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, AFLN23, 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, DAFS24], 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, AFLN23] , 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, and polylogarithmic-time verifier, 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 $\Omega(\log^2 \lambda)$ smaller 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	proof size
[BCS23]	M-SIS	$O\left(\log^6 m \cdot \lambda^2/\log \lambda\right)$
[CLM23]	vSIS	$O\left(\log^5 m \cdot \lambda^2/\log^2 \lambda\right)$
[FMN23]	PowerBASIS	$O\left(\log^5 m \cdot \lambda^2/\log^2 \lambda\right)$
[AFLN23]	M-SIS	$O\left(\log^5 m \cdot \lambda^2/\log^2 \lambda\right)$
[CMNW24]	SIS	$O(\log^3 m \cdot \lambda^2)$
This work	vSIS	$O\left(\log^3 m \cdot \lambda^2/\log^2 \lambda\right)$

Fig. 1. Asymptotic efficiency of our commitment opening proof (in bits) and comparison with prior interactive proofs 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] as well as our scheme, 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 highlight that for the sizes reported from [AFLN23, CMNW24], the knowledge extractor runs in expected super-polynomial time in m and λ.

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" defined as: $\tau(z) = \frac{1}{N}\mathsf{Trace}(\alpha \cdot z)$.

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 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).

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

[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.

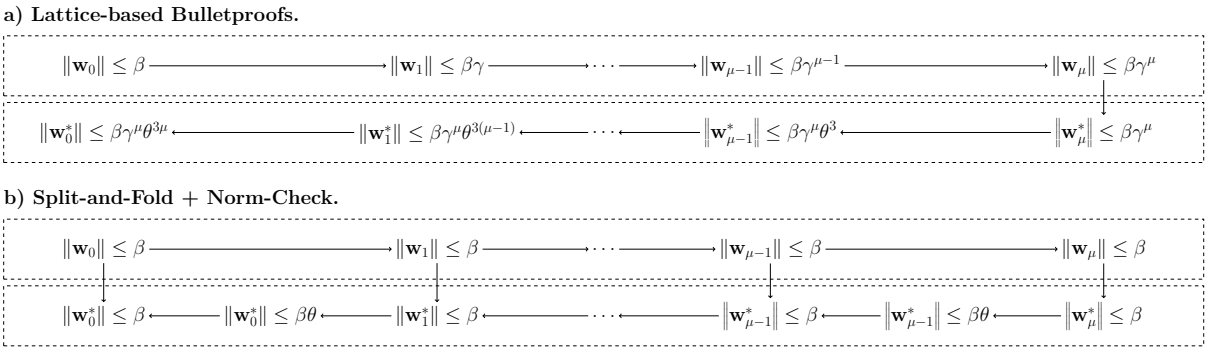

Fig. 2. Overview of the evolution of a prover witness $\mathbf{w}_0$ to an extracted witness $\mathbf{w}_0^*$ in lattice-based Bulletproofs and in Split-and-Fold + Norm-Check.

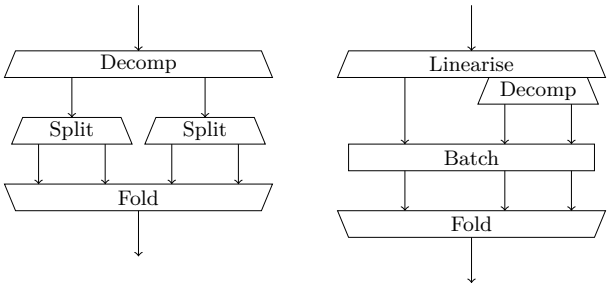

Fig. 3. Structures of Split-and-Fold (left) and Norm-Check (right) protocols.

$\beta^* \ggg \beta'$. 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. To eliminate correctness and soundness gaps, we propose two modular protocols called Split-and-Fold and Norm-Check, each of which is a composition of atomic elementary building blocks – (b-ary) Decomp, Split, Fold, Linearise, and Batch. The structures of Split-and-Fold and Norm-Check, designed to eliminate the correctness and soundness gaps resepectively, are depicted in Fig. 3. These protocols are designed to run in an interleaved manner to restrict the norm growth of the witness and to aid witness extraction. The lattice-based Bulletproofs protocol can be seen

[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].

as a repeated execution of the barebone Split protocol and Fold protocol without any norm-restricting mechanisms.

The Split-and-Fold and Norm-Check can be summarised as follows:

Split-and-Fold. The purpose of the Split-and-Fold protocol is to shrink the dimension of the relation to be proven without increasing the norm of the witness. On input a Ξ^{lin} instance $(\mathbf{F}, \mathbf{y})$ with witness $\mathbf{w}$ of norm at most β, run the Decomp protocol to decompose the witness into b-ary parts. This splits $(\mathbf{F}, \mathbf{y})$ into ℓ sub-instances $(\mathbf{F}, \mathbf{y}_i)$ each with witness $\mathbf{w}_i$. If b is small, each sub-witness $\mathbf{w}_i$ norms at most $b/2 \ll \beta$. Then, for each sub-instance $(\mathbf{F}, \mathbf{y}_i)$ with witness $\mathbf{w}_i$, run the Split protocol to peel off one tensor factor of $\mathbf{F}$, i.e. factor $\mathbf{F}$ into $\mathbf{F} = \mathbf{R} \bullet \widetilde{\mathbf{F}}$ and decompose $\mathbf{w}_i$ into $(\mathbf{w}_{i,j})_{j \in [d]}$. Each sub-instance is thus further split into finer sub-instances $(\widetilde{\mathbf{F}}, \mathbf{y}_{i,j})$ for some appropriate $\mathbf{y}_{i,j}$ with witnesses $\mathbf{w}_{i,j}$ still of norm at most $b/2$. Finally, the Fold protocol is run to merge all sub-instances into a single instance $(\widetilde{\mathbf{F}}, \widetilde{\mathbf{y}})$ with witness $\widetilde{\mathbf{w}}$. For appropriately chosen b, we should end up with a next-round witness $\widetilde{\mathbf{w}}$ of norm at most β again.

Note that the above suffices to eliminate the correctness gap, i.e. if the Split-and-Fold protocol were to be run recursively, the intermediate and hence final witnesses of the prover will remain to have norm at most β. However, given an extracted next-round witness $\widetilde{\mathbf{w}}$ of norm at most β, the knowledge extractor could only extract a candidate witness of norm at most $\approx \theta\beta$. To improve the norm bound to β, we need to interleave a Norm-Check protocol, as described below, between executions of Split-and-Fold.

Norm-Check. The purpose of the Norm-Check protocol is to upgrade a relaxed norm bound guarantee to a tight norm bound guarantee. On input a Ξ^{lin} instance $(\mathbf{F}, \mathbf{y})$ with witness $\mathbf{w}$ of norm at most β, the prover sends the value $t = \langle \mathbf{w}, \overline{\mathbf{w}} \rangle_{\mathcal{R}}$ to the verifier, who can check that $\mathsf{Trace}(t) \leq \beta^2$. If t is computed correctly, then $\mathsf{Trace}(t)$ is precisely the square of the canonical ℓ_2-norm of $\mathbf{w}$, which the verifier checks to be at most β^2. It thus remains for the prover to prove that t is computed correctly, along with all the other relations. To do so, the prover encodes $\mathbf{w}$ as the coefficients of a polynomial $g(X)$, and commit to the coefficients of the Laurent polynomial $L(X) = g(X) \cdot \bar{g}(X^{-1})$. 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, the prover runs the Decomp protocol (in a non-black-box manner) to shrink the coefficients of $L(X)$ back to norm β, at the cost of spawning new sub-instances. Second, the extra checks for L being computed correctly and $L(0) = t$ introduce more constraints which would translate to higher communication cost when handled naively. To tackle this, the parties run the Batch protocol to compress the newly added constraints with the existing ones. Finally, we use again the Fold protocol to merge all sub-instances into one.

In Fig. 2 part b) for "Split-and-Fold + Norm-Check", each horizontal arrow in the top row represents one "split-and-fold" step which replaces a Bulletproofs folding step. The effect of this is that the norm of $\mathbf{w}_i$ remains at most β for all i, i.e. the correctness gap is eliminated. Each vertical arrow from $\mathbf{w}_i$ to $\mathbf{w}_i^*$ for $i < \mu$ represents one "Norm-Check" which is used to prove that the norm of the current witness $\mathbf{w}_i$ is at most β. For this step, it is important that the norm function is chosen to be the ℓ_2-norm over the canonical embedding, so that it can be expressed in terms of the (complex) inner product which is natively supported by our protocol. This proof upgrades the bound $\|\mathbf{w}_i^*\| \leq \theta\beta$ guaranteed by a traditional witness extraction step to a tighter bound $\|\mathbf{w}_i^*\| \leq \beta$, i.e. the soundness gap is eliminated.

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 $\{0, 1\}$.

(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 6, 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.

[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 $\psi(x) = (x_i)_{i \in [\varphi]}$.

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

Motivated by the above, we present a succinct reduction of knowledge from checking the automorphism relation $\alpha(\mathbf{w}) = \mathbf{w}'$ to checking that $(\mathbf{w}, \mathbf{w}')$ satisfies some linear relations. Combined with the Split-and-Fold and Norm-Check protocols designed in Section 5, 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. In Section A.1, we recall some basics of algebraic number theory.

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.2.

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* $\mathcal{R}_q^{n \times \bigotimes_{i \in [\mu]} d_i}$ *for some dimensions* $n, d_0, \dots, d_{\mu-1}, \mu \in \mathbb{N}$*. The* $\mathsf{vSIS}_\mathsf{params}$ *assumption states that, for any PPT adversary* $\mathcal{A}$*, the advantage function satisfies*

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

For simplicity, in this work, we will consider the setting where the block sizes $d_0, \dots, d_{\mu-1}$ are identically set to some $d \in \mathbb{N}$, so that $\mathbf{F}$ can be factored into $\mathbf{F} = \mathbf{F}_{\mu-1} \bullet \dots \bullet \mathbf{F}_0$ 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.3 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 $(\mathsf{pp}, \mathsf{stmt})$).

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 $\forall\, c, x \in \mathcal{R},\ \|\sigma(c \cdot x)\|_2 \le \|\sigma(c)\|_\infty \cdot \|\sigma(x)\|_2$. We measure the quality of a subtractive set by its cardinality, expansion factor γ_S, and inverse-expansion factor θ_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*

$$\gamma_S := \max_{c \in S} \|\sigma(c)\|_\infty \qquad \text{and} \qquad \theta_S := \max_{c, c' \in S, c \neq c'} \left\| \sigma\left(\frac{1}{c - c'}\right) \right\|_\infty$$

respectively.

The following lemma often is handy for analysing inverse-expansion factors.

Lemma 1. *Let $K = \mathbb{Q}(\zeta)$ with ζ a primitive $\mathfrak{f}$-th root of unity such that $\mathfrak{f} > 4$. It holds that $\left\|\sigma\left(\frac{1}{1-\zeta}\right)\right\|_\infty \le \frac{\mathfrak{f}}{4\sqrt{2}}$. Furthermore, if ζ^i is a k-th root of unity, then $\left\|\sigma\left(\frac{1}{1-\zeta^i}\right)\right\|_\infty \le \frac{k}{4\sqrt{2}}$ and $\left\|\sigma\left(\frac{1}{1+\zeta^i}\right)\right\|_\infty \le \frac{k}{2\sqrt{2}}$.*

[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.

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 $\alpha \le \frac{1}{4}$ we have $\left|1-e^{\alpha 2\pi i}\right| = 2\cdot\sin(\frac{2\pi\alpha}{2}) \ge \alpha\cdot 4\sqrt{2}$. 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 $\left\|\sigma\left(\frac{1}{1-\zeta^i}\right)\right\|_\infty \le \frac{k}{4\sqrt{2}}$ follows for ζ^i being a k-th root of unity.

Next, we show that $\left\|\sigma\left(\frac{1}{1+\zeta^i}\right)\right\|_\infty \le \frac{k}{2\sqrt{2}}$ for ζ^i being a k-th root of unity. Observe that either $-\zeta^i$ is already a k-th root unity, then the proof concludes by applying the above bound for $\left\|\sigma\left(\frac{1}{1-(-\zeta^i)}\right)\right\|_\infty$. Otherwise, $-\zeta^i$ is a $2k$-th root unity. Thus $\left\|\sigma\left(\frac{1}{1+\zeta^i}\right)\right\|_\infty = \left\|\sigma\left(\frac{1}{1-(-\zeta^i)}\right)\right\|_\infty \le \frac{2k}{4\sqrt{2}} = \frac{k}{2\sqrt{2}}$. □

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 $S \coloneqq \{\mu_0, \ldots, \mu_{p-1}\} \subseteq_p \mathcal{O}_\mathcal{K}$ is subtractive, where $\mu_i = (\zeta^i - 1)/(\zeta - 1)$. Further, $\gamma_S \le p$ and $\theta_S \le \frac{\mathfrak{f}}{2\sqrt{2}}$.*

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.

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

$$\left\|\frac{1}{\mu_j - \mu_i}\right\| = \left\|\zeta^{-i}\cdot\frac{\zeta-1}{\zeta^{j-i}-1}\right\| \le \|\zeta-1\|\cdot\left\|\frac{1}{\zeta^{j-i}-1}\right\| \le 2\left\|\frac{1}{\zeta^{j-i}-1}\right\| \le \frac{\mathfrak{f}}{2\sqrt{2}},$$

where the last inequality follows from Lemma 1 and the rest are elementary. □

4.2 Non-Prime-Power Cyclotomics

A drawback of the subtractive set recalled above is its rather large expansion factor $\gamma_S \le p$. In some applications, e.g. Section 5, we would like γ_S to be constant. Below, we expose a subtractive set over non-prime-power cyclotomic rings with expansion factor $\gamma_S = 1$.

Theorem 2. *Let $\mathfrak{f}$ factor into $k \ge 2$ 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 $S \coloneqq \left\{1, \zeta, \zeta^2, \ldots, \zeta^{\mathfrak{f}/\hat{\mathfrak{f}}_{\max}-1}\right\} \subseteq_{\mathfrak{f}/\hat{\mathfrak{f}}_{\max}} \mathcal{O}_\mathcal{K}$, is subtractive. Furthermore, $\gamma_S = 1$ and $\theta_S \le \frac{\mathfrak{f}}{4\sqrt{2}}$.*

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 2. *Let $\mathcal{R} = \mathbb{Z}[\zeta_\mathfrak{f}]$ with a conductor $\mathfrak{f}$ having $k \ge 2$ coprime prime-power factors[12] $(\hat{\mathfrak{f}}_i)_{i\in[k]}$, i.e. $\mathfrak{f} = \prod_{i\in[k]}\hat{\mathfrak{f}}_i$. Write $\hat{\mathfrak{f}}_{\max} \coloneqq \max_{i\in[k]}\hat{\mathfrak{f}}_i$. For $j \in \left\{1, 2, \ldots, \frac{\mathfrak{f}}{\hat{\mathfrak{f}}_{\max}} - 1\right\}$, it holds that $1 - \zeta^j \in \mathcal{R}^\times$.*

[12] 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.

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 $j \in \left\{1, 2, \ldots, \frac{\mathfrak{f}}{\mathfrak{f}_{\mathsf{max}}} - 1\right\}$,

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

which implies $\mathfrak{f}_{\mathsf{max}} < \hat{\mathfrak{f}}_i^n$, 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.*

Finally, we state our proof of Theorem 2.

Proof. (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 2. The expansion factor satisfying $\gamma_S = 1$ is immediate. For the inverse-expansion factor, we have

$$\theta_S = \max_{i \neq j} \left\| \frac{1}{\zeta^i - \zeta_j} \right\|_\infty = \max_{i \neq j} \left\| \frac{1}{1 - \zeta_{i-j}} \right\|_\infty \leq \frac{\mathfrak{f}}{4\sqrt{2}}.$$

where the inequality is due to Lemma 1. □

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, $\gamma_S \leq p$ and $\theta_S \leq \frac{\mathfrak{f}^2}{8}$

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 $\gamma_S \leq p$. For θ_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 1,

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

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, $\gamma_S \leq 2$ and $\theta_S \leq \frac{\mathfrak{f}^2}{32}$.

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 $\gamma_S \leq 2$ is immediate. Write $\|\cdot\| = \|\sigma(\cdot)\|_\infty$. For θ_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 Succinct Arguments 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 mainly use the canonical 2-norm, and simply write $\|\cdot\| \coloneqq \|\sigma(\cdot)\|_2$, unless specified. We use the shorthand notation $\mathcal{R}_q^{n\times d^{\otimes\mu}} \coloneqq ((\mathcal{R}_q^{1\times d})^{\otimes\mu})^n$ 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. $\mathcal{R}_q^\times$).

We construct succinct arguments for proving that a short vector $\mathbf{w}$ satisfy:

- $\mathcal{R}_q$-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 $\|\mathbf{w}\| \leq \beta$.

In Table 1 on page 24, we provide an overview of parameters for correctness and relaxed knowledge soundness.

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

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

where we *always assume* that $\mathbf{H}$ has the block structure[13]

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

[13] 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.

Similarly, 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 $\mathbf{y}$.

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.

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 $\mathbf{F}\mathbf{w} = \mathbf{y}$ 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] (Definition 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 $\mathbf{H}\mathbf{F}$ and $\mathbf{y}$ as $\mathbf{H}\mathbf{y}$. 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 tensor 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) $\Pi^{b\text{-decomp}}$: Reduce one instance with bound β to many with bound $b \ll \beta$.
(ii) Π^{split}: Reduce one instance with witness dimension m to many with $m' = m/d$.
(iii) Π^{fold}: Reduce many instances with bounds β_i to one with $\beta' = \gamma \sum_i \beta_i$.
(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 $\Pi^{b\text{-decomp}}$: b-ary Decomposition Knowledge Reduction

Let $b \geq 1$ be an integer. The protocol $\Pi^{b\text{-decomp}}$ (Fig. 4) 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 $\mathbf{v}_i \in \mathcal{R}_b$ (hence $\|\mathbf{v}_i\|_\infty \leq b/2$) and $\ell = \lceil \log_b(2\beta+1) \rceil$. Then, appropriate claims $\mathbf{z}_i = \mathbf{H}\mathbf{F}\mathbf{v}_i$ 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 $((\mathbf{H}, \mathbf{F}, \mathbf{z}_i), \mathbf{v}_i)_{i\in[\ell]}$.

Remark 2. In protocol $\Pi^{b\text{-decomp}}$, 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 $\mathbf{y} = \sum_{i\in[\ell]} b^i \mathbf{z}_i$.

Lemma 3. *The protocol $\Pi^{b\text{-decomp}}$ is a reduction from $\Xi^{\mathsf{lin}}_{\beta}$ to $(\Xi^{\mathsf{lin}}_{\frac{1}{2}\sqrt{m}\varphi^{3/2}b})^\ell$ for $\ell = \lceil \log_b(2\beta+1) \rceil$. It is perfectly correct. It is perfectly relaxed knowledge sound from $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{\frac{b^\ell-1}{b-1}\cdot\beta'}$ to $(\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{\beta'})^\ell$ if $\frac{b^\ell-1}{b-1} \cdot \beta' \leq \beta^{\mathsf{sis}}$.*

Note that $\frac{b^\ell-1}{b-1} \cdot \beta' \leq 2b^{\ell-1}\beta' \leq 2\beta\beta'$.

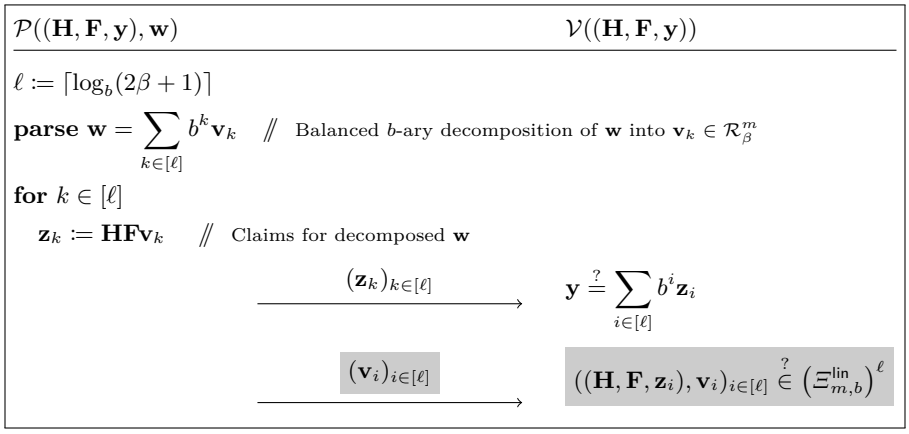

Fig. 4. Protocol $\Pi^{b\text{-decomp}}$, a reduction of $\Xi^{\mathsf{lin}}_{m,\beta}$ to $(\Xi^{\mathsf{lin}}_{m,b})^\ell$ for $\ell = \lceil \log_b(2\beta+1) \rceil$. As a *proof* of knowledge, Send the marked parts; as a *reduction* not. $\Pi^{b\text{-decomp}}$ sends the marked parts only as a *proof* (but not *reduction*) of knowledge.

Proof. Perfect correctness of $\Pi^{b\text{-decomp}}$ from $\Xi^{\mathsf{lin}}_{\beta}$ to $(\Xi^{\mathsf{lin}}_{\frac{1}{2}\sqrt{m}\varphi^{3/2}b})^\ell$ is easy to see: By construction, each $\mathbf{v}_i$ has $\|\psi(\mathbf{v}_i)\|_\infty \leq b/2$, and by Lemma 9 it follows that $\|\mathbf{v}_i\| \leq \frac{1}{2}\sqrt{m}\varphi^{3/2}b$ as claimed. The linear equations $\mathbf{H}\mathbf{F}\mathbf{v}_i = \mathbf{z}_i$ hold by definition for $i \neq 0$ and by linearity for $i = 0$.

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

$$\|\mathbf{w}\| \leq \sum_{i=0}^{\ell-1} b^i \|\mathbf{v}_i\| \leq \frac{b^\ell - 1}{b-1} \cdot \beta'.$$

The OR-branch in $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$ is handled by letting the $\mathbf{v}_i$ with $\overline{\mathbf{H}}\mathbf{F}\mathbf{v}_i = \mathbf{0}$ be the extracted witness. If $\frac{b^\ell-1}{b-1} \cdot \beta' \leq \beta^{\mathsf{sis}}$, this is a witness for $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$.

For the note, we observe that by geometric series $\frac{b^\ell-1}{b-1} \leq \frac{b^\ell}{b\cdot(1-1/b)} = \frac{b}{b-1}b^{\ell-1} \leq 2b^{\ell-1}$. Moreover, by definition of ℓ, we have $b^{\ell-1} < 2\beta+1 \leq b^\ell$. This yields the bounds. □

5.3 Π^{split}: Witness Splitting Knowledge Reduction

In Fig. 5 we describe protocol Π^{split} which takes a claim from $\Xi^{\mathsf{lin}}_{m,\beta}$ and splits it into d claims in $\Xi^{\mathsf{lin}}_{m/d,\beta}$. That is, the number of claims in $\Xi^{\mathsf{lin}}_{m,\beta}$ (i.e. 1) grows by d-fold to d, but the witness dimension of each claim shrinks by d-fold to m/d. We explain the idea and correctness of the protocol below.

To split the witness, interpret $\mathcal{R}^m$ as $\mathcal{R}^{d^{\otimes\mu}}$, and split $\mathbf{w} \in \mathcal{R}^m \cong \mathcal{R}^{d^{\otimes\mu}}$ into $\mathbf{w} = (\mathbf{w}_i)_{i\in[\mu]} = \sum_{i=0}^{\mu-1} \mathbf{e}_i \otimes \mathbf{w}_i$ where $\mathbf{w}_i \in \mathcal{R}^{m/d} \cong \mathcal{R}^{d^{\otimes(\mu-1)}}$ 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

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

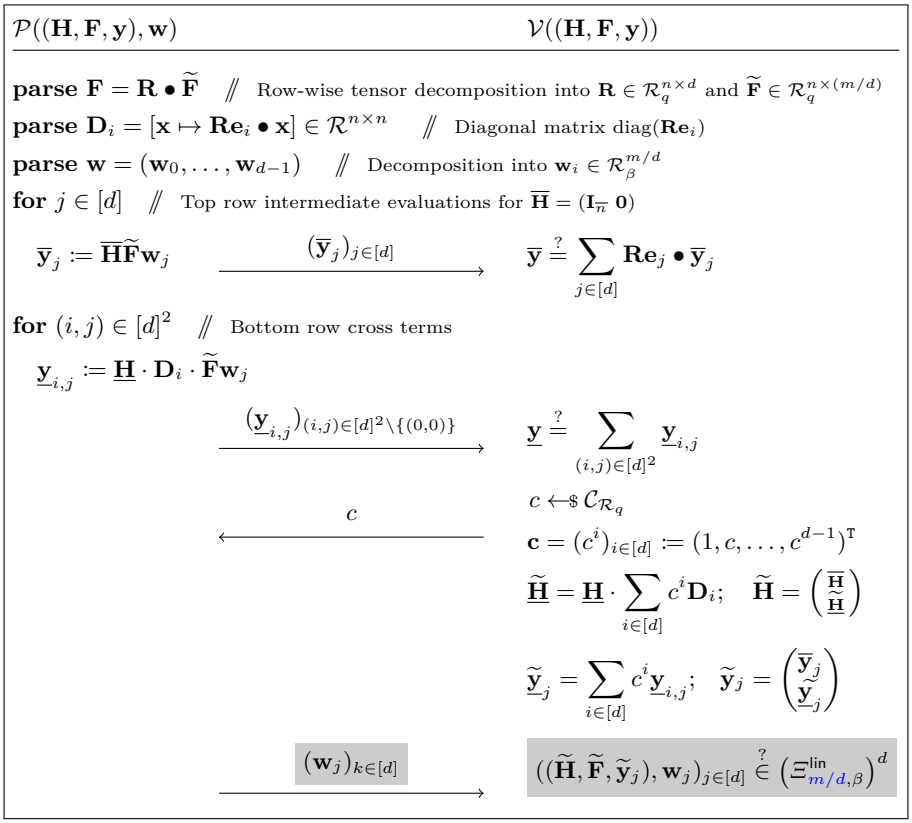

Fig. 5. Protocol Π^{split}, a reduction from $\Xi^{\mathsf{lin}}_{m,\beta}$ to $\left(\Xi^{\mathsf{lin}}_{m/d,\beta}\right)^d$. Π^{split} sends the marked parts only as a *proof* (but not *reduction*) of knowledge.

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[14] of $\mathbf{R}$ and $\widetilde{\mathbf{F}}$ as seen in Eq. (3). In this notation,

$$\mathbf{F}\cdot\mathbf{w} = (\mathbf{R}\bullet\widetilde{\mathbf{F}})\cdot\left(\sum_{i=0}^{\mu-1}\mathbf{e}_i\otimes\mathbf{w}_i\right) = \sum_i \underbrace{(\mathbf{Re}_i)}_{\in\mathcal{R}_q^n} \odot \underbrace{(\widetilde{\mathbf{F}}\mathbf{w}_i)}_{=\mathbf{y}_i\in\mathcal{R}_q^n} \tag{4}$$

where we use the Hadamard product to multiply the vector $\mathbf{Re}_i$ with $\mathbf{y}_i$ componentwise. Moreover, 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 $\widetilde{\mathbf{y}}_i = \widetilde{\mathbf{F}}\mathbf{w}_i$ for new statements $(\mathbf{H},\widetilde{\mathbf{F}},\widetilde{\mathbf{y}}_i)$ and witnesses $(\mathbf{w}_i)_{i\in[d]}$.

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. 5) 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.

Remark 3. In protocol Π^{split}, we could apply the optimisation of not sending $\overline{\mathbf{y}}_j$ (assuming that $\mathbf{Re}_j \neq \mathbf{0}$) and $\underline{\mathbf{y}}_{0,0}$, and instead let the verifier compute the unique accepting $\overline{\mathbf{y}}_j$ and $\underline{\mathbf{y}}_{0,0}$, i.e. such that $\overline{\mathbf{y}} = \sum_{j\in[d]} \mathbf{Re}_j \bullet \overline{\mathbf{y}}_j$ and $\underline{\mathbf{y}} = \sum_{(i,j)\in[d]^2} \underline{\mathbf{y}}_{i,j}$.

[14] (and more general forms, as block Kronecker product and Khatri–Rao product).

Lemma 4. *The protocol Π^{split} is a reduction of knowledge from $\Xi^{\mathsf{lin}}_{m,\mu,\beta}$ to $\left(\Xi^{\mathsf{lin}}_{m/d,\mu-1,\beta}\right)^d$. It is perfectly correct. It is relaxed knowledge sound from $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{m,\mu,\sqrt{d}\beta}$ to $\left(\Xi^{\mathsf{lin}}_{m/d,\mu-1,\beta}\right)^d$ with d-special sound extraction and knowledge error $\kappa = (d-1)/\left|\mathcal{C}_{\mathcal{R}_q}\right|$ if $2\beta \leq \beta^{\mathsf{sis}}$.*

Proof. Perfect correctness of Π^{split} from $\Xi^{\mathsf{lin}}_{m,\mu,\beta}$ to $\Xi^{\mathsf{lin}}_{m/d,\mu-1,\beta}$ 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

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

For relaxed knowledge soundness, we argue through d-special soundness. So we have d related accepting transcripts for challenge vectors $\mathbf{c}^{(k)}$ with witnesses $\mathbf{w}_j^{(k)}$ which satisfy $((\widetilde{\mathbf{H}}, \widetilde{\mathbf{F}}, \widetilde{\mathbf{y}}_j), \mathbf{w}_j) \in \Xi^{\mathsf{lin}\vee\mathsf{sis}}_{m/d,\mu-1}$.

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

$$\mathbf{w}^{(k)} = \begin{pmatrix} \mathbf{w}_0^{(k)} \\ \vdots \\ \mathbf{w}_{d-1}^{(k)} \end{pmatrix} \quad \text{satisfies} \quad \overline{\mathbf{H}}\mathbf{F}\mathbf{w}^{(k)} = \overline{\mathbf{H}} \sum_i \mathbf{D}_i \widetilde{\mathbf{F}} \mathbf{w}_i^{(k)} = \sum_i \overline{\mathbf{y}}_i = \overline{\mathbf{y}}$$

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, $\|\mathbf{w}\|_2 \leq \sqrt{d}\beta'$.

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 most 2β, such that $\overline{\mathbf{H}}\widetilde{\mathbf{H}}\mathbf{v}_j = \mathbf{0}$, and thus $\overline{\mathbf{H}}\mathbf{F}(\mathbf{v}_j, \mathbf{0}, \ldots, \mathbf{0}) = \mathbf{0}$ 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 $\overline{\mathbf{H}}\widetilde{\mathbf{H}}\mathbf{v}_j = \mathbf{0}$, we simply let $\mathbf{v}_j$ be the extracted witness and note that $\|\mathbf{v}_j\| \leq \beta \leq \beta^{\mathsf{sis}}$. □

Remark 4. 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 n^{out} $\mathcal{R}_q$-elements of communication. Moreover, whenever $\underline{\mathbf{H}}$ has structured rows (e.g., contains (permuted) identity submatrices, etc.), application specific optimisations may apply.

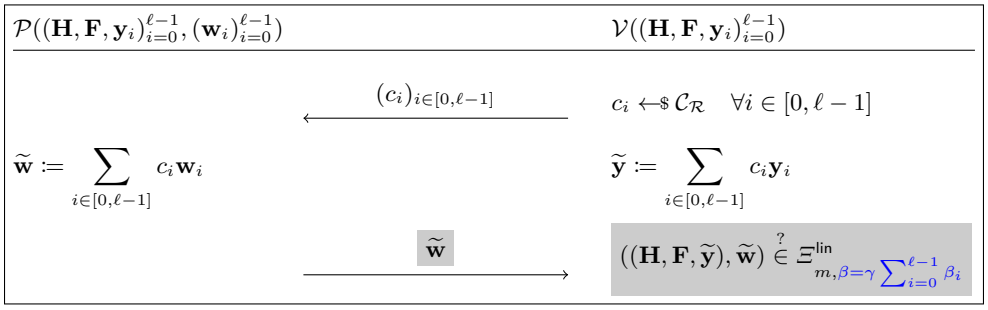

Fig. 6. Protocol Π^{fold} folds multiple instances of Ξ^{lin} with the same $(\mathbf{H},\mathbf{F})$ into one. Π^{fold} sends the marked parts only as a *proof* (but not *reduction*) of knowledge.

5.4 Π^{fold}: Fold Knowledge Reduction

In Fig. 6, we present the protocol Π^{fold}, which is a simple batch verification for many statements of the same type. It takes ℓ instances of $((\mathbf{H},\mathbf{F},\mathbf{y}_i),\mathbf{w}_i)_{i\in[\ell]}$ of $\Xi^{\mathsf{lin}}_{m,\beta}$ with the same $(\mathbf{H},\mathbf{F})$, and produces a random linear combination $((\mathbf{H},\mathbf{F},\mathbf{y}),\widetilde{\mathbf{w}})$ as output, with increased norm bounds.

Lemma 5. *The protocol Π^{fold} is a reduction of knowledge from $\left(\Xi^{\mathsf{lin}}_{\beta}\right)^{\ell}$ to $\Xi^{\mathsf{lin}}_{\gamma\ell\beta}$. It is perfectly correct. It is ℓ-CWSS and relaxed knowledge sound from $\left(\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{2\theta\beta'}\right)^{\ell}$ to $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{\beta'}$ if $\beta' \le \beta^{\mathsf{sis}}$.*

Proof. For perfect correctness, it is clear that $\widetilde{\mathbf{w}}$ satisfies $\mathbf{HF}\widetilde{\mathbf{w}} = \widetilde{\mathbf{y}}$ by construction. Moreover, $\|\widetilde{\mathbf{w}}\| \le \sum_{i=1}^{\ell-1}\|c_i\mathbf{w}_i\| \le \sum_{i=1}^{\ell-1}\gamma\|\mathbf{w}_i\| \le \gamma\ell\beta$, thus the norm is also within bounds and correctness follows.

For relaxed knowledge soundness, through ℓ-CWSS we are given $\ell+1$ accepting transcripts $\widetilde{\mathbf{w}}_0,\dots,\widetilde{\mathbf{w}}_\ell$, for challenges $(c_0^{(i)},\dots,c_{\ell-1}^{(i)})$ where $\mathbf{c}^{(i)}$ and $\mathbf{c}^{(\ell)}$ differ exactly in coordinate $i \in \{0,\dots\ell-1\}$. We can now subtract the accepting equations to obtain

$$\mathbf{HF}(\widetilde{\mathbf{w}}_i - \widetilde{\mathbf{w}}_\ell) = (\mathbf{c}^{(i)} - \mathbf{c}^{(\ell)})\mathbf{y} = (c_i^{(i)} - c_i^{(\ell)})\mathbf{y}_i$$

and thus, setting

$$\mathbf{w}_i \coloneqq \frac{1}{c_i^{(i)} - c_i^{(\ell)}}(\widetilde{\mathbf{w}}_i - \widetilde{\mathbf{w}}_\ell)$$

is a witness in $\mathcal{R}_q^m$ satisfying $\mathbf{HF}\mathbf{w}_i = \mathbf{y}_i$; Here, we use the subtractive set property of $\mathcal{C}_{\mathcal{R}}$. Moreover, we have

$$\|\mathbf{w}_i\| = \left\|\frac{1}{c_i^{(i)} - c_i^{(\ell)}}(\widetilde{\mathbf{w}}_i - \widetilde{\mathbf{w}}_\ell)\right\| \le \theta\cdot 2\cdot\beta'$$

by definition of the inverse-expansion factor for $\mathcal{C}_{\mathcal{R}}$ and the bounds on all $\widetilde{\mathbf{w}}_j$.

Finally, the OR-branch in $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$ is handled by letting $\mathbf{w}_i$ equal $\widetilde{\mathbf{w}}$ for all i; obviously, $\overline{\mathbf{H}}\mathbf{F}\mathbf{w}_i = \mathbf{0}$ holds and $\|\mathbf{w}_i\| \le \beta' \le \beta^{\mathsf{sis}}$. This completes the proof. □

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

The protocol Π^{batch} (Fig. 7) 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 $\Xi^{\mathsf{lin}}_{m,\beta}$ to an instance $((\widetilde{\mathbf{H}},\mathbf{F},\widetilde{\mathbf{y}}),\mathbf{w})$, where the dimension 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 6. *The protocol Π^{batch} is a reduction of knowledge from $\Xi^{\mathsf{lin}}_{n^{\mathsf{out}},\beta}$ to $\Xi^{\mathsf{lin}}_{\overline{n}+1,\beta}$. It is perfectly correct. It is relaxed knowledge sound from $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{n^{\mathsf{out}},\beta'}$ to $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{\overline{n}+1,\beta'}$ with knowledge error $\kappa = \frac{\underline{n}-1}{|\mathcal{C}_{\mathcal{R}_q}|} \le \frac{\underline{n}}{|\mathcal{C}_{\mathcal{R}_q}|}$ if $2\beta' \le \beta^{\mathsf{sis}}$. Extraction requires two uniformly distributed transcripts (Section 3.2).*

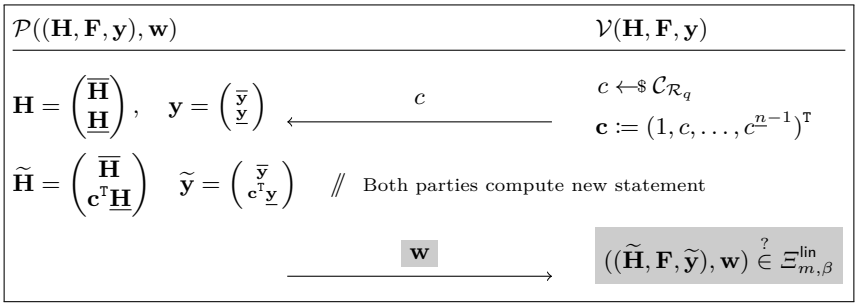

Fig. 7. Protocol Π^{batch} reduces an instance of Ξ^{lin} to another in instance with fewer rows by batching to last $\underline{n}$ of $\mathbf{H}$. Π^{batch} sends the marked parts only as a *proof* (but not *reduction*) of knowledge.

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 $d \cdot \left|\mathcal{C}_{\mathcal{R}_q}\right|$. 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 $\kappa = \frac{\underline{n}-1}{\left|\mathcal{C}_{\mathcal{R}_q}\right|}$ challenges can be accepting for $\mathbf{w}_1$ (by Schwartz–Zippel). 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.[15] Now, $\mathbf{v} = \mathbf{w}_1 - \mathbf{w}_2$ is a non-zero preimage $\overline{\mathbf{H}}\mathbf{F}\mathbf{v} = \mathbf{0}$ of norm at most $2\beta' \leq \beta^{\mathsf{sis}}$, i.e. the OR-branch of $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$. □

5.6 $\Pi^{\mathsf{split\&fold}}$: Split-and-Fold

We describe protocol $\Pi^{\mathsf{split\&fold}}$ which reduces the size of the witness. While it is perfectly correct, the extracted relation suffers a (small) growth in norm, i.e. it is only relaxed knowledge sound. The protocol $\Pi^{\mathsf{split\&fold}}$ proceeds as follows:

(1) $\Xi^{\mathsf{lin}}_{m,\beta_0} \xrightarrow{\Pi^{b\text{-decomp}}} (\Xi^{\mathsf{lin}}_{m,b})^\ell$ (Reduce norms by b-ary decomposition.)

(2) $(\Xi^{\mathsf{lin}}_{m,b})^\ell \xrightarrow{\Pi^{\mathsf{split}}} (\Xi^{\mathsf{lin}}_{m/d,b})^{\ell d}$ (Split statements into smaller ones.)

(3) $(\Xi^{\mathsf{lin}}_{m/d,b})^{\ell d} \xrightarrow{\Pi^{\mathsf{fold}}} \Xi^{\mathsf{lin}}_{m/d,\beta_3}$ (Fold statements into single one.)

The goal behind $\Pi^{\mathsf{split\&fold}}$ is to reduce the statement size by applying Π^{split} and then Π^{fold}. However, doing just this increases the norm the folded witness. To avoid this, $\Pi^{\mathsf{split\&fold}}$ first applies a b-ary decomposition which decreases the norm of $\mathbf{w}$ sufficiently, with proper parameters, we get $\beta_3 = \beta_0$ again. Correctness of this protocol is straightforward to show. Relaxed knowledge soundness also follows from relaxed soundness of the building blocks. Unfortunately, the norm of the extracted witness *does* grow, no matter how the parameters are chosen.

Remark 5. When applying Π^{split} to the product relation $(\Xi^{\mathsf{lin}}_{m,b})^\ell$, it is crucial that a single challenge is reused among all instances. This ensures that if all $(\mathbf{H}_i,\mathbf{F}_i)$ were identical before splitting, they still are identical after splitting.

Theorem 5. *Let $\beta_0, b \in \mathbb{N}$ (suitably chosen). Then $\Pi^{\mathsf{split\&fold}}$ is perfectly correct from $\Xi^{\mathsf{lin}}_{m,\beta_0}$ to $\Xi^{\mathsf{lin}}_{m/d,\beta_3}$, where $\beta_3 = \gamma\ell d\frac{1}{2}\sqrt{m}\varphi^{3/2}\beta_0$ with $\ell = \lceil \log_b(2\beta_0+1)\rceil$. It is relaxed knowledge sound from $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{m,\beta'_0}$ to $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{m/d,\beta'_3}$ if $\beta'_0 = 4\sqrt{d}\theta\beta_0\beta'_3 \leq \beta^{\mathsf{sis}}$. The knowledge error is at most $\frac{\ell d}{|\mathcal{C}_{\mathcal{R}}|} + \frac{d-1}{\left|\mathcal{C}_{\mathcal{R}_q}\right|}$.*

Proof. For correctness, the change in dimensions are clear. We write β_i for the bound after the i-th step, starting out with β_0 and finishing with β_3. Let $\ell = \lceil \log_b(2\beta_0+1)\rceil$ as in Lemma 3. We have

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

– $\beta_1 = \frac{1}{2}\sqrt{m}\varphi^{3/2}b$ by Lemma 3.
– $\beta_2 = \beta_1$ by Lemma 4.
– $\beta_3 = \gamma\ell d\beta_2$ by Lemma 5.

Overall, we find

$$\beta_3 = \gamma\ell d\frac{1}{2}\sqrt{m}\varphi^{3/2}b.$$

For knowledge soundness, we work in the opposite direction. Starting out from a witness in $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{m/d,\beta'_3}$, we follow the bounds back to the start. We have

– $\beta'_2 = 2\theta\beta'_3$ by Lemma 5.
– $\beta'_1 = \sqrt{d}\beta'_2$ by Lemma 4.
– $\beta'_0 = 2\beta_0\beta'_1$ by Lemma 3.

where β_0 is the fixed choice for $\Pi^{b\text{-}\mathsf{decomp}}$ (w.r.t. correctness). This yields overall

$$\beta'_0 = 4\sqrt{d}\theta\beta_0\beta'_3.$$

The knowledge error is $\frac{d-1}{|\mathcal{C}_{\mathcal{R}_q}|}$ for $\Pi^{b\text{-}\mathsf{decomp}}$ and Π^{split}, and it is $\frac{\ell d}{|\mathcal{C}_{\mathcal{R}}|}$ for Π^{fold}. (For completeness, we note that since β'_0 is the largest bound, all intermediate bounds satisfy $\beta'_i \leq \beta^{\mathsf{sis}}$ as well, so the extractors work.) □

5.7 Π^{norm}, $\Pi^{\mathsf{norm}+}$, Π^{ip}, $\Pi^{\mathsf{ip}+}$: Norm and Inner Product Checks

To restrain the norm growth of the extracted witness, we introduce norm checks. First we present the "core" norm check protocol Π^{norm}, which handles the interesting part of the norm check by reducing the norm relation Ξ^{norm}, to multiple Ξ^{lin} relations. We then compose Π^{norm} with Π^{batch} and Π^{fold} to yield the "full" norm check protocol $\Pi^{\mathsf{norm}+}$, which reduces the norm relation to a single Ξ^{lin} relation. At the core of Π^{norm} is a mechanism for checking the trace of an inner product. By removing the trace operation, we obtain similar protocols Π^{ip} and $\Pi^{\mathsf{ip}+}$ for proving inner product relations Ξ^{ip}. 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.

$$\Xi^{\mathsf{norm}\vee\mathsf{sis}}_{\substack{\mathcal{R},q,m,n^{\mathsf{out}},\\ \mu,\beta,\beta^{\mathsf{sis}}}} := \left\{ \begin{array}{l} ((\mathbf{H},\mathbf{F},\mathbf{y},\nu),\mathbf{w}): \\ \quad \mathbf{H} \in \mathcal{R}_q^{n^{\mathsf{out}}\times n};\ \mathbf{F} \in \mathcal{R}_q^{n\times d^{\otimes\mu}} \subseteq \mathcal{R}_q^{n\times m};\ \mathbf{y} \in \mathcal{R}_q^{n^{\mathsf{out}}}, \nu \leq \beta \\ \left\{ \begin{array}{c} \|\mathbf{w}\| \leq \nu \\ \mathbf{HFw} = \mathbf{Hy} \bmod q \end{array} \right\} \text{ or } \left\{ \begin{array}{c} \|\mathbf{w}\| \leq \beta^{\mathsf{sis}} \\ \overline{\mathbf{H}}\mathbf{F}\mathbf{w} = \mathbf{0}_{\overline{n}} \bmod q \end{array} \right\} \end{array} \right\},$$

$$\Xi^{\mathsf{ip}\vee\mathsf{sis}}_{\substack{\mathcal{R},q,m,n^{\mathsf{out}},\\ \mu,\beta,\beta^{\mathsf{sis}},\alpha}} := \left\{ \begin{array}{l} ((\mathbf{H},\mathbf{F},\mathbf{y},t),\mathbf{w}): \\ \quad \mathbf{H} \in \mathcal{R}_q^{n^{\mathsf{out}}\times n};\ \mathbf{F} \in \mathcal{R}_q^{n\times d^{\otimes\mu}} \subseteq \mathcal{R}_q^{n\times m};\ \mathbf{y} \in \mathcal{R}_q^{n^{\mathsf{out}}}, t \in \mathcal{R}_q \\ \left\{ \begin{array}{c} \|\mathbf{w}\| \leq \beta \\ \mathbf{HFw} = \mathbf{Hy} \bmod q \\ \langle \mathbf{w}, \alpha(\mathbf{w})\rangle_{\mathcal{R}} = t \bmod q \end{array} \right\} \text{ or } \left\{ \begin{array}{c} \|\mathbf{w}\| \leq \beta^{\mathsf{sis}} \\ \overline{\mathbf{H}}\mathbf{F}\mathbf{w} = \mathbf{0}_{\overline{n}} \bmod q \end{array} \right\} \end{array} \right\}.$$

Note that, compared to Ξ^{lin}, the norm relation Ξ^{norm} differs in that the witness norm bound $\nu \leq \beta$ is given as part of the statement, and a stricter norm relation $\|\mathbf{w}\| \leq \nu$ 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 an 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.

The core 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. That is, for $\mathbf{w}$, we define the polynomials $g(X) = \sum_{i\in[m]} w_i X^i$ and $\bar{g}(X) = \sum_{i\in[m]} \bar{w}_i X^i$, and observe that the Laurent polynomial $L(X) = \sum_{i\in\pm[m]} v_i X^i :=$

$g(X) \cdot \bar{g}(X^{-1})$ has constant coefficient $\sum_{i\in[m]} w_i \bar{w}_i$, which is the inner product $\langle \mathbf{w}, \bar{\mathbf{w}} \rangle_{\mathcal{R}}$. 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 $|k| \geq m$. We exploit this symmetry to commit to $L(X)$ by committing only to $(v_0, \ldots, v_{m-1})$. Setting $h(X) = \sum_{i\in[m]} v_i X^i$ and $\bar{h}(X) = \sum_{i\in[m]} \bar{v}_i X^i$, we see that $L(X) = h(X) + \bar{h}(X^{-1}) - v_0$. We use this equality to prove the polynomial identity $L(X) = g(X)\bar{g}(X^{-1})$ by evaluating $g, \bar{g}, h, \bar{h}$ at a random point $\xi \leftarrow\!\!\$\, \mathcal{C}_{\mathcal{R}_q}$ (and checking if $v_0 = t$).

However, A problem with the soundness occurs if the approach is used naively: The terms v_i have norm bounded by β^2, so $\|\mathbf{v}\|$ which may be beyond the threshold for which the commitment is binding.

A natural approach is to run $\Pi^{b\text{-}\mathsf{decomp}}$ to counteract this problem. However, doing so modularly runs into problems and comes at the cost of a suboptimal relaxed knowledge guarantee. We can tighten our analysis if we treat the composition with $\Pi^{b\text{-}\mathsf{decomp}}$ 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 $\Pi^{b\text{-}\mathsf{decomp}}$ incurs a large factor of norm growth, however, in Π^{ip}, we do not care about the auxiliary commitment to $\mathbf{v}$ (which norms up to $\approx \sqrt{m}\beta^2$). Thus, we can argue directly for the decomposition of $\mathbf{v}$ into smaller $\mathbf{v}_i$ of norm at most β, which are binding. This avoids the need for recovering recover $\mathbf{v}$ via relaxed knowledge for $\Pi^{b\text{-}\mathsf{decomp}}$, which significantly improves the parameters. This optimised protocol is presented in Fig. 8

Lemma 7 (Security of Π^{norm} and Π^{ip}). *The protocol Π^{norm} (resp. Π^{ip}) is a reduction of knowledge from $\Xi^{\mathsf{norm}}_{n^{\mathsf{out}}}$ (resp. $\Xi^{\mathsf{ip}}_{n^{\mathsf{out}}}$) to $\left(\Xi^{\mathsf{lin}}_{n^{\mathsf{out}}+3}\right)^{\ell+1}$, where $b_{\mathsf{ip}} \leq 2\beta/(\sqrt{m}\varphi^{3/2})$ into $\ell \geq \log_{b_{\mathsf{ip}}}(2\beta^2+1)$ for $b_{\mathsf{ip}}, \ell \in \mathbb{N}$. It is perfectly correct. It is relaxed knowledge sound from $\Xi^{\mathsf{norm}\vee\mathsf{sis}}_{n^{\mathsf{out}}}$ (resp. $\Xi^{\mathsf{ip}\vee\mathsf{sis}}_{n^{\mathsf{out}}}$) to $\left(\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{n^{\mathsf{out}}+3}\right)^{\ell}$ if $2\beta \leq \beta^{\mathsf{sis}}$. Extraction requires two uniformly distributed transcripts (Section 3.2) and has knowledge error $\kappa \leq \frac{2m}{|\mathcal{C}_{\mathcal{R}_q}|}$.*

Proof. To show that $\|\mathbf{v}_i\|_2 \leq \beta$, it suffices to recall the bound for $\Pi^{b\text{-}\mathsf{decomp}}$ from Lemma 3. Indeed, we set b_{ip} such that $\beta \geq \frac{1}{2}\sqrt{m}\varphi^{3/2} b_{\mathsf{ip}}$ holds by definition. That the linear relations hold was already explained above. Thus, the reduction of knowledge is perfectly complete.

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

$$L(X) \neq h(X) + \bar{h}(X^{-1}) - v_0 \quad \text{but} \quad L(\xi) \neq h(\xi) + \bar{h}(\xi^{-1}) - v_0 \tag{6}$$

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 6.) However, this is a soundness argument for *fixed* polynomials, but the vector $\mathbf{v} = \sum_{i=0}^{\ell-1} b_{\mathsf{ip}}^i \mathbf{v}_i$ is only determined in the last step, Now, we argue analogous to Lemma 6 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), where $\widetilde{\mathbf{y}}^{(0)} = \sum_{i=0}^{\ell} b_{\mathsf{ip}}^i \widetilde{\mathbf{y}}_i'^{(0)}$ denotes the non-decomposed linear claim for $\mathbf{v}$. Observe that only fraction κ of challenges can satisfy (6) for any fixed $\mathbf{v}^{(0)}$. Thus, if the adversary succeeds with probability ϵ, then with probability $\epsilon - \kappa$, the 2-transcript extractor gives transcripts with[16] $\mathbf{v}_i^{(0)} \neq \mathbf{v}_i^{(1)}$ for some i^*. Let $\mathbf{u} = \mathbf{v}_{i^*}^{(0)} - \mathbf{v}_{i^*}^{(1)}$. Then we have $\mathbf{HFu} = 0$, as $\mathbf{HFv}_{i^*}^{(j)} = \mathbf{y}'_i$ holds for both $j = 0, 1$. This yields an OR-branch witness for $\Xi^{\mathsf{ip}\vee\mathsf{sis}}$ with norm at most $2\beta \leq \beta^{\mathsf{sis}}$.

To conclude the proof, we note that we do not care about the auxiliary $\mathbf{v}_i$, and they do not play a role in the bound β (aside from the OR-branch), which is for $\mathbf{w}$ (which is merely passed through by Π^{norm} and Π^{ip}). □

We recall the technical peculiarity in the construction and proof, that since the auxiliary commitment to $\mathbf{v}$ in Π^{norm} and Π^{ip} is not extracted through $\Pi^{b\text{-}\mathsf{decomp}}$, but we directly committed to the decomposed $\mathbf{v}_i$. For this reason, relaxed soundness incurs almost no norm growth. If we had considered the intermediate relation for $\mathbf{v}$, then technically $\mathbf{v}$ becomes part of the witness which the reduction of knowledge for $\Pi^{b\text{-}\mathsf{decomp}}$ must then extract. This blows up the norm bound (by β).

The full protocols $\Pi^{\mathsf{norm}+}$ and $\Pi^{\mathsf{ip}+}$. The full protocol $\Pi^{\mathsf{norm}+}$ (resp. $\Pi^{\mathsf{ip}+}$) simply reduces the $\ell+1$ statements after Π^{norm} (resp. Π^{ip}) back to a single one with the minimal number of rows, by applying Π^{batch} and Π^{fold}. That is, $\Pi^{\mathsf{norm}+}$ (and analogously $\Pi^{\mathsf{ip}+}$) work as follows:

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

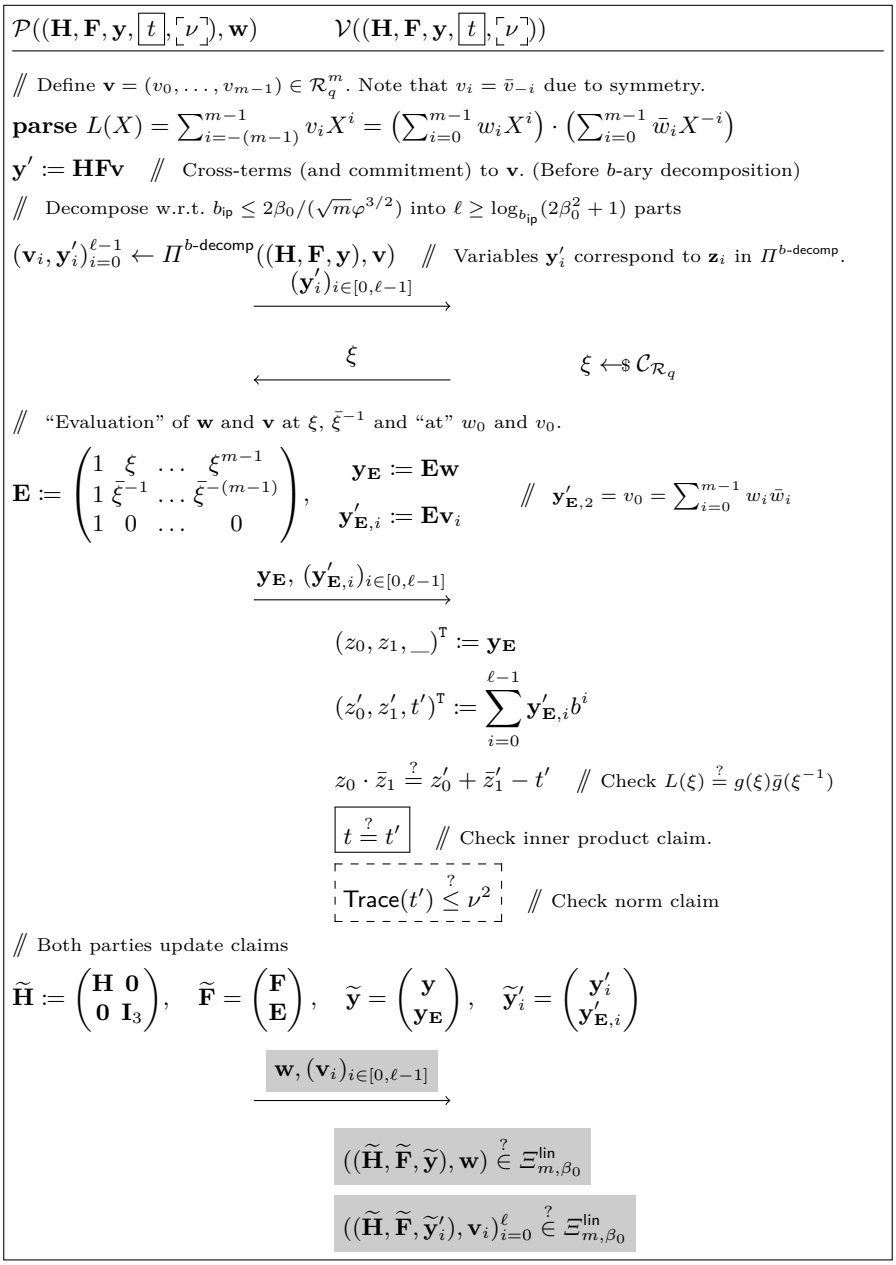

Fig. 8. Protocols $\boxed{\Pi^{\mathsf{norm}}}$ for Ξ^{norm} and $\boxed{\Pi^{\mathsf{ip}}}$ for Ξ^{ip} where $\alpha=\overline{\mathsf{id}}$. The case $\alpha=\mathsf{id}$ can be obtained by removing all conjugates. Π^{norm} and Π^{ip} send the marked parts only as a *proof* (but not *reduction*) of knowledge.

(1) $\Xi^{\mathsf{norm}}_{n^{\mathsf{out}},\beta} \xrightarrow{\Pi^{\mathsf{norm}}} (\Xi^{\mathsf{lin}}_{n^{\mathsf{out}}+3,\beta})^{\ell+1}$ (Reduce to claims in $(\Xi^{\mathsf{lin}})^{\ell+1}$.)

(2) $(\Xi^{\mathsf{lin}}_{n^{\mathsf{out}}+3,\beta})^{\ell+1} \xrightarrow{\Pi^{\mathsf{batch}}} (\Xi^{\mathsf{lin}}_{\overline{n}+1,\beta})^{\ell+1}$ (Reduce number of rows.)

(3) $(\Xi^{\mathsf{lin}}_{\overline{n}+1,\beta})^{\ell+1} \xrightarrow{\Pi^{\mathsf{fold}}} \Xi^{\mathsf{lin}}_{\overline{n}+1,(\ell+1)\gamma\beta}$ (Reduce to one claim with $\beta' = (\ell+1)\gamma\beta$.)

In words, first $\Pi^{\mathsf{norm}+}$ (resp. $\Pi^{\mathsf{ip}+}$) runs Π^{norm} (resp. Π^{ip}) to obtain the commitments to $\mathbf{v}_i$. Next it runs Π^{batch} to reduce to $\overline{n}+1$ rows again (i.e. to $\underline{n}=1$). Finally, the claims are folded into one by using Π^{fold}. The full norm protocol Π^{norm} and its security guarantees are analogous.

Remark 6. When applying Π^{batch} to the product relation $(\Xi^{\mathsf{lin}}_{n^{\mathsf{out}}+3,\beta})^{\ell+1}$, it is crucial a single challenge is reused among all instances. This ensures that if all $(\mathbf{H}_i, \mathbf{F}_i)$ were identical before batching, they still are identical after batching.

Theorem 6 (Security of $\Pi^{\mathsf{norm}+}$ and $\Pi^{\mathsf{ip}+}$). *The protocol $\Pi^{\mathsf{norm}+}$ (resp. $\Pi^{\mathsf{ip}+}$) is a reduction of knowledge $\Xi^{\mathsf{norm}}_{n^{\mathsf{out}},\beta_0}$ (resp. $\Xi^{\mathsf{ip}}_{n^{\mathsf{out}},\beta_0}$) to $(\Xi^{\mathsf{lin}}_{\overline{n}+1,\beta_3})^{\ell+1}$, where $\ell \geq \log_{b_{\mathsf{ip}}}(2\beta^2+1)$ for $b_{\mathsf{ip}} \leq 2\beta/(\sqrt{m}\varphi^{3/2})$ and $\beta_3 = (\ell+1)\beta_0$ for $b_{\mathsf{ip}}, \ell \in \mathbb{N}$. It is perfectly correct. It is relaxed knowledge sound from $\Xi^{\mathsf{norm}\vee\mathsf{sis}}_{n^{\mathsf{out}},\beta'_0}$ (resp. $\Xi^{\mathsf{ip}\vee\mathsf{sis}}_{n^{\mathsf{out}},\beta'_0}$) to $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{\overline{n}+1,\beta'_3}$ if $\beta'_0 \leq \beta^{\mathsf{sis}}$, where $\beta'_0 = 4\theta\beta'_3$. The knowledge error is $\frac{\ell+1}{|\mathcal{C}_{\mathcal{R}}|} + \frac{4(\ell+1)+2m}{|\mathcal{C}_{\mathcal{R}_q}|}$.*

Proof. For correctness, let β_i denote the norm bound after the i-th protocol was applied. Then we get We have

- $\beta_1 = \beta_0$ by Lemma 7.
- $\beta_2 = \beta_1$ by Lemma 6.
- $\beta_3 = \gamma(\ell+1)\beta_2$ by Lemma 5.

Overall, we find

$$\beta_3 = \gamma(\ell+1)\beta_0.$$

For knowledge soundness, we work in the opposite direction. Starting out from a witness in $\Xi^{\mathsf{lin}}_{m/d,\beta'_3}$, we follow the bounds back to the start. We have

- $\beta'_2 = 2\theta\beta'_3$ by Lemma 5.
- $\beta'_1 = 2\beta'_2$ by Lemma 6.
- $\beta'_0 = 2\beta'_1$ by Lemma 7.

This yields overall

$$\beta'_0 = 8\theta\beta'_3.$$

The knowledge error for Π^{ip} is $\frac{2m}{|\mathcal{C}_{\mathcal{R}_q}|}$, For Π^{batch} it is $\frac{4}{|\mathcal{C}_{\mathcal{R}_q}|}$ per instance (as 4 rows are batched, namely $\underline{n}+3$), so $\frac{4(\ell+1)}{|\mathcal{C}_{\mathcal{R}_q}|}$ overall. For Π^{fold} it is $\frac{\ell+1}{|\mathcal{C}_{\mathcal{R}}|}$ Taken together, we arrive at the claimed knowledge error. Note that the case of a vSIS break is handled implicitly in the knowledge reductions. □

Norm Checks to Upgrade Relaxed Soundness. Currently, our norm check is defined for the relation Ξ^{norm}, which inherits the parameter β from Ξ^{lin}, and also contains an explicit ν as a norm *statement*. For convenience, we now define the $\Pi^{\mathsf{norm}_\beta}$ protocol, which is a reduction of knowledge from Ξ^{lin}_β to Ξ^{lin}_β, that works as follows: $\Pi^{\mathsf{norm}_\beta}$ runs Π^{norm} on (implicit) input $\nu = \beta$ for both parties.

Analogous to $\Pi^{\mathsf{norm}+}$, we define $\Pi^{\mathsf{norm}^+_\beta} : \Xi^{\mathsf{lin}}_\beta \to \Xi^{\mathsf{lin}}_\beta$. Clearly, $\Pi^{\mathsf{norm}_\beta}$ (resp. and $\Pi^{\mathsf{norm}^+_\beta}$) directly inherits all correctness and security guarantees of Π^{norm} (resp. $\Pi^{\mathsf{norm}+}$). We introduce $\Pi^{\mathsf{norm}_\beta}$ and $\Pi^{\mathsf{norm}^+_\beta}$ solely for compositional reasons: They start and end with a claim(s) in Ξ^{lin}_β. Let us stress that the $\Pi^{\mathsf{norm}_\beta}$ protocol *upgrades* the (previously relaxed) bound on the norm of $\mathbf{w}$ to $\|\mathbf{w}\| \leq \beta$. We capture this in following corollary.

Corollary 1. *Adopt the setting of Theorem 6. Then $\Pi^{\mathsf{norm}^+_\beta}$ is relaxed knowledge sound from $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_\beta$ to $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{\beta'}$ if $2\beta' \leq \beta^{\mathsf{sis}}$, with the same knowledge error as Theorem 6.*

Proof. Unless the OR-branch occurs in $\Xi^{\mathsf{lin}\vee\mathsf{sis}}$, the norm of the starting witness $\mathbf{w}$ is *proven* to be at most $\nu = \beta$ by Π^{norm}. Thus, the extracted witness has norm at most β *or* it a witness for the OR-branch.

5.8 Π^{sfn}: Split-and-Fold with Norm Checks

We describe our split-and-fold protocol with intermediate norm check Π^{sfn}. It first runs the norm check $\Pi^{\mathsf{norm}^+_\beta}$ to upgrade the relaxed norm bound to a strict one. Then it splits-and-folds to reduce the witness size. If parameters are set correctly, then $\Pi^{\mathsf{split\&fold}} \colon \Xi^{\mathsf{lin}}_{m,\beta} \to \Xi^{\mathsf{lin}}_{m/d,\beta}$ is reduction of knowledge with relaxed knowledge soundness $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{m,\beta} \to \Xi^{\mathsf{lin}\vee\mathsf{sis}}_{m/d,\beta'}$. Crucially, the bound β is guaranteed exactly after extraction (unless the OR-branch, i.e. a vSIS break is extracted).

(1) $\Xi^{\mathsf{lin}}_{\beta_0} \xrightarrow{\Pi^{\mathsf{norm}^+_\beta}} \Xi^{\mathsf{lin}}_{\beta_1}$: Run a norm check for β_0.
(2) $\Xi^{\mathsf{lin}}_{m,\beta_1} \xrightarrow{\Pi^{\mathsf{split\&fold}}} \Xi^{\mathsf{lin}}_{m/d,\beta_2}$: Run the split-and-fold to reduce witness size.

Theorem 7. *Let*

- *Let* $b_{\mathsf{ip}} \le 2\beta/(\sqrt{m}\varphi^{3/2})$ *and* $\ell_0 \ge \log_{b_{\mathsf{ip}}}(2\beta^2+1)$.
- *Let* $b_1 \ge 1$ *be arbitrary and* $\ell_1 = \lceil \log_{b_1}(2\beta_0+1) \rceil$.
- *Let* $\beta_1 = \gamma(\ell_0+1)\beta_0$.
- *Let* $\beta_2 = \gamma\ell_1 d\frac{1}{2}\sqrt{m}\varphi^{3/2}b_1$.

Then Π^{sfn} *is perfectly correct from* $\Xi^{\mathsf{lin}}_{m,\beta_0}$ *to* $\Xi^{\mathsf{lin}}_{m/d,\beta_2}$.
If $\beta'_0 \le \beta^{\mathsf{sis}}$, *where*

- $\beta'_1 = 4\sqrt{d}\theta\beta_0\beta'_2$
- $\beta'_0 = 8\theta\beta'_1$

then Π^{sfn} *is relaxed knowledge sound from* $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{m,\beta_0}$ *to* $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{m/d,\beta'_2}$. *It has knowledge error* $\frac{2\ell+1}{|\mathcal{C}_\mathcal{R}|} + \frac{3(\ell_0+1)+2m+d-1}{|\mathcal{C}_{\mathcal{R}_q}|}$

Proof. Correctness follows since β_1 is the correctness assertion in Corollary 1 (or rather Theorem 6), and β_2 is the is the correctness assertion in Theorem 5. Similarly, the relaxed knowledge soundness follows by plugging in the respective relaxed knowledge soundness claims. Finally, the knowledge error is the sum of the knowledge errors. □

Π	β_1	β'_0	κ	#tr	Other
$\Pi^{b\text{-decomp}}$	$\frac{1}{2}\sqrt{m}\varphi^{3/2}b$	$2\beta_0\beta'_1$	0	1	$\ell = \lceil \log_b(2\beta_0+1) \rceil$
Π^{split}	β_0	$\sqrt{d}\beta'_1$	$(d-1)/\lvert\mathcal{C}_\mathcal{R}\rvert$	d	Need $2\beta'_0 \le \beta^{\mathsf{sis}}$
Π^{fold}	$\gamma\ell\beta_0$	$2\theta\beta'_1$	$\ell/\lvert\mathcal{C}_\mathcal{R}\rvert$	$\ell+1$	ℓ from $(\Xi^{\mathsf{lin}})^\ell$
Π^{batch}	β_0	β'_1	$\underline{n}/\lvert\mathcal{C}_{\mathcal{R}_q}\rvert$	2	Need $2\beta'_0 \le \beta^{\mathsf{sis}}$
$\Pi^{\mathsf{norm}}/\Pi^{\mathsf{ip}}$	β_0	β'_1	$2m/\lvert\mathcal{C}_{\mathcal{R}_q}\rvert$	2	
$\Pi^{\mathsf{split\&fold}}$	$\frac{1}{2}\gamma\ell d\sqrt{m}\varphi^{3/2}b\beta_0$	$4\sqrt{d}\theta\beta_0\beta'_1$	$\ell/\lvert\mathcal{C}_\mathcal{R}\rvert + (d-1)/\lvert\mathcal{C}_{\mathcal{R}_q}\rvert$	$d(\ell+1)$	$\ell = \ell_{\Pi^{b\text{-decomp}}}$
$\Pi^{\mathsf{norm}+}/\Pi^{\mathsf{ip}+}$	$\gamma(\ell+1)\beta_0$	$8\theta\beta'_1$	$\frac{\ell+1}{\lvert\mathcal{C}_\mathcal{R}\rvert} + \frac{4(\ell+1)+2m}{\lvert\mathcal{C}_{\mathcal{R}_q}\rvert}$	$4(\ell+2)$	See caption.

Table 1. Parameters of protocols. Expressed as $\beta_1 = f(\beta_0)$ for correctness, $\beta'_0 = g(\beta'_1)$ and $\beta'_0 \le \beta^{\mathsf{sis}}$ for relaxed soundness, knowledge error κ, number of transcripts to extract, other variables. For $\Pi^{\mathsf{ip}+}$, we have $\ell = \log_{b_{\mathsf{ip}}}(2\beta_0^2+1)$ for $b_{\mathsf{ip}} \le 2\beta/(\sqrt{m}\varphi^{3/2})$.

5.9 Asymptotic Communication Complexity

We now compute the proof size for the split-and-fold with norm checks protocol in Section 5.8. Having non-interactive proof systems in mind, we only count prover messages. Let $m = d^\mu$ where $d = O(1)$ and $\mu = O(\log m)$. The other parameters are chosen according to Table 1. As shown in Section 4, we can pick $\mathsf{f} = \mathsf{poly}(\lambda)$ and a subtractive set $\mathcal{C}_\mathcal{R}$ such that $\gamma = O(1)$ and $\theta = \mathsf{poly}(\lambda)$.

Recall the prover executes two sub-protocols:

(1) $\Xi^{\mathsf{lin}}_{\beta_0} \xrightarrow{\Pi^{\mathsf{norm}^+_\beta}} \Xi^{\mathsf{lin}}_{\beta_1}$: Run a norm check for β_0.

(2) $\Xi^{\mathsf{lin}}_{m,\beta_1} \xrightarrow{\Pi^{\mathsf{split\&fold}}} \Xi^{\mathsf{lin}}_{m/d,\beta_2}$: Run the split-and-fold to reduce witness size.

First, we turn to the $\Pi^{\mathsf{norm}^+_\beta}$ protocol in Figure 8. To prove relation $\Xi^{\mathsf{lin}}_{\beta_0}$, the prover starts with

$$\Xi^{\mathsf{norm}}_{n^{\mathsf{out}},\beta} \xrightarrow{\Pi^{\mathsf{norm}}} (\Xi^{\mathsf{lin}}_{n^{\mathsf{out}}+3,\beta})^{\ell+1}$$

where it sends all $\mathbf{y}'_i$ which in total have size $\ell_0 n^{\mathsf{out}}$ ring elements. After receiving the challenge ξ, the prover outputs $\mathbf{y}_E$ and $\mathbf{y}'_{E,i}$ of size 3 each – thus in total $3(\ell_0+1)$ ring elements. Then the prover runs

$$(\Xi^{\mathsf{lin}}_{n^{\mathsf{out}}+3,\beta})^{\ell+1} \xrightarrow{\Pi^{\mathsf{batch}}} (\Xi^{\mathsf{lin}}_{\overline{n}+1,\beta})^{\ell+1} \xrightarrow{\Pi^{\mathsf{fold}}} \Xi^{\mathsf{lin}}_{\overline{n}+1,(\ell+1)\gamma\beta}$$

where neither Π^{batch} nor Π^{fold} incur any no communication from the prover.

Next, we now move on to the $\Pi^{\mathsf{split\&fold}}$ protocol. The prover wants to give a proof for relation $\Xi^{\mathsf{lin}}_{m,\beta_1}$. The prover starts by running the $\Pi^{b\text{-}\mathsf{decomp}}$ protocol

$$\Xi^{\mathsf{lin}}_{m,\beta_1} \xrightarrow{\Pi^{b\text{-}\mathsf{decomp}}} (\Xi^{\mathsf{lin}}_{m,b_1})^{\ell_1}$$

where the prover sends ℓ_1 vectors $(\mathbf{z}_k)$ of size n^{out} elements in $\mathcal{R}_q$. Next, it runs Π^{split}

$$(\Xi^{\mathsf{lin}}_{m,b_1})^{\ell_1} \xrightarrow{\Pi^{\mathsf{split}}} (\Xi^{\mathsf{lin}}_{m/d,b_1})^{\ell_1 d}$$

and outputs (using the optimization in Remark 5): (i) $d-1$ intermediate "top-part" evaluations $\overline{\mathbf{y}}_j$ of size $\overline{n}$ elements in $\mathcal{R}_q$ and (ii) d^2-1 "bottom-part" evaluations $\underline{\mathbf{y}}_{i,j}$ of size $\underline{n}$ elements in $\mathcal{R}_q$. Finally the prover executes the Π^{fold} protocol

$$(\Xi^{\mathsf{lin}}_{m/d,b_1})^{\ell_1 d} \xrightarrow{\Pi^{\mathsf{fold}}} \Xi^{\mathsf{lin}}_{m/d,\beta_2}$$

where no prover message is sent. All in all, in each of $\mu = O(\log m)$ iterations of split-and-fold with norm checks, the prover sends

$$\big(\ell_0 n^{\mathsf{out}} + 3(\ell_0+1)\big) + (\ell_1 n^{\mathsf{out}} + (d-1)\overline{n} + (d^2-1)\underline{n})$$

elements in $\mathcal{R}_q$.

Simple example: vSIS opening proof. When proving knowledge of a vSIS commitment opening, we can set $n, n^{\mathsf{out}} \in O(1)$. Due to the polynomial challenge space for subtractive sets, we need to repeat $O(\lambda/\log\lambda)$ times to ensure $\approx 2^{-\lambda}$ soundness error. Finally, we can pick b_1 such that $\ell_1 = O(1)$. In total, the proof size in the number of ring elements is simply bounded by $O\left(\lambda \frac{\log m}{\log \lambda}\right)$.

Next, we turn to setting the bound required for the (v)SIS problem to be hard. From Theorem 7 we need to set $\beta'_2 = \beta_2$ and

$$\beta^{\mathsf{sis}} = \beta'_0 = 32\sqrt{d}\theta^2\beta_0\beta'_2 = 16\gamma\ell_1 d^{3/2}\sqrt{m}\varphi^{3/2}\theta^2\beta_0 b_1 = \mathsf{poly}(m,\lambda).$$

The next step is to estimate an asymptotic size of a ring element in $\mathcal{R}_q$.

Hardness of SIS. To measure 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 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]. That is, 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. [BDGL15]). 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 $\beta^\star = \beta^{\mathsf{sis}} \cdot \mathsf{poly}(\lambda)$ is hard when $\beta^\star < \min\left(2^{2\sqrt{\varphi \log q \log \delta_{\mathsf{rhf}}}}, q\right)$. By rearranging, we get that $\varphi \log q > {}^{\log^2 \beta^\star}/_{4\log\delta_{\mathsf{rhf}}}$. Note that

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

Finally, using the fact that $\beta^\star = \mathsf{poly}(m,\lambda)$, size of a single $\mathcal{R}_q$ element is asymptotically

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

Therefore, we deduce that the total proof size in bits is $O\left(\log^3 m \cdot {}^{\lambda^2}/_{\log^2\lambda}\right)$.

6 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 $\langle \mathbf{x}, \mathbf{y} \rangle_\mathbb{Z} = \tau\left(\psi_\mathbf{b}^{-1}(\mathbf{x}) \cdot_\mathcal{R} \overline{\psi_\mathbf{b}^{-1}(\mathbf{y})}\right)$.*

6.1 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 [LNP22].

Theorem 8. *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 $i \in \pm[\varphi]$.[17] 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 $x \coloneqq \psi^{-1}(\mathbf{x})$ and $\overline{y} \coloneqq \overline{\psi^{-1}(\mathbf{y})}$ can be expressed as

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

respectively. Note that their product $z \coloneqq x \cdot \overline{y}$ satisfies

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

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

Remark 7. The constant term map $\mathsf{ct}(x)$ from Theorem 8 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 may 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.

[17] 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 $\mathsf{ct}(\zeta^{-1}) = -1$.

6.2 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 9. *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 $\overline{y} = \overline{\psi_{\mathbf{b}^+}^{-1}(\mathbf{y})}$ for some $\mathbf{x}, \mathbf{y} \in \mathbb{Z}^\delta$. We have

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

□

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 $\gamma_S \leq \mathfrak{f}$.

6.3 Tensor of Prime Real Cyclotomics

To allow more fine-grained parameter selection, we extend the result in Sections 6.1 and 6.2 by constructing larger rings using the tensor product, inspired by [BFOV04, Proposition 6]. Concretely, we construct subtractive sets for rings $\mathcal{R} = \mathcal{O}_{\mathcal{K}_{2^d}} \otimes \mathcal{O}_{\mathcal{K}^+_{\mathfrak{f}_0}} \otimes \ldots \otimes \mathcal{O}_{\mathcal{K}^+_{\mathfrak{f}_{k-1}}}$ for distinct odd primes $\mathfrak{f}_0, \ldots, \mathfrak{f}_{k-1}$. Note that $\mathcal{R}$ has conductor $\mathfrak{f} = 2^d \cdot \prod_{i\in[k]} \mathfrak{f}_i$ and degree $\delta = 2^d \cdot \prod_{i\in[k]} (\mathfrak{f}_i - 1)$. 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 10. *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 9. 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 7 and Theorem 9 $\mathcal{R}_i$ 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:

$$
\begin{aligned}
\mathfrak{h}_{i+1} &\coloneqq \mathfrak{h}_{i+1} \cdot \widetilde{\mathfrak{h}}_i & \widetilde{t}_{i+1} &\coloneqq t_i \cdot \widetilde{t}_{i+1} \\
\widetilde{\alpha}_{i+1} &\coloneqq \alpha_i \cdot \widetilde{\alpha}_{i+1} & \widetilde{\mathcal{R}}_{i+1} &\coloneqq \mathcal{R}_{i+1} \otimes \widetilde{\mathcal{R}}_i \\
\widetilde{\alpha}_{i+1} &\coloneqq \alpha_{i+1} \cdot \widetilde{\alpha}_i & \widetilde{\mathbf{b}}_{i+1} &\coloneqq \mathbf{b}_{i+1} \otimes \widetilde{\mathbf{b}}_i
\end{aligned}
$$

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. □

6.4 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 Section 6.3, 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 need a 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. A precise statement is given in Lemma 8.

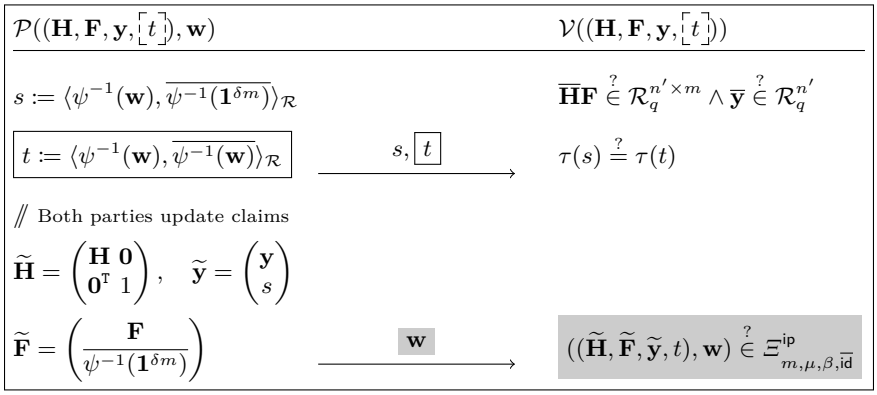

Fig. 9. Protocol $\boxed{\Pi^{\mathsf{lin\text{-}bin}}_{\tau}}$ or $\boxed{\Pi^{\mathsf{ip\text{-}bin}}_{\tau}}$, a reduction from $\boxed{\Xi^{\mathsf{lin}}_{m,\mu,\beta} \cap \Xi^{\mathsf{bin}}_{m}}$ or $\boxed{\Xi^{\mathsf{ip}}_{m,\mu,\beta,\overline{\mathsf{id}}} \cap \Xi^{\mathsf{bin}}_{m}}$ to $\Xi^{\mathsf{ip}}_{m,\mu,\beta,\overline{\mathsf{id}}}$. 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.

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

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

In Fig. 9, 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 11, whose proof relies on Lemma 8 stated immediately after.

Theorem 11. *Let $\mathcal{R} = \mathcal{O}_{\mathcal{K}_{\mathfrak{g}}} \otimes \mathcal{O}_{\mathcal{K}^+_{p_0}} \otimes \ldots \otimes \mathcal{O}_{\mathcal{K}^+_{p_{k-1}}}$, $\mathfrak{g} = 2^d$ for some $d \in \mathbb{N}$, and $\mathfrak{f}_0,\ldots,\mathfrak{f}_{k-1}$ distinct odd primes. Let 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 $\Xi^{\mathsf{lin}}_{m,\mu,\beta} \cap \Xi^{\mathsf{bin}}_{m}$ (resp. $\Xi^{\mathsf{ip}}_{m,\mu,\beta,\overline{\mathsf{id}}} \cap \Xi^{\mathsf{bin}}_{m}$) over $\mathcal{R}$ to $\Xi^{\mathsf{ip}}_{m,\mu,\beta,\overline{\mathsf{id}}}$ over $\mathcal{O}_{\mathcal{K}}$. There exists a constant $c_{\mathfrak{f}}$ such that it is relaxed knowledge sound from $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{m,\mu,\beta} \cap \Xi^{\mathsf{bin}}_{m}$ (resp. $\Xi^{\mathsf{ip}\vee\mathsf{sis}}_{m,\mu,\beta,\overline{\mathsf{id}}} \cap \Xi^{\mathsf{bin}}_{m}$) over $\mathcal{R}$ to $\Xi^{\mathsf{ip}}_{m,\mu,\beta,\overline{\mathsf{id}}}$ over $\mathcal{O}_{\mathcal{K}}$ if $2^k c_{\mathfrak{f}} \cdot \varphi^{5/2} \cdot \beta \leq \beta^{\mathsf{sis}}$.*

Proof. For perfect completeness, consider $((\mathbf{H},\mathbf{F},\mathbf{y}),\mathbf{w}) \in \Xi^{\mathsf{lin}}_{m,\mu,\beta} \cap \Xi^{\mathsf{bin}}_{m}$ over $\mathcal{R}$. We have $\psi(\mathbf{w}) \in \{0,1\}^{m\delta}$. By Proposition 2 and the discussion immediately after, it holds that $\langle\mathbf{w},\mathbf{1}^{m\delta}\rangle_{\mathbb{Z}} = \langle\mathbf{w},\mathbf{w}\rangle_{\mathbb{Z}}$. Since τ is an inner-product embedding over $\mathcal{R}$, it holds that

$$\begin{aligned}\tau(s) &= \tau(\langle\psi^{-1}(\mathbf{w}),\overline{\psi^{-1}(\mathbf{1}^{\delta m})}\rangle_{\mathcal{R}}) = \langle\mathbf{w},\mathbf{1}^{m\delta}\rangle_{\mathbb{Z}}, \text{ and}\\ \tau(t) &= \tau(\langle\psi^{-1}(\mathbf{w}),\overline{\psi^{-1}(\mathbf{w})}\rangle_{\mathcal{R}}) = \langle\mathbf{w},\mathbf{w}\rangle_{\mathbb{Z}},\end{aligned}$$

and thus $\tau(s) = \tau(t)$. Furthermore, since $((\mathbf{H},\mathbf{F},\mathbf{y}),\mathbf{w}) \in \Xi^{\mathsf{lin}}_{m,\mu,\beta}$, we have $s = \langle\psi^{-1}(\mathbf{w}),\overline{\psi^{-1}(\mathbf{1}^{\delta m})}\rangle_{\mathcal{R}}$, and $t = \langle\psi^{-1}(\mathbf{w}),\overline{\psi^{-1}(\mathbf{w})}\rangle_{\mathcal{R}}$. Therefore, $((\widetilde{\mathbf{H}},\widetilde{\mathbf{F}},\widetilde{\mathbf{y}},t),\mathbf{w}) \stackrel{?}{\in} \Xi^{\mathsf{ip}}_{m,\mu,\beta,\overline{\mathsf{id}}}$ over $\mathcal{R}$. Since $\mathcal{R} \subseteq \mathcal{O}_{\mathcal{K}}$, the claim follows.

For perfect relaxed knowledge soundness, suppose that $((\widetilde{\mathbf{H}},\widetilde{\mathbf{F}},\widetilde{\mathbf{y}},t),\mathbf{w}) \stackrel{?}{\in} \Xi^{\mathsf{ip}}_{m,\mu,\beta,\overline{\mathsf{id}}}$ over $\mathcal{O}_{\mathcal{K}}$. If $\mathbf{w} \in \mathcal{R}^m$, hen we have $s = \langle\psi^{-1}(\mathbf{w}),\overline{\psi^{-1}(\mathbf{1}^{\delta m})}\rangle_{\mathcal{R}}$ and $t = \langle\psi^{-1}(\mathbf{w}),\overline{\psi^{-1}(\mathbf{w})}\rangle_{\mathcal{R}}$. Since $\tau(s) = \tau(t)$, reversing the above argument gives $\psi(\mathbf{w}) \in \{0,1\}^{m\delta}$. Thus $((\mathbf{H},\mathbf{F},\mathbf{y}),\mathbf{w}) \in \Xi^{\mathsf{lin}}_{m,\mu,\beta} \cap \Xi^{\mathsf{bin}}_{m}$ over $\mathcal{R}$ as desired.

If $\mathbf{w} \in \mathcal{O}^m_{\mathcal{K}} \setminus \mathcal{R}^m$, then we can express $\mathbf{w}$ 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 8, we have $\|\sigma(\hat{\mathbf{w}})\|_2 \leq c_{\mathfrak{f}} \cdot \sqrt{m} \cdot \varphi^{5/2} \cdot \beta \leq \beta^{\mathsf{sis}}$, 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 8. *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 $\bigotimes_{i\in[k]}(1, \zeta_{\mathfrak{f}_i})$. There exists a constant $c_{\mathfrak{f}}$ (dependent only on $\mathfrak{f}$), such that if $\|\sigma(\mathbf{x})\|_2 \le \beta$, then each $\mathcal{R}$-coefficient $\hat{\mathbf{x}}$ of $\mathbf{x}$ satisfies $\|\sigma(\hat{\mathbf{x}})\|_2 \le \beta'$, where $\beta' \le 2^k c_{\mathfrak{f}} \cdot \varphi^{5/2} \cdot \beta$.

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

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

where $\widetilde{x}_{i,j} \in \mathbb{Z}^m$. Consider $\|\sigma(x_i)\|_2 = \beta_i$, by applying the same argument as in [DPSZ11, Theorem 7] but with the basis $\mathbf{b}$, we conclude that there exists a constant $c_{\mathfrak{f}}$ (dependent only on $\mathfrak{f}$) such that $\|\psi(x_i)\|_\infty \le c_{\mathfrak{f}} \beta_i$. Corollary 2 discusses the contant $c_{\mathfrak{f}}$.

Let $\hat{x}_{i,\ell} = \sum_{j\in[\varphi]} \widetilde{x}_{i,j}(s_{j,\ell} + t_{j,\ell}\zeta)$. Then,

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

We observe that as $\|\psi(\widetilde{x}_{i,j})\|_\infty \le c_{\mathfrak{f}}\beta_i$, then $\|\psi(\hat{x}_{i,\ell})\|_\infty \le 2\varphi c_{\mathfrak{f}}\beta_i$ and $\|\sigma(\hat{x}_{i,\ell})\|_2 \le 2\varphi^{5/2} c_{\mathfrak{f}}\beta_i$ due to the norm conversion.

Eventually, consider the norm of $\widetilde{\mathbf{x}}_j$, i.e.

$$\|\sigma(\widetilde{\mathbf{x}}_j)\|_2 \le \sqrt{\sum_{i\in[m]} \left(\beta_i 2\varphi^{5/2} c_{\mathfrak{f}}\right)^2} = 2\varphi^{5/2} c_{\mathfrak{f}} \sqrt{\sum_{i\in[m]} \beta_i^2} = 2\varphi^{5/2} c_{\mathfrak{f}} \|\sigma(\mathbf{x})\|_2 \le 2\varphi^{5/2} c_{\mathfrak{f}} \beta.$$

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:

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

Let $\hat{x}_{i,\ell^{(1)},\ell^{(2)}} = \sum_{j^{(1)} \in [\varphi^{(1)}], j^{(2)} \in [\varphi^{(2)}]} \widetilde{x}_{i,j^{(1)},j^{(2)}} (s^{(1)}_{j^{(1)},\ell^{(1)}} + t^{(1)}_{j^{(1)},\ell^{(1)}} \cdot \zeta^{(1)})(s^{(2)}_{j^{(2)},\ell^{(2)}} + t^{(2)}_{j^{(2)},\ell^{(2)}} \cdot \zeta^{(2)})$. Then, We observe that as $\left\| \psi\left(\widetilde{x}_{i,j^{(1)},j^{(2)}}\right) \right\|_\infty \leq c_{\mathfrak{f}} \beta_i$, then $\left\| \psi\left(\hat{x}_{i,\ell^{(1)},\ell^{(2)}}\right) \right\|_\infty \leq 4\varphi c_{\mathfrak{f}} \beta_i$ and continue the reasoning as in the base case. Clearly, the argument extends for terson rings of more than two prime rings.

□

7 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 $\mathcal{R}$.

7.1 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 $\mathcal{R} = \mathbb{Z}[\zeta_{\mathfrak{f}}]$ 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 $(\mathbf{x}, \mathbf{y}) = (\mathsf{CRT}_p(\widehat{\mathbf{x}}), \mathsf{CRT}_p(\widehat{\mathbf{y}})) \bmod p$).

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

$$\widetilde{\mathbf{x}} = \psi^{-1}(\mathsf{CRT}_p(\widehat{\mathbf{x}})) + p \cdot \mathbf{r}, \tag{7}$$
$$\widetilde{\mathbf{y}} = \psi^{-1}(\mathsf{CRT}_p(\widehat{\mathbf{y}})) + p \cdot \widetilde{\mathbf{s}}. \tag{8}$$

– 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. (7) and (8) hold.

From Eqs. (7) and (8), 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. (7) and (8), using tools developed in Section 5.

7.3 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 12, 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 a few succinct linear combinations of automorphisms in $\mathsf{Gal}(\mathcal{K}/\mathbb{Q})$ with $\mathcal{R}$ coefficients. To prove this theorem, we will make use of two elementary lemmas. In Lemma 11, we prove an elementary fact that, if L/K is a Galois extension, then any K-linear map $f : L \to L$ can be expressed as an L-linear combination of $\mathsf{Gal}(L/K)$. Then, in Lemma 12, we prove an analogous lemma for $\mathcal{O}_K/p\mathcal{O}_K$-linear map $f : \mathcal{O}_L/p\mathcal{O}_L \to \mathcal{O}_L/p\mathcal{O}_L$, if L is cyclotomic and has conductor less than p, where p is a rational prime. Using these two results, we arrive at the following theorem.

Theorem 12. *Let $\mathcal{R}$ be a cyclotomic ring*[18] *with a w-smooth conductor $\mathfrak{f}$. Then, the transformation $\psi^{-1} \circ \mathsf{CRT}_p : \mathcal{R} \to \mathcal{R}$ can be expressed as the succinct composition of $\mathcal{R}$-linear combinations of at most w automorphisms over $\mathcal{R}$. Formally, there exists $t \in O(\log \mathfrak{f})$, $h_i \leq w$, $s_{i,j} \in \mathcal{R}$, 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, $\mathcal{K}/\mathbb{Q}$ can be decomposed into a tower of $t \leq O(\log \mathfrak{f})$ Galois extensions where each step of the extension is of degree at most $h_i \leq w$. Let L/K denote the i-th step of the tower of extensions. Correspondingly, the map $(\psi^{-1} \circ \mathsf{CRT}_p \bmod p)$ can be decomposed as a composition

$$\psi^{-1} \circ \mathsf{CRT}_p = \bigcirc_{i\in[t]} \widehat{f}_i$$

where $\widehat{f}_i$ is obtained by lifting an $\mathcal{O}_K/p\mathcal{O}_K$-linear map $f_i : \mathcal{O}_L/p\mathcal{O}_L \to \mathcal{O}_L/p\mathcal{O}_L$ to $\mathcal{O}_\mathcal{K}/p\mathcal{O}_\mathcal{K}$. By Lemma 12, f_i can be expressed as an $\mathcal{O}_L/p\mathcal{O}_L$-linear combination of $\mathsf{Gal}(L/K)$ which contains at most $h_i \leq w$ elements. Correspondingly, $\widehat{f}_i$ can be expressed as an $\mathcal{O}_\mathcal{K}/p\mathcal{O}_\mathcal{K}$-linear combination of $\mathsf{Gal}(\mathcal{K}/\mathbb{Q})$ which contains at most $h_i \leq w$ elements. The theorem thus follows. □

7.4 $\Pi^{\mathsf{eip}+}$: Extended Inner-product relation

The Ξ^{ip} relation defined in Section 5.7 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 7, 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.

$$\Xi^{\mathsf{eip}\vee\mathsf{sis}}_{m,n^{\mathsf{out}},\beta,\beta^{\mathsf{sis}}} := \left\{ \begin{array}{l} ((\iota_{\mathsf{ip}}, \iota_{\mathsf{ip\text{-}in}}, \mathbf{H}, \mathbf{F}, \mathbf{y}, \mathbf{t}), \mathbf{w} = (\mathbf{w}_k)_{k\in[n^{\mathsf{blk}}]}) : \\ \quad \mathbf{H} \in \mathcal{R}_q^{n^{\mathsf{out}}\times n};\ \mathbf{F} \in \mathcal{R}_q^{n\times(d^{\otimes\mu}\cdot n^{\mathsf{blk}})} \subseteq \mathcal{R}_q^{n\times m};\ \mathbf{y} \in \mathcal{R}_q^{n^{\mathsf{out}}};\ \mathbf{t} \in \mathcal{R}_q^{n^{\mathsf{ip}}} \\ \quad \left\{ \begin{array}{l} \iota_{\mathsf{ip}} \in \{1,-1\}^{[n^{\mathsf{ip}}]};\ \iota_{\mathsf{ip\text{-}in}} \in [n^{\mathsf{blk}}]^{[n^{\mathsf{ip}}]} \\ \|\mathbf{w}\| \leq \beta;\ \mathbf{HFw} = \mathbf{Hy} \bmod q \\ \langle \mathbf{w}_{\iota_{\mathsf{ip\text{-}in}}(k)}, \alpha_{\iota_{\mathsf{ip}}(k)}(\mathbf{w}_{\iota_{\mathsf{ip\text{-}in}}(k)})\rangle = t_k \bmod q \quad \forall k \in [n^{\mathsf{ip}}] \end{array} \right\} \text{ or } \left\{ \begin{array}{c} \|\mathbf{w}\| \leq \beta^{\mathsf{sis}} \\ \overline{\mathbf{H}}\mathbf{F}\mathbf{w} = \mathbf{0}_{\overline{n}} \bmod q \end{array} \right\} \end{array} \right\}.$$

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

$$\langle \mathbf{w}_{\iota_{\mathsf{ip\text{-}in}}(k)}, \alpha_{\iota_{\mathsf{ip}}(k)}(\mathbf{w}_{\iota_{\mathsf{ip\text{-}in}}(k)})\rangle = t_k \bmod q \quad \forall k \in [n^{\mathsf{ip}}].$$

Recall that $\Pi^{\mathsf{ip}+}$ is a reduction of knowledge from Ξ^{ip} to Ξ^{lin}. Similarly, we can also construct a reduction of knowledge $\Pi^{\mathsf{eip}+} : \Xi^{\mathsf{eip}} \to \Xi^{\mathsf{lin}}$, by modifying the Π^{ip} (Fig. 8) part of $\Pi^{\mathsf{ip}+}$ into Π^{eip}.

Since Π^{ip} and $\Pi^{\mathsf{ip}+}$ are already quite notation heavy, and their generalisations to Π^{eip} and $\Pi^{\mathsf{eip}+}$ are 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 $\hat{\mathbf{v}}_j$) is replaced by $\hat{\mathbf{w}}_j := \mathbf{e}_{\iota_{\mathsf{ip\text{-}in}}(j)} \otimes \mathbf{w}_{\iota_{\mathsf{ip\text{-}in}}(j)}$.

[18] 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.

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

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

(ii) The new claims both parties compute are:

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

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

where $\mathbf{y}_{\hat{\mathbf{E}}_j} \coloneqq \hat{\mathbf{E}}_j\mathbf{w}$ and $\mathbf{y}'_{\hat{\mathbf{E}}_j,i} \coloneqq \hat{\mathbf{E}}_j\hat{\mathbf{v}}_{j,i}$.

(iii) Further, the verifications are replaced by

$$((\widetilde{\mathbf{H}}, \widetilde{\mathbf{F}}, \widetilde{\mathbf{y}}), \mathbf{w}) \overset{?}{\in} \Xi^{\mathsf{lin}}_{m,\beta_0}$$

$$\left((\widetilde{\mathbf{H}}, \widetilde{\mathbf{F}}, \widetilde{\mathbf{y}}'_i), \sum_{j\in[n^{\mathsf{ip}}]} \mathbf{e}_{\iota_{\mathsf{ip\text{-}in}}(j)} \otimes \hat{\mathbf{v}}_{j,i}\right)_{(i)\in[\ell]} \overset{?}{\in} \Xi^{\mathsf{lin}}_{m,\beta_0}$$

for the same parameters as in Fig. 8.

In sum, we obtain the reduction $\Xi^{\mathsf{eip}}_{n^{\mathsf{out}},\beta} \xrightarrow{\Pi^{\mathsf{eip}}} (\Xi^{\mathsf{lin}}_{n^{\mathsf{out}}+3n^{\mathsf{ip}},\beta})^{n^{\mathsf{ip}}\cdot(\ell+1)}$ for the same parameters as the Π^{ip} from Section 5.7. Next, we outline the full protocol $\Pi^{\mathsf{eip}+}$ analogous to $\Pi^{\mathsf{ip}+}$.

The full protocol The full protocol $\Pi^{\mathsf{eip}+}$ reduces the $n^{\mathsf{ip}}\cdot(\ell+1)$ statements after Π^{eip} back to a single one with the minimal number of rows, by applying Π^{batch} and Π^{fold}. That is, $\Pi^{\mathsf{eip}+}$ works as follows:

(1) $\Xi^{\mathsf{ip}}_{n^{\mathsf{out}},\beta} \xrightarrow{\Pi^{\mathsf{eip}}} (\Xi^{\mathsf{lin}}_{n^{\mathsf{out}}+3\cdot n^{\mathsf{ip}},\beta})^{n^{\mathsf{ip}}\cdot(\ell+1)}$ (Reduce to claims in $(\Xi^{\mathsf{lin}})^{\ell+1}$.)

(2) $(\Xi^{\mathsf{lin}}_{n^{\mathsf{out}}+3\cdot n^{\mathsf{ip}},\beta})^{n^{\mathsf{ip}}\cdot(\ell+1)} \xrightarrow{\Pi^{\mathsf{batch}}} (\Xi^{\mathsf{lin}}_{\overline{n}+1,\beta})^{n^{\mathsf{ip}}\cdot(\ell+1)}$ (Reduce number of rows again.)

(3) $(\Xi^{\mathsf{lin}}_{\overline{n}+1,\beta})^{n^{\mathsf{ip}}\cdot(\ell+1)} \xrightarrow{\Pi^{\mathsf{fold}}} \Xi^{\mathsf{lin}}_{\overline{n}+1,n^{\mathsf{ip}}\cdot(\ell+1)\gamma\beta}$ (Reduce to claim $\beta' = (n^{\mathsf{ip}}\cdot(\ell+1))\gamma\beta$.)

We do not provide explicit proof of Theorem 13 below as it completely analogous to Theorem 6. We denote by $\Xi^{\mathsf{eip}\vee\mathsf{sis}}$ the usual relaxed relation.

Theorem 13 (Security of $\Pi^{\mathsf{eip}+}$). *The protocol $\Pi^{\mathsf{eip}+}$ is a reduction of knowledge $\Xi^{\mathsf{eip}}_{m,n^{\mathsf{out}},\beta_0}$ to $(\Xi^{\mathsf{lin}}_{m,n^{\mathsf{out}}+3\cdot n^{\mathsf{ip}},\beta'_3=n^{\mathsf{ip}}\cdot(\ell+1)\beta_0})^{n^{\mathsf{ip}}\cdot(\ell+1)}$, where $\ell \geq \log_{b_{\mathsf{ip}}}(2\beta^2+1)$ for $b_{\mathsf{ip}} \leq 2\beta/(\sqrt{m}\varphi^{3/2})$ for $b_{\mathsf{ip}}, \ell \in \mathbb{N}$. It is perfectly correct. It is relaxed knowledge sound from $\Xi^{\mathsf{eip}\vee\mathsf{sis}}_{m,n^{\mathsf{out}},\beta'_0}$ to $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{m,n^{\mathsf{out}}+3\cdot n^{\mathsf{ip}},\beta'_3}$ if $\beta'_0 \leq \beta^{\mathsf{sis}}$ where $\beta'_0 = 4\theta\beta'_3$. The knowledge error is $n^{\mathsf{ip}}\cdot(\ell+1)\cdot(|\mathcal{C}_{\mathcal{R}}|^{-1} + 4|\mathcal{C}_{\mathcal{R}_q}|^{-1}) + 2m|\mathcal{C}_{\mathcal{R}_q}|^{-1}$.*

7.5 Π^{aut}: Automorphism Check

For our CRT-based protocol (Section 7), we need to efficiently check automorphism relations between different blocks of the witness. Formally, we define the automorphism check relation below.

$$\Xi^{\mathsf{aut}\vee\mathsf{sis}}_{m,n^{\mathsf{out}},\beta,\beta^{\mathsf{sis}}} \coloneqq \left\{ \begin{array}{l} \big((\iota_{\mathsf{ip}}, \iota_{\mathsf{ip\text{-}in}}, \iota_{\mathsf{aut}}, \iota_{\mathsf{aut\text{-}in}}, \iota_{\mathsf{aut\text{-}out}}, \mathbf{H}, \mathbf{F}, \mathbf{y}, \mathbf{t}),\ \mathbf{w} = (\mathbf{w}_k)_{k\in[n^{\mathsf{blk}}]}\big) : \\ \quad \mathbf{H} \in \mathcal{R}_q^{n^{\mathsf{out}}\times n};\ \mathbf{F} \in \mathcal{R}_q^{n\times(d^{\otimes\mu}\cdot n^{\mathsf{blk}})} \subseteq \mathcal{R}_q^{n\times m};\ \mathbf{y} \in \mathcal{R}_q^{n};\ \mathbf{t} \in \mathcal{R}_q^{n^{\mathsf{ip}}} \\ \quad \left\{ \begin{array}{l} \iota_{\mathsf{ip}} \in \{1,-1\}^{[n^{\mathsf{ip}}]};\ \iota_{\mathsf{ip\text{-}in}} \in [n^{\mathsf{blk}}]^{[n^{\mathsf{ip}}]} \\ \iota_{\mathsf{aut}} \in [\varphi]^{[n^{\mathsf{aut}}]};\ \iota_{\mathsf{aut\text{-}in}}, \iota_{\mathsf{aut\text{-}out}} \in [n^{\mathsf{blk}}]^{[n^{\mathsf{aut}}]} \\ \|\mathbf{w}\| \leq \beta;\ \mathbf{H}\cdot\mathbf{F}\mathbf{w} = \mathbf{H}\mathbf{y} \bmod q \\ \langle \mathbf{w}_{\iota_{\mathsf{ip\text{-}in}}(k)}, \alpha_{\iota_{\mathsf{ip}}(k)}(\mathbf{w}_{\iota_{\mathsf{ip\text{-}in}}(k)})\rangle = t_k \bmod q \quad \forall k \in [n^{\mathsf{ip}}] \\ \alpha_{\iota_{\mathsf{aut}}(k)}(\mathbf{w}_{\iota_{\mathsf{aut\text{-}in}}(k)}) = \mathbf{w}_{\iota_{\mathsf{aut\text{-}out}}(k)} \bmod q \quad \forall k \in [n^{\mathsf{aut}}] \end{array} \right\} \\ \quad \text{or} \quad \left\{ \begin{array}{l} \|\mathbf{w}\| \leq \beta^{\mathsf{sis}} \\ \overline{\mathbf{H}}\mathbf{F}\mathbf{w} = \mathbf{0}_{\overline{n}} \bmod q \end{array} \right\} \end{array} \right\}.$$

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. 10.

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

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

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

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

This reduces the automorphism check to verifying the linear relations

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

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. The proof is omitted due to its similarity with the previous proofs in this section.

Theorem 14 (Security of Π^{aut}). *The protocol Π^{aut} is a perfectly correct reduction of knowledge from $\Xi^{\mathsf{aut}}_{m,n^{\mathsf{out}},\beta}$ to $\Xi^{\mathsf{eip}}_{m,n^{\mathsf{out}}+2n^{\mathsf{aut}},\beta}$. It is relaxed knowledge sound from $\Xi^{\mathsf{aut}\vee\mathsf{sis}}_{m,n^{\mathsf{out}},\beta'_0}$ to $\Xi^{\mathsf{aut}\vee\mathsf{sis}}_{m,n^{\mathsf{out}}+2n^{\mathsf{aut}},\beta'_1}$ if $2\beta_1 \leq \beta^{\mathsf{sis}}$, where $\beta'_0 = \beta'_1$. The knowledge error is $n^{\mathsf{aut}}/|\mathcal{C}_{\mathcal{R}_q}|$.*

7.6 Reducing CRT-based Binariness Check to Automorphism Check

Equipped with Theorems 12 and 14, in Fig. 11, 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. 11 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:

(i) Compute $\widehat{\mathbf{w}}_0 := \mathsf{CRT}_p^{-1}(\psi(\mathbf{w}))$ and let $\mathbf{r}_0 := \mathbf{0}$.
(ii) Compute $y' := \langle \widehat{\mathbf{w}}_0, \mathsf{CRT}^{-1}(\mathbf{1}^{\delta m})\rangle_{\mathcal{R}}$, $\widetilde{t}_0 := \langle \widehat{\mathbf{w}}_0, \widehat{\mathbf{w}}_0\rangle_{\mathcal{R}}$, and $\widetilde{t}_1 := \langle \mathbf{w}, \overline{\mathbf{w}}\rangle_{\mathcal{R}}$.
(iii) For each $i \in [t]$: (i) Compute the i-th step of the CRT decomposition as described in Theorem 12, i.e. $\mathbf{w}_{i,j} := \alpha_{i,j}(\widehat{\mathbf{w}}_i)$ for all $j \in [h_i]$. (ii) Compute the p-ary representative $\widehat{\mathbf{w}}_{i+1} := \sum_{j\in[h_i]} s_{i,j}\mathbf{w}_{i,j} \bmod p$. (iii) Find the quotient $\mathbf{r}_{i+1} \in \mathcal{R}^m$ satisfying $\widehat{\mathbf{w}}_{i+1} - \sum_{j\in[h_i]} s_{i,j} \cdot \mathbf{w}_{i,j} + p \cdot \mathbf{r}_{i+1} = \mathbf{0}^m \pmod q$.
(iv) Concatenate all intermediate witnesses as

$$\widetilde{\mathbf{w}}^{\mathsf{T}} := ((\mathbf{w}_{i,j}^{\mathsf{T}})_{(i,j)\in[t,h_i]}, (\widehat{\mathbf{w}}_i^{\mathsf{T}}, \mathbf{r}_i^{\mathsf{T}})_{i\in[t+1]}) \in \mathcal{R}^{\widetilde{m}}$$

where $\widetilde{m} := m \cdot n^{\mathsf{blk}}$ and $n^{\mathsf{blk}} := w \log \mathfrak{f} + 2t + 2$.[19]
(v) Parse the witness as a block vector $\widetilde{\mathbf{w}} = (\widetilde{\mathbf{w}}_k)_{k\in[n^{\mathsf{blk}}]}$ where $\widetilde{\mathbf{w}}_k \in \mathcal{R}^m$.
(vi) Concatenate all linear map images as $\widetilde{\mathbf{y}}^{\mathsf{T}} := (\mathbf{y}^{\mathsf{T}}, \mathbf{0}^{tm}, c) \in \mathcal{R}_q^{\widetilde{n}}$ where $\widetilde{n} := n + tm + 1$.
(vii) Concatenate all inner product images as $\widetilde{\mathbf{t}}^{\mathsf{T}} := (\widetilde{t}_0, \widetilde{t}_1) \in \mathcal{R}^2$.

[19] 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 $h_i \leq w$ for all i and $t \leq \log \mathfrak{f}$, we pad $\widetilde{\mathbf{w}}$ to dimension $\widetilde{m} = m \cdot (w \log \mathfrak{f} + 2t + 2)$ for simplicity.

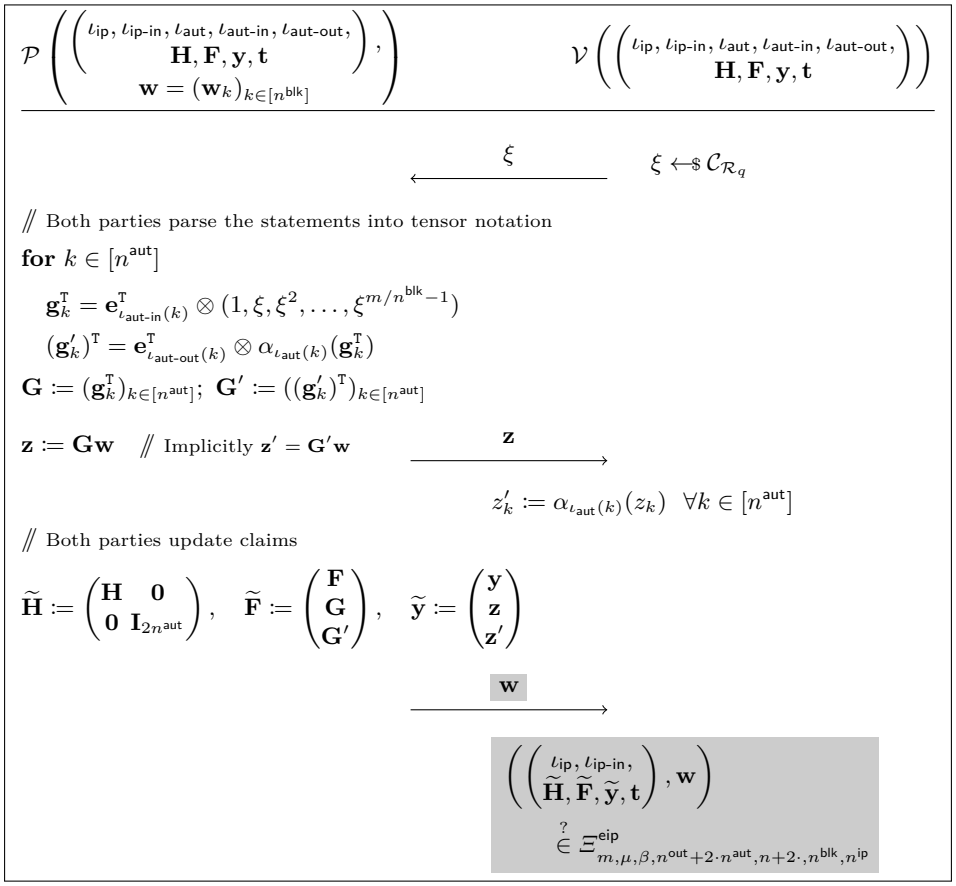

Fig. 10. Protocol Π^{aut}, a reduction of knowledge from $\Xi^{\mathsf{aut}}_{m,\mu,\beta,n^{\mathsf{out}},n,n^{\mathsf{blk}},n^{\mathsf{ip}},n^{\mathsf{aut}}}$ to $\Xi^{\mathsf{eip}}_{m,\mu,\beta,n^{\mathsf{out}}+2\cdot n^{\mathsf{aut}},n+2\cdot,n^{\mathsf{blk}},n^{\mathsf{ip}}}$. Π^{aut} sends the marked parts only as a *proof* (but not *reduction*) of knowledge.

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

$$\begin{aligned}\mathbf{F}\widehat{\mathbf{w}}_t &= \mathbf{y} \pmod q,\\ \langle \mathsf{CRT}^{-1}(\mathbf{1}^{\delta m}), \widehat{\mathbf{w}}_0\rangle_{\mathcal{R}} &= y' \pmod q,\\ \widehat{\mathbf{w}}_{i+1} - \sum_{j\in[h_i]} s_{i,j}\cdot\mathbf{w}_{i,j} + p\cdot\mathbf{r}_{i+1} &= \mathbf{0}^m \pmod q\ \ \forall i\in[t].\end{aligned}$$

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

(x) Let $n^{\mathsf{ip}} \coloneqq 2$ 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

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

(xi) Let $n^{\mathsf{aut}} \coloneqq w\log\mathfrak{f}$ 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

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

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

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

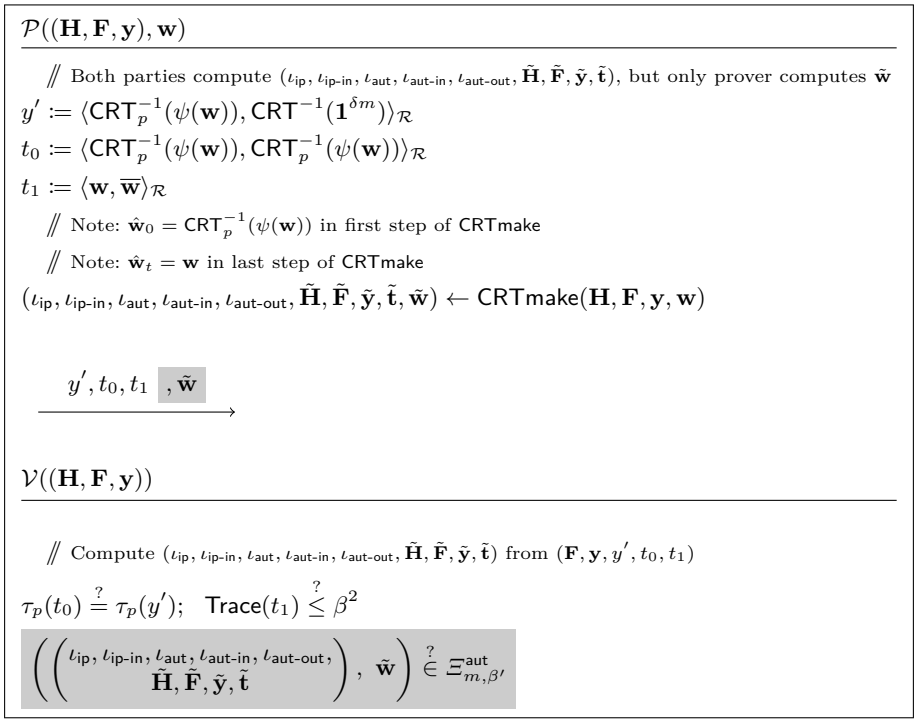

Fig. 11. Protocol $\Pi^{\mathsf{lin\text{-}bin}}_{\mathsf{crt},p}$, a reduction from $\Xi^{\mathsf{lin}} \cap \Xi^{\mathsf{bin}}$ to Ξ^{aut}, where β, β', and p are set as in Theorem 15. See the text for the definition of the subroutine CRTmake. The marked parts are only sent / checked when $\Pi^{\mathsf{lin\text{-}bin}}_{\mathsf{crt},p}$ is used as a proof of knowledge. As a reduction of knowledge, they are omitted.

Theorem 15. *The protocol $\Pi^{\mathsf{lin\text{-}bin}}_{\mathsf{crt},p}$ is a perfectly correct from $\Xi^{\mathsf{lin}}_{m,n^{\mathsf{out}},\beta} \cap \Xi^{\mathsf{bin}}_{m,n^{\mathsf{out}}}$ to $\Xi^{\mathsf{aut}}_{m,n^{\mathsf{out}},\beta'}$, and knowledge sound reduction of knowledge from $\Xi^{\mathsf{lin}\vee\mathsf{sis}}_{m,n^{\mathsf{out}},\beta} \cap \Xi^{\mathsf{bin}\vee\mathsf{sis}}_{m,n^{\mathsf{out}}}$ to $\Xi^{\mathsf{aut}\vee\mathsf{sis}}_{m,n^{\mathsf{out}},\beta'}$, where $n^{\mathsf{blk}} = w(1+t)+2t+2$, $n^{\mathsf{ip}} = 2$, $n^{\mathsf{aut}} = w\log\mathfrak{f}$, $\beta' = \frac{p}{4}\cdot w\cdot\gamma_{\mathcal{R}}\cdot\varphi^{3/2}\cdot\sqrt{m\cdot n^{\mathsf{blk}}}$, $\beta = \varphi^{3/2}\sqrt{m}$, $p/2 > c_{\mathfrak{f}}^2\cdot\beta^2\cdot m\cdot\varphi$, $c_{\mathfrak{f}}$ is a constant from Corollary 2, $n^{\mathsf{out}\prime} = n^{\mathsf{out}} + tm + 1$ and conductor $\mathfrak{f}$ of the ring is w-smooth.*

Proof. Completeness. For perfect completeness, we consider the statement $((\mathbf{H},\mathbf{F},\mathbf{y}),\mathbf{w}) \in \Xi^{\mathsf{lin}}_{m,\beta} \cap \Xi^{\mathsf{bin}}_{m}$. Clearly, we have $\psi(\mathbf{w}) \in \{0,1\}^{m\delta}$. Therefore,

$$\langle \psi(\mathbf{w}), \mathbf{1}^{m\delta}\rangle_{\mathbb{Z}} = \langle \psi(\mathbf{w}), \mathbf{w}\rangle_{\mathbb{Z}}$$

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

$$\tau_p(y') = \langle \psi(\mathbf{w}), \mathbf{1}^{m\delta}\rangle_{\mathbb{Z}} \bmod p, \text{ and}$$
$$\tau_p(t_0) = \langle \psi(\mathbf{w}), \psi(\mathbf{w})\rangle_{\mathbb{Z}} \bmod p,$$

As a consequence, $\tau_p(t_0) = \tau_p(y')$ and the first check of the verifier passes.

Further, as $\|\psi(\mathbf{w})\|_\infty \leq 1$, and by Corollary 2 $\|\sigma(\mathbf{w})\|_2 \leq \varphi^{3/2}\sqrt{m} = \beta$. 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 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.

In the first case, due to reduction modulo p,

$$\left\|\psi\left((\widehat{\mathbf{w}}_i)_{i\in[t+1]}\right)\right\|_\infty \leq p/2.$$

After translating the norm, we consider the canonical 2-norm of each element of $(\widehat{\mathbf{w}}_i)_{i\in[t+1]}$.

$$\|\sigma(\widehat{w})\|_2 \leq p/2 \cdot \varphi^{3/2} \cdot \sqrt{m\cdot(1+t)} \qquad \forall\ \widehat{w}\ \in (\widehat{\mathbf{w}}_i)_{i\in[t+1]}.$$

In the second case,

$$\|\sigma(w)\|_2 \leq p/2 \cdot \varphi^{3/2} \cdot \sqrt{m\cdot(1+t)} \qquad \forall\ w\ \in (\mathbf{w}_{i,j})_{(i,j)\in[t,h_j]}.$$

as $\mathbf{w}_{i,j} = \alpha_{i,j}(\widehat{\mathbf{w}}_i)$ and the automorphisms $\alpha_{i,j}$ do not impact the canonical norm of a ring element.
In the third case,

$$\|\psi(r)\|_\infty \leq \left(\frac{p}{2}\cdot\frac{p}{2}\cdot w\cdot\gamma_{\mathcal{R}}\right)/p = \frac{p}{4}\cdot w\cdot\gamma_{\mathcal{R}} \qquad \forall\ r\ \in (\mathbf{r}_i)_{i\in[t+1]}$$

and after translating norms,

$$\|\sigma(r)\|_2 \leq \frac{p}{4}\cdot w\cdot\gamma_{\mathcal{R}}\cdot\varphi^{3/2} \qquad \forall\ r\ \in (\mathbf{r}_i)_{i\in[t+1]}.$$

To sum up, canonical 2-norm of individual elements of $\widetilde{\mathbf{w}}$ is upper-bounded by

$$\|\sigma(\widetilde{w})\|_2 \leq \frac{p}{4}\cdot w\cdot\gamma_{\mathcal{R}}\cdot\varphi^{3/2} \qquad \forall\ \widetilde{w}\ \in \widetilde{\mathbf{w}}$$

as $w \geq 2$. Last, we observe that $|(\widehat{\mathbf{w}}_i)_{i\in[t+1]}| = m\cdot(t+1)$, $|(\mathbf{w}_{i,j})_{(i,j)\in[t,h_j]}| \leq m\cdot(t+1)\cdot w$ and $|(\mathbf{r}_i)_{i\in[t+1]}| = m\cdot(t+1)$. As a result $|\widetilde{\mathbf{w}}| \leq m\cdot(w(1+t)+2t+2) = n^{\mathsf{blk}}$ so we derive

$$\|\sigma(\widetilde{\mathbf{w}})\|_2 \leq \frac{p}{4}\cdot w\cdot\gamma_{\mathcal{R}}\cdot\varphi^{3/2}\cdot\sqrt{m\cdot n^{\mathsf{blk}}} = \beta',$$

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 $\mathsf{CRTmake}$. We have

$$\tau_p(y') = \langle\psi(\mathbf{w}),\mathbf{1}^{m\delta}\rangle_{\mathbb{Z}} \bmod p \text{ and } \tau_p(t_0) = \langle\psi(\mathbf{w}),\psi(\mathbf{w})\rangle_{\mathbb{Z}} \bmod p.$$

Since $\tau_p(y') = \tau_p(t_0)$, we have

$$\langle\psi(\mathbf{w}),\psi(\mathbf{w})-\mathbf{1}^{m\delta}\rangle_{\mathbb{Z}} = 0 \bmod p.$$

Furthermore, $\mathsf{Trace}(t_1) \leq \beta^2$ implies that $\|\sigma(\mathbf{w})\|_2 \leq \beta$, thus by Corollary 2 $\|\psi(\mathbf{w})\|_\infty \leq c_{\mathfrak{f}}\cdot\beta$, where $c_{\mathfrak{f}}$ is dependent only on $\mathfrak{f}$. Therefore, as $c_{\mathfrak{f}}^2\cdot\beta^2\cdot m\cdot\varphi < p/2$, the inner product above holds without modulo p, and thus $\psi(\mathbf{w}) \in \{0,1\}^{m\delta}$. □

Remark 8. Theorem 15 trivially generalises to yield a reduction of knowledge from $\Xi^{\mathsf{ip}} \cap \Xi^{\mathsf{bin}}$ to Ξ^{aut}, analogous to Theorem 11. We omit the details for succinctness.

Remark 9. 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 10. 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 11. We discuss a possible optimisation which allows to pick smaller primes p. Recall that, for knowledge soundness, we required $p/2 > c_{\mathfrak{f}}^2 \cdot \beta^2 \cdot m \cdot \varphi$, 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 $\prod_{p \in P} p/2 > c_{\mathfrak{f}}^2 \cdot \beta^2 \cdot m \cdot \varphi$. If the conductor $\mathfrak{f}$ is smooth, it is possible to find many highly favourable primes (cf. Remark 10).

8 Parameter Selection

We propose concrete instantiations of our protocols for various values of m. For comparison with prior works, e.g. [BS23, AFLN23], we aim for 128-bit security. This corresponds to the root Hermite factor $\delta_{\mathsf{rhf}} \approx 1.0044$ (cf. Section 5.9).

8.1 Split-and-Fold with Norm Checks

We start with instantiating the split-and-fold with an intermediate norm check described in Section 5.8. We focus on the following simple goal: commit to a short vector $\mathbf{w} \in \mathbb{Z}_q^h$ of length h, such that $\|\psi(\mathbf{w})\|_\infty \leq \beta_{\mathsf{init}} = 2^5$ and prove knowledge of the commitment opening. To this end, we will pack $h = m \cdot \varphi$ integers into a vector $\mathbf{w} \in \mathcal{R}_q^m$ of m 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 $v, y \in \mathcal{R}_q$. When adapting this relation to the language of Ξ^{ip}, we would initially set $(n, n^{\mathsf{out}}) = (1, 1)$. Throughout the batching protocols, we set $\overline{n} = 1$. In our experiments, we hardwire $\mathfrak{f} = 5544$. Hence, $\varphi = 1440$ and $|\mathcal{C}_{\mathcal{R}}| = 504$. We fix $\ell = 2$ for the whole execution of the protocol. Then, given the norm bound β and ℓ, we can deduce the decomposition base b. Since in each iteration of the split-and-fold protocol, the norm β may change, then so can the base b. We note that (at least concretely) the current proof sizes are not optimal, reaching high orders of Megabytes. We highlight that we could achieve better sizes for larger moduli. However, conceptually this would be contradictory to the original intention of split-and-fold with norm checks; avoiding unnecessary stretch and large proof system modulus. We defer more fine-grained optimisation to a follow-up work.

Proof Composition. To further shrink communication, one could use standard proof composition, where instead of the verifier checking the verification conditions, the prover provides a 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].

witness length in $\mathbb{Z}$-elements	$\approx 2^{18}$	$\approx 2^{20}$	$\approx 2^{24}$
$\log q$	110	110	120
$\mathfrak{f}$	5544	5544	5544
m	2^{20}	2^{22}	2^{26}
$\|\psi(\cdot)\|_\infty$ of the witness	2^5	2^5	2^5
witness size	1080 MB	4320 MB	69120 MB
d	2	2	2
μ	2	3	4
# of repetitions	17	18	19
(unoptimized) proof size	258.4 MB	263.6 MB	458.6 MB

Table 2. Concrete parameters, together with proof sizes, for security level $\lambda = 128$.

Fiat-Shamir Transformation. As noted in [AFK22], transforming interactive proofs, which admit parallel repetition, to the non-interactive setting via Fiat-Shamir transformation incurs significant loss in the order of Q^μ, where Q is the number of random oracle queries made by an adversary. In the asymptotic sense, here we could simply set $d = O(\lambda)$, and thus $\mu = O(\frac{\log m}{\log \lambda})$ which is constant for $m = \mathsf{poly}(\lambda)$. Unfortunately, as seen in the experiments, the most optimal proof sizes are achieved for small values of d. In this section, we heuristically omit the security loss Q^μ and leave any potential improvements as future work.

Concrete parameters. In Table 2 we suggest concrete parameters with the estimated proof size. The results are obtained via a dedicated script[20] simulating protocol execution and measuring the communication cost.

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, Heidelberg, August 2022. 1, 45

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, Heidelberg, November 2022. 41

AFLN23. Martin R. Albrecht, Giacomo Fenzi, Oleksandra Lapiha, and Ngoc Khanh Nguyen. Slap: Succinct lattice-based polynomial commitments from standard assumptions. Cryptology ePrint Archive, Paper 2023/1469, 2023. https://eprint.iacr.org/2023/1469. 1, 2, 4, 40

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, Heidelberg. 2, 5, 10, 11

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, Heidelberg, October / November 2016. 1

[20] The script and the output are available at https://github.com/russell-lai/rok-paper-sissors-estimator/blob/camera-ready/rok-estimator.ipynb

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, Heidelberg, August 2023. 2, 4

BDGL15. Anja Becker, Léo Ducas, Nicolas Gama, and Thijs Laarhoven. New directions in nearest neighbor searching with applications to lattice sieving. Cryptology ePrint Archive, Report 2015/1128, 2015. https://eprint.iacr.org/2015/1128. 25

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. 8, 27, 28

BL17. Carsten Baum and Vadim Lyubashevsky. Simple amortized proofs of shortness for linear relations over polynomial rings. Cryptology ePrint Archive, Report 2017/759, 2017. https://eprint.iacr.org/2017/759. 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, Heidelberg, 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, Heidelberg, August 2023. 1, 2, 8, 32, 40

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, Heidelberg, August 2023. 1, 3, 4, 5, 9, 14, 40, 45

CMNW24. Valerio Cini, Giulio Malavolta, Ngoc Khanh Nguyen, and Hoeteck Wee. Polynomial commitments from lattices: Post-quantum security, fast verification and transparent setup. Cryptology ePrint Archive, Paper 2024/281, 2024. https://eprint.iacr.org/2024/281. 1, 4

DAFS24. Thomas Debris-Alazard, Pouria Fallahpour, and Damien Stehlé. Quantum oblivious lwe sampling and insecurity of standard model lattice-based snarks. Cryptology ePrint Archive, Paper 2024/030, 2024. https://eprint.iacr.org/2024/030. 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

DPSZ11. I. Damgard, V. Pastro, N.P. Smart, and S. Zakarias. Multiparty computation from somewhat homomorphic encryption. Cryptology ePrint Archive, Report 2011/535, 2011. https://eprint.iacr.org/2011/535. 31, 44

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, Heidelberg, August 2012. 44

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, 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, 10, 45, 46

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, Heidelberg, May / June 2022. 2

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, Heidelberg, August 2023. 9, 10, 45, 46

Len76. H. W. Lenstra. Euclidean number fields of large degree. *Inventiones mathematicae*, 38(3):237–254, 1976. 2, 5, 10, 11

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, Heidelberg, August 2023. 7

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, Heidelberg, 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, Heidelberg, August 2022. 1, 2, 3, 26, 29

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. 8, 32

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, Heidelberg, 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, Heidelberg, May 2013. 44

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

MR09. Daniele Micciancio and Oded Regev. *Lattice-based Cryptography*, pages 147–191. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009. 4, 25

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, Heidelberg, May 2013. 3

Was97. Lawrence C. Washington. *Introduction to cyclotomic fields*, volume 83. Springer, 1997. 12

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, Heidelberg, December 2023. 1

A Extended Preliminaries

A.1 Algebraic Number Theory

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 $\mathbf{b} = (1, \zeta, \ldots, \zeta^{\varphi-1})$ is the standard power basis, we may omit $\mathbf{b}$ from the subscript of $\psi_{\mathbf{b}}$. 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 $\|\sigma(\cdot)\|_2$. 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 $\|\psi(x)\|_\infty \le p/2$. Note that $\|\sigma(\mathbf{x})\|_2^2 = \mathsf{Trace}(\mathbf{x}^\mathsf{T}\overline{\mathbf{x}})$, where $\bar{\cdot}$ denotes the complex conjugate.

Lemma 9. *Let $x \in \mathcal{K}$. It holds that $\|\sigma(x)\|_2 \le \varphi^{3/2}\|\psi(x)\|_\infty$*

Proof. Let $x = \sum_{i\in[\varphi]} x_i\zeta^i$, with $x_i \in \mathbb{Q}$. Then we have $\|\sigma(x)\|_2 = \left\|\sigma\left(\sum_{i\in[\varphi]} x_i\zeta^i\right)\right\|_2 \le \sum_{i\in[\varphi]} x_i \left\|\sigma\left(\zeta^i\right)\right\|_2 \le \varphi\cdot\max_i(|x_i|)\cdot\sqrt{\varphi} = \varphi^{3/2}\|\psi(x)\|_\infty$.

Lemma 10 ([DPSZ11, Theorem 7], [DPSZ12]). *For any conductor $\mathfrak{f}$, there exists a constant $c_\mathfrak{f}$, such that for any $x \in \mathcal{K}$, it holds that $\|\psi(x)\|_\infty \le c_\mathfrak{f}\cdot\|\sigma(x)\|_\infty$.*

Corollary 2. *For any conductor $\mathfrak{f}$, there exists a constant $c_\mathfrak{f}$, such that for any $x \in \mathcal{K}$, it holds that $\|\psi(x)\|_\infty \le c_\mathfrak{f}\cdot\|\sigma(x)\|_2$. More generally, for any $\mathbf{x} \in \mathcal{K}^m$, it holds that $\|\psi(\mathbf{x})\|_\infty \le c_\mathfrak{f}\cdot\|\sigma(\mathbf{x})\|_2$.*

In [DPSZ11, Appendix C.2], it was discussed that $c_\mathfrak{f} \le 8.6$ for any square-free $\mathfrak{f} \le 400$. However, note that in Lemma 10 the second inequality is trivial and usually not tight. If we only concern about bounding $\|\psi(x)\|_\infty$ in terms of $\|\sigma(x)\|_2$, it seems that $\|\psi(x)\|_\infty \le \|\sigma(x)\|_2$ always holds. We verified this empirically for $100 \le \mathfrak{f} \le 500$ with 100 random elements for each conductor.

Lemma 11. *If L/K is a Galois extension, then any K-linear map $f : L \to L$ can be expressed as an L-linear combination of $\mathsf{Gal}(L/K)$.*

Proof. Let L/K be of degree φ. Let $\mathbf{e} = (e_i)_{i\in[\varphi]}$ be a $\mathcal{O}_K$-basis of $\mathcal{O}_L$, and $\mathbf{e}^\vee = (e_i^\vee)_{i\in[\varphi]}$ be a $\mathcal{O}_K$-basis of $\mathcal{O}_L^\vee$. We show that there exists a vector $\mathbf{a} = (a_\tau)_{\tau\in\mathsf{Gal}(L/K)} \in (\mathcal{O}_L^\vee)^\varphi$ such that, for any $x \in L$, we can write

$$f(x) = \sum_{\tau\in\mathsf{Gal}(L/K)} a_\tau\cdot\tau(x).$$

To construct $\mathbf{a}$, we first construct another vector $\mathbf{b} = (b_i)_{i\in\varphi} \in L^\varphi$ by setting $b_i \coloneqq f(e_i)$ for all $i \in [\varphi]$. We then define $a_\tau \coloneqq \mathbf{b}^\mathsf{T}\cdot\tau(\mathbf{e}^\vee)$ for all $\tau \in \mathsf{Gal}(L/K)$, where the automorphism τ is applied component-wise. For any $x = \sum_{i\in[\varphi]} x_i e_i \in L$ where $x_i \in K$, we observe that

$$\begin{aligned}
\sum_{\tau\in\mathsf{Gal}(L/K)} a_\tau\cdot\tau(x) &= \sum_{\tau\in\mathsf{Gal}(L/K)} \mathbf{b}^\mathsf{T}\tau(\mathbf{e}^\vee)\cdot\tau(x)\\
&= \sum_{\tau\in\mathsf{Gal}(L/K)}\sum_{j\in[\varphi]} b_j\tau(e_j^\vee)\cdot\tau(\sum_{i\in[\varphi]} x_i e_i)\\
&= \sum_{\tau\in\mathsf{Gal}(L/K)}\sum_{i,j\in[\varphi]} b_j\tau(e_j^\vee e_i)x_i\\
&= \sum_{i,j\in[\varphi]} b_j\mathsf{Trace}_{L/K}(e_j^\vee e_i)x_i\\
&= \sum_{i\in[\varphi]} b_i x_i = \sum_{i\in[\varphi]} f(e_i)x_i = f(x).
\end{aligned}$$

□

Lemma 12. *Let L be a cyclotomic field, L/K a Galois extension, $\mathfrak{f}_L$ be the conductor of L, and p be a rational prime with $\mathfrak{f}_L < p$. It holds that any $\mathcal{O}_K/p\mathcal{O}_K$-linear map $f : \mathcal{O}_L/p\mathcal{O}_L \to \mathcal{O}_L/p\mathcal{O}_L$ can be expressed as an $\mathcal{O}_L/p\mathcal{O}_L$-linear combination of $\mathsf{Gal}(L/K)$.*

Proof. In the proof of Lemma 11, we can see that any $\mathcal{O}_K$-linear map $f : \mathcal{O}_L \to \mathcal{O}_L$ can be expressed as an $\mathcal{O}_L^\vee$-linear combination of $\mathsf{Gal}(L/K)$. Since L is cyclotomic, it is known (see e.g. [LPR13]) that

$$\mathcal{O}_L \subseteq \mathcal{O}_L^\vee \subseteq \mathfrak{f}_L^{-1}\mathcal{O}_L.$$

Taking quotients, we have

$$\frac{\mathcal{O}_L}{p\mathcal{O}_L} \subseteq \frac{\mathcal{O}_L^\vee}{p\mathcal{O}_L} \subseteq \frac{\mathfrak{f}_L^{-1}\mathcal{O}_L}{p\mathcal{O}_L}.$$

However, since $\mathfrak{f}_L < p$ and p is prime, we have

$$\frac{\mathcal{O}_L}{p\mathcal{O}_L} \subseteq \frac{\mathcal{O}_L^\vee}{p\mathcal{O}_L} \subseteq \frac{\mathfrak{f}_L^{-1}\mathcal{O}_L}{p\mathcal{O}_L} = \frac{\mathcal{O}_L}{p\mathcal{O}_L},$$

forcing $\frac{\mathcal{O}_L^\vee}{p\mathcal{O}_L} = \frac{\mathcal{O}_L}{p\mathcal{O}_L}$. The claim thus follows. □

A.2 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.3 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.[21] 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].

Remark 12. 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*.*

[21] See also Remark 12.

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.4). 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.4 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 $(\mathbf{x}_i, y_i)_{0 \le i \le \ell}$ such that $V(\mathbf{x}_i, y_i) = 1$ for all $0 \le i \le \ell$ and $\{\mathbf{x}_0, \ldots, \mathbf{x}_\ell\} \in \Gamma(S, \ell)$.