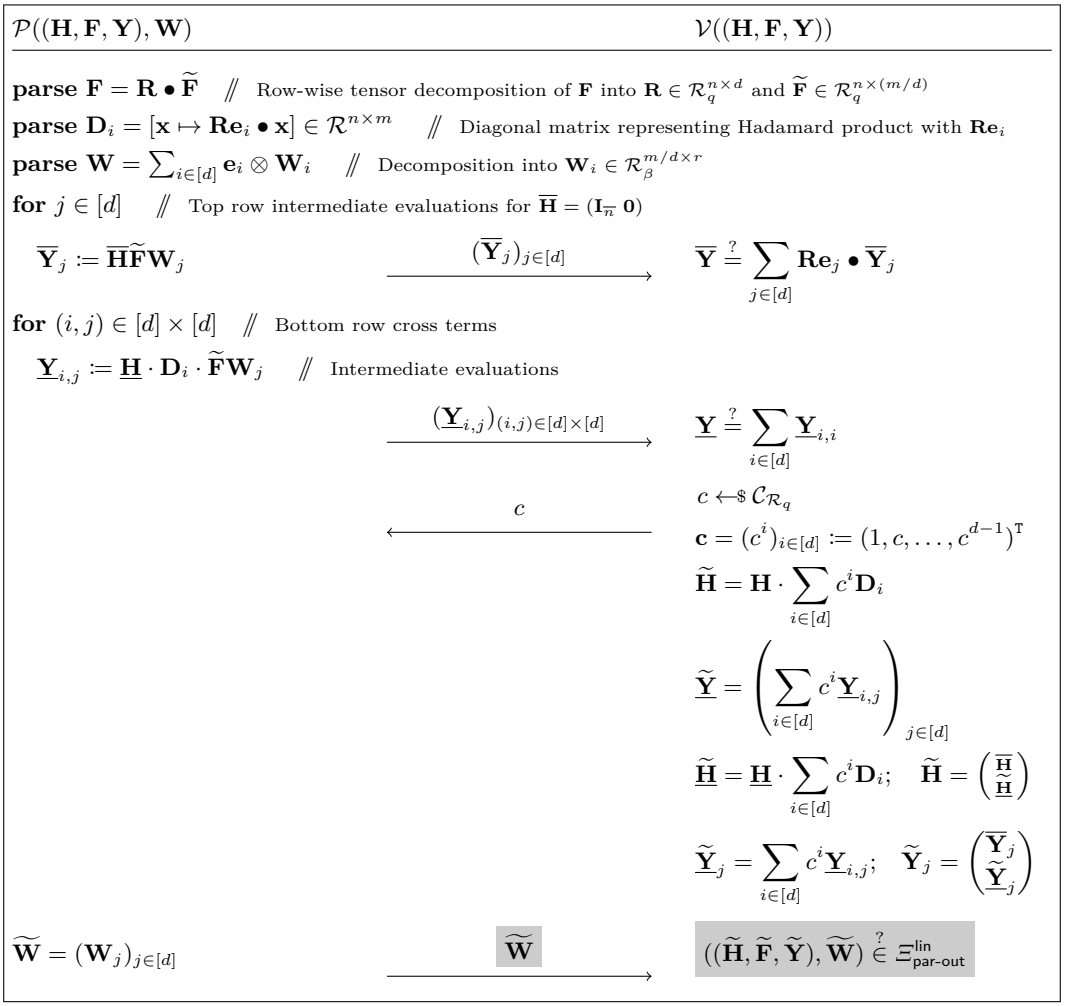

Fig. 4. Protocol Π^{split}, a reduction from $\Xi^{\mathsf{lin}}_{\mathsf{par\text{-}in}}$ to $\Xi^{\mathsf{lin}}_{\mathsf{par\text{-}out}}$ with par-in, par-out specified in Lemma 5. Π^{split} sends the marked parts only as a *proof* (but not *reduction*) of knowledge.

5.3 Π^{split}: Witness Splitting Knowledge Reduction

In Fig. 4 we describe protocol Π^{split} which takes a claim from $\Xi^{\mathsf{lin}}_{m,r,\beta}$ and splits it into claim in $\Xi^{\mathsf{lin}}_{m/d,r\cdot d,\beta}$. We explain the idea and correctness of the protocol below.

To split the witness, interpret $\mathcal{R}^{m\cdot r}$ as $\mathcal{R}^{d^{\otimes\mu}\times r}$, and split $\mathbf{W} \in \mathcal{R}^{m\times r} \cong \mathcal{R}^{d^{\otimes\mu}\times r}$ into $\mathbf{W} = \sum_{i=0}^{\mu-1} \mathbf{e}_i \otimes \mathbf{W}_i$ where $\mathbf{W}_i \in \mathcal{R}^{m/d\times r} \cong \mathcal{R}^{d^{\otimes(\mu-1)}\times r}$ and $\mathbf{e}_i \in \{0,1\}^d$ is the i-th standard unit vector. Splitting $\mathbf{W}$ like this is compatible with the row-wise tensor structure of $\mathbf{F}$. Let us take a closer look at this. /

For simplicity, first consider a single row $\mathbf{f} \in \mathcal{R}_q^{1\times d^{\otimes\mu}}$ of $\mathbf{F}$. By the elementary tensor structure of the row-vector $\mathbf{f}$, we can write it as $\mathbf{f} = \mathbf{r} \otimes \widetilde{\mathbf{f}} = (r_0 \cdot \widetilde{\mathbf{f}}, \ldots, r_{d-1} \cdot \widetilde{\mathbf{f}}) = (\mathbf{f}_0, \ldots, \mathbf{f}_{d-1})$ where $\widetilde{\mathbf{f}} \in \mathcal{R}_q^{1\times d^{\otimes(\mu-1)}}$, $\mathbf{r} = (r_0, \ldots, r_{d-1}) \in \mathcal{R}_q^{1\times d}$, and $\mathbf{f}_i = r_i \cdot \widetilde{\mathbf{f}}_i$. Therefore, $\mathbf{f} \cdot \mathbf{W} = \sum_{i\in[d]} \mathbf{f}_i \mathbf{W}_i = \sum_{i\in[d]} (\widetilde{\mathbf{f}} \cdot \mathbf{W}_i) \cdot (\mathbf{r} \cdot \mathbf{e}_i^\intercal) = \sum_{i\in[d]} r_i \widetilde{\mathbf{f}} \cdot \mathbf{W}_i$.

Now, consider any matrix $\mathbf{F}$ with row-wise tensor structure and n rows, as in Ξ^{lin}. That is, $\mathbf{F} \in \mathcal{R}_q^{n\times d^{\otimes\mu}}$. Observe that

$$\mathbf{F} = \begin{pmatrix} \mathbf{F}_{0,\bullet} \\ \vdots \\ \mathbf{F}_{n-1,\bullet} \end{pmatrix} = \begin{pmatrix} \mathbf{F}_{0,0} & \ldots & \mathbf{F}_{0,d-1} \\ \vdots & & \vdots \\ \mathbf{F}_{n-1,0} & \ldots & \mathbf{F}_{0,d-1} \end{pmatrix} = \begin{pmatrix} \mathbf{r}_0 \otimes \widetilde{\mathbf{f}}_0 \\ \vdots \\ \mathbf{r}_{n-1} \otimes \widetilde{\mathbf{f}}_{n-1} \end{pmatrix} \tag{3}$$

where $\mathbf{F}_{i,\bullet}$ denotes the i-th row of $\mathbf{F}$, and $\mathbf{F}_{i,j} \in \mathcal{R}_q^{1\times d^{\otimes\mu}}$ the block of rows (the analogue of $(\mathbf{f}_0,\dots,\mathbf{f}_{d-1})$ of the single-row case), and $\widetilde{\mathbf{f}}_i \in \mathcal{R}_q^{1\times d^{\otimes(\mu-1)}}$ and $\mathbf{r}_i \in \mathcal{R}_q^{1\times d}$ are the analogues of $\mathbf{r}$ and $\widetilde{\mathbf{f}}$ of the single-row case respectively. To ease notation, we define $\mathbf{R} = (\mathbf{r}_i^\intercal)_{i\in[n]} \in \mathcal{R}_q^{n\times d}$ and $\widetilde{\mathbf{F}} = (\widetilde{\mathbf{f}}_i)_{i\in[n]} \in \mathcal{R}_q^{n\times d^{\otimes(\mu-1)}}$, and we write $\mathbf{F} = \mathbf{R} \bullet \widetilde{\mathbf{F}}$ for the row-wise tensor product[16] of $\mathbf{R}$ and $\widetilde{\mathbf{F}}$ as seen in Eq. (3). In this notation,

$$\mathbf{F}\cdot\mathbf{W} = (\mathbf{R}\bullet\widetilde{\mathbf{F}})\cdot\left(\sum_{i=0}^{\mu-1}\mathbf{e}_i\otimes\mathbf{W}_i\right) = \sum_i \underbrace{(\mathbf{R}\mathbf{e}_i)}_{\in\mathcal{R}_q^n} \bullet \underbrace{(\widetilde{\mathbf{F}}\mathbf{W}_i)}_{=\mathbf{Y}_i\in\mathcal{R}_q^{n\times r}}. \tag{4}$$

Note that for the vector $\mathbf{r}_i = \mathbf{R}\mathbf{e}_i$, row-wise tensoring $\mathbf{r}_i \bullet \mathbf{W}_i$ is just componentwise multiplication with $r_{i,j}$ in the j-th row of $\mathbf{W}_i$. Thus, with $\mathbf{D}_i \coloneqq \mathrm{diag}(\mathbf{r}_i)$, we can rewrite (4) as

$$\mathbf{F}\cdot\mathbf{W} = \sum_i \mathbf{D}_i\cdot\widetilde{\mathbf{F}}\mathbf{W}_i. \tag{5}$$

With the above, we have derived a splitting protocol for the special case where $\mathbf{H} = \mathbf{I}_n$ is the identity matrix: Simply send $\mathbf{Y}_i = \mathbf{D}_i\cdot\widetilde{\mathbf{F}}\mathbf{W}_i$ and set $\mathbf{Y} = (\mathbf{Y}_0,\dots,\mathbf{Y}_{d-1})$ for new statement $(\mathbf{H},\widetilde{\mathbf{F}},\mathbf{Y})$ and witness $\widetilde{\mathbf{W}} = (\mathbf{W}_i)_{i\in[d]}$.

When $\mathbf{H}$ is not necessarily the identity, we must also handle the bottom part $\underline{\mathbf{H}}$ of $\mathbf{H}$. To do so, our protocol (cf. Fig. 4) additionally sends cross terms, namely $\mathbf{D}_i\cdot\widetilde{\mathbf{F}}\mathbf{W}_j$ for $i,j\in[d]$, which are then randomly recombined.

Lemma 5 (Split). *Let $m,d,r,\mu\in\mathbb{N}$ where $d|m$ and $0\le\beta\le\beta^{\mathsf{sis}}\le q$. Protocol Π^{split} is a perfectly correct self-reduction of knowledge for Ξ^{lin} with parameters*

$$(m,\,r,\,\mu,\,\beta)\mapsto(m/d,\,r\cdot d,\,\mu-1,\,\beta).$$

It is a perfectly relaxed knowledge sound self-reduction for $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$ with parameters

$$(m,\,r,\,\mu,\,\beta,\,\beta^{\mathsf{sis}})\leftrightarrow(m/d,\,r\cdot d,\,\mu-1,\,\beta,\,\beta^{\mathsf{sis}})$$

with d-special sound extraction and knowledge error $\kappa = (d-1)/|\mathcal{C}_{\mathcal{R}_q}|$ if $2\beta\le\beta^{\mathsf{sis}}$.

Proof. Perfect correctness of Π^{split} is straightforward for the top rows: Since $\overline{\mathbf{H}} = (\mathbf{I}_{\overline{n}}\ \mathbf{0})$, we have $\overline{\mathbf{F}} = \overline{\mathbf{H}}\mathbf{F}$ are just the $\overline{n}$ top rows of $\mathbf{F}$, and similar for $\widetilde{\mathbf{F}}$, and thus by our discussion before and some renaming (using $\overline{\mathbf{F}}$ instead of $\mathbf{F}$ makes $\overline{\mathbf{H}}$ the identity) we know that the top part is perfectly correct. For the bottom rows, correctness follows essentially from Eqs. (4) and (5) which asserts that

$$\underline{\widetilde{\mathbf{H}}}\widetilde{\mathbf{F}}\mathbf{W}_j = \sum_{i\in[d]} c_i\underline{\mathbf{H}}\cdot\mathbf{D}_i\cdot\widetilde{\mathbf{F}}\mathbf{W}_j = \sum_{i\in[d]} c_i\underline{\mathbf{Y}}_{i,j} = \underline{\widetilde{\mathbf{Y}}}_j.$$

For relaxed knowledge soundness, we argue through d-special soundness. So we have d related accepting transcripts for challenge vectors $\mathbf{c}^{(k)}$ with witness $\widetilde{\mathbf{W}}^{(k)} = (\widetilde{\mathbf{W}}_i^{(k)})_{i\in[d]}$ which satisfies $((\widetilde{\mathbf{H}}^{(k)},\widetilde{\mathbf{F}}^{(k)},\widetilde{\mathbf{Y}}^{(k)}),\widetilde{\mathbf{W}}^{(k)})\in\Xi^{\mathsf{lin}\vee\mathsf{sis}}$.

Step 1 (top rows): Let us first consider the top rows (and any single transcript): Here, it is straightforward to see that

$$\mathbf{W}^{(k)} = \underbrace{\widetilde{\mathbf{W}}_i^{(k)}}_{i\in[d]} \quad\text{satisfies}\quad \overline{\mathbf{H}}\mathbf{F}\mathbf{W}^{(k)} = \overline{\mathbf{H}}\sum_i\mathbf{D}_i\widetilde{\mathbf{F}}\mathbf{W}_i^{(k)} = \sum_i\overline{\mathbf{Y}}_i = \overline{\mathbf{Y}}$$

for all $k\in[d]$ by construction (and using $\overline{\mathbf{H}} = (\mathbf{I}_{\overline{n}}\ \mathbf{0})$). Thus, we trivially and unconditionally find a witness for the top rows. Clearly, $\|\mathbf{W}\| = \beta$ as the witness is simply rearranged.

Moreover, by looking at the top rows, we see that: Either, there is a unique $\mathbf{W}_j$ over all transcripts, i.e. $\mathbf{W}_j^{(k)} = \mathbf{W}_j^{(k')}$ for all $k,k'\in[d]$. Or, there is a *non-zero* difference $\mathbf{V}_j = \mathbf{W}_j^{(k)} - \mathbf{W}_j^{(k')}$ of norm at

[16] This row-wise tensor product is known under several names, e.g. row-wise Kronecker product, "face-splitting product", "transposed Khatri–Rao product" (and more general forms, as block Kronecker product and Khatri–Rao product).

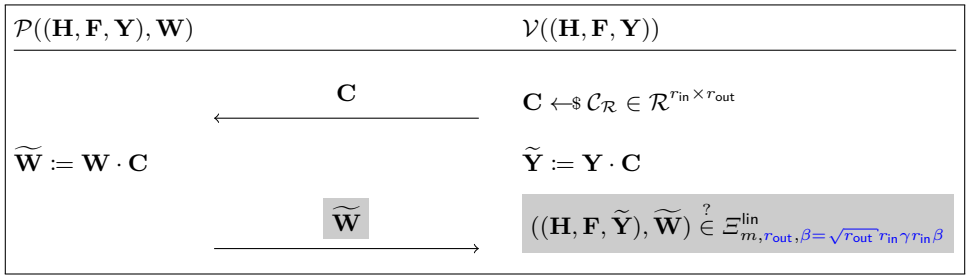

Fig. 5. Protocol Π^{fold} folds instance of $\Xi^{\mathsf{lin}}_{\mathsf{par\text{-}in}}$ into $\Xi^{\mathsf{lin}}_{\mathsf{par\text{-}out}}$ with par-in, par-out specified in Lemma 6. Π^{fold} sends the marked parts only as a *proof* (but not *reduction*) of knowledge.

most 2β. Consider a non-zero column $\mathbf{v}_j$, such that $\overline{\mathbf{H}}\widetilde{\mathbf{F}}\mathbf{v}_j = \mathbf{0}$, and thus $\overline{\mathbf{H}}\mathbf{F}\mathbf{v}_j = \mathbf{0}$ is a witness for the OR-branch in $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$ (of norm at most 2β). Hence, from now on, we assume all transcripts contain the same $\mathbf{W}_j = \mathbf{W}_j^{(k)}$ for all $k \in [d]$.

Step 2 (bottom rows): Now, we consider the bottom row, with an arbitrary $\underline{\mathbf{H}}$. Towards showing d-special soundness, define for $i \in [0, \mu-1]$ and $j \in [0, d-1]$ the shorthand

$$\mathbf{Z}_{i,j} = \mathbf{D}_i \widetilde{\mathbf{F}} \mathbf{W}_j.$$

As the first step, we show that $\underline{\mathbf{H}}\mathbf{Z}_{i,j} = \underline{\mathbf{Y}}_{i,j}$ for all i, j. Towards this, we rewrite the verifier's checks as

$$\underline{\mathbf{H}} \cdot \mathbf{Z}_j \cdot \mathbf{c}^{(k)} = \sum_i \underline{\mathbf{H}} \cdot (\mathbf{Z}_{i,j})_i c_i^{(k)} = \sum_i (\underline{\mathbf{Y}}_{i,j})_i c_i^{(k)} = \underline{\mathbf{Y}}_j \cdot \mathbf{c}^{(k)}$$

where $\mathbf{Z}_j = (\mathbf{Z}_{0,j}, \ldots, \mathbf{Z}_{d-1,j})$ and likewise for $\underline{\mathbf{Y}}_j$. From d distinct challenges, we assemble a (Vandermonde) matrix $\mathbf{C} = (\mathbf{c}^{(0)}, \ldots, \mathbf{c}^{(d-1)})$. Since $\mathcal{C}_{\mathcal{R}_q}$ has the invertibility of differences property, $\mathbf{C}$ is invertible over $\mathcal{R}_q$, and therefore

$$\underline{\mathbf{H}} \cdot \mathbf{Z}_j \cdot \mathbf{C} = \underline{\mathbf{Y}}_j \cdot \mathbf{C} \quad \Longrightarrow \quad \underline{\mathbf{H}} \cdot \mathbf{Z}_j = \underline{\mathbf{Y}}_j.$$

Thus $\underline{\mathbf{H}}\mathbf{Z}_{i,j} = \underline{\mathbf{Y}}_{i,j}$ for all i, j as claimed. Then we see that

$$\underline{\mathbf{Y}} = \sum_i \underline{\mathbf{Y}}_{i,i} = \underline{\mathbf{H}} \sum_i \mathbf{Z}_{i,i} = \underline{\mathbf{H}} \sum_i \mathbf{D}_i \widetilde{\mathbf{F}} \mathbf{W}_i = \underline{\mathbf{H}}\mathbf{F}\mathbf{W}$$

and therefore, $\mathbf{W}$ (as assembled in Step 1) is a witness for the bottom rows as well.

Step 3 (OR-branch): Finally consider the OR-branch in $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$. If $\overline{\mathbf{H}}\widetilde{\mathbf{F}}\mathbf{v}_j = \mathbf{0}$, we simply let $\mathbf{v}_j$ be the extracted witness note that $\|\mathbf{v}_j\| \le \beta \le \beta^{\mathsf{sis}}$. □

Remark 6. Protocol Π^{split} can be optimised. For example, suppose that $\mathbf{r}_0 = \mathbf{R}\mathbf{e}_0$ has no zero component. Then instead of sending $\overline{\mathbf{Y}}_0$, we can compute it as $\mathbf{D}_0^{-1} \sum_{i \in [d]\setminus\{0\}} \mathbf{D}_i \overline{\mathbf{Y}}_i$ because no other choice satisfies the verifier's check. Similarly, we can omit $\underline{\mathbf{Y}}_{0,0}$. For arbitrary $\mathbf{R}$, in each row there is some i such that $\mathbf{R}\mathbf{e}_i$ is not zero (else $\mathbf{F}$ has a zero row, which is useless), so a more complex variant of this optimization always applies, saving d $\mathcal{R}_q$-elements of communication. Moreover, whenever $\underline{\mathbf{H}}$ has structured rows (e.g. contains (permuted) identity submatrices, etc.), application specific optimisations may apply.

5.4 Π^{fold}: Fold Knowledge Reduction

In Fig. 5, we present the protocol Π^{fold}, which is a simple batching technique which reduces the number r of columns in $\mathbf{W}$. It takes an instance of $((\mathbf{H},\mathbf{F},\mathbf{Y}),\mathbf{W})$ of $\Xi^{\mathsf{lin}}_{m,r_{\mathsf{in}}}$, and produces a random linear combination $((\mathbf{H},\mathbf{F},\mathbf{Y}),\widetilde{\mathbf{W}})$ in $\Xi^{\mathsf{lin}}_{m,r_{\mathsf{out}}}$ as output, with increased norm bounds.

Lemma 6 (Fold). *Let $m, r_{\mathsf{in}}, r_{\mathsf{out}} \in \mathbb{N}$, $\beta \ge 0$, and $0 \le \beta' \le \beta^{\mathsf{sis}} \le q$. Protocol Π^{fold} is a perfectly correct self-reduction of knowledge for Ξ^{lin} with parameters*

$$(r_{\mathsf{in}}, \beta) \mapsto \left(r_{\mathsf{out}}, \boxed{\sqrt{r_{\mathsf{out}}}\, r_{\mathsf{in}} \gamma_{\|\sigma(\cdot)\|_2} \beta}_{\|\sigma(\cdot)\|_2} \left(\text{resp.}, \boxed{r_{\mathsf{in}} \gamma_{\|\psi(\cdot)\|_\infty} \beta}_{\|\psi(\cdot)\|_\infty} \right) \right).$$

It is a relaxed knowledge sound self-reduction for $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$ *with parameters*

$$\left(r_{\mathsf{in}}, \boxed{2\sqrt{r_{\mathsf{in}}}\theta_{\|\sigma(\cdot)\|_2}\beta'}_{\|\sigma(\cdot)\|_2} \left(\text{resp.}, \boxed{2\theta_{\|\psi(\cdot)\|_\infty}\beta'}_{\|\psi(\cdot)\|_\infty}\right), \beta^{\mathsf{sis}}\right) \leftarrow\!\shortmid \left(r_{\mathsf{out}}, \beta', \beta^{\mathsf{sis}}\right)$$

with r_{in}*-CWSS extraction, and hence knowledge error* $\kappa \leq \frac{r_{\mathsf{in}}}{|\mathcal{C}_{\mathcal{R}}|^{r_{\mathsf{out}}}}$.

Proof. For perfect correctness, it is clear that $\widetilde{\mathbf{W}}$ satisfies $\mathbf{HF}\widetilde{\mathbf{W}} = \widetilde{\mathbf{Y}}$ by construction. Moreover, for $\mathbf{W} = (\mathbf{w}_i)_{i\in[r_{\mathsf{in}}]}$ and $\widetilde{\mathbf{W}} = (\widetilde{\mathbf{w}}_i)_{i\in[r_{\mathsf{out}}]}$, $\|\sigma\left(\widetilde{\mathbf{W}}\right)\|_2 \leq \sqrt{r_{\mathsf{out}}}\max_{i\in[r_{\mathsf{out}}]}\|\sigma\left(\widetilde{\mathbf{w}}_i\right)\|_2 \leq \sqrt{r_{\mathsf{out}}}\sum_{j\in[r_{\mathsf{out}}]}\|\sigma\left(c_{j,i}\mathbf{w}_j\right)\|_2 \leq \sqrt{r_{\mathsf{out}}}\sum_{j\in[r_{\mathsf{in}}]}\gamma_2\|\sigma\left(\mathbf{w}_j\right)\|_2 \leq \sqrt{r_{\mathsf{out}}}\gamma_2 r_{\mathsf{in}}\beta$, and with a similar reasoning $\|\psi\left(\widetilde{\mathbf{W}}\right)\|_\infty \leq \gamma_\infty r_{\mathsf{in}}\beta$, thus the norm is also within bounds and correctness follows.

For relaxed knowledge soundness, through r_{in}-CWSS we are given $r_{\mathsf{in}}+1$ accepting transcripts $\widetilde{\mathbf{W}}_0, \ldots, \widetilde{\mathbf{W}}_{r_{\mathsf{in}}}$, for challenges $\mathbf{C}^{(i)} = \sum_{k\in[r_{\mathsf{in}}]}\mathbf{e}_k \otimes \mathbf{c}_k^{\mathsf{T}(i)}$ where $\mathbf{C}^{(i)}$ and $\mathbf{C}^{(r_{\mathsf{in}})}$ differ exactly in row $i \in \{0, \ldots r_{\mathsf{in}}-1\}$. We can now subtract the accepting equations to obtain

$$\mathbf{HF}(\widetilde{\mathbf{W}}_i - \widetilde{\mathbf{W}}_{r_{\mathsf{in}}}) = \mathbf{Y}(\mathbf{C}^{(i)} - \mathbf{C}^{(r_{\mathsf{in}})}) = \mathbf{y}_i(\mathbf{c}_i^{(i)} - \mathbf{c}_i^{(r_{\mathsf{in}})})^{\mathsf{T}}.$$

Let $j \in [r_{\mathsf{out}}]$ be selected such that $c_{i,j}^{(i)} \neq c_{i,j}^{(r_{\mathsf{in}})}$. Let $(\widetilde{\mathbf{W}}_i - \widetilde{\mathbf{W}}_{r_{\mathsf{in}}}) = \widehat{\mathbf{W}}_i = (\widehat{\mathbf{w}}_{i,j})_{j\in[r_{\mathsf{out}}]}$ Thus, setting

$$\mathbf{w}_i \coloneqq \frac{1}{c_{i,j}^{(i)} - c_{i,j}^{(r_{\mathsf{in}})}}\widehat{\mathbf{w}}_{i,j}$$

is a column of witness in $\mathcal{R}_q^m$, where we use the subtractive set property of $\mathcal{C}_{\mathcal{R}}$ to ensure division in $\mathcal{R}$. The recovered witness satisfies $\mathbf{HFW} = \mathbf{Y}$ by construction. Moreover, we have

$$\|\sigma\left(\mathbf{W}\right)\|_2 \leq \sqrt{r_{\mathsf{in}}}\cdot\max_{i\in[r_{\mathsf{in}}]}\|\sigma\left(\mathbf{w}_i\right)\|_2 = \sqrt{r_{\mathsf{in}}}\cdot\|\sigma\left(\frac{1}{c_i^{(i)} - c_i^{(r_{\mathsf{in}})}}\widehat{\mathbf{w}}_{i,j}\right)\|_2 \leq \sqrt{r_{\mathsf{in}}}\cdot\theta_2\cdot 2\cdot\beta_1'$$

and by the same reasoning

$$\|\psi\left(\mathbf{W}\right)\|_\infty \leq \max_{i\in[r_{\mathsf{in}}]}\|\sigma\left(\mathbf{w}_i\right)\|_2 \leq \theta_\infty\cdot 2\cdot\beta_1'$$

by definition of the inverse-expansion factor for $\mathcal{C}_{\mathcal{R}}$. For the knowledge error, we use r_{in}-CWSS: The challenge space per coordinate is $\mathcal{C}_{\mathcal{R}}^{r_{\mathsf{out}}}$, and we need to extract r_{in} coordinates, hence $\kappa \leq \frac{r_{\mathsf{in}}}{|\mathcal{C}_{\mathcal{R}}|^{r_{\mathsf{out}}}}$.

Finally, the OR-branch in $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$ is handled by letting $\mathbf{w}$ equal to a non-zero columns of $\widehat{\mathbf{w}}$; obviously, $\overline{\mathbf{H}}\mathbf{F}\mathbf{w} = \mathbf{0}$ holds and $\|\mathbf{w}\| \leq \beta' \leq \beta^{\mathsf{sis}}$. This completes the proof.

5.5 Π^{batch}: Batch-Rows Knowledge Reduction

The protocol Π^{batch} (Fig. 6) is a protocol to batch the claims along multiple rows into fewer rows of claims. This is done by a random linear combination of the rows in question. This protocol maps an instance $((\mathbf{H}, \mathbf{F}, \mathbf{Y}), \mathbf{W})$ of $\Xi_{m,\beta}^{\mathsf{lin}}$ to an instance $((\widetilde{\mathbf{H}}, \mathbf{F}, \widetilde{\mathbf{Y}}), \mathbf{W})$, where the height of $\widetilde{\mathbf{Y}}$ is smaller. We describe it in more detail: Let $n^{\mathsf{out}} = \overline{n} + \underline{n}$. Then Π^{batch} keeps the top $\overline{n}$ rows $\overline{\mathbf{y}}$ of $\mathbf{Y}$ (resp. $\overline{\mathbf{H}}$ of $\mathbf{H}$, and thus of $\mathbf{HF}$) unchanged. But the bottom $\underline{n}$ rows are linearly combined into a single row. For this, $\mathbf{H}$ and $\mathbf{Y}$ are split into top and bottom half, and the bottom half is multiplied by a vector $\mathbf{c}$ consisting of powers of $c \leftarrow\!\!\$\, \mathcal{C}_{\mathcal{R}_q}$. Both parties then update the statement suitably.

Lemma 7 (Batch). *Let* $n^{\mathsf{out}}, \overline{n} \in \mathbb{N}$ *and* $0 \leq \beta \leq \beta^{\mathsf{sis}} \leq q$. *Protocol* Π^{batch} *is a perfectly correct self-reduction of knowledge for* Ξ^{lin} *with parameters*

$$\left(n^{\mathsf{out}}, \beta\right) \mapsto \left(\overline{n}+1, \beta\right).$$

It is a relaxed knowledge sound self-reduction for $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$ *with knowledge error* $\kappa = \frac{n^{\mathsf{out}}-\overline{n}-1}{|\mathcal{C}_{\mathcal{R}_q}|} \leq \frac{r\cdot\underline{n}}{|\mathcal{C}_{\mathcal{R}_q}|}$ *if* $2\beta' \leq \beta^{\mathsf{sis}}$.

$$\left(n^{\mathsf{out}}, \beta', \beta^{\mathsf{sis}}\right) \leftarrow\!\shortmid \left(\overline{n}+1, \beta', \beta^{\mathsf{sis}}\right).$$

Extraction requires two uniformly distributed transcripts (Footnote 11).

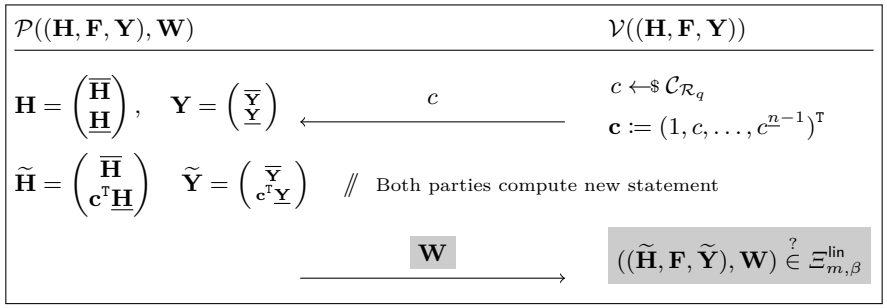

Fig. 6. Protocol Π^{batch} reduces an instance of $\Xi^{\mathsf{lin}}_{\mathsf{par\text{-}in}}$ to $\Xi^{\mathsf{lin}}_{\mathsf{par\text{-}out}}$ with par-in, par-out specified in Lemma 4, with fewer rows by batching to the last $\underline{n}$ of $\mathbf{H}$. Π^{batch} sends the marked parts only as a *proof* (but not *reduction*) of knowledge.

Proof. The correctness of this protocol is straightforward by linearity. For (knowledge) soundness, we rely on the Schwartz–Zippel lemma over $\mathcal{C}_{\mathcal{R}_q}$, which we recall is almost a subfield $\mathbb{F}$ of $\mathcal{R}_q$ except that 0 is missing. The lemma states that, for any degree-d non-zero polynomial over $\mathbb{F}$, the probability that the polynomial evaluates to zero at a uniformly random point chosen from $\mathcal{C}_{\mathcal{R}_q}$ is at most $d/\left|\mathcal{C}_{\mathcal{R}_q}\right|$. To translate this upper bound into a *knowledge* error, observe the following: If $\mathcal{A}$ succeeds for 2 challenges, then the first transcript fixes some $\mathbf{W}_1$ which satisfies $((\widetilde{\mathbf{H}}_1,\mathbf{F},\widetilde{\mathbf{Y}}_1),\mathbf{W}_1) \in \Xi^{\mathsf{lin}}$. Suppose $((\mathbf{H},\mathbf{F},\mathbf{Y}),\mathbf{W}_1) \notin \Xi^{\mathsf{lin}}$, i.e. $\mathbf{W}_1$ is not a witness for the original statement. Then we observe that *at most* a fraction of $\kappa = \frac{n^{\mathsf{out}}-\overline{n}-1}{\left|\mathcal{C}_{\mathcal{R}_q}\right|}$ challenges can be accepting for $\mathbf{W}_1$ (by Schwartz–Zippel and union bound). In other words, if $\mathcal{A}$ succeeds with probability ϵ, then with probability at least $\epsilon - \kappa$ the 2-transcript extractor successfully outputs two transcripts where the responses $\mathbf{W}_1$ and $\mathbf{W}_2$ differ.[17] Now, $\mathbf{V} = \mathbf{W}_1 - \mathbf{W}_2$ is a non-zero preimage with a non-zero column $\mathbf{v}$, s.t. $\overline{\mathbf{H}}\mathbf{F}\mathbf{v} = \mathbf{0}$ of norm at most $2\beta' \leq \beta^{\mathsf{sis}}$, i.e. the OR-branch of $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$.

Remark 7. The knowledge-error can improved by issuing $t > 1$ challenges which yields $\kappa = \frac{n^{\mathsf{out}}-\overline{n}-1}{\left|\mathcal{C}_{\mathcal{R}_q}\right|^t} \leq \left(\frac{r\cdot\underline{n}}{\left|\mathcal{C}_{\mathcal{R}_q}\right|}\right)^t$. The protocol remains the same with the exception that instead of a vector $\mathbf{c}$, the protocol uses a matrix $\mathbf{C} \in \mathcal{R}_q^{\underline{n}\times t}$, where i-th row is a series of consecutive powers of challenge ξ_i for $i \in [t]$. The protocol is a perfectly correct self-reduction of knowledge for Ξ^{lin} with parameters

$$\left(n^{\mathsf{out}},\, \beta\right) \mapsto \left(\overline{n}+t,\, \beta\right).$$

It is a relaxed knowledge sound self-reduction for $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$ with knowledge error κ if $2\beta' \leq \beta^{\mathsf{sis}}$.

$$\left(n^{\mathsf{out}},\, \beta',\, \beta^{\mathsf{sis}}\right) \leftarrow\!\shortmid \left(\overline{n}+t,\, \beta',\, \beta^{\mathsf{sis}}\right).$$

Similarly, the extraction requires two uniformly distributed transcripts.

5.6 Π^{norm}, Π^{ip}: Weighted Norm and Inner Product Checks

To restrain the norm growth of the extracted witness, we introduce norm checks. We present the norm check protocol Π^{norm}, which handles reducing the norm relation Ξ^{norm}, to multiple Ξ^{lin} relations. The relations Ξ^{norm} and Ξ^{ip}, as well as their variants $\Xi^{\mathsf{norm}\vee\mathsf{sis}}$ and $\Xi^{\mathsf{ip}\vee\mathsf{sis}}$, are defined as follows.

$$\Xi^{\mathsf{norm}\vee\mathsf{sis}}_{\mathcal{R},q,m,n^{\mathsf{out}},r,\mu,\beta,\beta^{\mathsf{sis}}} \coloneqq \left\{ \begin{array}{l} ((\mathbf{H},\mathbf{F},\mathbf{Y},\mathbf{c},\nu),\mathbf{W} \text{ or } \mathbf{w}): \\ \quad \mathbf{H} \in \mathcal{R}_q^{n^{\mathsf{out}}\times\underline{n}};\ \mathbf{F} \in \mathcal{R}_q^{\underline{n}\times d^{\otimes\mu}} \subseteq \mathcal{R}_q^{\underline{n}\times m};\ \mathbf{Y} \in \mathcal{R}_q^{n^{\mathsf{out}}\times r};\ \mathbf{c} \in \mathcal{R}_q^r \text{ s.t. } \overline{\mathbf{c}} = \mathbf{c}; \\ \quad 0 \leq \nu \leq \beta;\ \mathbf{W} = (\mathbf{w}_i)_{i=0}^r \in \mathcal{R}^{m\times r};\ \mathbf{w} \in \mathcal{R}^m \\ \quad \left\{ \begin{array}{l} \sum_{i=0}^r c_i\|\mathbf{w}_i\| \leq \nu \\ \mathbf{H}\mathbf{F}\mathbf{W} = \mathbf{Y} \bmod q \end{array} \right\} \text{ or } \left\{ \begin{array}{l} \|\mathbf{w}\| \leq \beta^{\mathsf{sis}} \\ \overline{\mathbf{H}}\mathbf{F}\mathbf{w} = \mathbf{0}_{\overline{n}} \bmod q \end{array} \right\} \end{array} \right\},$$

[17] We exploit that the challenges are *uniformly* distributed (conditioned on accepting).

$$\Xi^{\mathsf{ip}\vee\mathsf{sis}}_{\mathcal{R},q,m,n^{\mathsf{out}},r,\mu,\beta,\beta^{\mathsf{sis}},\alpha} := \left\{ \begin{array}{l} ((\mathbf{H},\mathbf{F},\mathbf{Y},\mathbf{c},t),\mathbf{W} \text{ or } \mathbf{w}): \\ \quad \mathbf{H} \in \mathcal{R}_q^{n^{\mathsf{out}}\times n};\ \mathbf{F} \in \mathcal{R}_q^{n\times d^{\otimes\mu}} \subseteq \mathcal{R}_q^{n\times m};\ \mathbf{Y} \in \mathcal{R}_q^{n^{\mathsf{out}}\times r};\ \mathbf{c} \in \mathcal{R}_q^r \text{ s.t. } \alpha(\mathbf{c}) = \mathbf{c} \\ \quad t \in \mathcal{R};\ \mathbf{W} = (\mathbf{w}_i)_{i=0}^r \in \mathcal{R}^{m\times r};\ \mathbf{w} \in \mathcal{R}^m \\ \quad \left\{ \begin{array}{c} \|\mathbf{W}\| \le \beta \\ \mathbf{HFW} = \mathbf{Y} \bmod q \\ \sum_{i=0}^r c_i \langle \mathbf{w}_i, \alpha(\mathbf{w}_i)\rangle_{\mathcal{R}} = t \bmod q \end{array} \right\} \text{ or } \left\{ \begin{array}{c} \|\mathbf{w}\| \le \beta^{\mathsf{sis}} \\ \overline{\mathbf{H}}\mathbf{F}\mathbf{w} = \mathbf{0}_{\overline{n}} \bmod q \end{array} \right\} \end{array} \right\}.$$

Note that, compared to Ξ^{lin}, the norm relation Ξ^{norm} differs in that a witness norm bound $\nu \le \beta$ is given as part of the statement, and a stricter weighted norm relation $\sum_{i=0}^r c_i\|\mathbf{w}_i\| \le \nu$ is checked. Similarly, Ξ^{ip} differs from Ξ^{lin} in that the statement additionally includes an inner product value t, and the witness additionally satisfies a weighted inner product relation. Furthermore, we note that Ξ^{ip} is parametrised by $\alpha \in \{\mathsf{id}, \overline{\mathsf{id}}\}$ which is either the identity or complex conjugate, controlling which type of inner product is being considered. We require that the weights are invariant under α, i.e. $\alpha(\mathbf{c}) = \mathbf{c}$.

The protocols Π^{norm} and Π^{ip}. The protocols Π^{norm}, Π^{ip} for $\alpha = \mathsf{id}$ and Π^{ip} for $\alpha = \overline{\mathsf{id}}$ are very similar. In the following description we focus on Π^{ip} for $\alpha = \overline{\mathsf{id}}$. Removing all conjugates yields the protocol Π^{ip} for $\alpha = \mathsf{id}$. The protocol Π^{norm} can be obtained by letting the verifier compute the trace of the alleged inner product.

Our approach is based on polynomial identities. For $\mathbf{w} \in \mathcal{R}^m$, define the polynomials

$$g_{\mathbf{w}}(X) = \sum_{j\in[m]} w_j X^j \quad \text{resp.} \quad \bar{g}_{\mathbf{w}}(X) = \sum_{j\in[m]} \bar{w}_j X^j \tag{6}$$

and observe that $\bar{g}_{\mathbf{w}} = g_{\bar{\mathbf{w}}}$ and that the Laurent polynomial

$$L(X) = \sum_{i\in\pm[m]} v_i X^i := g_{\mathbf{w}}(X) \cdot \bar{g}_{\mathbf{w}}(X^{-1}) \tag{7}$$

has constant coefficient $\langle \mathbf{w}, \bar{\mathbf{w}}\rangle_{\mathcal{R}}$. Also, observe that

$$v_k = \sum_{i-j=k} v_i\bar{v}_j = \overline{\mathsf{id}}\left(\sum_{i-j=k} \bar{v}_i v_j\right) = \overline{\mathsf{id}}\left(\sum_{j-i=k} \bar{v}_j v_i\right) = \bar{v}_{-k}$$

where $v_k := 0$ if $|k| \ge m$.We exploit this symmetry to commit to $L(X)$ by committing only to $(v_0, \ldots, v_{m-1})$. Setting

$$h(X) = \sum_{i\in[m]} v_i X^i \quad \text{resp.} \quad \bar{h}(X) = \sum_{i\in[m]} \bar{v}_i X^i$$

we see that

$$L(X) = h(X) + \bar{h}(X^{-1}) - v_0.$$

We use this equality to prove the polynomial identity in Eq. (7) between $\mathbf{v}$ and $\mathbf{W}$ by evaluating $g, \bar{g}, h, \bar{h}$ at a random point $\xi \leftarrow\!\!\$\ \mathcal{C}_{\mathcal{R}_q}$ (and checking if $v_0 = t$).

To generalise the above to the weighted inner product $v_0 = \sum_{j\in[r]} c_j\langle \mathbf{w}_j, \overline{\mathbf{w}}_j\rangle_{\mathcal{R}}$ for a matrix witness $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_r) \in \mathcal{R}^{m\times r}$ with weights $\mathbf{c}$ that satisfy $\mathbf{c} = \overline{\mathbf{c}}$, we apply the above approach component-wise, and then compute the weighted sum. Consequently, we set

$$\begin{aligned} L(X) = \sum_{i\in\pm[m]} v_i X^i &:= \sum_{j\in[rep]} c_j L_j(X) \\ \text{where} \quad L_j(X) = \sum_{i\in\pm[m]} v_{j,i} X^i &= g_{\mathbf{w}_j}(X) \cdot \bar{g}_{\mathbf{w}_j}(X^{-1}) \end{aligned} \tag{8}$$

and observe that the constant coefficient v_0 of $L(X)$ is now $\sum_{j\in[r]} c_j\langle \mathbf{w}_j, \bar{\mathbf{w}}_j\rangle_{\mathcal{R}}$. Since we require $\mathbf{c} = \alpha(\mathbf{c})$ from the weights $\mathbf{c}$, we still have $v_k = \bar{v}_{-k}$. Thus, we can again define

$$h(X) = \sum_{i\in[m]} v_i X^i \quad \text{resp.} \quad \bar{h}(X) = \sum_{i\in[m]} \bar{v}_i X^i \tag{9}$$

and obtain again the equality

$$L(X) = h(X) + \bar{h}(X^{-1}) - v_0. \tag{10}$$

where $v_0 = \sum_{j\in[r]} c_j \langle \mathbf{w}_j, \bar{\mathbf{w}}_j \rangle_{\mathcal{R}}$. Thus, we extended the check from a vector $\mathbf{w}$ to a matrix $\mathbf{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_r)$, by considering $g_{\mathbf{w}_j}, \bar{g}_{\mathbf{w}_j}$ and L_j first component-wise, and then summing up those components with weights c to obtain L (and the symmetry decomposition h of L).

If the above check based on polynomial identities is used naively, a problem occurs: The terms v_i have the norm of the individual coefficients bounded by β^2, so $\|\mathbf{v}\|$ may be beyond the threshold for which the commitment is binding.

A natural approach is to run $\Pi^{b\text{-decomp}}$ to counteract this problem. However, doing so modularly comes at the cost of a suboptimal relaxed knowledge guarantee. We can tighten our analysis if we treat the composition with $\Pi^{b\text{-decomp}}$ as *within* the protocol Π^{ip}, i.e. we immediately send the decomposed (and binding) commitments. The reason is a technical artefact of relaxed knowledge soundness and reductions of knowledge: Relaxed soundness in $\Pi^{b\text{-decomp}}$ incurs a large factor of norm growth when extracting the witness. However, in Π^{ip}, the auxiliary commitment to $\mathbf{v}$ is *not* part of the witness (yet), and we need not extract it. Thus, we argue directly for the decomposition of $\mathbf{v}$ into smaller $\mathbf{V} = (\mathbf{v}_i)_i$ of norm $\approx \beta$. This avoids treating $\mathbf{v}$ as a witness in $\Pi^{b\text{-decomp}}$, which significantly improves the parameters. This optimised protocol is presented in Fig. 7

Lemma 8 (Norm and Inner Product). *Let $m, r, \ell, b_{\mathsf{ip}} \in \mathbb{N}$, $0 \le \beta \le \beta^{\mathsf{sis}} \le q$, and $0 \le 2\beta' \le \beta^{\mathsf{sis}}$. Protocol Π^{ip} is a perfectly correct reduction of knowledge from Ξ^{ip} to Ξ^{lin}*

$$\left(m, n^{\mathsf{out}}, r, \beta\right) \mapsto \left(m, n^{\mathsf{out}\prime}, r+\ell, \beta_{\mathsf{out}}\right)$$

where $\beta_{\mathsf{out}} = \sqrt{\beta^2 + {\beta_{\mathbf{V}}}^2}$ (resp. $\beta_{\mathsf{out}} = \max\{\beta, \beta_{\mathbf{V}}\}$), and knowledge sound reduction of knowledge from Ξ^{ip} to Ξ^{lin} with parameters

$$\left(m, n^{\mathsf{out}}, r, \beta', \beta^{\mathsf{sis}}\right) \leftarrow\!\shortmid \left(m, n^{\mathsf{out}\prime}, r+\ell, {\beta_{\mathsf{out}}}', \beta^{\mathsf{sis}}\right),$$

where $n^{\mathsf{out}\prime} = n^{\mathsf{out}} + 3$, b_{ip} and $\ell \ge \log_{b_{\mathsf{ip}}}(2{\beta_{\mathbf{V}}}^2 + 1)$ is such that $b_{\mathsf{ip}} \le 2\beta_{\mathbf{V}} / (\sqrt{\ell m}\sqrt{\hat{\mathsf{f}}\varphi})$ (resp. $b_{\mathsf{ip}} = 2\beta_{\mathbf{V}} + 1$ and $\ell \ge \log_{b_{\mathsf{ip}}}(2\beta_{\mathbf{V}} + 1)$). Extraction requires two uniformly distributed transcripts (Footnote 11) and has knowledge error $\kappa \le \frac{2m}{|\mathcal{C}_{\mathcal{R}_q}|}$.

Let $\beta'^2 < q/2$. For Π^{norm}, the analogous statements hold and additionally Π^{norm} is a knowledge sound reduction of knowledge from Ξ^{norm} to Ξ^{lin} with parameters

$$\left(m, n^{\mathsf{out}}, r, \boxed{\nu}_{\|\sigma(\cdot)\|_2} \left(\text{resp.}, \boxed{\sqrt{\hat{\mathsf{f}}\varphi}\sqrt{mr}\,\nu}_{\|\psi(\cdot)\|_\infty}\right), \beta^{\mathsf{sis}}\right) \leftarrow\!\shortmid \left(m, n^{\mathsf{out}\prime}, r+\ell, \beta', \beta^{\mathsf{sis}}\right).$$

Proof. For the norm $\|\sigma\left(\widetilde{\mathbf{W}}\right)\|_2$, observe that $\|\sigma\left(\widetilde{\mathbf{W}}\right)\|_2^2 = \|\sigma\left((\mathbf{V}, \mathbf{W})\right)\|_2^2 = \|\sigma\left(\mathbf{V}\right)\|_2^2 + \|\sigma\left(\mathbf{W}\right)\|_2^2$, where $\|\sigma\left(\mathbf{W}\right)\|_2^2 \le \beta^2$ by assumption on $\mathbf{W}$ and $\|\sigma\left(\mathbf{V}\right)\|_2^2 \le {\beta_{\mathbf{V}}}^2$ by definition of b_{ip} and the bounds for $\Pi^{b\text{-decomp}}$ from Lemma 4. (We set b_{ip} such that $\frac{1}{2}\sqrt{\ell m}\sqrt{\hat{\mathsf{f}}\varphi}\, b_{\mathsf{ip}} \le \beta_{\mathbf{V}}$ holds by definition.)

For the norm $\|\psi\left(\widetilde{\mathbf{W}}\right)\|_\infty$, observe that $\|\psi\left(\widetilde{\mathbf{W}}\right)\|_\infty = \|\psi\left((\mathbf{V}, \mathbf{W})\right)\|_\infty = \max\{\|\psi\left(\mathbf{V}\right)\|_\infty, \|\psi\left(\mathbf{W}\right)\|_\infty\}$, where $\|\psi\left(\mathbf{W}\right)\|_\infty \le \beta$ by assumption on $\mathbf{W}$ and $\|\psi\left(\mathbf{V}\right)\|_\infty \le \beta_{\mathbf{V}}$ by definition of $b_{\mathsf{ip}} = 2\beta_{\mathbf{V}} + 1$ and the bounds for $\Pi^{b\text{-decomp}}$ from Lemma 4.

For correctness, observe that $\mathbf{EW}$ contains the evaluations $z_{+1,j} = \sum_i w_{i,j}\xi^i = g_{\mathbf{w}_j}(\xi)$ and $z_{-1,j} = \sum_i w_{i,j}\xi^{-i} = \overline{g_{\mathbf{w}_j}(\xi^{-1})}$, and the component $w_{0,j}$ which we ignore. We can thus compute

$$\sum_{j\in[r]} c_j g_{\mathbf{w}_j}(\xi) \cdot \bar{g}_{\mathbf{w}_j}(\xi^{-1}) = \mathbf{c}^{\mathsf{T}}(\mathbf{z}_{+1} \odot \bar{\mathbf{z}}_{-1}).$$

Similarly, $\mathbf{EV}(1, b, \ldots, b^{\ell-1}) = \mathbf{Ev}$ contains $z'_{+1} = \sum_i v_i\xi^i$ and $z'_{-1} = \sum_i v_i\bar{\xi}^{-i}$ and v_0. By the symmetry property $L(\xi) = \sum_{i\in[m]} v_i\xi^i - v_0 =$ we recover $L(\xi) = z'_{+1} + \bar{z}'_{-1} - t'$. Thus, correctness holds by the polynomial identities explained above, specifically Eqs. (8) to (10). As we already showed that the norm bounds are respected for the output, we have shown that the reduction of knowledge is perfectly complete.

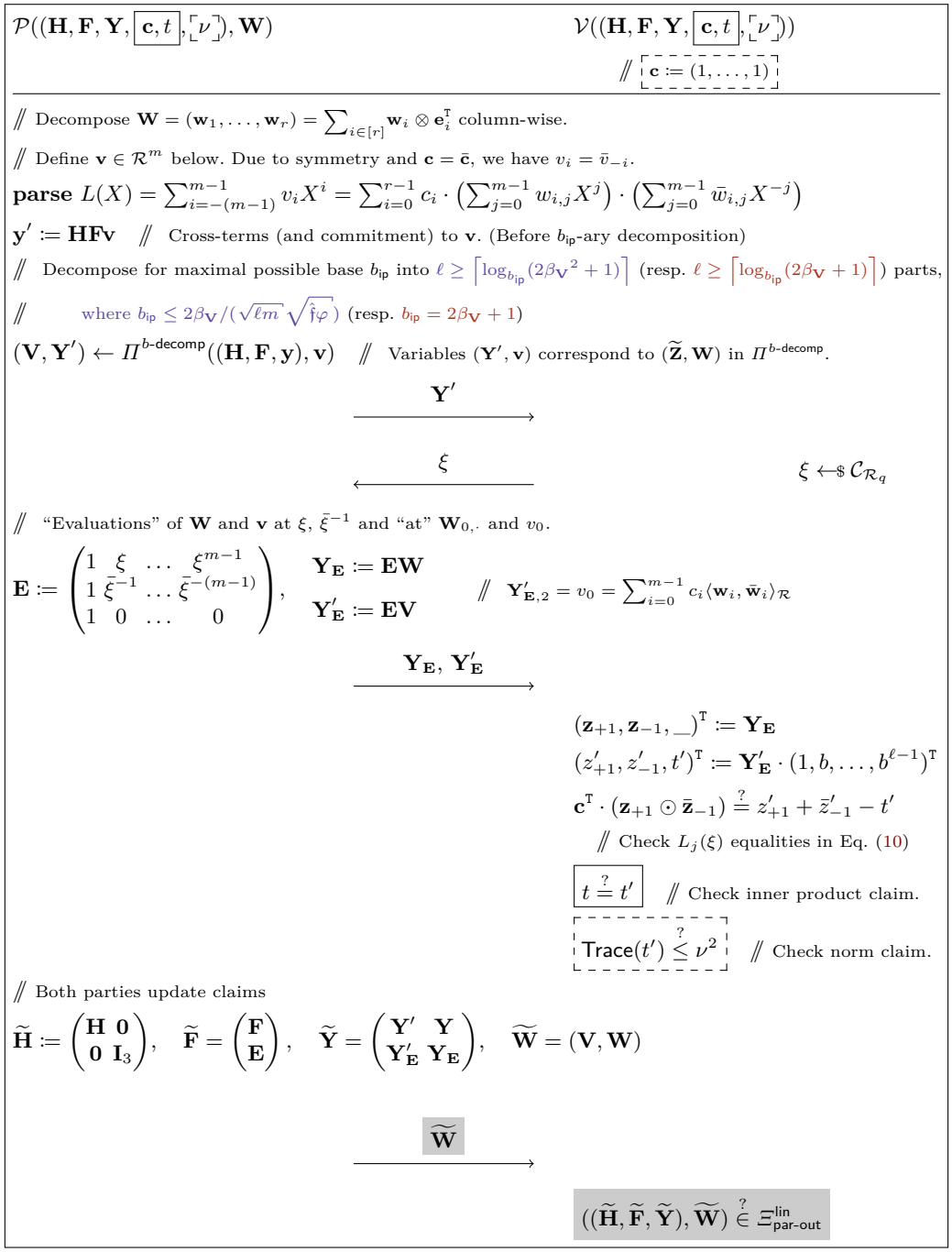

Fig. 7. Protocol $\boxed{\Pi^{\mathsf{ip}}}$ or $\lceil\Pi^{\mathsf{norm}}\rfloor$, a reduction of $\boxed{\Xi^{\mathsf{ip}}_{\mathsf{par\text{-}in}}}$ or $\lceil\Xi^{\mathsf{norm}}_{\mathsf{par\text{-}in}}\rfloor$ to $\Xi^{\mathsf{lin}}_{\mathsf{par\text{-}out}}$ with par-in, par-out specified in Lemma 8, with optimisation to directly include $\Pi^{b\text{-decomp}}$, presented for $\alpha = \overline{\mathsf{id}}$. To obtain $\alpha = \mathsf{id}$ all conjugates are removed. Π^{ip} sends the marked parts only as a *proof* (but not *reduction*) of knowledge.

Now, we show relaxed knowledge soundness. First, observe that, as argued above, given fixed $\mathbf{v}$ which defines $h(X) = \sum_{i=0}^{m-1} v_i X^i$, then the probability that

$$L(X) \neq h(X) + \bar{h}(X^{-1}) - v_0 \quad \text{but} \quad L(\xi) \neq h(\xi) + \bar{h}(\xi^{-1}) - v_0 \tag{11}$$

holds, is bounded by $\frac{2m-1}{|\mathcal{C}_{\mathcal{R}_q}|} \leq \frac{2m}{|\mathcal{C}_{\mathcal{R}_q}|} = \kappa$, where $\xi \leftarrow\!\!\$\, \mathcal{C}_{\mathcal{R}_q}$. (By the lemma of Schwartz–Zippel, analogous to Lemma 7.) However, this is a soundness argument for *fixed* polynomials, but the vector $\mathbf{v} = \mathbf{V} \cdot (1, \ldots, b_{\mathsf{ip}}^{\ell-1})$ is only determined in the last step, Now, we argue analogous to Lemma 7 to turn soundness into knowledge soundness: Let $\mathbf{V}^{(0)}$ with $\widetilde{\mathbf{H}}\widetilde{\mathbf{F}}\mathbf{V}^{(0)} = \widetilde{\mathbf{Y}}'^{(0)}$ denote the (first) preimage (from the two accepting transcripts). Let $\widetilde{\mathbf{y}}^{(0)} = \widetilde{\mathbf{Y}}'^{(0)} \cdot (1, \ldots, b_{\mathsf{ip}}^{\ell-1})$ denote the non-decomposed linear claim for $\mathbf{V}$. Observe that, unless the polynomial identity holds, only fraction κ of challenges can satisfy (11) for any fixed $\mathbf{V}^{(0)}$. Thus, if the adversary succeeds with probability ϵ, then with probability $\epsilon - \kappa$, the 2-transcript extractor gives two transcripts where[18] either one $\mathbf{V}^b$ satisfies the polynomial identity (and we're done), or $\mathbf{V}^{(0)} \neq \mathbf{V}^{(1)}$. Let $\mathbf{U} = \mathbf{V}^{(0)} - \mathbf{V}^{(1)}$. Then we have $\mathbf{HFU} = 0$, as $\mathbf{HFV}^{(j)} = \mathbf{Y}'_i$ holds for both $j = 0, 1$. This yields a witness for the OR-branch of $\Xi^{\mathsf{ip} \vee \mathsf{sis}}$ with norm at most $2\beta' \leq \beta^{\mathsf{sis}}$. (Note here that the vector $\mathbf{v}$ (which we decompose as $\mathbf{V}$) is *not* part of the initial witness, so its norm is of no concern during extraction and we avoid the large growth of β'_0 for decomposition in Lemma 4.)

To handle the norm protocol, we note that $\sum_i x_i \bar{x}_i = \|\sigma(\mathbf{x})\|_2 \bmod q$ and $(\sum_i x_i \bar{x}_i)^2 = \|\sigma(\mathbf{x})\|_2^2 \bmod q$ Further, as $\beta'^2 < q/2$, $\sum_i x_i \bar{x}_i = \|\sigma(\mathbf{x})\|_2$ holds with no modulus wrap-around and $\|\psi(\mathbf{x})\|_\infty \leq \sqrt{\hat{\mathsf{f}}\varphi}\sqrt{mr}\|\sigma(\mathbf{x})\|_2$ for $\mathbf{x} \in \mathcal{R}^{mr}$ by Lemma 1 and the standard inequality between ∞-norm and 2-norm. □

π	$m \mapsto m'$	$r_{\mathsf{in}} \mapsto r_{\mathsf{out}}$	$\beta_0 \mapsto \beta_1$	$\beta'_1 \mapsto \beta'_0$	κ	#tr	Condition	Reference
$\Pi^{b\text{-decomp}}$	1	ℓr_{in}	$(0, \sqrt{\ell r_{\mathsf{in}} m}\sqrt{\hat{\mathsf{f}}\varphi}\, b/2)$	$\frac{b^\ell - 1}{b-1}$	0	1	$\ell = \lceil \log_b(2\beta_0 + 1) \rceil$	4
Π^{split}	$1/d$	dr_{in}	β_0	1	0	1		5
Π^{fold}	1	r_{out}	$(\sqrt{r_{\mathsf{out}} r_{\mathsf{in}}}\gamma, 0)$	$2\sqrt{r_{\mathsf{in}}}\theta$	$\frac{r_{\mathsf{in}}}{\lvert\mathcal{C}_{\mathcal{R}}\rvert^{r_{\mathsf{out}}}}$	$r_{\mathsf{in}} + 1$		6
Π^{batch}	1	r_{in}	β_0	1	$rn/\lvert\mathcal{C}_{\mathcal{R}_q}\rvert$	2	$2\beta'_1 \leq \beta^{\mathsf{sis}}$	7
Π^{ip}	1	$\ell + r_{\mathsf{in}}$	$\sqrt{\beta_0^2 + \beta_{\mathbf{v}}^2}$	1	$2m/\lvert\mathcal{C}_{\mathcal{R}_q}\rvert$	2	$2\beta'_0 \leq \beta^{\mathsf{sis}}$	8
Π^{norm}	1	$\ell + r_{\mathsf{in}}$	$\sqrt{\beta_0^2 + \beta_{\mathbf{v}}^2}$	1	ν/β'_1	2	$2\beta'_0 \leq \beta^{\mathsf{sis}}$	8

Table 1. Parameters of protocols expressed in the canonical 2-norm. Expressed as $\beta_1 = f(\beta_0)$ for correctness when starting from β_0, and as $\beta'_0 = g(\beta'_1)$ and $\beta'_0 \leq \beta^{\mathsf{sis}}$ for relaxed soundness when guaranteed β'_1, knowledge error κ, number $\#tr$ of transcripts to extract, and other variables or important constraints. Full details in are in the respective theorems.

6 Succinct Arguments for Bounded-Norm Satisfiability: Composition

In this section, we discuss how to compose the atomic protocols constructed in Section 5 to obtain asymptotically and concrete efficient succinct arguments for the principal relation Ξ^{lin} which does not have any correctness and soundness gaps. We begin in Section 6.1 by overviewing the functionality of each atomic protocol $\Pi^{b\text{-decomp}}, \Pi^{\mathsf{split}}, \Pi^{\mathsf{fold}}, \Pi^{\mathsf{batch}}$ and Π^{norm} and discussing considerations when composing them.[19] We also introduce a dummy protocol Π^{finish} which implements the trivial step of sending the witness in plain. We then provide an example composition Section 6.2 which, although not necessarily optimal in terms of concrete efficiency, serves as a baseline for compositions which we consider reasonable.

6.1 General Composition Strategy

We discuss how to compose the atomic protocols $\Pi^{b\text{-decomp}}, \Pi^{\mathsf{split}}, \Pi^{\mathsf{fold}}, \Pi^{\mathsf{batch}}$ and Π^{norm} to obtain succinct arguments for Ξ^{lin} (or more precisely $\Xi^{\mathsf{lin} \vee \mathsf{sis}}$, see below).

[18] We exploit that the challenges are *uniformly* distributed (conditioned on accepting).

[19] We note that Π^{ip} is not necessary for obtaining succinct arguments for Ξ^{lin} without correctness and soundness gaps, but is instead used in Sections 7 and 8 for more complex relations.

Bird-eye view of principal relation. Recall that in Ξ^{lin} a statement $(\mathbf{H}, \mathbf{F}, \mathbf{Y})$ and a witness $\mathbf{W}$ satisfy the relation

$$\mathbf{HFW} = \mathbf{Y} \bmod q \qquad \text{and} \qquad \|\mathbf{W}\| \leq \beta$$

where $\mathbf{H} \in \mathcal{R}_q^{n^{\mathsf{out}} \times n}$ consists of a top part $\overline{\mathbf{H}} = (\mathbf{I}_{\overline{n}}\ \mathbf{0}) \in \mathcal{R}_q^{\overline{n} \times n}$ and a bottom part $\underline{\mathbf{H}} \in \mathcal{R}_q^{\underline{n} \times n}$, $\mathbf{F} \in \mathcal{R}_q^{n \times d^{\otimes \mu}} \subseteq \mathcal{R}_q^{n \times m}$, $\mathbf{Y} \in \mathcal{R}_q^{n^{\mathsf{out}} \times r}$ and $\mathbf{W} \in \mathcal{R}^{m \times r}$. Splitting $\mathbf{F}$ and $\mathbf{Y}$ into a top part $(\overline{\mathbf{F}}, \overline{\mathbf{Y}}) \in \mathcal{R}_q^{\overline{n} \times d^{\otimes \mu}} \times \mathcal{R}_q^{\overline{n} \times r}$ and a bottom part $(\underline{\mathbf{F}}, \underline{\mathbf{Y}}) \in \mathcal{R}_q^{\underline{n} \times d^{\otimes \mu}} \times \mathcal{R}_q^{\underline{n} \times r}$, we can equivalently write the above relation as

$$_{\overline{n}}\{\overbrace{\overline{\mathbf{F}}}^{m=d^{\otimes\mu}}\overbrace{\mathbf{W}}^{r} = \overbrace{\overline{\mathbf{Y}}}^{r} \bmod q, \quad _{\underline{n}}\{\overbrace{\underline{\mathbf{H}}}^{n}\overbrace{\mathbf{F}}^{m=d^{\otimes\mu}}\overbrace{\mathbf{W}}^{r} = \overbrace{\underline{\mathbf{Y}}}^{r} \bmod q, \quad \text{and} \quad _{m\,=\,d^{\otimes\mu}}\{\overbrace{\|\mathbf{W}\|}^{r} \leq \beta. \tag{12}$$

To catch the exception of a malicious prover succeeding in violating soundness through solving vSIS, the OR-branch of $\Xi^{\mathsf{lin} \vee \mathsf{sis}}$ allows an alternative witness $\mathbf{w} \in \mathcal{R}^m$ satisfying

$$_{\overline{n}}\{\overbrace{\overline{\mathbf{F}}}^{m=d^{\otimes\mu}}\mathbf{w} = \mathbf{0} \bmod q \quad \text{and} \quad _{m\,=\,d^{\otimes\mu}}\{\|\mathbf{w}\| \leq \beta^{\mathsf{sis}}. \tag{13}$$

Since all atomic protocols $\Pi^{b\text{-}\mathsf{decomp}}, \Pi^{\mathsf{split}}, \Pi^{\mathsf{fold}}, \Pi^{\mathsf{batch}}$ and Π^{norm} (and also Π^{ip}) preserve the norm bound β^{sis} for the OR-branch, when choosing parameters it suffices to ensure the hardness of finding $\mathbf{w}$ satisfying Eq. (13). We therefore omit the discussion about the OR-branch and the parameter β^{sis} below.

Keeping track of composition costs. The basic idea of obtaining a succinct argument protocol for Ξ^{lin} is to compose the self-reductions of knowledge $\Pi^{\mathsf{split}}, \Pi^{\mathsf{fold}}$ and Π^{batch} to reduce the witness dimensions $m \times r$ so that the resulting witness is small enough in description size to be sent in plain. We denote this last step as Π^{finish}. As highlighted in Sections 1 and 2 this would result in an argument with both correctness and soundness gaps. To recall, the correctness gap refers to the growth of the norm bound β of the running witness, while the soundness gap refers to the norm of a witness extracted by the knowledge extractor. We denote the latter by β^{ext}.

To eliminate the correctness gap, the idea is to throw the self-reduction of knowledge $\Pi^{b\text{-}\mathsf{decomp}}$ into the mix, so that the norm bound β of the running witness is controlled throughout the composition. It remains to remove the soundness gap. For this, we let the prover send out explicit norm claims ν for the running witness from time to time, which expands the running Ξ^{lin} relation into a Ξ^{norm} relation, and run Π^{norm} to reduce this Ξ^{norm} relation back to a Ξ^{lin} relation. We will assume that the prover is rational and always sets $\nu := \beta$, i.e. making the tightest claim possible about the norm of the running witness. The effect of this procedure is, therefore, to insert a "checkpoint" into the composition, so that the witness extracted at this step of the composition is "reset" to $\beta^{\mathsf{ext}} = \beta$ (assuming that the norm of the running witness does not exceed the allowed boundary in subsequent protocols).

When composing the above atomic RoKs, we have to keep track of the following for each atomic RoK:

- The changes to the parameters $\underline{n}, n, m, r, \beta$ so that the hypothetical norm bound β^{ext} of the extracted witness (and hence the norm bound β of the running witness) do not exceed the allowed budget β^{sis}, i.e. to ensure $(\beta \leq)\beta^{\mathsf{ext}} < \beta^{\mathsf{sis}}/2$.
- The soundness cost, i.e. how much soundness error does a RoK add to the overall composition, so that the cumulative soundness cost does not exceed the allowed budget, say 2^{-80}.
- The communication cost.

We note that, somewhat confusingly, the hypothetical norm bound β^{ext} of the extracted witness is not a function of the parameters of preceding protocols in the chain of composition, but rather a function of the parameters of subsequent protocols. In other words, as we insert more protocols into the composition, the β^{ext} values of all previous protocols may change.

The overall communication cost of the composition, which is a natural target for optimisation, is the sum of the communication costs of all instances of atomic protocols involved plus the size of the final witness. We note, however, that each atomic protocol also has different prover and verifier time costs which are harder to keep track. In the following, we focus only on minimising the communication cost for simplicity.

Composition strategy for minimising communication. We next discuss a natural (but not necessarily optimal) strategy for minimising the overall communication cost. To aid reasoning, we first give intuitive descriptions of the functionality of each atomic RoK. We will omit mention of soundness costs since parameters can be easily set to make them negligible.

- $\Pi^{b\text{-decomp}}$: Shrink β but grow r and β^{ext}. Cost $(\ell-1)n^{\mathsf{out}}r$ of $\mathcal{R}_q$ communication.
- Π^{split}: Shrink m but grow r. Cost $r((d^2-1)\cdot(n^{\mathsf{out}}-\overline{n})+(d-1)\cdot\overline{n})$ of $\mathcal{R}_q$ communication.
- Π^{fold}: Reset r to some fixed (e.g. the initial) value but grow β and β^{ext}. Cost 0 communication.
- Π^{batch}: Shrink $\underline{n}$ to 1. Cost 0 communication.
- Π^{norm}: Reset β^{ext} to β but grow $\underline{n}$, r and β (for the next round). Cost $\ell n^{\mathsf{out}}+3\cdot(r+\ell)$ of $\mathcal{R}_q$ communication.
- Π^{finish}: Finish the composition. Cost $mr\varphi\log\beta$ communication.

We make several observations. Each protocol except Π^{batch} and Π^{finish} shrinks one of the parameters at the cost of growing others. The Π^{batch} protocol is essentially free (ignoring soundness cost)[20], shrinking $\underline{n}$ while costing no parameter growth nor communication, and can therefore always be run immediately after Π^{norm} to suppress the growth of $\underline{n}$ there. The Π^{split} protocol trades m for r and thus does not reduce the witness size, i.e. the communication cost of Π^{finish}.

With the above observations, a natural composition strategy is to split the protocol into 2 phases – looping and finishing. We first define the 2 phases and then provide an explanation.

(i) Looping phase: Repeat the sequence

$$(\Pi^{b\text{-decomp}}\rightarrow)\Pi^{\mathsf{norm}}\rightarrow\Pi^{\mathsf{batch}}\rightarrow\Pi^{\mathsf{split}}\rightarrow\Pi^{\mathsf{fold}},$$

where the optional step is specified in parenthesis. After each loop, check what would be the overall communication cost if Π^{finish} is run now. Exit the loop if the overall communication cost does not decrease if another loop is executed.

(ii) Finishing phase: Execute Π^{finish}.

In the beginning of the looping phase, we start with Π^{norm} (as the $\Pi^{b\text{-decomp}}$ is unnecessary) to create a checkpoint for $\beta^{\mathsf{ext}}=\beta$, so that the final extracted witness is guaranteed to be of norm at most β, i.e. without soundness gap, given that the witness norm does not blow up in subsequent protocols. As observed above, Π^{norm} should be followed by Π^{batch} to negate the growth of $\underline{n}$. We run Π^{split} to trade m for r, followed by Π^{fold} to shrink r to the initial value at the cost of growing β and β^{ext}. At this point, the norm of the running witness is possibly quite large. Therefore, we potentially insert a $\Pi^{b\text{-decomp}}$ step at the beginning of the next loop to control the norm β of the running witness[21].

At the end of each loop, if the hypothetical exiting cost does not decrease in further loops, i.e. the overall communication cost if Π^{finish} is run now is not higher than running it later, there is no reason to continue looping. We therefore execute Π^{finish} to finish the protocol.

6.2 Asymptotic Complexity

For the asymptotic parameters, we assume a slightly different composition than suggested in Section 6.1. The looping phase is now defined with the following sequence:

$$\Pi^{\mathsf{norm}}\rightarrow\Pi^{\mathsf{batch}}\rightarrow\Pi^{b\text{-decomp}}\rightarrow\Pi^{\mathsf{split}}\rightarrow\Pi^{\mathsf{fold}},$$

repeated μ times. For simplicity, $\Pi^{b\text{-decomp}}$ is included in each round. Such ordering, although not optimal, is easier to analyse as we can assume that the final and the initial norm are identical. However, it yields slightly worse concrete proof sizes (cf. Section 9). Conveniently, all the bounds are tracked according to the canonical 2-norm. The parameters are chosen as shown in Table 2 and argued below.

[20] We highlight that in some cases, it might be beneficial to omit Π^{batch}, or in other words, to include "bottom" rows into "top" rows. This is because Π^{split} communicates $O(d^2)$ elements for each "bottom" row but only $O(d)$ elements for each "top" row (cf. Fig. 4).

[21] The amount that β shrinks depends on the parameter choice of $\Pi^{b\text{-decomp}}$. We note that it is not always optimal (in communication cost) to shrink β all the way back to the initial witness norm, as this may incur high communication costs.

Parameters	description	instantiation
λ	security parameter	–
m	height of the witness matrix $\mathbf{W}$	–
q	argument system modulus	$\mathsf{poly}(\lambda, m)$
β^{sis}	norm of the (presumably hard) vSIS instance	$\mathsf{poly}(\lambda, m)$
$\mathfrak{f}$	conductor of the ring	$O(\lambda \log m / \log \lambda)$
φ	ring dimension	$\varphi(\mathfrak{f})$
r_4	soundness amplification factor	$O(\lambda / \log \lambda)$
r	width of the witness matrix $\mathbf{W}$	r_4
n	height of the matrix $\mathbf{F}$	1
μ	total number of invocations of the protocol	$\log_d m$
$\overline{n}$	number of top rows of $\mathbf{H}$	1
$\underline{n}$	number of bottom rows of $\mathbf{H}$	1
n^{out}	number of rows of $\mathbf{H}$	$\overline{n} + \underline{n}$
e	number of irreducible factors	$\omega(1)$
β_0	initial norm bound	$\mathsf{poly}(\lambda, m)$
r_0	initial width of the witness $\mathbf{W}$	$O(\lambda / \log \lambda)$
b_{n}	decomposition base for the coeffs. of the Laurent poly.	β_1
ℓ_{n}	length of the decomposition basis w.r.t.	$O(1)$
b_{d}	decomposition base	$O\left(\left(m\hat{\mathfrak{f}}^2(\lambda/\log\lambda)^{5/2}\right)^{1/\Theta(1)}\right)$
ℓ_{d}	length of the decomposition basis w.r.t. b_{d}	$O(1)$
d	folding factor	$O(1)$

Table 2. Parameter instantiation for the protocol.

Hardness of SIS. To measure the hardness of vSIS, we heuristically assume that it is as hard as the plain SIS problem for the dimension $\varphi = \varphi(\mathfrak{f})$. To measure the hardness of SIS, we first translate the canonical norm $\|\sigma(\cdot)\|_2$ into the Euclidean norm $\|\psi(\cdot)\|_2$, and then follow the heuristic methodology from [MR09]. Let $b = O(\lambda)$ be the block size of the BKZ algorithm to find a short vector in the corresponding q-ary lattice for SIS (cf. [BDGL16]). Define the root Hermite factor as

$$\delta_{\mathsf{rhf}} = \left(\frac{b(\pi b)^{1/b}}{2\pi e}\right)^{1/(2(b-1))}.$$

Then, SIS with matrix dimensions $\varphi \times \varphi m$ and Euclidean norm β^{sis} is hard when

$$\beta^{\mathsf{sis}} < \min\left(2^{2\sqrt{\varphi \log q \log \delta_{\mathsf{rhf}}}}, q\right).$$

By rearranging, we get that

$$\varphi \log q > \frac{\log^2 \beta^{\mathsf{sis}}}{4 \log \delta_{\mathsf{rhf}}}.$$

Note that

$$\log \delta_{\mathsf{rhf}} = \frac{1}{2(b-1)} \log\left(\frac{b(\pi b)^{1/b}}{2\pi e}\right) = \Theta\left(\frac{\log b}{b}\right) = \Theta\left(\frac{\log \lambda}{\lambda}\right).$$

This means that the size of a single $\mathcal{R}_q$ element is asymptotically

$$\Omega\left(\frac{\lambda \cdot (\log m + \log \lambda)^2}{\log \lambda}\right) = \Omega\left(\frac{\lambda \cdot \log^2 m}{\log \lambda}\right).$$

We will assume this size of an $\mathcal{R}_q$ in the analysis below.

Round Communication Complexity. To aid discussion, we introduce auxiliary variables for keeping track of how the parameters of the Ξ^{lin} relation change throughout a single loop of the protocol.

$$(m_0, n_0, r_0, \beta_0) \xrightarrow{\Pi^{\mathsf{norm}}} (m_0, n_1, r_1, \beta_1) \xrightarrow{\Pi^{\mathsf{batch}}} (m_0, n_0, r_1, \beta_1)$$

$$\xrightarrow{\Pi^{b\text{-decomp}}} (m_0, n_0, r_2, \beta_2) \xrightarrow{\Pi^{\mathsf{split}}} (m_1, n_0, r_3, \beta_2) \xrightarrow{\Pi^{\mathsf{fold}}} (m_1, n_0, r_4, \beta_3).$$

We will use the parameters for analysing the communication complexity. For simple recursion, we assume that parameters are chosen so that $\beta_0 = \beta_3$.

For Π^{norm}, we set $\beta_1 = O(\sqrt{mr_1\hat{\mathsf{f}}\varphi}\,\beta_0)$. To argue that the norm grows by that factor, we observe that the coefficient ∞-norm of $\mathbf{v}$ (cf. Lemma 8) is at most β_1^2 (cf. Corollary 1), and therefore the matrix of such norm can be decomposed into matrix of coefficient ∞-norm β_0 with a constant decomposition basis length. Translation back to the canonical 2-norm incurrs the additional $\sqrt{mr_1\varphi\hat{\mathsf{f}}}$ factor (cf. Corollary 1). Furthermore, $\beta_2 = \sqrt{mr_2\hat{\mathsf{f}}\varphi}\,b_{\mathsf{d}}$ and $\beta_3 = \beta_2\sqrt{r_4}\,r_3\gamma$, where γ is the expansion factor of the subtractive set (assumed constant).

To argue about the feasibility of the setting, we need to establish that there exists b_{d} such that $\ell_{\mathsf{d}} = O(1)$ and $\beta_3 = \beta_0$. We observe that $\beta_3 = O(r_3\sqrt{r_4}\sqrt{mr_1\hat{\mathsf{f}}\varphi}\,\sqrt{mr_2\hat{\mathsf{f}}\varphi}\,b_{\mathsf{d}})$. After substituting $\beta_3 = \beta_0$, $\beta_0 = b_{\mathsf{d}}^{\Theta(1)}$ and $(r_i)_{i\in[5]}$ params with asymptotic expressions, we establish the condition for b_{d} as

$$b_{\mathsf{d}} = O\left(\left(m\hat{\mathsf{f}}^2(\lambda/\log\lambda)^{5/2}\right)^{1/\Theta(1)}\right).$$

We analyse the necessary size of the proof system modulus q. For vSIS hardness, we need $q > \beta^{\mathsf{sis}}$, where the bound on β^{sis} is computed as follows. By setting $\beta_3' = \beta_3 (= \beta_0)$, we get

$$\beta^{\mathsf{sis}} \geq 2\left(2\beta_0\sqrt{r_3}\theta_{\|\sigma(\cdot)\|_2}\right)^{\ell_{\mathsf{d}}} = \mathsf{poly}(m, \lambda).$$

as $\ell_{\mathsf{d}} = O(1)$, resulting in a polynomial-sized modulus q.

To discuss this in further detail, we assume a witness of norm β_0 and analyse RoKs in the "extraction direction". The extractor for Π^{fold} extracts a witness of norm $2\beta_0\sqrt{r_3}\theta_{\|\sigma(\cdot)\|_2}$. The extractor for Π^{split} does not alter the norm. For $\Pi^{b\text{-decomp}}$, the recomposed witness norm becomes $\left(2\beta_0\sqrt{r_3}\theta_{\|\sigma(\cdot)\|_2}\right)^{\ell_{\mathsf{d}}}$. In Π^{batch}, the extractor retrieves the unaltered witness or produces a vSIS break with norm $2\left(2\beta_0\sqrt{r_3}\theta_{\|\sigma(\cdot)\|_2}\right)^{\ell_{\mathsf{d}}}$. Similarly, for Π^{norm}, the extractor extracts a witness of norm β_0 or results in a vSIS break with norm $2\left(2\beta_0\sqrt{r_3}\theta_{\|\sigma(\cdot)\|_2}\right)^{\ell_{\mathsf{d}}}$.

Finally, we are ready to estimate the proof size. Considering the non-interactive setting, we will only count the prover messages. We track the communication cost for atomic RoKs:

- Π^{norm} involves sending $\ell_{\mathsf{n}} = O(1)$ $\mathcal{R}_q$ elements,
- Π^{batch} incurs no communication cost,
- $\Pi^{b\text{-decomp}}$ involves sending $\ell_{\mathsf{n}} = O(r_2) = O(r) = O(\lambda/\log\lambda)$ $\mathcal{R}_q$ elements,
- Π^{split} involves sending $\ell_{\mathsf{n}} = O(r_2) = O(r) = O(\lambda/\log\lambda)$ $\mathcal{R}_q$ elements,
- Π^{fold} incurs no communication cost.

We conclude that the total communication cost for a single round is $O(\lambda/\log\lambda)$ $\mathcal{R}_q$ elements, expressed in bits as

$$O\left(\frac{\lambda^2}{\log^2\lambda}\cdot\log^2 m\right).$$

Total Communication Complexity. After $\mu = O(\log m)$ rounds, the height of the witness is $O(1)$ and the width becomes $O(r_0)$, consisting of $\mathcal{R}_q$ elements. The size (in bits) of such a witness is

$$O\left(\frac{\lambda}{\log\lambda}\cdot\frac{\lambda\cdot\log^2 m}{\log\lambda}\right) = O\left(\frac{\lambda^2}{\log^2\lambda}\cdot\log^2 m\right).$$

The total communication cost across all rounds and the final witness sent is

$$O\left(\frac{\lambda^2}{\log^2\lambda}\cdot\log^3 m\right).$$

7 Packed $\mathbb{Z}$-Inner Products via Twisted Trace Maps

We propose an abstract framework based on "twisted trace maps" that reduces $\mathbb{Z}$-inner products to $\mathcal{R}$-inner products over various choices of $\mathcal{R}$. In a nutshell, for a fixed choice of $\mathcal{R}$, we would like to construct a twisted trace map $\tau : \mathcal{R} \to \mathbb{Z}$ of the form shown below, where $N \in \mathbb{N}$ is some normalisation factor and $\alpha \in \mathcal{R}$ is called a "twist" element, such that the following diagram commutes:

$$\begin{array}{ccc} \mathbb{Z}^\delta \times \mathbb{Z}^\delta & \xrightarrow{\langle\cdot,\cdot\rangle_\mathbb{Z}} & \mathbb{Z} \\ {\scriptstyle \psi^{-1}(\cdot)\times\overline{\psi^{-1}(\cdot)}}\downarrow & & \uparrow{\scriptstyle \tau} \\ \mathcal{R} \times \mathcal{R} & \xrightarrow{\cdot_\mathcal{R}} & \mathcal{R} \end{array} \qquad \text{where} \qquad \tau : z \mapsto \frac{1}{N} \cdot \mathsf{Trace}(\alpha \cdot z).$$

Definition 5 (Inner-Product Embedding). *Let $\mathcal{R} \subset \mathcal{O}_\mathcal{K}$ be a subring identified by a $\mathbb{Z}$-basis $\mathbf{b} \in \mathcal{R}^\delta$ of δ elements. We say that a tuple τ is an inner-product embedding over $\mathcal{R}$ if $\tau : \mathcal{R} \to \mathbb{Z}^\delta$ is a $\mathbb{Z}$-linear map and, for any $\mathbf{x}, \mathbf{y} \in \mathbb{Z}^\delta$, it holds that $\langle \mathbf{x}, \mathbf{y} \rangle_\mathbb{Z} = \tau\left(\psi_\mathbf{b}^{-1}(\mathbf{x}) \cdot_\mathcal{R} \overline{\psi_\mathbf{b}^{-1}(\mathbf{y})}\right)$.*

7.1 Power-of-Two Cyclotomics via Constant Term

As a simple concrete example, we recall a well-known folklore technique for computing the inner product over the coefficient embeddings of power-of-two cyclotomics.

Theorem 5. *Let $\mathcal{R} = \mathbb{Z}[\zeta_\mathfrak{f}]$ with a conductor $\mathfrak{f} = 2^k$ for some $k \in \mathbb{N}$, $\delta = \varphi = \varphi(\mathfrak{f}) = \mathfrak{f}/2$, $\tau(\cdot) = \mathsf{ct}(\cdot) = (\psi(\cdot))_0$, where ψ denotes the coefficient embedding and $\mathsf{ct}(\cdot)$ is the constant term of the coefficient embedding. Then τ is an inner-product embedding over $\mathcal{R}$.*

Proof. Write $\zeta = \zeta_\mathfrak{f}$. Observe that $\mathsf{ct}(\zeta^i) = (i \stackrel{?}{=} 0)$ for all $i \in \pm[\varphi]$.[22] Consider the vectors $\mathbf{x} = (x_0, \ldots, x_{\varphi-1}), \mathbf{y} = (y_0, \ldots, y_{\varphi-1}) \in \mathbb{Z}^\varphi$. The elements $x := \psi^{-1}(\mathbf{x})$ and $\bar{y} := \overline{\psi^{-1}(\mathbf{y})}$ can be expressed as

$$x = \psi^{-1}(\mathbf{x}) = \sum_{i \in [\varphi]} x_i \zeta^i \qquad \text{and} \qquad \bar{y} = \overline{\psi^{-1}(\mathbf{y})} = \sum_{i \in [\varphi]} y_i \zeta^{-i}$$

respectively. Note that their product $z := x \cdot \bar{y}$ satisfies

$$\begin{aligned} z = x \cdot \bar{y} &= \left(\sum_{i \in [\varphi]} x_i \zeta^i\right)\left(\sum_{j \in [\varphi]} y_j \zeta^{-j}\right) = \sum_{i,j \in [\varphi]} x_i y_j \zeta^{i-j} \\ &= \underbrace{\sum_{i \in [\varphi]} x_i y_i}_{\mathsf{ct}(z)} + \sum_{i,j \in [\varphi] : i \neq j} a_i b_j \zeta^{i-j}. \end{aligned}$$

Therefore, $\tau(\psi^{-1}(\mathbf{x}), \overline{\psi^{-1}(\mathbf{y})}) = \langle \mathbf{x}, \mathbf{y} \rangle_\mathbb{Z}$, as desired. □

Remark 8. The constant term map $\mathsf{ct}(x)$ from Theorem 5 can be expressed in terms of the Trace function as $\tau(x) = \frac{1}{\varphi}\mathsf{Trace}(x)$, where $\varphi = \mathfrak{f}/2$ since $\mathfrak{f}$ is a power of 2, and might be viewed as a twisted trace map $\tau(x) = \frac{1}{\varphi}\mathsf{Trace}(\alpha \cdot x)$ with $\alpha = 1$.

As pointed out in Section 4, power-of-two cyclotomic rings do not admit large subtractive sets, and are therefore ill-suited for certain applications, e.g. instantiating the succinct arguments presented in Section 5. This motivates the search for inner-product embeddings τ over other rings.

[22] This is true for power-of-2 cyclotomics since power-of-2 cyclotomic polynomials are of the form $\Phi_\mathfrak{f}(X) = X^\varphi + 1$. Note that this is false for non-power-of-2 conductors. For example, if $\mathfrak{f}$ is prime, then $\zeta^{-1} = \zeta^\varphi = -\sum_{i \in [\varphi]} \zeta^i$ with $\mathsf{ct}(\zeta^{-1}) = -1$.

7.2 Prime Real Cyclotomics via Twisted Trace

A natural class of rings to search for inner-product embeddings are cyclotomic rings with large prime conductors, since they admit large subtractive sets (cf. Section 4). Although we did not manage to design inner-product embeddings in those rings, we did so for its maximal real subring, adapting a result from lattice code theory [BFOV04, Proposition 1].

Theorem 6. *Let $\mathcal{K} = \mathbb{Q}(\zeta_\mathfrak{f})$ where $\mathfrak{f}$ is prime and $\mathcal{R} = \mathbb{Z}[\zeta_\mathfrak{f} + \zeta_\mathfrak{f}^{-1}]$ be identified by the $\mathbb{Z}$-basis $\mathbf{b}^+ = \left\{\sum_{i=[j+1]}(\zeta^{\varphi/2-i} + \zeta^{-(\varphi/2-i)})\right\}_{j\in[\varphi/2]}$. For $z \in \mathcal{R}$, let $\tau(z) = \frac{1}{2\mathfrak{f}}\mathsf{Trace}(\alpha z)$ be a twisted trace map for the twist element $\alpha = t \cdot \bar{t}$ where $t = \zeta^{-\varphi/2} - \zeta^{\varphi/2}$. Then τ is an inner product embedding over $\mathcal{R}$.*

Proof. Since $\mathfrak{f}$ is prime, we have $\varphi = \mathfrak{f} - 1$ and $\delta = \varphi/2 = (\mathfrak{f}-1)/2$. In the following, write $\mathsf{Trace} = \mathsf{Trace}$. Recall that $\mathsf{Trace}(1) = \sum_{j\in[\varphi]} 1 = \varphi = \mathfrak{f} - 1$. Furthermore, for $i \in \mathbb{Z}_\mathfrak{f}^\times$, we have $\mathsf{Trace}(\zeta^i) = \sum_{j\in\mathbb{Z}_\mathfrak{f}^\times} \zeta^{ij} = \sum_{j\in\mathbb{Z}_\mathfrak{f}^\times} \zeta^j = -1$.

As a starting point, we consider the following sequence.

$$\mathbf{b}^- = (b_i^-)_{i\in[\varphi/2]} = (\zeta^{i+1} - \zeta^{-(i+1)})_{i\in[\varphi/2]}.$$

Note that the sequence $\mathbf{b}^-$ is trace-orthogonal. Namely, the following function acts as Kronecker delta.

$$\frac{1}{2\mathfrak{f}} \cdot \mathsf{Trace}(\overline{b_i^-} \cdot b_j^-) = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}$$

To see this, let $i' = i + 1$ and $j' = j + 1$ and consider the trace of the expression below.

$$\begin{aligned}
\overline{b_i^-} \cdot b_j^- &= \overline{(\zeta^{i'} - \zeta^{-i'})} \cdot (\zeta^{j'} - \zeta^{-j'}) \\
&= (\zeta^{-i'} - \zeta^{i'}) \cdot (\zeta^{j'} - \zeta^{-j'}) = (\zeta^{-i'+j'} - \zeta^{i'+j'} - \zeta^{-i'-j'} + \zeta^{i'-j'}), \\
\mathsf{Trace}(b_i^- \cdot \overline{b_j^-}) &= \mathsf{Trace}(\zeta^{-i'+j'} - \zeta^{i'+j'} - \zeta^{-i'-j'} + \zeta^{i'-j'}) \\
&= \mathsf{Trace}(\zeta^{-(i'-j')}) - \mathsf{Trace}(\zeta^{i'+j'}) - \mathsf{Trace}(\zeta^{-(i'+j')}) + \mathsf{Trace}(\zeta^{i'-j'}).
\end{aligned}$$

Since $i, j \in [\varphi/2]$, we have $2 \leq i' + j' \leq \varphi$, meaning that $i' + j' \in \mathbb{Z}_\mathfrak{f}^\times$ and $-(i' + j') \in \mathbb{Z}_\mathfrak{f}^\times$. Furthermore, if $i \neq j$, then $i' - j' \in \pm[\varphi/2] \setminus \{0\}$, hence $i' - j' \in \mathbb{Z}_\mathfrak{f}^\times$ and $-(i' - j') \in \mathbb{Z}_\mathfrak{f}^\times$. Therefore, we conclude that

$$\begin{aligned}
(i = j) &\implies \mathsf{Trace}(b_i^- \cdot \overline{b_j^-}) = 2\mathfrak{f}, \\
(i \neq j) &\implies \mathsf{Trace}(b_i^- \cdot \overline{b_j^-}) = 0.
\end{aligned}$$

Although $\mathbf{b}^-$ is trace-orthogonal, it does not constitute a basis of any ring. It does, however, match in cardinality the degree of the maximal real subring $\mathcal{R}$, to which our attention now turns. Consider the "suffix-sum" basis for the maximal real subring as in the theorem statement:

$$\mathbf{b}^+ = (b_j^+)_{j\in[\varphi/2]} = \left\{\sum_{i\in[j+1]} (\zeta^{\varphi/2-i} + \zeta^{-(\varphi/2-i)})\right\}_{j\in[\varphi/2]}$$

From the identity $-\varphi/2 = \mathfrak{f} - \varphi/2 = \varphi + 1 - \varphi/2 = \varphi/2 + 1 \bmod \mathfrak{f}$, for each $j \in [\varphi/2]$, we observe that

$$b_j^- = b_j^+ \cdot \underbrace{(\zeta^{-\varphi/2} - \zeta^{\varphi/2})}_{t}$$

since

$$b_j^+ \cdot (\zeta^{-\varphi/2} - \zeta^{\varphi/2}) = \sum_{i\in[j+1]} (\zeta^{\varphi/2-i} + \zeta^{-(\varphi/2-i)}) \cdot (\zeta^{-\varphi/2} - \zeta^{\varphi/2})$$

$$
\begin{aligned}
&= \sum_{i\in[j+1]} (\zeta^{-i} - \zeta^{i} + \zeta^{-(\varphi-i)} - \zeta^{\varphi-i}) \\
&= \sum_{i\in[j+1]} (\zeta^{-i} - \zeta^{-(i+1)} + \zeta^{i+1} - \zeta^{i}) \\
&= \sum_{i\in[j+1]} (\zeta^{-i} - \zeta^{-(i+1)}) + \sum_{i\in[j+1]} (\zeta^{i+1} - \zeta^{i}) \\
&= \zeta^{j+1} - \zeta^{-j+1} = b_j^- .
\end{aligned}
$$

Therefore,

$$
\frac{1}{2\mathfrak{f}} \cdot \mathsf{Trace}(t \cdot b_i^+ \cdot \overline{t \cdot b_j^+}) = \frac{1}{2\mathfrak{f}} \cdot \mathsf{Trace}(\alpha \cdot b_i^- \cdot \overline{b_j^-}) = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases}
$$

Now, suppose $x = \psi_{\mathbf{b}^+}^{-1}(\mathbf{x})$ and $\bar{y} = \overline{\psi_{\mathbf{b}^+}^{-1}(\mathbf{y})}$ for some $\mathbf{x}, \mathbf{y} \in \mathbb{Z}^\delta$. We have

$$
\begin{aligned}
\tau\left(x \cdot \bar{y}\right) &= \frac{1}{2\mathfrak{f}} \mathsf{Trace}(\alpha x \cdot \bar{y}) = \sum_{i,j\in[\varphi/2]} x_i y_j \frac{1}{2\mathfrak{f}} \mathsf{Trace}(t \cdot b_i^+ \cdot \bar{t} \cdot \overline{b_i^+}) = \sum_{i\in[\varphi/2]} x_i y_i \\
&= \langle \mathbf{x}, \mathbf{y} \rangle.
\end{aligned}
$$

□

The above theorem constructs inner-product embeddings for $\mathcal{R} = \mathbb{Z}[\zeta_\mathfrak{f} + \zeta_\mathfrak{f}^{-1}]$ where $\mathfrak{f}$ is prime. This restricts the choice of $\mathcal{R}$ quite severely, especially considering that the subtractive set constructed in Section 4 for $\mathbb{Z}[\zeta_\mathfrak{f}]$ or $\mathbb{Z}[\zeta_\mathfrak{f} + \zeta_\mathfrak{f}^{-1}]$ for prime $\mathfrak{f}$ has a large expansion factor bound $\gamma_S \leq \mathfrak{f}$.

7.3 Tensor of Prime Real Cyclotomics

To allow more fine-grained parameter selection, we extend the result in Sections 7.1 and 7.2 by constructing larger rings using the tensor product, inspired by [BFOV04, Proposition 6]. Concretely, we construct subtractive sets for rings

$$
\mathcal{R} = \mathcal{O}_{\mathcal{K}_{2^d}} \otimes \mathcal{O}_{\mathcal{K}^+_{p_0}} \otimes \ldots \otimes \mathcal{O}_{\mathcal{K}^+_{p_{k-1}}} \tag{14}
$$

for distinct odd primes $p_0, \ldots, p_{k-1}$. Note that $\mathcal{R}$ has conductor $\mathfrak{f} = 2^d \cdot \prod_{i\in[k]} p_i$ and degree $\delta = 2^d \cdot \prod_{i\in[k]}(p_i - 1)$. It is contained in the ring $\mathcal{O}_{\mathcal{K}_\mathfrak{f}}$ which admits a subtractive set S of size $\mathfrak{f}/\mathfrak{f}_{\max}$ with expansion factor $\gamma_S = 1$ (cf. Section 4).

Theorem 7. *Let $\mathcal{R} = \mathcal{O}_{\mathcal{K}_\mathfrak{g}} \otimes \mathcal{O}_{\mathcal{K}^+_{\mathfrak{f}_0}} \otimes \ldots \otimes \mathcal{O}_{\mathcal{K}^+_{\mathfrak{f}_{k-1}}}$, $\mathfrak{g} = 2^d$ for some $d \in \mathbb{N}$, and $\mathfrak{f}_0, \ldots, \mathfrak{f}_{k-1}$ distinct odd primes. Let $\mathbf{b} = \mathbf{b}_\mathfrak{g} \otimes \left(\bigotimes_{i\in[k]} \mathbf{b}^+_{\mathfrak{f}_i}\right)$, where $\mathbf{b}_\mathfrak{g}$ is the power basis for $\mathcal{R}_\mathfrak{g}$ and $\mathbf{b}^+_{\mathfrak{f}_i}$ is a basis for $\mathcal{R}^+_{\mathfrak{f}_i}$ defined as in Theorem 6. Then, $\tau(\cdot) = \frac{1}{t} \cdot \mathsf{Trace}(\alpha \cdot (\cdot))$ is inner-product embedding for $\alpha = \prod_{i\in[k]} \alpha_{\mathfrak{f}_i}$, where $t = 2^k \varphi(\mathfrak{g}) \prod_{i\in[k]} \mathfrak{f}_i$.*

Proof. Write $\mathcal{R}_\mathfrak{g}$ for $\mathcal{O}_{\mathcal{K}_\mathfrak{g}}$ and $\mathcal{R}_i$ for $\mathcal{O}_{\mathcal{K}^+_{p_i}}$ for $i \in [k]$. Define $t_{\mathfrak{f}_i} = 2\mathfrak{f}_i$ and $t_\mathfrak{g} = \varphi(\mathfrak{g})$. We prove by induction on tensoring consecutive rings $\mathcal{R}_\mathfrak{g}$ and $\mathcal{R}_{\mathfrak{f}_i} \forall i \in [k]$, indexed as $\mathcal{R}_i = \mathcal{O}_{\mathcal{K}_i}$ for $i \in [h]$, where $h = k + 1$ or $h = k$ (if no power-of-two components). We use $\mathfrak{h}_i$ for $i \in [h]$ to iterate over coprime factors of the conductor.

By Remark 8 and Theorem 6 $\mathcal{R}_i$ has an inner-product embedding τ_i, i.e.

$$
\tau_i((b_i)_m \cdot (\bar{b}_i)_n) = \frac{1}{t_i} \mathsf{Trace}_{\mathcal{K}_i/\mathbb{Q}}(\alpha_i \cdot (b_i)_m \cdot (\bar{b}_i)_n) = \begin{cases} 1 & \text{if} \quad m = n, \\ 0 & \text{if} \quad m \neq n \end{cases} \quad \forall i \in [h].
$$

We devise a proof by induction.
First, we define the base case $(\widetilde{\mathfrak{h}}_0, \widetilde{t}_0, \widetilde{\alpha}_0, \widetilde{\mathcal{R}}_0, \widetilde{\alpha}_0, \widetilde{\mathbf{b}}_0) = (\mathfrak{h}_0, t_0, \alpha_0, \mathcal{R}_0, \alpha_0, \mathbf{b}_0)$.
Then, for $i \in [h-1]$, define the following inductive steps:

$$
\mathfrak{h}_{i+1} \coloneqq \mathfrak{h}_{i+1} \cdot \widetilde{\mathfrak{h}}_i \qquad\qquad \widetilde{t}_{i+1} \coloneqq t_i \cdot \widetilde{t}_{i+1}
$$

$$\widetilde{\alpha}_{i+1} \coloneqq \alpha_i \cdot \widetilde{\alpha}_{i+1} \qquad \widetilde{\mathcal{R}}_{i+1} \coloneqq \mathcal{R}_{i+1} \otimes \widetilde{\mathcal{R}}_i$$
$$\widetilde{\alpha}_{i+1} \coloneqq \alpha_{i+1} \cdot \widetilde{\alpha}_i \qquad \widetilde{\mathbf{b}}_{i+1} \coloneqq \mathbf{b}_{i+1} \otimes \widetilde{\mathbf{b}}_i$$

We want to show that, if $\widetilde{\mathcal{R}}_i$ has an inner-product embedding, then $\widetilde{\mathcal{R}}_{i+1}$ also has an inner-product embedding.

We write

$$\widetilde{\mathbf{b}}_i = \left\{\widetilde{b}_{i,0}, \ldots, \widetilde{b}_{i,\widetilde{\varphi}_i - 1}\right\},$$

and

$$\mathbf{b}_{i+1} = \left\{b_{i+1,0}, \ldots, b_{i+1,\varphi_{i+1}-1}\right\}.$$

Elements of a new basis $\widetilde{\mathbf{b}}_{i+1}$ are uniquely defined as a product of two elements from bases $\widetilde{\mathbf{b}}_i$ and $\mathbf{b}_{i+1}$. Consider elements $b_{i+1,m} \cdot \widetilde{b}_{i,r}$ and $b_{i+1,n} \cdot \widetilde{b}_{i,s}$ of a new basis $\widetilde{\mathbf{b}}_{i+1}$ Due to the coprimality of $\widetilde{\mathfrak{h}}_i$ and $\mathfrak{h}_{i+1}$, the tower structure of traces is interchangeable, thus

$$\begin{aligned}
&\widetilde{\tau}_{i+1}\left(b_{i+1,m}\widetilde{b}_{i,r} \cdot \overline{b_{i+1,n} \cdot \widetilde{b}_{i,s}}\right) \\
&= \frac{1}{\widetilde{t}_{i+1}} \mathsf{Trace}_{\widetilde{\mathcal{K}}_{i+1}/\mathbb{Q}}\left(\widetilde{\alpha}_{i+1} \cdot b_{i+1,m}\widetilde{b}_{i,r} \cdot \overline{b_{i+1,n} \cdot \widetilde{b}_{i,s}}\right) \\
&= \frac{1}{\widetilde{t}_i} \mathsf{Trace}_{\widetilde{\mathcal{K}}_i/\mathbb{Q}}\left(\widetilde{\alpha}_i \cdot \widetilde{b}_{i,r} \cdot \overline{\widetilde{b}}_{i,s}\right) \cdot \frac{1}{t_{i+1}} \mathsf{Trace}_{\mathcal{K}_{i+1}/\mathbb{Q}}\left(\alpha_{i+1} \cdot b_{i+1,m} \cdot \overline{b}_{i+1,n}\right) \\
&= \widetilde{\tau}_i\left(\widetilde{b}_{i,r} \cdot \overline{\widetilde{b}}_{i,s}\right) \cdot \tau_{i+1}\left(b_{i+1,m} \cdot \overline{b}_{i+1,n}\right) = \begin{cases} 1 & \text{if} \quad (m,r) = (n,s) \\ 0 & \text{if} \quad (m,r) \neq (n,s) \end{cases}
\end{aligned}$$

Finally,

$$\left(\mathfrak{h}, t, \alpha, \mathcal{R}, \alpha, \mathbf{b}\right) = \left(\widetilde{\mathfrak{h}}_{h-1}, \widetilde{t}_{h-1}, \widetilde{\alpha}_{h-1}, \widetilde{\mathcal{R}}_{h-1}, \widetilde{\alpha}_{h-1}, \widetilde{\mathbf{b}}_{h-1}\right),$$

which concludes the proof. □

7.4 Reducing Binariness to Bounded Norm

We show how to reduce the $\mathbb{Z}$-relation $\mathbf{x} \in \{0,1\}^{m\delta}$ to an $\mathcal{R}$-relation natively supported by the succinct arguments presented in Section 5, via the inner-product embedding framework. First, we recall the following elementary fact from [LNP22].

Proposition 2. *A vector $\mathbf{x} \in \mathbb{Z}^{m\delta}$ is binary if and and only if $\langle \mathbf{x}, 1^m - \mathbf{x}\rangle_{\mathbb{Z}} = 0$.*

Proof. To argue about the " $\Longrightarrow$ " direction this is enough to observe that, for each $j \in [m\delta]$, we have $x_j = 0$ or $1 - x_j = 0$. And therefore the sum satisfies $\sum_{j\in[m\delta]} x_j(1-x_j) = 0$. The " $\Longleftarrow$ " direction relies on the observation that, for each $j \in [m\delta]$, $x_j(1-x_j) \geq 0$ as $x_j \in \mathbb{Z}$. Also, if $x_j \notin \{0,1\}$, then $x_j(1-x_j) > 0$. Hence, if for some $j \in [m\delta]$, $x_j \notin \{0,1\}$, then $\sum_{j\in[m\delta]} x_j(1-x_j) > 0$, which is a contradiction. □

Next, we observe the following equivalence: $\langle \mathbf{x}, \mathbf{1}^{m\delta} - \mathbf{x}\rangle_{\mathbb{Z}} = 0 \iff \langle \mathbf{x}, \mathbf{1}^{m\delta}\rangle_{\mathbb{Z}} - \langle \mathbf{x}, \mathbf{x}\rangle_{\mathbb{Z}} = 0 \iff \langle \mathbf{x}, \mathbf{1}^{m\delta}\rangle_{\mathbb{Z}} = \langle \mathbf{x}, \mathbf{x}\rangle_{\mathbb{Z}}$. This suggests the following reduction:

(i) The prover sends two claimed values $s, t \in \mathcal{R}$ supposedly satisfying $\tau(t) = \tau(s)$
(ii) The prover then sends a succinct proof for $\langle \psi^{-1}(\mathbf{x}), \overline{\psi^{-1}(\mathbf{1}^{\delta m})}\rangle_{\mathcal{R}} = s$ and $\langle \psi^{-1}(\mathbf{x}), \overline{\psi^{-1}(\mathbf{x})}\rangle_{\mathcal{R}} = t$.

From the identity $\forall\ \mathbf{a}, \mathbf{b} \in \mathbb{Z}^{m\delta},\ \tau(\langle \psi^{-1}(\mathbf{a}), \overline{\psi^{-1}(\mathbf{b})}\rangle_{\mathcal{R}}) = \langle \mathbf{a}, \mathbf{b}\rangle_{\mathbb{Z}}$, the verifier would be convinced that $\mathbf{x}$ is indeed binary.

However, there is a subtle issue that, on one hand, the rings $\mathcal{R}$ considered in this section are of the form displayed in Eq. (14), which are not necessarily equal to $\mathcal{O}_{\mathcal{K}}$ or $\mathcal{O}_{\mathcal{K}^+}$ for any cyclotomic field $\mathcal{K}$. On the other hand, the succinct arguments constructed in Section 5 are over rings which admit large subtractive sets, for which we only know constructions in $\mathcal{O}_{\mathcal{K}}$ and $\mathcal{O}_{\mathcal{K}^+}$. We therefore need to lift the $\mathcal{R}$-relations that we want to prove to some $\mathcal{O}_{\mathcal{K}}$-relations (or $\mathcal{O}_{\mathcal{K}^+}$-relations, but we focus on the former) with $\mathcal{R} \subseteq \mathcal{O}_{\mathcal{K}}$, while ensuring that the prover cannot cheat by using a witness over $\mathcal{O}_{\mathcal{K}}$. To do this, we

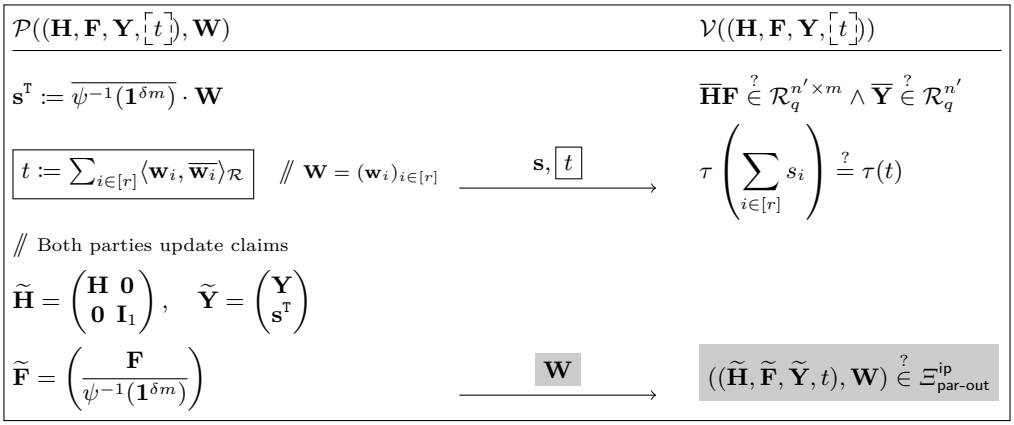

Fig. 8. Protocol $\boxed{\Pi^{\mathsf{lin\text{-}bin}}_{\tau}}$ or $\boxed{\Pi^{\mathsf{ip\text{-}bin}}_{\tau}}$, a reduction from $\boxed{\Xi^{\mathsf{lin}}_{\mathsf{par\text{-}in}} \cap \Xi^{\mathsf{bin}}_{m,r}}$ or $\boxed{\Xi^{\mathsf{ip}}_{\mathsf{par\text{-}in}} \cap \Xi^{\mathsf{bin}}_{m,r}}$ to $\Xi^{\mathsf{ip}}_{\mathsf{par\text{-}out}}$ with par-in, par-out specified in Theorem 8. The marked parts are only sent / checked when the protocol is used as a proof of knowledge. As a reduction of knowledge, they are omitted.

need the lemma which allows viewing $\mathcal{O}_{\mathcal{K}}$ as an $\mathcal{R}$-module in such a way that the geometry of $\mathcal{O}_{\mathcal{K}}$ is respected. We refer to Lemma 9 for a precise lemma with the proof.

We next formally define the binariness relation which ignores the statement and simply checks that the witness is a binary vector.

$$\Xi^{\mathsf{bin}}_{m,r} := \left\{ (\mathsf{stmt}, \mathbf{W}) \colon \mathsf{stmt} \in \{0,1\}^*;\ \mathbf{W} \in \mathcal{R}^{m\times r};\ \psi(\mathbf{W}) \in \{0,1\}^{mr\delta} \right\}.$$

In Fig. 8, we present two similar reductions of knowledge $\Pi^{\mathsf{lin\text{-}bin}}$ and $\Pi^{\mathsf{ip\text{-}bin}}$ from $\Xi^{\mathsf{lin}} \cap \Xi^{\mathsf{bin}}$ or $\Xi^{\mathsf{ip}} \cap \Xi^{\mathsf{bin}}$ to Ξ^{ip}, respectively. Note that, when reducing $\Xi^{\mathsf{ip}} \cap \Xi^{\mathsf{bin}}$ to Ξ^{ip}, the inner product $t = \langle \psi^{-1}(\mathbf{x}), \overline{\psi^{-1}(\mathbf{x})} \rangle_{\mathcal{R}}$ is already included as part of the statement, and thus the prover does not need to send it. The formal result is stated in Theorem 8, whose proof relies on Lemma 9 stated immediately after.

Theorem 8. *Let $m, r, \ell, b_{\mathsf{ip}} \in \mathbb{N}$ and $0 \le \beta \le \beta^{\mathsf{sis}} \le q$, $0 \le 2\beta' \le \beta^{\mathsf{sis}}$. Let τ be an inner-product embedding over $\mathcal{R}$. The protocol $\Pi^{\mathsf{lin\text{-}bin}}_{\tau}$ (resp. $\Pi^{\mathsf{ip\text{-}bin}}_{\tau}$) is a perfectly correct reduction of knowledge from $\Xi^{\mathsf{ip}} \cap \Xi^{\mathsf{bin}}$ to Ξ^{lin}*

$$\left(m, n^{\mathsf{out}}, r, \beta\right) \mapsto \left(m, n^{\mathsf{out}\prime}, r+\ell, \beta_{\mathsf{out}}\right)$$

and knowledge sound reduction of knowledge from $\Xi^{\mathsf{ip}} \cap \Xi^{\mathsf{bin}\vee\mathsf{sis}}$ to Ξ^{lin} with parameters

$$\left(m, n^{\mathsf{out}}, r, \beta', \beta^{\mathsf{sis}}\right) \leftarrow\!\shortmid \left(m, n^{\mathsf{out}\prime}, r+\ell, \beta_{\mathsf{out}}{}', \beta^{\mathsf{sis}}\right),$$

where $n^{\mathsf{out}\prime} = n^{\mathsf{out}} + 3$ for $\beta = 1$ if

$$\boxed{\sqrt{m} \cdot 2^k \cdot \varphi \sqrt{\hat{\mathsf{f}}\varphi} \cdot \beta \le \beta^{\mathsf{sis}}}_{\|\sigma(\cdot)\|_2} \left(\text{resp.}, \boxed{2^k \varphi \beta \le \beta^{\mathsf{sis}}}_{\|\psi(\cdot)\|_\infty} \right),$$

where k is defined as in Lemma 9.

Proof. For perfect completeness, consider $((\mathbf{H},\mathbf{F},\mathbf{Y}),\mathbf{W}) \in \Xi^{\mathsf{lin}}_{m,n^{\mathsf{out}},\mu,\beta} \cap \Xi^{\mathsf{bin}}_{m,r}$ over $\mathcal{R}$. We have $\psi(\mathbf{W}) \in \{0,1\}^{rm\delta}$. Clearly, $\beta = 1$, regardless if canonical 2-norm or coefficient ∞-norm is concerned. By Proposition 2 and the discussion immediately after, it holds that $\langle \psi(\mathbf{w}_i), \mathbf{1}^{m\delta} \rangle_{\mathbb{Z}} = \langle \psi(\mathbf{w}_i), \psi(\mathbf{w}_i) \rangle_{\mathbb{Z}}$. Since τ is an inner-product embedding over $\mathcal{R}$, it holds that

$$\tau\left(\sum_{i\in[r]} s_i\right) = \tau\left(\sum_{i\in[r]} \langle \mathbf{w}_i, \overline{\psi^{-1}(\mathbf{1}^{\delta m})} \rangle_{\mathcal{R}}\right) = \sum_{i\in[r]} \langle \psi(\mathbf{w}_i), \mathbf{1}^{m\delta} \rangle_{\mathbb{Z}}, \text{ and}$$

$$\tau(t) = \tau\left(\sum_{i\in[r]} \langle \mathbf{w}_i, \overline{\mathbf{w}_i} \rangle_{\mathcal{R}}\right) = \sum_{i\in[r]} \langle \psi(\mathbf{w}_i), \psi(\mathbf{w}_i) \rangle_{\mathbb{Z}},$$

and thus $\tau\left(\sum_{i\in[r]} s_i\right) = \tau(t)$. Furthermore, since $((\mathbf{H},\mathbf{F},\mathbf{Y}),\mathbf{W}) \in \Xi^{\mathsf{lin}}_{m,n^{\mathsf{out}},\mu,\beta}$, we have $\mathbf{s}^{\mathsf{T}} \coloneqq \overline{\psi^{-1}(\mathbf{1}^{\delta m})}\cdot\mathbf{W}$ and $t \coloneqq \sum_{i\in[r]}\langle \mathbf{w}_i, \overline{\mathbf{w}_i}\rangle_{\mathcal{R}}$. Therefore, $((\widetilde{\mathbf{H}},\widetilde{\mathbf{F}},\widetilde{\mathbf{Y}},t),\mathbf{W}) \in \Xi^{\mathsf{ip}}_{m,n^{\mathsf{out}}+1,\mu,\beta,\bar{\mathsf{id}}}$ over $\mathcal{R}$. Since $\mathcal{R}\subseteq\mathcal{O}_{\mathcal{K}}$, the claim follows.

For perfect relaxed knowledge soundness, suppose that $((\widetilde{\mathbf{H}},\widetilde{\mathbf{F}},\widetilde{\mathbf{Y}},t),\mathbf{W}) \in \Xi^{\mathsf{ip}}_{m,n^{\mathsf{out}}+1,\mu,\beta,\bar{\mathsf{id}}}$ over $\mathcal{O}_{\mathcal{K}}$. If $\mathbf{W}\in\mathcal{R}^{m\times r}$, then we have $\mathbf{s}^{\mathsf{T}} \coloneqq \overline{\psi^{-1}(\mathbf{1}^{\delta m})}\cdot\mathbf{W}$ and $t \coloneqq \sum_{i\in[r]}\langle \mathbf{w}_i, \overline{\mathbf{w}_i}\rangle_{\mathcal{R}}$. Since $\tau\left(\sum_{i\in[r]} s_i\right) = \tau(t)$, reversing the above argument gives $\psi(\mathbf{w})\in\{0,1\}^{m\delta r}$. Thus $((\mathbf{H},\mathbf{F},\mathbf{Y}),\mathbf{w}) \in \Xi^{\mathsf{lin}}_{m,n^{\mathsf{out}},\mu,\beta} \cap \Xi^{\mathsf{bin}}_{m,r}$ over $\mathcal{R}$ as desired.

If $\mathbf{W}\in\mathcal{O}_{\mathcal{K}}^m \setminus \mathcal{R}^m$, then we can express any column $\mathbf{w}$ as a linear combination of $\mathcal{R}$-vectors. Let $\hat{\mathbf{w}}$ be the coefficient of any basis element other than 1. Note that $\overline{\mathbf{H}}\mathbf{F}\hat{\mathbf{w}} = \mathbf{0}^{n'} \bmod q$. Moreover, by Lemma 9, we have

$$\boxed{\|\sigma(\hat{\mathbf{w}})\|_2 \le \sqrt{m}\cdot 2^k\cdot\varphi\sqrt{\hat{\mathfrak{f}}\varphi}\cdot\beta \le \beta^{\mathsf{sis}}} \left(\text{resp., } \boxed{\|\psi(\hat{\mathbf{w}})\|_\infty \le 2^k\varphi\beta \le \beta^{\mathsf{sis}}}\right)$$

i.e. a vSIS break.

The argument for $\Pi^{\mathsf{ip\text{-}bin}}_{\tau}$ is almost verbatim, except that t is given as part of the statement rather than being sent by the prover. □

Lemma 9. *If $\mathfrak{f}$ be an odd prime, then $\mathcal{O}_{\mathcal{K}}$ can be seen as an $\mathcal{O}_{\mathcal{K}^+}$-module with the basis $\{1,\zeta\}$. More generally, let $\mathcal{R} = \mathcal{O}_{\mathcal{K}_{\mathfrak{g}}}\otimes\mathcal{O}_{\mathcal{K}^+_{\mathfrak{f}_0}}\otimes\ldots\otimes\mathcal{O}_{\mathcal{K}^+_{\mathfrak{f}_{k-1}}}$ where $\mathfrak{g}=2^d$ for some $d\in\mathbb{N}$ and $\mathfrak{f}_0,\ldots,\mathfrak{f}_{k-1}$ are distinct odd primes. Let $\mathfrak{f}\coloneqq\mathfrak{g}\prod_{i\in[k]}\mathfrak{f}_i$. Then $\mathcal{O}_{\mathcal{K}_{\mathfrak{f}}}$ is an $\mathcal{R}$-module with the basis $\bigotimes_{i\in[k]}(1,\zeta_{\mathfrak{f}_i})$.*

Furthermore, let $\mathbf{x}\in\mathcal{O}^m_{\mathcal{K}_{\mathfrak{f}}}$ be expressed as a $\mathcal{R}^m$-combination of $\bigotimes_{i\in[k]}(1,\zeta_{\mathfrak{f}_i})$. If

$$\boxed{\|\sigma(\mathbf{x})\|_2 \le \beta} \left(\text{resp., } \boxed{\|\psi(\mathbf{x})\|_\infty \le \beta}\right),$$

then each $\mathcal{R}$-coefficient $\hat{\mathbf{x}}$ of $\mathbf{x}$ satisfies

$$\boxed{\|\sigma(\hat{\mathbf{x}})\|_2 \le 2^k\cdot\varphi\sqrt{\hat{\mathfrak{f}}\varphi}\cdot\beta} \left(\text{resp., } \boxed{\|\psi(\hat{\mathbf{x}})\|_\infty \le 2^k\varphi\beta}\right).$$

Proof. First part. We first consider the simple case where $\mathfrak{f}$ is an odd prime. Consider the $\mathbb{Z}$-basis $\mathbf{b}^+ = \{b^+_0,\ldots,b^+_{\varphi/2-1}\} = \{1,\zeta+\zeta^{-1},\ldots,\zeta^{\varphi/2-1}+\zeta^{1-\varphi/2}\}$ of $\mathcal{O}_{\mathcal{K}^+}$. We show that $(1,\zeta)\otimes\mathbf{b}^+$ is a $\mathbb{Z}$-basis of $\mathcal{O}_{\mathcal{K}}$, implying that $\mathcal{O}_{\mathcal{K}}$ is an $\mathcal{O}_{\mathcal{K}^+}$-module with the basis $\{1,\zeta\}$. Consider the "balanced power basis"

$$\begin{aligned}\mathbf{b} &= \{b_{-\varphi/2+1},\ldots,b_{-1},b_0,\ldots,b_{\varphi/2}\}\\ &= \{\zeta^{-\varphi/2+1},\ldots,\zeta^{-1},1,\ldots,\zeta^{\varphi/2}\}.\end{aligned}$$

We prove by induction that b_{-i} and b_{i+1} and can be expressed as a $\{-1,0,1\}$-combination of elements from $(1,\zeta)\otimes\mathbf{b}^+$ for all $i\in[\varphi/2]$.

For $i=0$, we observe that $b_0 = b^+_0$ and $b_1 = b^+_0\cdot\zeta$. Now, suppose the induction hypothesis holds for $i\le k$ for some $\ell\in[\varphi/2]$, i.e. b_{-i} and b_{i+1} are constructed for all $i\in[k]$. Our goal is to obtain b_{-k} and b_{k+1}. Observe that

$$\begin{aligned}b^+_k &= b_k + b_{-k},\\ b^+_k\cdot\zeta &= (b_k+b_{-k})\cdot\zeta = b_{k+1}+b_{-k+1},\\ b_{-k} &= b^+_k - b_k,\\ b_{k+1} &= b^+_k\cdot\zeta - b_{-k+1}.\end{aligned}$$

The claim then follows from the induction hypothesis. The above directly generalises for the tensor rings by arguing about each factor ring independently.

"Furthermore" part. For simplicity, we first consider the case where $\mathfrak{f}$ is prime. From the previous part of the proof, we know that there exists a $\mathbb{Z}$-basis $\mathbf{b}$ of $\mathcal{O}_{\mathcal{K}}$ which can be expressed as a $\{-1,0,1\}$-combination of elements in $(1,\zeta)\otimes\mathbf{b}^+$. We can write

$$b_i = \sum_{j\in[\varphi/2]} s_{i,j}b^+_j + \sum_{j\in[\varphi/2]} t_{i,j}b^+_j\cdot\zeta,$$

where $s_{i,j}, t_{i,j} \in \{-1, 0, 1\}$.

Consider $\mathbf{x} = (x_0, \dots x_{m-1})$ for any $\mathbf{x} \in \mathcal{O}^m_{\mathcal{K}_\mathfrak{f}}$. Then, we write

$$x_i = \sum_{j\in[\varphi]} \tilde{x}_{i,j} b_j = \sum_{j\in[\varphi],\ell\in[\varphi/2]} \tilde{x}_{i,j} \left(s_{j,\ell} + t_{j,\ell}\zeta\right) b^+_\ell,$$

where $\tilde{x}_{i,j} \in \mathbb{Z}^m$.

Let $\hat{x}_{\ell,i} = \sum_{j\in[\varphi]} \tilde{x}_{i,j}(s_{j,\ell} + t_{j,\ell}\zeta)$ and $\hat{\mathbf{x}}_\ell = (x_{\ell,0}, \dots x_{\ell,m-1})$. Then,

$$x_i = \sum_{j=\ell\in[\varphi/2]} \hat{x}_{\ell,i} b^+_j.$$

For canonical 2-norm, consider $\|\sigma(x_i)\|_2 = \beta_i$, by applying Corollary 1, we derive that $\|\psi(x_i)\|_\infty \le \beta_i$. We observe that as $\|\psi(\tilde{x}_{i,j})\|_\infty \le \beta_i$, then $\|\psi(\hat{x}_{\ell,i})\|_\infty \le 2\varphi\beta_i$ and $\|\sigma(\hat{x}_{\ell,i})\|_2 \le 2\varphi\sqrt{\hat{\mathfrak{f}}\varphi}\,\beta_i$ due to the norm conversion.

Eventaully, consider the norm of $\hat{\mathbf{x}}_\ell$, i.e.

$$\|\sigma(\hat{\mathbf{x}}_\ell)\|_2 \le \sqrt{\sum_{i\in[m]} \left(\beta_i 2\varphi\sqrt{\hat{\mathfrak{f}}\varphi}\right)^2} = 2\varphi\sqrt{\hat{\mathfrak{f}}\varphi}\sqrt{\sum_{i\in[m]}\beta_i^2} = 2\varphi\sqrt{\hat{\mathfrak{f}}\varphi}\,\|\sigma(\mathbf{x})\|_2 \le 2\varphi\sqrt{\hat{\mathfrak{f}}\varphi}\,\beta.$$

For coefficient ∞-norm, observe that $\|\psi(x_i)\|_\infty \le \beta$. Further, $\|\psi(\tilde{x}_{i,j})\|_\infty \le \beta$ so $\|\psi(\hat{x}_{\ell,i})\|_\infty \le 2\varphi\beta$ and $\|\psi(\hat{\mathbf{x}}_j)\|_\infty \le 2\varphi\beta$.

To argue about the composite case, we consider bases $\mathbf{b} = \mathbf{b}^{(1)} \otimes \mathbf{b}^{(2)}$, where $\mathbf{b}^{(1)}$ and $\mathbf{b}^{(2)}$ are bases of the cyclotomic rings with prime conductors. Let $b_{i,j} = b_i \cdot b_j$ and $b^+_{i,j} = b^+_i \cdot b^+_j$. Let $\varphi = \varphi^{(1)} \cdot \varphi^{(2)}$ and $\zeta = \zeta^{(1)} \cdot \zeta^{(2)}$ defined analogously. Then, we write:

$$\begin{aligned}
x_i &= \sum_{j^{(1)}\in[\varphi^{(1)}], j^{(2)}\in[\varphi^{(2)}]} \tilde{x}_{i,j^{(1)},j^{(2)}} b^{(1)}_{\ell^{(1)}} b^{(2)}_{\ell^{(2)}} \\
&= \sum_{\substack{j^{(1)}\in[\varphi^{(1)}]\\ j^{(2)}\in[\varphi^{(2)}]\\ \ell^{(1)}\in[\varphi^{(1)}/2]\\ \ell^{(2)}\in[\varphi^{(2)}/2]}} \tilde{x}_{i,j^{(1)},j^{(2)}} b^{(1)}_{\ell^{(1)}} b^{(2)}_{\ell^{(2)}} \\
&= \sum_{\substack{j^{(1)}\in[\varphi^{(1)}]\\ j^{(2)}\in[\varphi^{(2)}]\\ \ell^{(1)}\in[\varphi^{(1)}/2]\\ \ell^{(2)}\in[\varphi^{(2)}/2]}} \tilde{x}_{i,j^{(1)},j^{(2)}} (s^{(1)}_{j^{(1)},\ell^{(1)}} + t^{(1)}_{j^{(1)},\ell^{(1)}} \cdot \zeta^{(1)})(s^{(2)}_{j^{(2)},\ell^{(2)}} + t^{(2)}_{j^{(2)},\ell^{(2)}} \cdot \zeta^{(2)}) \cdot b^+_{\ell^{(1)}.\ell^{(2)}}
\end{aligned}$$

Let $\hat{x}_{i,\ell^{(1)},\ell^{(2)}} = \sum_{j^{(1)}\in[\varphi^{(1)}],j^{(2)}\in[\varphi^{(2)}]} \tilde{x}_{i,j^{(1)},j^{(2)}} (s^{(1)}_{j^{(1)},\ell^{(1)}} + t^{(1)}_{j^{(1)},\ell^{(1)}} \cdot \zeta^{(1)})(s^{(2)}_{j^{(2)},\ell^{(2)}} + t^{(2)}_{j^{(2)},\ell^{(2)}} \cdot \zeta^{(2)})$. Then, For canonical 2-norm, we observe that as $\|\psi(\tilde{x}_{i,j^{(1)},j^{(2)}})\|_\infty \le \beta_i$, then $\|\psi(\hat{x}_{i,\ell^{(1)},\ell^{(2)}})\|_\infty \le 4\varphi\beta_i$. For coefficient ∞-norm, we observe that as $\|\psi(\tilde{x}_{i,j^{(1)},j^{(2)}})\|_\infty \le \beta$, then $\|\psi(\hat{x}_{i,\ell^{(1)},\ell^{(2)}})\|_\infty \le 4\varphi\beta$. Continue the reasoning as in the base case. Clearly, the argument extends for terson rings of more than two prime rings. □

8 Packed $\mathbb{Z}$-Inner Products via CRT Embedding

The idea of embedding $\mathbb{Z}$-relations into $\mathcal{R}$-relations via the CRT embedding is well-established (e.g. [BS23, LNS20]). However, an obstacle to applying this to lattice-based succinct arguments is the lack of a succinct-verifier argument for proving the consistency of two vectors related via the coefficient and the CRT embeddings.

In this section, we first recall the method of embedding $\mathbb{Z}$-relations into $\mathcal{R}$-relations via the CRT embedding. Then, by exploiting the fine-grained tower structure of cyclotomic rings with smooth conductors, we provide a verifier-succinct argument for proving the consistency between the coefficient and the CRT embeddings. Throughout this section, we assume that $\mathcal{R} = \mathbb{Z}[\zeta_\mathfrak{f}]$ is a cyclotomic ring of degree φ, and $p \in \mathbb{N}$ is a rational prime which splits completely over $\mathcal{R}$.

8.1 Embedding $\mathbb{Z}_p$-inner products into $\mathcal{R}_p$-inner products

To begin, let us write $\mathsf{CRT}_p : \mathcal{R} \to \mathbb{Z}^\varphi$, for the invertible $\mathbb{Z}$-linear transform which maps a ring element $x \in \mathcal{R}$ to its Chinese remainder representation modulo each prime ideal dividing p. Note that we are viewing CRT_p as a $\mathbb{Z}$-linear map rather than a $\mathbb{Z}_p$-linear map, and we will write mod p explicitly when reducing modulo p. We extend the notation naturally to vectors, i.e. for $\mathbf{x} = (x_i)_{i\in[m]} \in \mathcal{R}^m$ we define $\mathsf{CRT}_p(\mathbf{x}) = (\mathsf{CRT}_p(x_i))_{i\in[m]}$.

It is well-known that addition and multiplication in the CRT domain is component-wise. Using this property, there exists a natural method of embedding $\mathbb{Z}_p$-inner products into $\mathcal{R}_p$-inner-products, as summarised in Proposition 3. The proof is trivial and thus omitted.

Proposition 3. *Let $\mathcal{R} = \mathbb{Z}[\zeta_\mathsf{f}]$ be a cyclotomic ring of degree φ and $p \in \mathbb{N}$ be a rational prime which fully splits over $\mathcal{R}$. Let $\tau_p : \mathcal{R} \to \mathbb{Z}$ be defined as $\tau_p(z) := \langle \mathbf{1}^\varphi, \mathsf{CRT}_p(z)\rangle$. For any $\mathbf{x} = (\mathbf{x}_i)_{i\in[m]}, (\mathbf{y}_i)_{i\in[m]} \in \mathbb{Z}^{m\varphi}$, it holds that*

$$\tau_p(\langle \mathsf{CRT}_p^{-1}(\mathbf{x}), \mathsf{CRT}_p^{-1}(\mathbf{y})\rangle_\mathcal{R}) = \langle \mathbf{x}, \mathbf{y}\rangle_\mathbb{Z} \bmod p.$$

Using Proposition 3, a prover is already able to succinctly prove that certain $\mathbb{Z}_p$-inner product relations hold using the succinct arguments provided in Section 5, provided that the application allows the witness vectors to be committed in their $\mathsf{CRT}_p^{-1}(\cdot)$ form. A bit more concretely, consider a toy example where the prover wishes to prove that $\langle \mathbf{x}, \mathbf{y}\rangle_\mathbb{Z} = z \bmod p$ for some public value $z \in \mathbb{Z}_p$, where p is sufficiently shorter than the modulus q used in the argument system. It performs the following procedures:

- Compute $\widehat{z} := \langle \mathsf{CRT}_p^{-1}(\mathbf{x}), \mathsf{CRT}_p^{-1}(\mathbf{y})\rangle_\mathcal{R} \bmod p$ and send it to the verifier.
- Let $\widehat{\mathbf{x}} := \mathsf{CRT}_p^{-1}(\mathbf{x})$ and $\mathsf{CRT}_p^{-1}(\mathbf{y}) := \mathsf{CRT}_p^{-1}(\mathbf{y}) \bmod p$.
- Find $\mathbf{r} \in \mathcal{R}^m$ such that $\widehat{z} = \langle \widehat{\mathbf{x}}, \widehat{\mathbf{y}}\rangle_\mathcal{R} + p \cdot \mathbf{r}$.
- Commit to $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}, \mathbf{r})$.
- Provide a proof that $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}, \mathbf{r})$ satisfies $\widehat{z} = \langle \widehat{\mathbf{x}}, \widehat{\mathbf{y}}\rangle_\mathcal{R} + p \cdot \mathbf{r}$.

In turn, the verifier checks that $\tau_p(\widehat{z}) = z \bmod p$ and the proof for $\widehat{z} = \langle \widehat{\mathbf{x}}, \widehat{\mathbf{y}}\rangle_\mathcal{R} + p \cdot \mathbf{r}$ is valid. If both checks go through, then by the soundness of the argument system the verifier would be convinced that $\langle \mathbf{x}, \mathbf{y}\rangle_\mathbb{Z} = z \bmod p$ (for some $\mathbf{x}, \mathbf{y}$ which satisfy $(\mathbf{x}, \mathbf{y}) = (\mathsf{CRT}_p(\widehat{\mathbf{x}}), \mathsf{CRT}_p(\widehat{\mathbf{y}})) \bmod p$).

8.2 Lifting to $\mathbb{Z}$ and $\mathcal{R}$

In case the prover wishes to prove that $\langle \mathbf{x}, \mathbf{y}\rangle_\mathbb{Z} = z$ without reduction modulo p, and/or the application postulates that the witness vectors are committed in ψ^{-1} form, then the prover needs to additionally perform the following:

- Write $\widetilde{\mathbf{x}} := \psi^{-1}(\mathbf{x})$ and $\widetilde{\mathbf{y}} := \psi^{-1}(\mathbf{y})$.
- Find $\mathbf{r}, \widetilde{\mathbf{s}} \in \mathcal{R}^m$ such that

$$\widetilde{\mathbf{x}} = \psi^{-1}(\mathsf{CRT}_p(\widehat{\mathbf{x}})) + p \cdot \mathbf{r}, \tag{15}$$

$$\widetilde{\mathbf{y}} = \psi^{-1}(\mathsf{CRT}_p(\widehat{\mathbf{y}})) + p \cdot \widetilde{\mathbf{s}}. \tag{16}$$

- Further commit to $(\widetilde{\mathbf{x}}, \widetilde{\mathbf{y}}, \mathbf{r}, \widetilde{\mathbf{s}})$.
- Provide a proof that $\|\sigma(\widetilde{\mathbf{x}})\|_2$ and $\|\sigma(\widetilde{\mathbf{y}})\|_2$ are short.
- Provide a proof that Eqs. (15) and (16) hold.

From Eqs. (15) and (16), the verifier would be convinced that

$$\psi(\widetilde{\mathbf{x}}) = \mathsf{CRT}_p(\widehat{\mathbf{x}}) \bmod p \qquad \text{and} \qquad \psi(\widetilde{\mathbf{y}}) = \mathsf{CRT}_p(\widehat{\mathbf{y}}) \bmod p.$$

Combined with the previous guarantee that $\langle \mathsf{CRT}_p(\widehat{\mathbf{x}}), \mathsf{CRT}_p(\widehat{\mathbf{y}})\rangle_\mathbb{Z} = z \bmod p$, the verifier would be convinced that

$$\langle \psi(\widetilde{\mathbf{x}}), \psi(\widetilde{\mathbf{y}})\rangle_\mathbb{Z} = z \bmod p.$$

With the proof of $\|\sigma(\widetilde{\mathbf{x}})\|_2$ and $\|\sigma(\widetilde{\mathbf{y}})\|_2$ being short, it must be the case that $\|\psi(\widetilde{\mathbf{x}})\|_\infty$ and $\|\psi(\widetilde{\mathbf{y}})\|_\infty$ are also short. Provided that these norms are small enough relative to p, the reduction modulo p has no effect, and thus we arrive at

$$\langle \psi(\widetilde{\mathbf{x}}), \psi(\widetilde{\mathbf{y}})\rangle_\mathbb{Z} = z.$$

Next, we discuss how to succinctly instantiate the above protocol, in particular the arguments for Eqs. (15) and (16), using tools developed in Section 5.

8.3 Computing CRT via Automorphisms

In the above, we established that the method of embedding $\mathbb{Z}$-inner products via CRT requires proving consistency between the $\mathsf{CRT}_p^{-1}(\cdot)$ and $\psi^{-1}(\cdot)$ encodings of the witness vectors. Specifically, we would like to design a succinct argument for arguing that

$$\widetilde{\mathbf{x}} = \psi^{-1}(\mathsf{CRT}_p(\widehat{\mathbf{x}})) \bmod p$$

where $\widetilde{\mathbf{x}}, \widehat{\mathbf{x}} \in \mathcal{R}^m$ are committed vectors. We note that existing protocols for proving such correspondence treat $\psi^{-1} \circ \mathsf{CRT}_p$ as a generic $\mathbb{Z}$-linear map and prove the correspondence as an unstructured system of linear equations over $\mathbb{Z}$. Instead, we would like to exploit the tensor structure of the $\psi^{-1} \circ \mathsf{CRT}_p$ map for carefully chosen rings $\mathcal{R}$, and the fact that any $\mathbb{Z}$-linear map can be expressed as a linear combination of automorphisms in $\mathsf{Gal}(\mathcal{K}/\mathbb{Q})$ with $\mathcal{R}$ coefficients.

Motivated by the above, the goal of this subsection is to prove Theorem 9, which states that, for $\mathcal{R}$ with a smooth conductor, the $\psi^{-1} \circ \mathsf{CRT}_p$ map can be expressed as the composition of $O(\log \mathfrak{f})$ layers, each being a succinct $\mathcal{R}$-linear combination of automorphisms in $\mathsf{Gal}(\mathcal{K}/\mathbb{Q})$. Our proof invokes the Galois decomposition framework of Peikert and Pepin [PP25]: the Kronecker factorisation of the CRT basis along the tower ([PP25, Definition 5.8, Lemma 5.11]), the top-down layered factorisation of the CRT map ([PP25, Section 4.2, Equation (4.3)]), and the support analysis of each layer ([PP25, Lemmas 4.1, 4.3, 5.13]). Combining these ingredients we arrive at the following theorem.

Theorem 9. *Let $\mathcal{R}$ be a cyclotomic ring[23] with a w-smooth conductor $\mathfrak{f}$, and let p be a rational prime which fully splits over $\mathcal{R}$. Then, the transformation $\psi^{-1} \circ \mathsf{CRT}_p : \mathcal{R}/p\mathcal{R} \to \mathcal{R}/p\mathcal{R}$ can be expressed as a composition of $O(\log \mathfrak{f})$ succinct $\mathcal{R}$-linear combinations of automorphisms over $\mathcal{R}$, where each layer uses at most w^2 automorphisms. Formally, there exist $t \in O(\log \mathfrak{f})$, $h_i \le w^2$, $s_{i,j} \in \mathcal{R}$, and $\alpha_{i,j} \in \mathsf{Gal}(\mathcal{K}/\mathbb{Q})$ for all $i \in [t]$ and $j \in [h_i]$, such that*

$$(\psi^{-1} \circ \mathsf{CRT}_p)(\cdot) = \bigcirc_{i \in [t]} \sum_{j \in [h_i]} s_{i,j} \alpha_{i,j}(\cdot) \bmod p,$$

where $\bigcirc$ denotes function composition.

Proof. Let $\mathcal{K}$ denote the $\mathfrak{f}$-th cyclotomic field. Since $\mathfrak{f}$ is w-smooth, we fix a tower of Galois extensions

$$\mathbb{Q} = K_0 \subset K_1 \subset \cdots \subset K_t = \mathcal{K},$$

with $t \in O(\log \mathfrak{f})$ steps and relative degrees $[K_{i+1} : K_i] \le w$ for every $i \in \{0, \ldots, t-1\}$.

By [PP25, Definition 5.8 and Lemma 5.11], the CRT basis of $\mathcal{O}_\mathcal{K}/p\mathcal{O}_\mathcal{K}$ (with respect to the prime ideals of p in $\mathcal{O}_\mathcal{K}$) factors as a Kronecker product down this tower with respect to a *consistent transversal* $T = T_0 \cdot T_1 \cdots T_{t-1}$, where each $T_i \subset \mathsf{Gal}(\mathcal{K}/K_i)$ is a set of coset representatives for $\mathsf{Gal}(\mathcal{K}/K_{i+1})$ in $\mathsf{Gal}(\mathcal{K}/K_i)$. In particular, $|T_i| = [K_{i+1} : K_i] \le w$. Applying the top-down decomposition of [PP25, Section 4.2, Equation (4.3)] recursively along the tower yields a layered factorisation

$$\psi^{-1} \circ \mathsf{CRT}_p \;\equiv\; L_0 \circ L_1 \circ \cdots \circ L_{t-1} \pmod p,$$

where, for each i, the *factor-converting* layer L_i is the $\mathcal{O}_{K_i}/p\mathcal{O}_{K_i}$-linear endomorphism of $\mathcal{O}_\mathcal{K}/p\mathcal{O}_\mathcal{K}$ that converts the K_{i+1}-factor representation of its input into the K_i-factor representation.

We bound the support of each L_i. By the quotient (ideal) form of [PP25, Lemma 4.1], the map L_i admits the expansion

$$L_i(x) \;=\; \sum_{\tau \in \mathsf{Gal}(\mathcal{K}/K_i)} s_{i,\tau} \cdot \tau(x) \bmod p,$$

with coefficients $s_{i,\tau} \in \mathcal{O}_\mathcal{K}/p\mathcal{O}_\mathcal{K}$ of the inner-product form $\langle \mathbf{c}_i, \tau(\mathbf{b}_i^\vee) \rangle$ for an appropriate K_i-basis $\mathbf{b}_i$ of the relevant factor and image vector $\mathbf{c}_i$. Both [PP25, Lemma 4.3] and the sparsity analysis of [PP25, Section 5.3] are stated for the map sending the structured basis $\mathbf{b}$ to the CRT basis $\mathbf{c}$, whereas L_i realises the inverse conversion. The analysis transfers since [PP25, Lemma 4.1] is symmetric in the roles of $\mathbf{b}$ and

[23] The technique should apply for the ring of integers of any Abelian number field with a known w-smooth tower structure, but a formal proof is out of scope.

$\mathbf{c}$. Concretely, writing the CRT basis in its Kronecker-factored form $\mathbf{c}_{\mathcal{K}/K_{i+1}} \otimes \mathbf{c}_{K_{i+1}/K_i}$ and $\mathbf{b}_{K_{i+1}/K_i}$ for the structured basis of the converted factor, the coefficient of τ in L_i is

$$\langle \mathbf{c}_{\mathcal{K}/K_{i+1}} \otimes \mathbf{b}_{K_{i+1}/K_i},\ \tau(\mathbf{c}^\vee_{\mathcal{K}/K_{i+1}} \otimes \mathbf{c}^\vee_{K_{i+1}/K_i})\rangle \ =\ \langle \mathbf{c}_{\mathcal{K}/K_{i+1}}, \tau(\mathbf{c}^\vee_{\mathcal{K}/K_{i+1}})\rangle \cdot \langle \mathbf{b}_{K_{i+1}/K_i}, \tau|_{K_{i+1}}(\mathbf{c}^\vee_{K_{i+1}/K_i})\rangle,$$

whose first factor is exactly the one constrained by [PP25, Lemma 5.13], so the support condition and the resulting bound below carry over unchanged. By [PP25, Lemma 4.3], this coefficient factors over the tower and *vanishes* whenever τ lies outside the set

$$T_i \cdot T_i^{-1} \ :=\ \{\, t' \circ t^{-1} \,:\, t, t' \in T_i \,\} \ \subseteq\ \mathsf{Gal}(\mathcal{K}/K_i) \ \subseteq\ \mathsf{Gal}(\mathcal{K}/\mathbb{Q}).$$

Finally, [PP25, Lemma 5.13] gives the cardinality bound

$$|T_i \cdot T_i^{-1}| \ \le\ |T_i|^2 \ =\ [K_{i+1} : K_i]^2 \ \le\ w^2.$$

Setting $h_i := |T_i \cdot T_i^{-1}| \le w^2$ and re-indexing the (at most h_i) nonzero terms of L_i as $\{(s_{i,j}, \alpha_{i,j})\}_{j \in [h_i]}$ with $s_{i,j} \in \mathcal{R}/p\mathcal{R}$ and $\alpha_{i,j} \in \mathsf{Gal}(\mathcal{K}/\mathbb{Q})$, we conclude

$$(\psi^{-1} \circ \mathsf{CRT}_p)(\cdot) \ =\ \bigcirc_{i \in [t]} \sum_{j \in [h_i]} s_{i,j} \cdot \alpha_{i,j}(\cdot) \bmod p,$$

with $t \in O(\log \mathfrak{f})$ and $h_i \le w^2$ for every $i \in [t]$, as claimed. □

Remark 9 (Direct-product tower steps). The step K_{i+1}/K_i is called a *direct-product split* when K_{i+1} is the compositum $K_{i+1} = L \cdot L'$ of two intermediate fields L, L' with $L \cap L' = K_i$ and coprime relative degrees, so that $\mathsf{Gal}(K_{i+1}/K_i) \cong \mathsf{Gal}(L/K_i) \times \mathsf{Gal}(L'/K_i)$. The canonical example is $K_i = \mathbb{Q}(\zeta_{\mathfrak{f}_i})$, $K_{i+1} = \mathbb{Q}(\zeta_{\mathfrak{f}_i \cdot \mathfrak{f}'})$ with $\gcd(\mathfrak{f}_i, \mathfrak{f}') = 1$, where the new layer factors off via $\mathbb{Q}(\zeta_{\mathfrak{f}_i \cdot \mathfrak{f}'}) \cong \mathbb{Q}(\zeta_{\mathfrak{f}_i}) \otimes_{\mathbb{Q}} \mathbb{Q}(\zeta_{\mathfrak{f}'})$. In this case the consistent transversal T_i is itself a subgroup, so $T_i \cdot T_i^{-1} = T_i$ and $h_i \le |T_i| = [K_{i+1} : K_i] \le w$ by [PP25, Case 1 of Lemma 5.13]. If every step of the tower is direct-product, then $h_i \le w$ uniformly and the total automorphism count is at most $w \cdot t \in O(w \log \mathfrak{f})$.

Remark 10 (Power-of-two cyclotomic rings). For the power-of-two cyclotomic ring $\mathcal{R} = \mathbb{Z}[\zeta_{2^k}]$ (and the totally-real power-of-two cyclotomic), the per-layer bound improves to $h_i \ \le\ 2w - 1$ by [PP25, Case 2 of Lemma 5.13], where these settings are explicitly identified as instances of Case 2. Taking the natural tower of degree-2 steps gives $w = 2$ and $h_i \le 2w - 1 = 3$ at every layer, so the total number of automorphisms across all $t \in O(\log \mathfrak{f})$ layers is in $O(\log \mathfrak{f})$.

Remarks 9 and 10 can be applied jointly, when some steps of the tower are direct-product and the remaining steps are power-of-two. It perfectly aligns with the settings of Section 4, where the conductor $\mathfrak{f}$ is a product of distinct small prime powers, and a power-of-two.

8.4 Π^{eip}: Extended Inner-product relation

The Ξ^{ip} relation defined in Section 5.6 asserts a single constraint on the self-inner-product of the entire witness. In preparation for our CRT-based embedding for $\mathbb{Z}$-inner-products to be presented in Section 8, we need a slightly extended relation which captures multiple inner-product relations between different blocks of the witness vector, which is now interpreted as a block vector. Formally, we define the "extended inner-product" relation Ξ^{eip} below.

$$\Xi^{\mathsf{eip} \vee \mathsf{sis}}_{\substack{\mathcal{R},q,m,n^{\mathsf{out}},r,\mu,\\ \beta,n^{\mathsf{blk}},n^{\mathsf{ip}},\alpha,\beta^{\mathsf{sis}}}} := \left\{ \begin{array}{l} ((\iota_{\mathsf{ip}}, \iota_{\mathsf{ip\text{-}in}}, \mathbf{H}, \mathbf{F}, \mathbf{Y}, \mathbf{c}, \mathbf{t}), \mathbf{W} \text{ or } \mathbf{w}): \\ \quad \mathbf{H} \in \mathcal{R}_q^{n^{\mathsf{out}} \times n};\ \mathbf{F} \in \mathcal{R}_q^{n \times d^{\otimes \mu}} \subseteq \mathcal{R}_q^{n \times m};\ \mathbf{Y} \in \mathcal{R}_q^{n^{\mathsf{out}} \times r}, \mathbf{c} \in \mathcal{R}_q^r, \mathbf{t} \in \mathcal{R}_q^{n^{\mathsf{ip}}} \\ \quad \iota_{\mathsf{ip}} \in \{1, -1\} \to [n^{\mathsf{ip}}];\ \iota_{\mathsf{ip\text{-}in}} \in [n^{\mathsf{blk}}] \to [n^{\mathsf{ip}}] \\ \quad \mathbf{W} = (\mathbf{w}_i)_{i \in [r]};\ \mathbf{w}_i^{\mathsf{T}} = (\mathbf{w}_{i,k}^{\mathsf{T}})_{k \in [n^{\mathsf{blk}}]} \\ \quad \left\{ \begin{array}{c} \|\mathbf{W}\| \le \beta \\ \mathbf{HFW} = \mathbf{Y} \bmod q \\ \alpha(\mathbf{c}) = \mathbf{c} \\ \sum_{i \in [r]} c_i \langle \mathbf{w}_{i,\iota_{\mathsf{ip\text{-}in}}(k)}, \alpha_{\iota_{\mathsf{ip}}(k)}(\mathbf{w}_{i,\iota_{\mathsf{ip\text{-}in}}(k)}) \rangle_{\mathcal{R}} = t_k \bmod q \\ \forall [k] \in [n^{\mathsf{ip}}] \end{array} \right\} \text{ or } \left\{ \begin{array}{c} \|\mathbf{w}\| \le \beta^{\mathsf{sis}} \\ \overline{\mathbf{H}}\mathbf{F}\mathbf{w} = \mathbf{0}_{\overline{n}} \bmod q \end{array} \right\} \end{array} \right\}.$$

Note that Ξ^{eip} implicitly has two additional parameters: number of blocks n^{blk} and number of inner-product relations n^{ip}. Compared to Ξ^{ip}, a statement in Ξ^{eip} contains additionally two index maps $\iota_{\mathsf{ip}} : [n^{\mathsf{ip}}] \to \{1, -1\}$ and $\iota_{\mathsf{ip\text{-}in}} : [n^{\mathsf{ip}}] \to [n^{\mathsf{blk}}]$, and the inner product image t is replaced by a vector $\mathbf{t} \in \mathcal{R}_q^{n^{\mathsf{ip}}}$. Furthermore, the single inner-product relation in Ξ^{ip} is replaced with weighted inner-product

$$\sum_{i\in[r]} c_i \langle \mathbf{w}_{i,\iota_{\mathsf{ip\text{-}in}}(k)}, \alpha_{\iota_{\mathsf{ip}}(k)}(\mathbf{w}_{i,\iota_{\mathsf{ip\text{-}in}}(k)}) \rangle_{\mathcal{R}} = t_k \bmod q \quad \forall k \in [n^{\mathsf{ip}}].$$

Since Π^{ip} is already quite notation heavy, and the generalisations to Π^{eip} is straightforward, we omit a formal description but instead highlight the differences between Π^{eip} and Π^{ip} below.

The main protocol The main protocol Π^{eip} differs from Π^{ip} in the following:

(i) The protocol runs in parallel for $j \in [n^{\mathsf{ip}}]$:

(i) The witness $\mathbf{W}$ used for obtaining $\mathbf{V}$ (now, denoted as $\widetilde{\mathbf{V}}_j$) is replaced by $\mathbf{W}_{\iota_{\mathsf{ip\text{-}in}}(j)}$. Let

$$\widehat{\mathbf{V}}_j \coloneqq \mathbf{e}_{\iota_{\mathsf{ip\text{-}in}}(j)} \otimes \widetilde{\mathbf{V}}_j.$$

(ii) Similarly, matrix $\mathbf{E}$ is replaced by

$$\widehat{\mathbf{E}}_j \coloneqq \mathbf{e}_{\iota_{\mathsf{ip\text{-}in}}(j)} \otimes \begin{pmatrix} 1 & \xi & \dots & \xi^{m/n^{\mathsf{blk}}-1} \\ 1 & \bar{\xi}^{-1} & \dots & \bar{\xi}^{-(m/n^{\mathsf{blk}}-1)} \\ 1 & 0 & \dots & 0 \end{pmatrix}.$$

(ii) The new claims both parties compute are:

$$\widetilde{\mathbf{H}} \coloneqq \begin{pmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{n^{\mathsf{aut}}\cdot 3} \end{pmatrix}, \quad \widetilde{\mathbf{F}} \coloneqq \begin{pmatrix} \mathbf{F} \\ \sum_{j\in[n^{\mathsf{ip}}]} \mathbf{e}_j \otimes \widehat{\mathbf{E}}_j \end{pmatrix},$$

$$\widetilde{\mathbf{Y}} \coloneqq \begin{pmatrix} \mathbf{Y} \\ \sum_{j\in[n^{\mathsf{ip}}]} \mathbf{e}_j \otimes \mathbf{Y}_{\widehat{\mathbf{E}}_j} \end{pmatrix}, \quad \widetilde{\mathbf{Y}}'_i \coloneqq \begin{pmatrix} \mathbf{Y}'_i \\ \sum_{j\in[n^{\mathsf{ip}}]} \mathbf{e}_j \otimes \mathbf{Y}'_{\widehat{\mathbf{E}}_j,i} \end{pmatrix},$$

where $\mathbf{Y}_{\widehat{\mathbf{E}}_j} \coloneqq \widehat{\mathbf{E}}_j \mathbf{W}$ and $\mathbf{Y}'_{\widehat{\mathbf{E}}_j,i} \coloneqq \widehat{\mathbf{E}}_j \widehat{\mathbf{V}}_{j,i}$.

(iii) Further, the verifications are replaced by

$$\overline{\mathbf{V}} = \sum_{[i,j]\in[\ell,n^{\mathsf{ip}}]} \mathbf{e}_i^{\mathsf{T}} \otimes \widehat{\mathbf{V}}_{j,i}$$

$$\overline{\mathbf{W}} = (\overline{\mathbf{V}}, \mathbf{W})$$

$$\overline{\mathbf{Y}} = \sum_{[i]\in[\ell]} \mathbf{e}_i^{\mathsf{T}} \otimes \widetilde{\mathbf{Y}}'_i + \mathbf{e}_\ell^{\mathsf{T}} \otimes \widetilde{\mathbf{Y}}$$

$$((\widetilde{\mathbf{H}}, \widetilde{\mathbf{F}}, \overline{\mathbf{Y}}), \overline{\mathbf{W}})_{j\in[n^{\mathsf{ip}}]} \overset{?}{\in} \Xi^{\mathsf{lin}}_{m, n^{\mathsf{out}}+3\cdot n^{\mathsf{ip}}, r\cdot n^{\mathsf{ip}}\cdot(\ell+1), \beta_0}$$

for the same parameters as in Fig. 7.

We do not provide explicit proof of Lemma 10 below as it completely analogous to Lemma 8.

Lemma 10 (Extended Norm and Inner Product). *Let $m, r, \ell, b_{\mathsf{ip}} \in \mathbb{N}$ and $0 \le \beta \le \beta^{\mathsf{sis}} \le q$, $0 \le 2\beta' \le \beta^{\mathsf{sis}}$. Protocol Π^{ip} is a perfectly correct reduction of knowledge from Ξ^{ip} to Ξ^{lin}*

$$\left(m, n^{\mathsf{out}}, r, n^{\mathsf{ip}}, \beta,\right) \mapsto \left(m, n^{\mathsf{out}\prime}, r + n^{\mathsf{ip}} \cdot \ell, \beta_{\mathsf{out}}\right)$$

where $\beta_{\mathsf{out}} = \sqrt{\beta^2 + \beta_{\overline{\mathbf{V}}}{}^2}$ (resp. $\beta_{\mathsf{out}} = \max\{\beta, \beta_{\overline{\mathbf{V}}}\}$), and knowledge sound reduction of knowledge from Ξ^{ip} to Ξ^{lin} with parameters

$$\left(m, n^{\mathsf{out}}, r, \beta', \beta^{\mathsf{sis}}\right) \leftarrow\!\shortmid \left(m, n^{\mathsf{out}\prime}, r + n^{\mathsf{ip}} \cdot \ell, \beta_{\mathsf{out}}{}', \beta^{\mathsf{sis}}\right),$$

where $n^{\mathsf{out}\prime} = n^{\mathsf{out}} + 3n^{\mathsf{ip}}$, b_{ip} and $\ell \ge \log_{b_{\mathsf{ip}}}(2\beta_{\overline{\mathbf{V}}}{}^2 + 1)$ is such that $b_{\mathsf{ip}} \le 2\beta_{\overline{\mathbf{V}}}/(\sqrt{\ell m}\sqrt{\hat{\mathfrak{f}}\varphi})$ (resp. $b_{\mathsf{ip}} = 2\beta_{\overline{\mathbf{V}}} + 1$ and $\ell \ge \log_{b_{\mathsf{ip}}}(2\beta_{\overline{\mathbf{V}}} + 1)$). Extraction requires two uniformly distributed transcripts (Footnote 11) and has knowledge error $\kappa \le \frac{2m}{|\mathcal{C}_{\mathcal{R}_q}|}$. For Π^{norm}, the analogous statements hold.

8.5 Π^{aut}: Automorphism Check

For our CRT-based protocol, we need to efficiently check automorphism relations between different blocks of the witness. Formally, we define the automorphism check relation below.

$$\Xi^{\mathsf{aut}\vee\mathsf{sis}}_{\substack{\mathcal{R},q,m,n^{\mathsf{out}},r,\mu,\\ \beta,n^{\mathsf{blk}},n^{\mathsf{ip}},n^{\mathsf{aut}},\alpha,\beta^{\mathsf{sis}}}} := \left\{ \begin{array}{l} ((\iota_{\mathsf{ip}},\iota_{\mathsf{ip\text{-}in}},\iota_{\mathsf{aut}},\iota_{\mathsf{aut\text{-}in}},\iota_{\mathsf{aut\text{-}out}},\mathbf{H},\mathbf{F},\mathbf{Y},\mathbf{c},\mathbf{t}),\mathbf{W} \text{ or } \mathbf{w}): \\ \quad \mathbf{H}\in\mathcal{R}_q^{n^{\mathsf{out}}\times n};\ \mathbf{F}\in\mathcal{R}_q^{n\times d^{\otimes\mu}}\subseteq\mathcal{R}_q^{n\times m};\ \mathbf{Y}\in\mathcal{R}_q^{n^{\mathsf{out}}\times r},\mathbf{c}\in\mathcal{R}_q^{r},\mathbf{t}\in\mathcal{R}_q^{n^{\mathsf{ip}}} \\ \quad \iota_{\mathsf{ip}}\in\{1,-1\}\to[n^{\mathsf{ip}}];\ \iota_{\mathsf{ip\text{-}in}}\in[n^{\mathsf{blk}}]\to[n^{\mathsf{ip}}] \\ \quad \iota_{\mathsf{aut}}\in[\varphi]\to[n^{\mathsf{aut}}];\ \iota_{\mathsf{aut\text{-}in}},\iota_{\mathsf{aut\text{-}out}}\in[n^{\mathsf{blk}}]\to[n^{\mathsf{aut}}] \\ \quad \mathbf{W}=(\mathbf{w}_i)_{i\in[r]};\ \mathbf{w}_i=\underbrace{\widetilde{\mathbf{w}_{i,k}}}_{k\in[n^{\mathsf{blk}}]};\ \widetilde{\mathbf{W}}_k=\underbrace{\widetilde{\mathbf{W}_k}}_{k\in[n^{\mathsf{blk}}]} \\ \quad \left\{ \begin{array}{c} \|\mathbf{W}\|\le\beta \\ \mathbf{HFW}=\mathbf{Y} \bmod q \\ \alpha(\mathbf{c})=\mathbf{c} \\ \sum_{i\in[r]} c_i\langle\mathbf{w}_{i,\iota_{\mathsf{ip\text{-}in}}(k)},\alpha_{\iota_{\mathsf{ip}}(k)}(\mathbf{w}_{i,\iota_{\mathsf{ip\text{-}in}}(k)})\rangle_{\mathcal{R}}=t_k \bmod q\forall[k]\in[n^{\mathsf{ip}}] \\ \alpha_{\iota_{\mathsf{aut}}(k)}(\widetilde{\mathbf{W}}_{i,\iota_{\mathsf{aut\text{-}in}}(k)})=\widetilde{\mathbf{W}}_{i,\iota_{\mathsf{aut\text{-}out}}(k)} \bmod q\ \ \forall k\in[n^{\mathsf{aut}}] \end{array} \right\} \\ \quad \text{or}\ \left\{ \begin{array}{c} \|\mathbf{w}\|\le\beta^{\mathsf{sis}} \\ \overline{\mathbf{H}}\mathbf{F}\mathbf{w}=\mathbf{0}_{\overline{n}} \bmod q \end{array} \right\} \end{array} \right\}.$$

Note that Ξ^{aut} implicitly has an additional parameter – the number of automorphism relations n^{aut}. An instance of Ξ^{aut} is almost identical to an instance of Ξ^{eip}, except that it additionally asserts that the automorphism relations $\alpha_{\iota_{\mathsf{aut}}(k)}(\mathbf{w}_{\iota_{\mathsf{aut\text{-}in}}(k)})=\mathbf{w}_{\iota_{\mathsf{aut\text{-}out}}(k)} \bmod q$ hold for all $k\in[n^{\mathsf{aut}}]$ for the index maps $\iota_{\mathsf{aut}},\iota_{\mathsf{aut\text{-}in}},\iota_{\mathsf{aut\text{-}out}}$ given as part of the statement. Thus, to see that Ξ^{aut} reduces to Ξ^{eip}, we only need to check that the automorphism relations can be reduced to linear relations. We outline the logic of this reduction below and give a formal description in Fig. 9.

Instead of checking that all automorphism relations hold, which would not be succinct, the verifier sends a random $\xi\leftarrow\$\,\mathcal{C}_{\mathcal{R}_q}\subseteq\mathcal{R}_q$. The prover then sends

$$\begin{aligned}\mathbf{z}_k &:= (1,\xi,\dots,\xi^{m-1})\cdot\widetilde{\mathbf{W}}_{\iota_{\mathsf{aut\text{-}in}}(k)} \bmod q\\ \mathbf{z}'_k &:= (1,\alpha_{\iota_{\mathsf{aut}}(k)}(\xi),\dots,\alpha_{\iota_{\mathsf{aut}}(k)}(\xi)^{m-1})\cdot\widetilde{\mathbf{W}}_{\iota_{\mathsf{aut\text{-}out}}(k)} \bmod q\end{aligned}$$

for $k\in[n^{\mathsf{aut}}]$.

In turn, the verifier checks that $\alpha_{\iota_{\mathsf{aut}}(k)}(\mathbf{z}_k)=\mathbf{z}'_k$ for $k\in[n^{\mathsf{aut}}]$. Completeness can be seen by observing

$$\begin{aligned}\alpha_{\iota_{\mathsf{aut}}(k)}(\mathbf{z}_k) &= \alpha_{\iota_{\mathsf{aut}}(k)}((1,\xi,\dots,\xi^{m-1})\cdot\widetilde{\mathbf{W}}_{\iota_{\mathsf{aut\text{-}in}}(k)})\\ &= (1,\alpha_{\iota_{\mathsf{aut}}(k)}(\xi),\dots,\alpha_{\iota_{\mathsf{aut}}(k)}(\xi)^{m-1})\cdot\alpha_{\iota_{\mathsf{aut}}(k)}(\widetilde{\mathbf{W}}_{\iota_{\mathsf{aut\text{-}in}}(k)})\\ &= (1,\alpha_{\iota_{\mathsf{aut}}(k)}(\xi),\dots,\alpha_{\iota_{\mathsf{aut}}(k)}(\xi)^{m-1})\cdot\widetilde{\mathbf{W}}_{\iota_{\mathsf{aut\text{-}out}}(k)}=\mathbf{z}'_k.\end{aligned}$$

This reduces the automorphism check to verifying the linear relations

$$\begin{aligned}\mathbf{z}_k &:= (1,\xi,\dots,\xi^{m-1})\cdot\widetilde{\mathbf{W}}_{\iota_{\mathsf{aut\text{-}in}}(k)} \bmod q\ \ \forall k\in[n^{\mathsf{aut}}],\\ \mathbf{z}'_k &:= (1,\alpha_{\iota_{\mathsf{aut}}(k)}(\xi),\dots,\alpha_{\iota_{\mathsf{aut}}(k)}(\xi)^{m-1})\cdot\widetilde{\mathbf{W}}_{\iota_{\mathsf{aut\text{-}out}}(k)} \bmod q\ \ \forall k\in[n^{\mathsf{aut}}]\end{aligned}$$

which is supported by Ξ^{eip}. Note that the reduction increases the number of rows significantly, and Π^{batch} could be run to batch them before further reductions.

The security of Π^{aut} is summarised below. Nevertheless, the proof is omitted due to its similarity with the previous proofs in this section.

Lemma 11 (Security of Π^{aut}). *Let $n^{\mathsf{out}},n^{\mathsf{ip}\prime},\in\mathbb{N}$ and $0\le\beta\le\beta^{\mathsf{sis}}\le q$. The protocol Π^{aut} is a perfectly correct reduction of knowledge from Ξ^{aut} to Ξ^{eip} with parameters*

$$(n^{\mathsf{ip}},n^{\mathsf{out}})\mapsto(n^{\mathsf{ip}\prime})$$

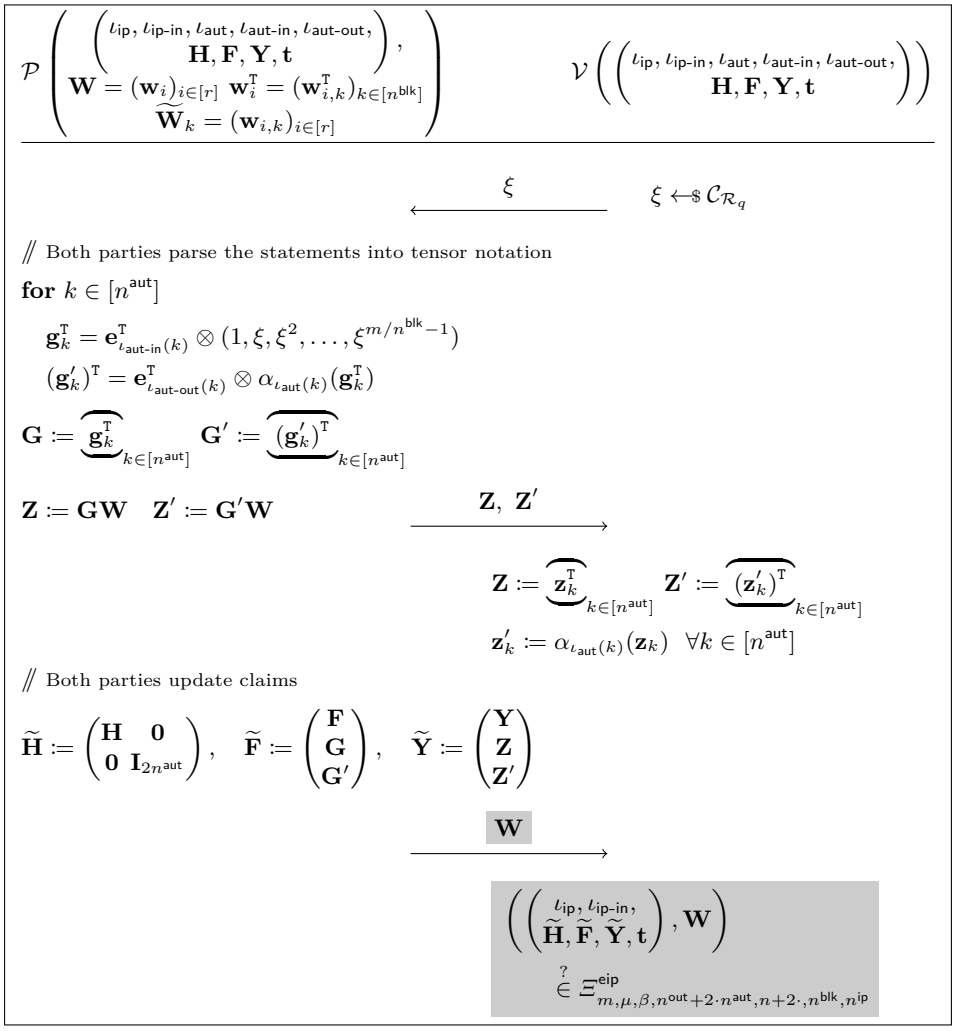

Fig. 9. Protocol Π^{aut}, a reduction of knowledge from $\Xi^{\mathsf{aut}}_{m,\mu,\beta,n^{\mathsf{out}},n,n^{\mathsf{blk}},n^{\mathsf{ip}},n^{\mathsf{aut}}}$ to $\Xi^{\mathsf{eip}}_{m,\mu,\beta,n^{\mathsf{out}}+2\cdot n^{\mathsf{aut}},n+2\cdot,n^{\mathsf{blk}},n^{\mathsf{ip}}}$. Π^{aut} sends the marked parts only as a *proof* (but not *reduction*) of knowledge.

and knowledge sound reduction of knowledge from Ξ^{aut} *to* Ξ^{eip}*, with parameters*

$$\left(n^{\mathsf{ip}},n^{\mathsf{out}},\beta^{\mathsf{sis}}\right)\leftarrow\!\shortmid\left(n^{\mathsf{ip}\prime},\beta^{\mathsf{sis}}\right),$$

where $n^{\mathsf{ip}\prime}=n^{\mathsf{out}}+n^{\mathsf{ip}}$ *The knowledge error is* $n^{\mathsf{aut}}/\left|\mathcal{C}_{\mathcal{R}_q}\right|$.

8.6 Reducing CRT-based Binariness Check to Automorphism Check

Equipped with Theorem 9 and Lemma 11, in Fig. 10, we construct a reduction of knowledge $\Pi^{\mathsf{lin\text{-}bin}}_{\mathsf{crt},p}$ using a CRT-based binariness check. For the ease of notation, the reduction of knowledge $\Pi^{\mathsf{lin\text{-}bin}}_{\mathsf{crt},p}$ in Fig. 10 makes use of the following subroutine $\mathsf{CRTmake}(\mathbf{H},\mathbf{F},\mathbf{Y},\mathbf{W})$ which maps a $\Xi^{\mathsf{lin}}\cap\Xi^{\mathsf{bin}}$ instance to a Ξ^{aut} instance:

$\mathsf{CRTmake}(\mathbf{H},\mathbf{F},\mathbf{Y},\mathbf{W})$

(i) Compute $\widehat{\mathbf{W}}_0\coloneqq\mathsf{CRT}_p^{-1}(\psi(\mathbf{W}))$ and let $\mathbf{R}_0\coloneqq\mathbf{0}^{m\times r}$.
(ii) Compute $\widehat{\mathbf{y}}^{\mathsf{T}}\coloneqq\left(\mathsf{CRT}^{-1}(\mathbf{1}^{\varphi m})\right)^{\mathsf{T}}\cdot\left(\mathsf{CRT}_p^{-1}(\psi(\mathbf{W}))\right)$, $t_0\coloneqq\sum_{i\in[r]}\langle\mathsf{CRT}_p^{-1}(\psi(\mathbf{w}_i)),\mathsf{CRT}_p^{-1}(\psi(\mathbf{w}_i))\rangle_{\mathcal{R}}$, and $t_1\coloneqq\sum_{i\in[r]}\langle\mathbf{w}_i,\overline{\mathbf{w}_i}\rangle_{\mathcal{R}}$.

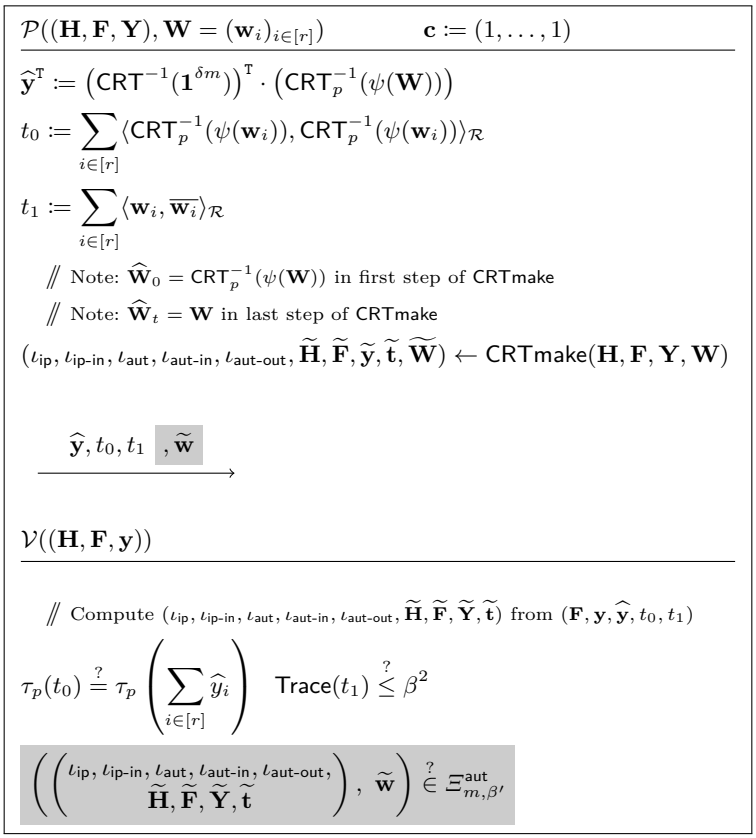

Fig. 10. Protocol $\Pi^{\mathsf{lin\text{-}bin}}_{\mathsf{crt},p}$, a reduction from $\Xi^{\mathsf{lin}} \cap \Xi^{\mathsf{bin}}$ to Ξ^{aut}, where β, β', and p are set as in Lemma 12. See the text for the definition of the subroutine CRTmake. The marked parts are only sent / checked when $\Pi^{\mathsf{lin\text{-}bin}}_{\mathsf{crt},p}$ is used as a proof of knowledge. As a reduction of knowledge, they are omitted.

(iii) For each $i \in [t]$:
  (i) Compute the i-th step of the CRT decomposition as described in Theorem 9, i.e. $\mathbf{W}_{i,j} \coloneqq \alpha_{i,j}(\widehat{\mathbf{W}}_i)$ for all $j \in [h_i]$.
  (ii) Compute the p-ary representative $\widehat{\mathbf{W}}_{i+1} \coloneqq \sum_{j\in[h_i]} s_{i,j}\mathbf{W}_{i,j} \bmod p$.
  (iii) Find the quotient $\mathbf{R}_{i+1} \in \mathcal{R}^m$ satisfying $\widehat{\mathbf{W}}_{i+1} - \sum_{j\in[h_i]} s_{i,j} \cdot \mathbf{W}_{i,j} + p \cdot \mathbf{R}_{i+1} = \mathbf{0}^{m\times r} \pmod q$.

(iv) Concatenate all intermediate witnesses as

$$\widetilde{\mathbf{W}} \coloneqq \begin{matrix} \overbrace{\underbrace{\widehat{\mathbf{W}}_{i,j}^{\mathsf{T}}}_{(i,j)\in[t,h_i]}}^{} \\ \overbrace{\underbrace{\widehat{\mathbf{W}}_{i}^{\mathsf{T}}}_{i\in[t+1]}}^{} \\ \underbrace{\overbrace{\mathbf{R}_{i}^{\mathsf{T}}}^{}_{i\in[t+1]}}_{} \end{matrix} \in \mathcal{R}^{\tilde{m}\cdot r}$$

where $\tilde{m} \coloneqq m \cdot n^{\mathsf{blk}}$ and $n^{\mathsf{blk}} \coloneqq w^2 \log \mathfrak{f} + 2t + 2$.[24]

(v) Parse the witness as a block vector $\widetilde{\mathbf{W}} = \overbrace{\widetilde{\mathbf{W}}_k}^{}_{k\in[n^{\mathsf{blk}}]}$ where $\widetilde{\mathbf{W}}_k \in \mathcal{R}^{m\cdot r}$.

(vi) Concatenate all linear map images as

$$\widetilde{\mathbf{Y}} \coloneqq \begin{matrix} \overbrace{\mathbf{Y}}^{} \\ \mathbf{0}^{tm\times r} \\ \underbrace{\widehat{\mathbf{y}}^{\mathsf{T}}}_{} \end{matrix} \in \mathcal{R}^{\tilde{n}\cdot r}$$

(vii) Concatenate all inner product images as $\widetilde{\mathbf{t}}^{\mathsf{T}} \coloneqq (\widetilde{t}_0, \widetilde{t}_1) \in \mathcal{R}^2$.

(viii) Define a matrix $\widetilde{\mathbf{F}} \in \mathcal{R}_q^{\tilde{n}\times\tilde{m}}$ such that $\widetilde{\mathbf{F}}\widetilde{\mathbf{W}} = \widetilde{\mathbf{Y}} \bmod q$ represented the following system of equations:

$$\begin{aligned} \mathbf{F}\widehat{\mathbf{W}}_t &= \mathbf{Y} \pmod q, \\ \left(\mathsf{CRT}^{-1}(\mathbf{1}^{\varphi m})\right)^{\mathsf{T}} \cdot \widehat{\mathbf{W}}_0 &= \widehat{\mathbf{y}}^{\mathsf{T}} \pmod q, \\ \widehat{\mathbf{W}}_{i+1} - \sum_{j\in[h_i]} s_{i,j} \cdot \mathbf{W}_{i,j} + p \cdot \mathbf{R}_{i+1} &= \mathbf{0}^{m\times r} \pmod q \quad \forall i \in [t]. \end{aligned}$$

(ix) Write $\{\sigma_k\}_{k\in\mathbb{Z}_{\mathfrak{f}}^\times} = \mathsf{Gal}(\mathcal{K}/\mathbb{Q})$.

(x) Let $\widehat{\mathbf{W}} = \left(\overbrace{\widehat{\mathbf{w}}_{k,i}}^{}_{k\in[n^{\mathsf{blk}}]}\right)_{i\in[r]}$. Let $n^{\mathsf{ip}} \coloneqq 2$ and define the index maps $\iota_{\mathsf{ip}} : [n^{\mathsf{ip}}] \to \{1, -1\}$ and $\iota_{\mathsf{ip\text{-}in}} : [n^{\mathsf{ip}}] \to [n^{\mathsf{blk}}]$ for inner products such that

$$\forall k \in [n^{\mathsf{ip}}],\ t_k = \sum_{i\in[r]} \langle \widetilde{\mathbf{w}}_{\iota_{\mathsf{ip\text{-}in}}(k),i}, \sigma_{\iota_{\mathsf{ip}}(k)}(\widetilde{\mathbf{w}}_{\iota_{\mathsf{ip\text{-}in}}(k),i}) \rangle \iff \widetilde{t}_0 = \sum_{i\in[r]} \langle \widehat{\mathbf{w}}_{0,i}, \widehat{\mathbf{w}}_{0,i} \rangle_{\mathcal{R}} \ \wedge\ \widetilde{t}_1 = \sum_{i\in[r]} \langle \widehat{\mathbf{w}}_{t,i}, \overline{\widehat{\mathbf{w}}_{t,i}} \rangle_{\mathcal{R}}$$

(xi) Let $\widehat{\mathbf{W}} = \overbrace{\widehat{\mathbf{W}}_k}^{}_{k\in[n^{\mathsf{blk}}]}$. Let $n^{\mathsf{aut}} \coloneqq w^2 \log \mathfrak{f}$ and define the index maps $\iota_{\mathsf{aut}} : [n^{\mathsf{aut}}] \to \mathbb{Z}_{\mathfrak{f}}^\times$ and $\iota_{\mathsf{aut\text{-}in}} : [n^{\mathsf{aut}}] \to [n^{\mathsf{blk}}]$, and $\iota_{\mathsf{aut\text{-}out}} : [n^{\mathsf{aut}}] \to [n^{\mathsf{blk}}]$ for automorphisms so that

$$\forall k \in [n^{\mathsf{aut}}],\ \widetilde{\mathbf{W}}_{\iota_{\mathsf{aut\text{-}out}}(k)} = \sigma_{\iota_{\mathsf{aut}}(k)}(\widetilde{\mathbf{W}}_{\iota_{\mathsf{aut\text{-}in}}(k)}) \iff \forall i \in [t],\ j \in [h_i],\ \mathbf{W}_{i,j} \coloneqq \alpha_{i,j}(\widehat{\mathbf{W}}_i).$$

(xii) Set $\widetilde{\mathbf{H}} \coloneqq \begin{pmatrix} \mathbf{H} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{tm+1} \end{pmatrix}$.

[24] Recall that $\mathfrak{f}$ is the conductor of $\mathcal{R}$ and is w-smooth. Technically, $\widetilde{\mathbf{w}}$ has dimension $m \cdot (\sum_{i\in[t]} h_i + 2t + 2)$. Since $h_i \le w^2$ for all i by Theorem 9 and $t \le \log \mathfrak{f}$, we pad $\widetilde{\mathbf{w}}$ to dimension $\tilde{m} = m \cdot (w^2 \log \mathfrak{f} + 2t + 2)$ for simplicity.

(xiii) Output $(\iota_{\mathsf{ip}}, \iota_{\mathsf{ip\text{-}in}}, \iota_{\mathsf{aut}}, \iota_{\mathsf{aut\text{-}in}}, \iota_{\mathsf{aut\text{-}out}}, \widetilde{\mathbf{H}}, \widetilde{\mathbf{F}}, \widetilde{\mathbf{Y}}, \widetilde{\mathbf{t}}, \widetilde{\mathbf{W}})$.

Lemma 12. *Let $m, n^{\mathsf{out}}, r, \in \mathbb{N}$, $0 \leq \beta \leq \beta^{\mathsf{sis}} \leq q$. The protocol $\Pi^{\mathsf{lin\text{-}bin}}_{\mathsf{crt},p}$ is a perfectly correct reduction of knowledge from $\Xi^{\mathsf{lin}} \cap \Xi^{\mathsf{bin}}$ to Ξ^{aut} with parameters*

$$\left(m, n^{\mathsf{out}}, \beta\right) \mapsto \begin{pmatrix} m' & n^{\mathsf{out}\prime}, & \beta', \\ n^{\mathsf{ip}} = 2, & n^{\mathsf{aut}} = w^2 \log \mathfrak{f}, & n^{\mathsf{blk}} = w^2 \log \mathfrak{f} + 2t + 2 \end{pmatrix}.$$

and knowledge sound reduction of knowledge from $\Xi^{\mathsf{lin} \vee \mathsf{sis}} \cap \Xi^{\mathsf{bin} \vee \mathsf{sis}}$ to $\Xi^{\mathsf{aut} \vee \mathsf{sis}}$, with parameters

$$\left(m, n^{\mathsf{out}}, \beta, \beta^{\mathsf{sis}}\right) \leftarrow\!\shortmid \left(m'\ n^{\mathsf{out}\prime}, \beta', \beta^{\mathsf{sis}}\right).$$

where

$$\begin{aligned}
\beta &= \boxed{\sqrt{\hat{\mathfrak{f}}\varphi} \cdot \sqrt{mr}}_{\|\sigma(\cdot)\|_2} \left(\text{resp.}, \boxed{1}_{\|\psi(\cdot)\|_\infty}\right) \\
m' &= m \cdot (w^2 \log \mathfrak{f} + 2t + 2), \\
\beta' &= \boxed{\frac{p}{4} \cdot w^2 \cdot \gamma_{\mathcal{R}} \cdot \sqrt{\hat{\mathfrak{f}}\varphi} \cdot \sqrt{mr \cdot n^{\mathsf{blk}}}}_{\|\sigma(\cdot)\|_2} \left(\text{resp.}, \boxed{\frac{p}{4}\left(2 \cdot \sqrt{\hat{\mathfrak{f}}\varphi} + w^2 \cdot \gamma_{\mathcal{R}}\right)}_{\|\psi(\cdot)\|_\infty}\right), \\
n^{\mathsf{out}\prime} &= n^{\mathsf{out}} + tm + 1, \\
p/2 &> \beta^2 \cdot m \cdot r \cdot \varphi.
\end{aligned}$$

Proof. Completeness. For perfect completeness, we consider the statement $((\mathbf{H}, \mathbf{F}, \mathbf{Y}), \mathbf{W}) \in \Xi^{\mathsf{lin}}_{m,\beta} \cap \Xi^{\mathsf{bin}}_{m,r}$. Let $\mathbf{W} = (\mathbf{w}_i)_{i \in [r]}$. Clearly, we have $\psi(\mathbf{w}_i) \in \{0,1\}^{m\varphi}$. Therefore,

$$\langle \psi(\mathbf{w}_i), \mathbf{1}^{m\varphi} \rangle_{\mathbb{Z}} = \langle \psi(\mathbf{w}_i), \psi(\mathbf{w}_i) \rangle_{\mathbb{Z}} \quad \forall i \in [r]$$

by Proposition 2 (and the discussion immediately after). Since CRT_p is a τ_p-embedding over $\mathcal{R}$ and $\widehat{\mathbf{w}}_t = \mathbf{w}$ due to the construction it holds that

$$\begin{aligned}
\tau_p\left(\sum_{i \in [r]} \hat{y}_i\right) &= \left(\sum_{i \in [r]} \langle \psi(\mathbf{w}_i), \mathbf{1}^{m\varphi} \rangle_{\mathbb{Z}}\right) \bmod p, \text{ and} \\
\tau_p(t_0) &= \left(\sum_{i \in [r]} \langle \psi(\mathbf{w}_i), \psi(\mathbf{w}_i) \rangle_{\mathbb{Z}}\right) \bmod p,
\end{aligned}$$

As a consequence, $\tau_p(t_0) = \tau_p\left(\sum_{i \in [r]} \hat{y}_i\right)$ and the first check of the verifier passes.

Further, as $\|\psi(\mathbf{W})\|_\infty \leq 1$, and by Corollary 1 $\|\sigma(\mathbf{W})\|_2 \leq \sqrt{\hat{\mathfrak{f}}\varphi}\sqrt{mr} = \beta$. We observe that $\|\sigma(\mathbf{W})\|_2^2 = \mathsf{Trace}(d)$ and hence the second verifier's check passes.

Next, we show that

$$\left(\left(\iota_{\mathsf{ip}}, \iota_{\mathsf{ip\text{-}in}}, \iota_{\mathsf{aut}}, \iota_{\mathsf{aut\text{-}in}}, \iota_{\mathsf{aut\text{-}out}}, \widetilde{\mathbf{H}}, \widetilde{\mathbf{F}}, \widetilde{\mathbf{y}}, \widetilde{\mathbf{t}}\right),\ \widetilde{\mathbf{w}}\right) \in \Xi^{\mathsf{aut}}_{m,\mu,\beta',\widetilde{n^{\mathsf{out}}},\widetilde{n},n^{\mathsf{blk}},n^{\mathsf{ip}},n^{\mathsf{aut}}}.$$

The linear part of the relation Ξ^{aut} above holds due to Item (viii) of the $\mathsf{CRTmake}$ subroutine. Similarly, the correctness of all steps of the CRT transformation implies that all automorphism relations hold. For the inner-product type of relation, they are correct due to the honest computation of t_0 and t_1.

Finally, it remains to argue about the correctness of the bound β' for the new witness. Observe that the witness $\widetilde{\mathbf{W}}$ is a concatenation of three types of blocks. Particularly:

- the steps of the CRT transformation, $\widehat{\mathbf{W}}_i$,
- the remainders, $\mathbf{R}_i$,
- and $\mathbf{W}_{i,j}$ used for automorphisms-based relations.

For canonical 2-norm, due to reduction modulo p, $\|\psi\left((\widehat{\mathbf{W}}_i)_{i\in[t+1]}\right)\|_\infty \le p/2$ After translating the norm, $\|\sigma\left((\widehat{\mathbf{W}}_i)_{i\in[t+1]}\right)\|_2 \le p/2\cdot\sqrt{\hat{\mathfrak{f}}\varphi}\cdot\sqrt{mr\cdot(1+t)}$ with the canonical 2-norm of an individual element bounded by $p/2\cdot\sqrt{\hat{\mathfrak{f}}\varphi}$. Thereby, $\|\sigma\left((\mathbf{W}_{i,j})_{(i,j)\in[t,h_j]}\right)\|_2 \le p/2\cdot\sqrt{\hat{\mathfrak{f}}\varphi}\cdot\sqrt{mr\cdot(1+t)\cdot w^2}$ as the automorphisms do not impact the canonical norm of a ring element. However, $\|\psi\left((\mathbf{R}_i)_{i\in[t+1]}\right)\|_\infty \le (\frac{p}{2}\cdot\frac{p}{2}\cdot w^2\cdot\gamma_{\mathcal{R}})/p = \frac{p}{4}\cdot w^2\cdot\gamma_{\mathcal{R}} = \widehat{p}$. After translating norms, $\|\sigma\left((\mathbf{R}_i)_{i\in[t+1]}\right)\|_2 \le \widehat{p}\cdot\sqrt{\hat{\mathfrak{f}}\varphi}\cdot\sqrt{mr\cdot(1+t)}$. To sum up, $\|\sigma\left(\widetilde{\mathbf{W}}\right)\|_2 \le \frac{p}{4}\cdot w^2\cdot\gamma_{\mathcal{R}}\cdot\sqrt{\hat{\mathfrak{f}}\varphi}\cdot\sqrt{mr\cdot n^{\mathsf{blk}}}$, which concludes the proof of the perfect correctness.

For coefficient ∞-norm, due to reduction modulo p, $\|\psi\left((\widehat{\mathbf{W}}_i)_{i\in[t+1]}\right)\|_\infty \le p/2$. After translating the norm the canonical 2-norm of an individual element bounded by $p/2\cdot\sqrt{\hat{\mathfrak{f}}\varphi}$. Thereby, $\|\psi\left((\mathbf{W}_{i,j})_{(i,j)\in[t,h_j]}\right)\|_\infty \le p/2\cdot\sqrt{\hat{\mathfrak{f}}\varphi}$ as the automorphisms do not impact the canonical norm of a ring element. However, $\|\psi\left((\mathbf{R}_i)_{i\in[t+1]}\right)\|_\infty \le (\frac{p}{2}\cdot\frac{p}{2}\cdot w^2\cdot\gamma_{\mathcal{R}})/p = \frac{p}{4}\cdot w^2\cdot\gamma_{\mathcal{R}}$. To sum up, $\|\psi\left(\widetilde{\mathbf{W}}\right)\|_\infty \le \frac{p}{2}\cdot\sqrt{\hat{\mathfrak{f}}\varphi} + \frac{p}{4}\cdot w^2\cdot\gamma_{\mathcal{R}} \le \frac{p}{4}\left(2\cdot\sqrt{\hat{\mathfrak{f}}\varphi} + w^2\cdot\gamma_{\mathcal{R}}\right)$, which concludes the proof of the perfect correctness.

Soundness. For perfect knowledge soundness, suppose that

$$\left(\left(\iota_{\mathsf{ip}},\iota_{\mathsf{ip\text{-}in}},\iota_{\mathsf{aut}},\iota_{\mathsf{aut\text{-}in}},\iota_{\mathsf{aut\text{-}out}},\widetilde{\mathbf{H}},\widetilde{\mathbf{F}},\widetilde{\mathbf{Y}},\widetilde{\mathbf{t}}\right),\ \widetilde{\mathbf{W}}\right) \in \Xi^{\mathsf{aut}}_{m,\mu,\beta',\widetilde{n^{\mathsf{out}}},\widetilde{n},n^{\mathsf{blk}},n^{\mathsf{ip}},n^{\mathsf{aut}}},$$

meaning that $\left(\iota_{\mathsf{ip}},\iota_{\mathsf{ip\text{-}in}},\iota_{\mathsf{aut}},\iota_{\mathsf{aut\text{-}in}},\iota_{\mathsf{aut\text{-}out}},\widetilde{\mathbf{H}},\widetilde{\mathbf{F}},\widetilde{\mathbf{Y}},\widetilde{\mathbf{t}}\right)$ takes the form as constructed in CRTmake. We have

$$\tau_p\left(\sum_{i\in[r]}\hat{y}_i\right) = \left(\sum_{i\in[r]}\langle\psi(\mathbf{w}_i),\mathbf{1}^{m\varphi}\rangle_{\mathbb{Z}}\right) \bmod p, \text{ and}$$
$$\tau_p(t_0) = \left(\sum_{i\in[r]}\langle\psi(\mathbf{w}_i),\psi(\mathbf{w}_i)\rangle_{\mathbb{Z}}\right) \bmod p,$$

where $\mathbf{W} = (\mathbf{w}_i)_{i\in[r]}$. Since $\tau_p\left(\sum_{i\in[r]}\hat{y}_i\right) = \tau_p(t_0)$, we have

$$\sum_{i\in[r]}\langle\psi(\mathbf{w}_i),\psi(\mathbf{w}_i)-\mathbf{1}^{m\varphi}\rangle_{\mathbb{Z}} = 0 \bmod p.$$

Furthermore, $\mathsf{Trace}(t_1) \le \beta^2$ implies that $\|\sigma(\mathbf{W})\|_2 \le \beta$, thus by Corollary 1 $\|\psi(\mathbf{W})\|_\infty \le \beta$. Therefore, as $\beta^2\cdot m\cdot r\cdot\varphi < p/2$, the inner product above holds without modulo p, and thus $\psi(\mathbf{W}) \in \{0,1\}^{m\varphi r}$. Hence, $\|\sigma(\mathbf{W})\|_2 \le \sqrt{\hat{\mathfrak{f}}\varphi}\sqrt{mr}$ and $\|\psi(\mathbf{W})\|_\infty \le 1$. □

Remark 11. Lemma 12 trivially generalises to yield a reduction of knowledge from $\Xi^{\mathsf{ip}} \cap \Xi^{\mathsf{bin}}$ to Ξ^{aut}, analogous to Theorem 8. We omit the details for succinctness.

Remark 12. Note that the reduction of knowledge $\Pi^{\mathsf{lin\text{-}bin}}_{\mathsf{crt},p}$ increases the number of linear relations from n to $n+tm+1$, where we recall that m is the dimension for each block of the witness. Nevertheless, we observe that the matrices $\widetilde{\mathbf{H}}$ and $\widetilde{\mathbf{F}}$ in the statement are sparse and highly structured. Therefore, when applying the chain of reductions in Section 5, the verifier time can still be polylogarithmic in m as long as succinct representations of $\widetilde{\mathbf{H}}$ and $\widetilde{\mathbf{F}}$ are used when running Π^{batch} for batching.

Remark 13. The CRT-based embedding method discussed in this section requires the prime p to split completely, i.e. $p = 1 \bmod \mathfrak{f}$, which immediately implies that $\mathfrak{f}$ must be even. When $\mathfrak{f} = 2^k$, we observe that choices of p are limited, and they tend to be significantly larger than $\mathfrak{f}$ – the smallest values of p for $\mathfrak{f} \in (256, 512, 1024)$ are $p \in (267, 7681, 12289)$. For general $\mathfrak{f}$, choices appear to be more flexible – the smallest values of p for $\mathfrak{f} \in (598, 1102, 2926)$ are $p \in (599, 1103, 2927)$.

Remark 14. We discuss a possible optimisation which allows to pick smaller primes p. Recall that, for knowledge soundness, we required $p/2 > \beta^2 \cdot m \cdot r \cdot \varphi$, which makes p linear in the length of the witness length. Towards picking smaller primes p, we suggest using multiple primes $p_1, p_2, \ldots$ such that all of which fully split over $\mathcal{R}$. To be concrete, consider the case with two primes, i.e. p_1 and p_2. After repeating the protocol twice for p_1 and p_2 respectively (or even better, running a merged version of the protocol), the verifier should be convinced that:

$$\left.\begin{aligned} \langle \mathbf{x}, \mathbf{y} \rangle &= c \bmod p_1 \\ \langle \mathbf{x}, \mathbf{y} \rangle &= c \bmod p_2 \\ (p_1, p_2) &= 1 \end{aligned}\right\} \implies \langle \mathbf{x}, \mathbf{y} \rangle = c \bmod p_1 \cdot p_2.$$

Obviously, this generalises to having a set P of arbitrarily many primes. Consequently, for knowledge soundness, we would only require $\prod_{p \in P} p/2 > \beta^2 \cdot m \cdot r \cdot \varphi$. If the conductor $\mathfrak{f}$ is smooth, it is possible to find many highly favourable primes (cf. Remark 13).

9 Parameters Selection

We propose concrete instantiations of our protocols for various values of m and initial norm. For comparison with prior works, e.g. [BS23, AFLN24], we aim for 128-bit security and knowledge error $\kappa = 2^{-80}$. This corresponds to the root Hermite factor $\delta_{\mathsf{rhf}} \approx 1.0044$ (cf. Section 6.2).

We instantiate protocol as described in Section 6 combining atomic RoKs. We focus on the following simple goal: commit to a short vector $\psi(\mathbf{W}) \in \mathbb{Z}_q^h$, such that $\|\psi(\mathbf{W})\|_\infty \leq \beta$ and prove knowledge of the commitment opening. To this end, we will pack $h = m \cdot \varphi \cdot r$ integers into a matrix $\mathbf{W} \in \mathcal{R}_q^{m \times r}$ of $m \cdot r$ ring elements employing standard coefficient embedding. Then, we will use the vSIS commitment scheme on $\mathbf{W}$.

The relation of our interest is a proof of vSIS commitment opening [CLM23], i.e. the polynomial evaluation equation $\mathbf{w}(v) = y \pmod q$ for public ring elements $v, y \in \mathcal{R}_q$, where $\mathbf{w}$ represents a column of $\mathbf{W}$. When adapting this relation to the language of Ξ^{ip}, we would initially set $(\overline{n}, n^{\mathsf{out}}) = (\overline{n}_0, \overline{n}_0)$, where $\overline{n}_0$ is defined so that (module-)vSIS is hard (we adapt reasoning from Section 6). Throughout the batching protocols, we set $\underline{n} = 1$.

Concrete parameters. In Table 3 we suggest concrete parameters with the estimated proof size. The results are obtained via a dedicated script[25] simulating protocol execution and measuring the communication cost.

The script also includes the details of the composition and fine-grained parameters selection. From the high-level perspective, the most optimal composition for each selection is described as

$$\left((\Pi^{b\text{-}\mathsf{decomp}} \to) \Pi^{\mathsf{norm}} \to \Pi^{\mathsf{batch}} \to \Pi^{\mathsf{split}} \to \Pi^{\mathsf{fold}}\right)_{i \in [\mu]} \to \Pi^{\mathsf{finish}},$$

where $\Pi^{b\text{-}\mathsf{decomp}}$ is performed in (roughly) 2/3 of rounds.

Proof Composition. To further shrink communication, one could use standard proof composition, where instead of the verifier checking the verification conditions, the prover provides proof of knowledge of the input for which the verification holds. To this end, we can directly apply the LaBRADOR proof system [BS23]. Note that this approach is different from running [BS23] for the original relation because now the statement/witness size for LaBRADOR is only of size $\mathsf{poly}(\lambda, \log m)$, and thus we do maintain succinct verification. Hence, we estimate the final communication size to be $\approx$ 100KB based on performance described in [BS23].

Fiat-Shamir Transformation. As noted in [AFK22], transforming interactive proofs, which admit parallel repetition, to the non-interactive setting via Fiat-Shamir transformation can incur a significant loss in the order of Q^μ, where Q is the number of random oracle queries made by an adversary. Following [CMNW24], we designed our protocols with the "bundling approach" for parallel repetition, i.e.,

[25] The script and the output are available at https://github.com/russell-lai/rok-paper-sissors-estimator/blob/v2/example-params.ipynb

witness length in $\mathbb{Z}$-elements	$\approx 2^{30}$	$\approx 2^{30}$	$\approx 2^{32}$
$\log q$	64	64	64
β	1	1023	1023
β^{sis}	35.6	38.8	39.9
$\mathfrak{f}$	60	60	60
m	$\approx 2^{21.4}$	$\approx 2^{21.4}$	$\approx 2^{23.4}$
r	25	25	25
μ	8	9	11
rounds with $\Pi^{b\text{-decomp}}$	$\{2,3,4,5,7\}$	$\{1,2,4,5,7,8\}$	$\{1,2,4,5,7,9,11\}$
$\overline{n}_0$	49	59	62
witness size	128 MB	1280 MB	5120 MB
proof size	5.3 MB	5.7 MB	7.1 MB

Table 3. Concrete parameters, together with proof sizes, for security level $\lambda = 128$ and $\kappa = 2^{-80}$.

we don't treat the parallel repetitions (the columns of $\mathbf{Y}$ and $\mathbf{W}$) separately, but mix them together. In particular, Π^{fold} randomly combines the parallel threads and is $(r_{\mathsf{in}} + 1)$-special sound (Table 1), where $r_{\mathsf{in}} \in \mathcal{O}(d\lambda) = \mathcal{O}(\lambda)$ if $d = \mathcal{O}(1)$. The other protocols either require a single transcript or require two (suitable) transcripts, which behave like 2-special soundness. Hence, we can heuristically assume that the tree-special soundness extractors from [AFK22] are applicable to our protocols. Overall every round is $\mathcal{O}(\lambda)$-special sound, and thus extracting from $\mu \leq \log(m) \in \mathcal{O}(\log(\lambda))$ rounds requires a tree of $\mathcal{O}(\lambda)^{\mathcal{O}(\log(\lambda))}$ transcripts. Hence, heuristically, there is an extractor for the Fiat–Shamir-compiled protocol whose running time is in the order of $Q \cdot \mathcal{O}(\lambda)^{\mathcal{O}(\log(\lambda))}$, where Q denotes the number of random oracle queries.

Acknowledgments

The work of R.L. and M.O. was supported by the Research Council of Finland project No. 358951. This work was supported by the Helsinki Institute for Information Technology (HIIT) and conducted while M.K. was affiliated with Aalto University. N.K.N. was supported by the Protocol Labs RFP-013: Cryptonet network grant.

References

ACL+22. Martin R. Albrecht, Valerio Cini, Russell W. F. Lai, Giulio Malavolta, and Sri Aravinda Krishnan Thyagarajan. Lattice-based SNARKs: Publicly verifiable, preprocessing, and recursively composable - (extended abstract). In Yevgeniy Dodis and Thomas Shrimpton, editors, *CRYPTO 2022, Part II*, volume 13508 of *LNCS*, pages 102–132. Springer, Cham, August 2022. 1, 52

AFK22. Thomas Attema, Serge Fehr, and Michael Klooß. Fiat-shamir transformation of multi-round interactive proofs. In Eike Kiltz and Vinod Vaikuntanathan, editors, *TCC 2022, Part I*, volume 13747 of *LNCS*, pages 113–142. Springer, Cham, November 2022. 49, 50

AFLN24. Martin R. Albrecht, Giacomo Fenzi, Oleksandra Lapiha, and Ngoc Khanh Nguyen. SLAP: Succinct lattice-based polynomial commitments from standard assumptions. In Marc Joye and Gregor Leander, editors, *EUROCRYPT 2024, Part VII*, volume 14657 of *LNCS*, pages 90–119. Springer, Cham, May 2024. 1, 2, 4, 49

AL21. Martin R. Albrecht and Russell W. F. Lai. Subtractive sets over cyclotomic rings - limits of Schnorr-like arguments over lattices. In Tal Malkin and Chris Peikert, editors, *CRYPTO 2021, Part II*, volume 12826 of *LNCS*, pages 519–548, Virtual Event, August 2021. Springer, Cham. 2, 5, 13, 14

BC24. Dan Boneh and Binyi Chen. Latticefold: A lattice-based folding scheme and its applications to succinct proof systems. Cryptology ePrint Archive, Paper 2024/257, 2024. https://eprint.iacr.org/2024/257. 3

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, Berlin, Heidelberg, October / November 2016. 1

BCS23. Jonathan Bootle, Alessandro Chiesa, and Katerina Sotiraki. Lattice-based succinct arguments for NP with polylogarithmic-time verification. In Helena Handschuh and Anna Lysyanskaya, editors, *CRYPTO 2023, Part II*, volume 14082 of *LNCS*, pages 227–251. Springer, Cham, August 2023. 2, 4

BDGL16. Anja Becker, Léo Ducas, Nicolas Gama, and Thijs Laarhoven. New directions in nearest neighbor searching with applications to lattice sieving. In Robert Krauthgamer, editor, *27th SODA*, pages 10–24. ACM-SIAM, January 2016. 30

BFOV04. E. Bayer-Fluckiger, F. Oggier, and E. Viterbo. New algebraic constructions of rotated z/sup n/-lattice constellations for the rayleigh fading channel. *IEEE Transactions on Information Theory*, 50(4):702–714, 2004. 9, 33, 34

BL17. Carsten Baum and Vadim Lyubashevsky. Simple amortized proofs of shortness for linear relations over polynomial rings. Cryptology ePrint Archive, Report 2017/759, 2017. 2

BLNS20. Jonathan Bootle, Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Gregor Seiler. A non-PCP approach to succinct quantum-safe zero-knowledge. In Daniele Micciancio and Thomas Ristenpart, editors, *CRYPTO 2020, Part II*, volume 12171 of *LNCS*, pages 441–469. Springer, Cham, August 2020. 2

BS23. Ward Beullens and Gregor Seiler. LaBRADOR: Compact proofs for R1CS from module-SIS. In Helena Handschuh and Anna Lysyanskaya, editors, *CRYPTO 2023, Part V*, volume 14085 of *LNCS*, pages 518–548. Springer, Cham, August 2023. 1, 2, 9, 38, 49, 54

CLM23. Valerio Cini, Russell W. F. Lai, and Giulio Malavolta. Lattice-based succinct arguments from vanishing polynomials - (extended abstract). In Helena Handschuh and Anna Lysyanskaya, editors, *CRYPTO 2023, Part II*, volume 14082 of *LNCS*, pages 72–105. Springer, Cham, August 2023. 1, 3, 4, 5, 11, 17, 49, 52

CMNW24. Valerio Cini, Giulio Malavolta, Ngoc Khanh Nguyen, and Hoeteck Wee. Polynomial commitments from lattices: Post-quantum security, fast verification and transparent setup. In Leonid Reyzin and Douglas Stebila, editors, *CRYPTO 2024, Part X*, volume 14929 of *LNCS*, pages 207–242. Springer, Cham, August 2024. 1, 4, 49

DFS24. Thomas Debris-Alazard, Pouria Fallahpour, and Damien Stehlé. Quantum oblivious LWE sampling and insecurity of standard model lattice-based SNARKs. In Bojan Mohar, Igor Shinkar, and Ryan O'Donnell, editors, *56th ACM STOC*, pages 423–434. ACM Press, June 2024. 1

DKL+18. Léo Ducas, Eike Kiltz, Tancrède Lepoint, Vadim Lyubashevsky, Peter Schwabe, Gregor Seiler, and Damien Stehlé. Crystals-dilithium: A lattice-based digital signature scheme. *IACR Transactions on Cryptographic Hardware and Embedded Systems*, 2018(1):238–268, Feb. 2018. 2

DPSZ12. Ivan Damgård, Valerio Pastro, Nigel P. Smart, and Sarah Zakarias. Multiparty computation from somewhat homomorphic encryption. In Reihaneh Safavi-Naini and Ran Canetti, editors, *CRYPTO 2012*, volume 7417 of *LNCS*, pages 643–662. Springer, Berlin, Heidelberg, August 2012. 11, 12

EKS+21. Muhammed F. Esgin, Veronika Kuchta, Amin Sakzad, Ron Steinfeld, Zhenfei Zhang, Shifeng Sun, and Shumo Chu. Practical post-quantum few-time verifiable random function with applications to algorand. In Nikita Borisov and Claudia Díaz, editors, *FC 2021, Part II*, volume 12675 of *LNCS*, pages 560–578. Springer, Berlin, Heidelberg, March 2021. 2

FMN23. Giacomo Fenzi, Hossein Moghaddas, and Ngoc Khanh Nguyen. Lattice-based polynomial commitments: Towards asymptotic and concrete efficiency. Cryptology ePrint Archive, Paper 2023/846, 2023. https://eprint.iacr.org/2023/846. 1, 4, 11, 52, 53, 54

GHL22. Craig Gentry, Shai Halevi, and Vadim Lyubashevsky. Practical non-interactive publicly verifiable secret sharing with thousands of parties. In Orr Dunkelman and Stefan Dziembowski, editors, *EUROCRYPT 2022, Part I*, volume 13275 of *LNCS*, pages 458–487. Springer, Cham, May / June 2022. 2, 54

HKR19. Max Hoffmann, Michael Klooß, and Andy Rupp. Efficient zero-knowledge arguments in the discrete log setting, revisited. In Lorenzo Cavallaro, Johannes Kinder, XiaoFeng Wang, and Jonathan Katz, editors, *ACM CCS 2019*, pages 2093–2110. ACM Press, November 2019. 6

KP23. Abhiram Kothapalli and Bryan Parno. Algebraic reductions of knowledge. In Helena Handschuh and Anna Lysyanskaya, editors, *CRYPTO 2023, Part IV*, volume 14084 of *LNCS*, pages 669–701. Springer, Cham, August 2023. 11, 52, 53

Len76. H. W. Lenstra. Euclidean number fields of large degree. *Inventiones mathematicae*, 38(3):237–254, 1976. 2, 5, 13, 14

LM23. Russell W. F. Lai and Giulio Malavolta. Lattice-based timed cryptography. In Helena Handschuh and Anna Lysyanskaya, editors, *CRYPTO 2023, Part V*, volume 14085 of *LNCS*, pages 782–804. Springer, Cham, August 2023. 8, 56

LN17. Vadim Lyubashevsky and Gregory Neven. One-shot verifiable encryption from lattices. In Jean-Sébastien Coron and Jesper Buus Nielsen, editors, *EUROCRYPT 2017, Part I*, volume 10210 of *LNCS*, pages 293–323. Springer, Cham, April / May 2017. 2

LNP22. Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Maxime Plançon. Lattice-based zero-knowledge proofs and applications: Shorter, simpler, and more general. In Yevgeniy Dodis and Thomas Shrimpton, editors, *CRYPTO 2022, Part II*, volume 13508 of *LNCS*, pages 71–101. Springer, Cham, August 2022. 1, 2, 3, 35

LNS20. Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Gregor Seiler. Practical lattice-based zero-knowledge proofs for integer relations. In Jay Ligatti, Xinming Ou, Jonathan Katz, and Giovanni Vigna, editors, *ACM CCS 2020*, pages 1051–1070. ACM Press, November 2020. 9, 38

LNS21. Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Gregor Seiler. Shorter lattice-based zero-knowledge proofs via one-time commitments. In Juan Garay, editor, *PKC 2021, Part I*, volume 12710 of *LNCS*, pages 215–241. Springer, Cham, May 2021. 2

LPR13. Vadim Lyubashevsky, Chris Peikert, and Oded Regev. A toolkit for ring-LWE cryptography. In Thomas Johansson and Phong Q. Nguyen, editors, *EUROCRYPT 2013*, volume 7881 of *LNCS*, pages 35–54. Springer, Berlin, Heidelberg, May 2013. 11, 12

Lyu12. Vadim Lyubashevsky. Lattice signatures without trapdoors. In David Pointcheval and Thomas Johansson, editors, *EUROCRYPT 2012*, volume 7237 of *LNCS*, pages 738–755. Springer, Berlin, Heidelberg, April 2012. 1, 2

MR09. Daniele Micciancio and Oded Regev. *Lattice-based Cryptography*, pages 147–191. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. 4, 30

PP25. Chris Peikert and Zachary Pepin. Vive galois! Part 1: Optimal SIMD packing and packed bootstrapping for FHE. In Benny Applebaum and Huijia (Rachel) Lin, editors, *TCC 2025, Part I*, volume 16268 of *LNCS*, pages 185–219. Springer, Cham, December 2025. 40, 41

PSTY13. Charalampos Papamanthou, Elaine Shi, Roberto Tamassia, and Ke Yi. Streaming authenticated data structures. In Thomas Johansson and Phong Q. Nguyen, editors, *EUROCRYPT 2013*, volume 7881 of *LNCS*, pages 353–370. Springer, Berlin, Heidelberg, May 2013. 3

Was97. Lawrence C. Washington. *Introduction to cyclotomic fields*, volume 83. Springer, 1997. 14

WW23. Hoeteck Wee and David J. Wu. Lattice-based functional commitments: Fast verification and cryptanalysis. In Jian Guo and Ron Steinfeld, editors, *ASIACRYPT 2023, Part V*, volume 14442 of *LNCS*, pages 201–235. Springer, Singapore, December 2023. 1

A Extended Preliminaries

A.1 Vanishing Short Integer Solution Assumption

To prove the soundness of some of the argument systems proposed in this work, we rely on the vanishing short integer solution (vSIS) assumption, proposed in [CLM23] as a structured variant of the standard SIS assumption, and can be seen as a variant of the kRISIS assumption [ACL+22] without hints. To recall, the vSIS assumption is in fact a family of assumptions parametrised by a ring $\mathcal{R}$, a modulus q, a number of points n, a number of variables μ, a norm bound β (for an implicit norm function), and a family $\mathcal{G}$ of μ-variate (possibly Laurent) polynomials over $\mathcal{R}$. It states that, given n randomly sampled evaluation points $\mathbf{x}_i \leftarrow\!\!\$\ (\mathcal{R}_q^\times)^\mu$ for $i \in [n]$, it is infeasible to find a polynomial $g \in \mathcal{G}$ satisfying $g(\mathbf{x}_i) = 0 \bmod q$ for all $i \in [n]$ and whose coefficient vector has norm at most β.

Currently, the best known approach to solve a vSIS problem is to solve it as an unstructured SIS problem, except for extreme cases, e.g. when g is allowed to have a degree close to q. We refer to [CLM23] for more discussion about the conjectured hardness of vSIS.

Referring to the formulation of the vSIS assumption given in Definition 1, setting χ to be the uniform distribution over $\mathcal{R}_q^{n\times m}$ recovers the standard SIS assumption. More interestingly, by instantiating the distribution χ differently, we can recover various variants of the vSIS assumption stated in the style of [CLM23]. For example, setting χ to the uniform distribution of vectors of the form $\bigotimes_{i\in[\mu]}(1, x_i)$ for $x_i \in \mathcal{R}_q^\times$, we recover the single-point multilinear variant. Another example is to set χ to the uniform distribution of vectors of the form $\bigotimes_{i\in[\mu]}(1, x^{2^i})$, which corresponds to the single-point univariate variant.

A.2 Reduction of Knowledge

In this paper we consider ternary relations $\Xi \subseteq \{0,1\}^* \times \{0,1\}^* \times \{0,1\}^*$, where a tuple $(\mathsf{pp}, \mathsf{stmt}, \mathsf{wit}) \in \Xi$ consists of public parameters pp, statement stmt and witness wit. For presentation, we omit including pp when it is known from the context. We consider a modified and simplified definition of a reduction of knowledge [KP23] for the following reasons: All of our protocols are *public coin* and *(coordinate-wise) special sound* [FMN23] or similar.[26] Thus, public reducibility is automatic and we have (super-constant) sequential composition results due to known (tree) black-box extractors, whereas composition in [KP23] is limited a constant number of protocols. Lastly, we define a *relaxed* knowledge soundness notion which is not present in [KP23].

[26] See also Remark 15.

Remark 15. To turn soundness errors of probabilistic tests (such as Schwartz–Zippel) into knowledge errors, we merely need two uniformly random accepting transcripts. These are produced by (CW)SS extractors for example. We call such extractors 2-transcript extractors. We note that the protocols themselves are *not* (CW)SS, as not *any* pair of transcripts (with distinct challenges) suffices for extraction. However, given two transcripts (where the challenges are uniformly distributed conditioned on accepting), we can nonetheless bound the knowledge error. This occurs when extracting a reduction of knowledge whose soundness relies on a Schwartz–Zippel-type argument. All common extractors for (CW)SS satisfy the required distribution of transcripts, hence we can use these extractors as 2-transcript extractors. Importantly, we can "pretend" to deal with (CW)SS in terms of extracting two transcripts. Thus, the tree-special soundness is still applicable and the running time is bounded in the tree size (for state of the art extractors). Our definition of knowledge error is simply additive for sequential compositions of extractors. Hence we can compose as many extractors as we need, and the resulting extractor is efficient if the extracted tree of transcripts is remains polynomial in size.

Definition 6 (Reduction of Knowledge (modified)). *Let Ξ_0, Ξ_1 be ternary relations. A* reduction of knowledge *Π from Ξ_0 to Ξ_1, short $\Pi\colon \Xi_0 \to \Xi_1$, is defined by two PPT algorithms $\Pi = (\mathcal{P}, \mathcal{V})$, the prover $\mathcal{P}$, and the verifier $\mathcal{V}$, with the following interface:*

- $\mathcal{P}(\mathsf{pp}, \mathsf{stmt}_1, \mathsf{wit}_1) \to (\mathsf{stmt}_2, \mathsf{wit}_2)$*: Interactively reduce the input statement* $(\mathsf{pp}, \mathsf{stmt}, \mathsf{wit}) \in \Xi_0$ *to a new statement* $(\mathsf{pp}, \widetilde{\mathsf{stmt}}, \widetilde{\mathsf{wit}}) \in \Xi_1$ *or* $\perp$.
- $\mathcal{V}(\mathsf{pp}, \mathsf{stmt}) \to \widetilde{\mathsf{stmt}}$*: Interactively reduce the task of checking the input statement* $(\mathsf{pp}, \mathsf{stmt})$ *w.r.t* Ξ_0 *to checking a new statement* $(\mathsf{pp}, \widetilde{\mathsf{stmt}})$ *w.r.t.* Ξ_1.

Let $\langle \mathcal{P}, \mathcal{V} \rangle$ denote the interaction between $\mathcal{P}$ and $\mathcal{V}$, as a function that takes as input $(\mathsf{pp}, \mathsf{stmt}, \mathsf{wit})$ and runs the prover $\mathcal{P}$ (resp. verifier $\mathcal{V}$) on input $(\mathsf{pp}, \mathsf{stmt}, \mathsf{wit})$ (resp. $(\mathsf{pp}, \mathsf{stmt})$). At the end of the interaction, $\langle \mathcal{P}, \mathcal{V} \rangle$ outputs the verifier's statement $\widetilde{\mathsf{stmt}}$ and prover's witness $\widetilde{\mathsf{wit}}$. We define following properties.

Definition 7 (Correctness). *Let $\Pi = (\mathcal{P}, \mathcal{V})$ be a reduction of knowledge from Ξ_0 to Ξ_1. We say Π has correctness error $\gamma(\cdot)$, if for all* $(\mathsf{pp}, \mathsf{stmt}, \mathsf{wit}) \in \Xi_0$

$$\mathsf{Pr}[(\mathsf{pp}, \widetilde{\mathsf{stmt}}, \widetilde{\mathsf{wit}}) \in \Xi_1 \mid (\widetilde{\mathsf{stmt}}, \widetilde{\mathsf{wit}}) \leftarrow \langle \mathcal{P}, \mathcal{V} \rangle (\mathsf{pp}, \mathsf{stmt}, \mathsf{wit})] \geq 1 - \gamma(\mathsf{pp}, \mathsf{stmt}).$$

If $\gamma \equiv 0$, we call Π perfectly correct.

Our definitions of knowledge soundness and error are tailored to knowledge extractors for (coordinate-wise) special soundness (cf. Section A.3). Unlike [KP23], we require black-box extraction and ignore efficiency of the adversary.

Definition 8 ((Black-Box) Knowledge Soundness). *Let $\Pi = (\mathcal{P}, \mathcal{V})$ be a reduction of knowledge. We say that Π is relaxed knowledge sound from Ξ_0^{KS} to Ξ_1^{KS} with knowledge error $\kappa(\mathsf{pp}, \mathsf{stmt})$ if there exists a* black-box *expected polynomial-time extractor $\mathcal{E}$ such that: For all* $\mathsf{pp}, \mathsf{stmt}$*, and every (unbounded) malicious prover $\mathcal{P}^*$, we have*

$$\begin{aligned} &\mathsf{Pr}[(\mathsf{pp}, \mathsf{stmt}, \mathsf{wit}) \in \Xi_1^{KS} \mid \mathsf{wit} \leftarrow \mathcal{E}^{\mathcal{P}^*}(\mathsf{pp}, \mathsf{stmt})] \\ \geq\ &\mathsf{Pr}[(\mathsf{pp}, \widetilde{\mathsf{stmt}}, \widetilde{\mathsf{wit}}) \in \Xi_0^{KS} \mid (\widetilde{\mathsf{stmt}}, \widetilde{\mathsf{wit}}) \leftarrow \langle \mathcal{P}^*, \mathcal{V} \rangle (\mathsf{pp}, \mathsf{stmt})] - \kappa(\mathsf{pp}, \mathsf{stmt}). \end{aligned}$$

If Π is both correct and relaxed knowledge sound from Ξ_0^{KS} to Ξ_1^{KS}, then we say that Π is knowledge sound from Ξ_0^{KS} to Ξ_1^{KS}.

We assert no composition theorem. Instead we appeal to the fact that all of our protocols are (coordinate-wise) special sound, so we eventually require are tree-CWSS extractor. These are known to exist, even for the Fiat–Shamir transformed protocol, see e.g. [FMN23]. Formally, we must translate reductions of knowledge to proofs of knowledge to apply special soundness. But this is a triviality:

Definition 9. *Let $\Pi = (\mathcal{P}, \mathcal{V})$ be a reduction of knowledge from Ξ_0 to Ξ_1. We define the* induced proof of knowledge *$\widehat{\Pi} = (\widehat{\mathcal{P}}, \widehat{\mathcal{V}})$ of Π, where the prover $\widehat{\mathcal{P}}$ sends an additional (final) protocol message $\widehat{\mathsf{wit}}$, and the verifier outputs the bit* $(\mathsf{pp}, \widehat{\mathsf{stmt}}, \widehat{\mathsf{wit}}) \in \Xi_1$.

By considering the induced proof of knowledge for $\Pi\colon \Xi_0 \to \Xi_1$, our for $\Pi_{\mathsf{KS}}\colon \Xi_0^{\mathsf{KS}} \to \Xi_1^{\mathsf{KS}}$ in case of relaxed knowledge soundness, we see that the notion of correctness and (relaxed) knowledge soundness is equivalent to the respective notion for reduction of knowledge, with the same knowledge error. Hence, we can indeed apply all our tools for (coordinate-wise special sound) proofs of knowledge.

A.3 Coordinate-Wise Special Soundness

We recall the notion of coordinate-wise special soundness (CWSS) from [FMN23] in a simple form. Let S be a finite set and $\ell \geq 1$. For $i \in [\ell]$ define the following relation $\equiv_i$ for two vectors $\mathbf{x}, \mathbf{y} \in S^\ell$:

$$\mathbf{x} \equiv_i \mathbf{y} \iff x_i \neq y_i \quad \text{and} \quad x_j = y_j \quad \forall j \in [\ell] \backslash \{i\}.$$

This means that $\mathbf{x}$ and $\mathbf{y}$ differ in exactly the i-th coordinate. Next, we consider the following set:

$$\Gamma(S, \ell) \coloneqq \left\{ (\mathbf{x}_0, \ldots, \mathbf{x}_\ell) \subseteq (S^\ell)^{\ell+1} \ : \ \forall i \in [\ell], \mathbf{x}_0 \equiv_i \mathbf{x}_i \, . \right\}$$

In other words, $(\mathbf{x}_0, \ldots, \mathbf{x}_\ell) \in \Gamma(S, \ell)$ if for every coordinate $i \in [\ell]$, there exists exactly one vector $\mathbf{x}_i$ that differs from $\mathbf{x}_0$ in exactly (and only) the i-th coordinate. One can think of $\mathbf{x}_0$ as the "central" vector.

Intuitively, coordinate-wise special soundness for three-round interactive proofs says that given $\ell + 1$ valid transcripts with challenges $\mathbf{x}_0, \ldots, \mathbf{x}_\ell$ which satisfy $(\mathbf{x}_0, \ldots, \mathbf{x}_\ell) \in \Gamma(S, \ell)$, one can efficiently extract the witness. In this paper, we will call such protocols ℓ-CWSS. To argue knowledge soundness from coordinate-wise special soundness in the context of reduction of knowledge, we will use the following lemma from [FMN23].

Lemma 13 (CWSS). *Let $\ell \in \mathbb{N}$, and S be a finite set of cardinality N. Let $\mathcal{C} := S^\ell$ and take a verification function $\mathcal{V} : \mathcal{C} \times \{0,1\}^* \to \{0,1\}$. Then there exists an extractor algorithm $\mathcal{E}$, which given oracle access to a probabilistic algorithm $\mathcal{A}$ such that*

$$\varepsilon := \mathsf{Pr}\left[\mathcal{V}(\mathbf{x}, \mathcal{A}(\mathbf{x})) = 1\right],$$

where the probability is over the choice of $\mathbf{x} \leftarrow \mathcal{C}$ and random coins of $\mathcal{A}$, it makes an expected number of at most $\ell + 1$ queries to $\mathcal{A}$ and with probability at least

$$\varepsilon - \ell / N$$

outputs $\ell + 1$ pairs $(\mathbf{x}_i, y_i)_{0 \leq i \leq \ell}$ such that $V(\mathbf{x}_i, y_i) = 1$ for all $0 \leq i \leq \ell$ and $\{\mathbf{x}_0, \ldots, \mathbf{x}_\ell\} \in \Gamma(S, \ell)$.

B Random projections

B.1 Informal introduction

Random projection serves as another way of performing the "norm upgrade" in the extraction direction. With the "exact" norm check, the update helps to extract a "short" witness (under the vSIS assumption) from the "short-ish" witness so that the extracted statement aligns with the original proved statement.

The problem with this approach is that the initial witness needs to be "short-ish" so that stretch might impact only the numerator, not the denominator. With the random projection, we start from a "short-ish over short" witness so that the upgrade eliminates the growth in both the numerator and denominator.

From an even more high-level perspective, the hope is to avoid the usage of the subtractive set. Then, the challenge set could be large enough (i.e. it could consist of simply short elements) and repetitions avoided, i.e. $r = 1$. Then, the expectation is to have the proof size subquadratic, i.e. $O(\lambda \cdot \mathsf{polylog}(\lambda m))$.

Currently, in the communication cost, the quadratic dependency on λ comes from: (i) the linear dependency on λ of size of single ring element $\mathcal{R}_q$, (ii) the linear dependency on λ of r.

Attempt I Let $\mathcal{R}$ be some power-of-two cyclotomic ring, e.g. $\mathcal{R} = \mathbb{Z}[\zeta_{256}]$ with $\mathfrak{f} = 256$ so that $\varphi = 128$. Assume for the non-module settings, i.e. $\overline{n} = 1$ and the description of $\mathcal{R}_q$ consists of $\Theta\left(\frac{\lambda \cdot \log^2 m}{\log \lambda}\right)$ bits and $q = O(\log(m))$ (cref. Section 6.2).

Let $\mathbf{J} \in \mathcal{R}_q^{k \times u}$ be a matrix sampled from the distribiution of $\{-1, 0, 1\}$ with the probabilities $\{1/4, 1/2, 1/4\}$ denoted as $\chi'^{k \times u}$, where $k = O(\lambda)$ [BS23, GHL22]. To make the matrix foldable we impose an additional structure:

$$\mathbf{J} = \mathbf{J}' \bullet \begin{pmatrix} 1 & \zeta^{i_{0,0}} \\ & \vdots \\ 1 & \zeta^{i_{k-1,0}} \end{pmatrix} \bullet \ldots \bullet \begin{pmatrix} 1 & \zeta^{i_{0,u-1}} \\ & \vdots \\ 1 & \zeta^{i_{k-1,u-1}} \end{pmatrix},$$

where $\mathbf{J} \leftarrow\!\$\ \chi'$, which seems to have enough randomness and each entry is sampled from $\{-1, 0, 1\}$ with the probabilities $\{1/4, 1/2, 1/4\}$ as in χ'. The key observation is that multiplying by ζ^i might be viewed as a sign permutation of coefficients in the power-of-two cyclotomic rings. We denote this structured distribution as $\chi^{k\times m}$.

The complete RoK Π^{JL1} contains of

(i) claiming a coefficient 2-norm norm μ of the witness $\mathbf{w}$ by the prover,
(ii) sending a challenge $\mathbf{J}$ by the verifier,
(iii) sending a response $\mathbf{s} = \mathbf{J}\mathbf{w}$ by the verifier,
(iv) verifying if $\sqrt{30}\mu \leq \|\mathsf{ct}(\mathbf{s})\|_2 \leq \sqrt{337}\mu$,

where $\mathsf{ct} : \mathcal{R}_q^t \to \mathbb{Z}_q^t$ maps the ring elements to their constant terms.

The correctness holds as for $b, c \in \mathcal{R}_q$ the constant term of $a = b \cdot c$ is the inner product of $\langle \psi(b), \overline{\psi(c)} \rangle_{\mathbb{Z}}$ and $\chi = \bar{\chi}$ exploiting that $\mathcal{R}$ is a power-of-two cyclotomic ring.

Although the random projections serve as an extraction upgrade, $\mathbf{s}$ has a description of $\Theta\left(\frac{\lambda^2 \cdot \log^2 m}{\log \lambda}\right)$ bits, which is not satisfactory with the objective subquadratic proof size.

Attempt II The crucial observation is that $\mathbf{s}$ is not "interesting" to the verifier as a complete vector and sending $\mathsf{ct}(\mathbf{s})$ is sufficient for the sake of norm verification. Unfortunately, splitting and folding are no longer possible. Luckily, we still can batch the new rows with the challenge $\mathbf{c}^\intercal = (c, c^2, \dots c^k)$, where $c \in \mathbb{Z}_q$. After batching, the complete element $s' = \mathbf{s}^\intercal \cdot \mathbf{c}$ is sent over and the protocol continues.

In more detail, the complete RoK Π^{JL2} contains of

(i) claiming a coefficient 2-norm norm μ of the witness $\mathbf{w}$ by the prover,
(ii) sending a challenge $\mathbf{J}$ by the verifier,
(iii) sending a response $\mathsf{ct}(\mathbf{s})$, where $\mathbf{s} = \mathbf{J}\mathbf{w}$ by the verifier,
(iv) verifying if $\sqrt{30}\mu \leq \|\mathsf{ct}(\mathbf{s})\|_2 \leq \sqrt{337}\mu$,
(v) sending a challenge $\mathbf{c}^\intercal = (c, c^2, \dots c^k)$, where $c \leftarrow\!\$\ \mathbb{Z}_q$,
(vi) batching $s' = \mathbf{s}^\intercal \cdot \mathbf{c}$ and sending s' by the verifier,
(vii) verifying if $\mathsf{ct}(s') = \mathsf{ct}(\mathbf{s})^\intercal \cdot \mathbf{c}$.

The verifier sends two messages. First, $\mathsf{ct}(\mathbf{s})$ has a description of $\Theta(\lambda \cdot \log m)$ bits. Second, s' has a description of $\Theta\left(\frac{\lambda \cdot \log^2 m}{\log \lambda}\right)$ bits. The foldable structure is maintained and correctness holds immediately.

For the knowledge-soundness under the Schwartz-Zippel lemma, we require $\mathbb{Z}_q$ to have the largest subfield of cardinality at least $\approx 2^\kappa$. Concretely, it may require setting a large prime modulus q. To workaround this limitation, it is enough to batch not into a single row, but e.g. into $g = O(1)$ (or even $g = \mathsf{polylog}(\lambda)$) rows with a structured $\mathbf{C} \in \mathcal{R}_q^{k\times g}$ challenge so that $2^\kappa \geq q^g$.

Perspectrives

(i) Can we use an arbitrary cyclotomic ring? I expect that it is possible as we have the inner product trick generalised via trace and $\mathbf{J}$ might be sampled from $\mathcal{R}_q^\vee$ ($= \mathcal{R}_q$ assuming q is not very smooth). The tensor structure seems not to work then as multiplying by ζ^i is no longer a sign permutation.
(ii) Can we prove that this structured projection is still equally secure?

C 2nd attempt to Random projections

C.1 Informal introduction

It seems that imposing the tensor structure on random projection does not bring us many benefits due to the impossibility results. In this attempt, we explore something easier. We assume the ring variant of the JL lemma.

Lemma 14. *Let $\mathcal{C}$ be a distribution on $\{-1, 0, 1\}$ with $\mathsf{Pr}[\mathcal{C} = 0] = \frac{1}{2}$, and $\mathsf{Pr}[\mathcal{C} = 1] = \mathsf{Pr}[\mathcal{C} = -1] = \frac{1}{4}$. Then, for every vector $\mathbf{w} \in \mathbb{Z}^d$, we have*

$$\mathsf{Pr}\left[\frac{\|\Pi\mathbf{w}\|_2}{\|\mathbf{w}\|_2} \notin [\mu, \gamma] \mid \Pi \leftarrow\!\$\, \mathcal{C}^{256\times d}\right] \leq 2^{128}$$

for some $\mu, \gamma = O(1)$. Further, let $\mathbf{v} = \psi^{-1}(\mathbf{w}) \in \mathcal{R}^{d/\delta}$, where $\mathcal{R}$ is some cyclotomic ring.

$$\mathsf{Pr}\left[\frac{\|\Pi^{\mathsf{T}}\mathbf{v}\|_2}{\|\mathbf{v}\|_2} \notin [\mu', \gamma'] \mid \Pi \leftarrow\!\$\, \psi^{-1}(\mathcal{C}^{t\times d}) \in \mathcal{R}^{t\times d/\delta}\right] \leq 2^{128}$$

for some $\mu', \gamma', t = O(1)$.

We present the RoK. We assume the correctness VDF from [LM23] adjusted to our folding structure utilising the inner product norm upgrade (S&P paper).

(i) The initial relation is $\mathbf{f}^{\mathsf{T}}\mathbf{w} = c_{\mathbf{w}} \bmod q$.
(ii) Verifier sends two challanges $y_0, y_0' \leftarrow\!\$\, \mathcal{R}_q$.
(iii) Prover computes

$$\begin{pmatrix} \mathbf{f}^{\mathsf{T}} \\ \begin{bmatrix}\mathbf{G}\\ \mathbf{A}\end{bmatrix}_{\searrow \bar{m}} \end{pmatrix} \cdot \begin{pmatrix}\mathbf{w} & \mathbf{x} & \mathbf{x}'\end{pmatrix} = \begin{pmatrix} c_{\mathbf{w}} & c_{\mathbf{x}} & c_{\mathbf{x}'} \\ - & y_0 & y_0' \\ & \vdots & \vdots \\ - & y_{\bar{m}} & y_{\bar{m}}' \end{pmatrix}$$

and sends $y_{\bar{m}}, y_{\bar{m}}'$.
(iv) Verifier sends structured challange $\mathbf{c}$.
(v) Prover computes

$$\begin{pmatrix} \mathbf{f}^{\mathsf{T}} \\ \underline{\mathbf{f}}^{\mathsf{T}} \end{pmatrix} \cdot \begin{pmatrix}\mathbf{w} & \mathbf{x} & \mathbf{x}'\end{pmatrix} = \begin{pmatrix} c_{\mathbf{w}} & c_{\mathbf{x}} & c_{\mathbf{x}'} \\ \pi_{\mathbf{w}} & \pi_{\mathbf{x}} & \pi_{\mathbf{x}'} \end{pmatrix},$$

where $\underline{\mathbf{f}}^{\mathsf{T}} = \mathbf{c}^{\mathsf{T}} \left(\begin{bmatrix}\mathbf{G}\\ \mathbf{A}\end{bmatrix}_{\searrow \bar{m}}\right)$ and sends $\pi_{\mathbf{w}}$. Verifier computes $\pi_{\mathbf{x}}$, $\pi_{\mathbf{x}'}$ herself.
(vi) Prover uses the regular tricks to prove the following:
- Coefficients of $\mathbf{x}$ and $\mathbf{x}'$ are binary
- $\langle \mathbf{x}, \mathbf{w}\rangle = t$ and $\langle \mathbf{x}, \mathbf{w}'\rangle = t'$.

As the distributions of $\mathbf{x}$ and $\mathbf{x}'$ are close to uniform due to the decision variant of SIS, the distribution of $\psi(\mathbf{x} - \mathbf{x}')$ is computationally close to the one from Lemma 14 and the norm of $t - t'$ is close to the norm of the $\mathbf{w}$.