Basefold in the List Decoding Regime

Ulrich Haböck
uhaboeck@polygon.technology,

October 5, 2024

Abstract

In this writeup we discuss the soundness of the Basefold multilinear polynomial commitment scheme [ZCF23] applied to Reed-Solomon codes, and run with proximity parameters up to the Johnson list decoding bound. Our security analysis relies on a generalization of the celebrated correlated agreement theorem from [BCI+20] to linear *subcodes* of Reed-Solomon codes, which turns out a by-product of the Guruswami-Sudan decoder analysis from [BCI+20].

We further highlight a non-linear variant of the subcode correlated agreement theorem, which is flexible enough to apply to Basefold-like protocols such as the optimization [Dia24b] of FRI-Binius [DP24], and which we believe sufficient for proving the security of a recent multilinear version of STIR [ACFY24b] in the list-decoding regime.

Contents

1 Introduction

A *polynomial commitment scheme (PCS)* is a cryptographic primitive, in which one party (the prover) is given the possibility for providing *evaluation proofs* on previously committed polynomials, i.e. cryptographic proofs of their value at arbitrary query points. Polynomial commitment schemes are a key component of modern argument systems, which essentially are polynomial *interactive oracle proofs*, that are then turned into a cryptographic argument by instantiating the ideal PCS (the oracle) with the concrete scheme.[1] Polynomial commitment schemes come in two flavors, serving either univariate polynomials, or multivariate, most importantly *multilinear*, polynomials. Their constructions rely on various mathematical structures, for example pairings [KZG10, PST13], general elliptic curves [BCC$^+$16, BBB$^+$18], lattices [AFLN23, CMNW24, NS24], and groups of unknown order [BFS20, SB23]. An alternative line of constructions uses error-correcting codes, foremost Reed-Solomon codes and their siblings from algebraic geometry, for instance [AHIV17, BBHR18, BGKS20, BLNR22, DP23] or more recently FRI-Binius [DP24] and [HLP24]. (This list is not complete, and neglects constructions from other types of codes.) The constructions of univariate and multilinear schemes are often closely related, and there are general transformations for turning univariate schemes into multilinear ones, see [ZXZS20, BCHO22, CBBZ22, KT23] and [PH23].

Basefold [ZCF23] is a multivariate polynomial commitment scheme built from *foldable codes*. Foldable codes, as introduced by Zeilberger, Chen and Fisch, are error-correcting codes over arbitrary fields, which possess an encoding procedure similar to the Fast Fourier Transform (FFT), and hence admit a proximity test similar to the *Fast Reed Solomon Code Interactive Proof of Proximity (FRI)* [BBHR18]. On very high level, the protocol works as follows. Multilinear polynomials are mapped to code words so that folding of their "even" and "odd" parts (in the generalized FFT sense) translates to partial evaluation of the multilinear. With this parallelism in mind, the value at a query is then proven via the multivariate sumcheck protocol [LFKN92], intertwined with the proof of proximity, round by round, sharing the same verifier challenges.

Although random foldable codes have acceptable metrics [ZCF23], the most important use cases remain Reed-Solomon codes (in all facets, over binary fields [DP24], over elliptic curves [BCKL21, BCKL22] or the circle curve [HLP24]), due to their amenability for a tighter and more extensive soundness analysis. This brings us to the main purpose of this write-up: We specialise the soundness analysis from [ZCF23] to the case of Reed-Solomon codes, and extend it to the list decoding regime, allowing proximity parameters up to the Guruswami-Sudan-Johnson bound $1 - \sqrt{\rho}$, where ρ is the rate of the code. Concretely,

[1] We are aware that this simplistic view falls short of many technical aspects of code based constructions, in particular in the list decoding regime. However, for the sake of brevity we keep with the term PCS.

we make use of the celebrated correlated agreement theorem from Ben-Sasson, et al. [BCI+20] to establish a soundness proof in a round-by-round manner, similar to that of FRI therein. In fact, in the unique decoding regime there are no additional difficulties beyond the increased complexity of the protocol itself (in comparison to FRI). In the list decoding regime however, things are slightly more delicate. Correlated agreement alone is not sufficient to link a successful folding with the sumcheck specialization, and we need to take a closer look at the Guruswami-Sudan decoder analysis from [BCI+20] in order to generalize the statement accordingly. This leads to two strengthenings of the correlated agreement theorem:

- One for linear *subcodes* of Reed-Solomon codes, sufficient for Basefold and FRI-Binius [DP24] with its specifically structured evaluation inner product; and
- a non-linear generalization, which turns out useful in the analysis of Basefold for more expressive inner products, such as multi-query evaluation proofs or the recent optimizations of FRI-Binius [Dia24a] and [Dia24b].

The obtained soundness error of the complete protocol is essentially the sum of the round-wise errors imposed by the (strengthened) correlated agreement theorem, yielding the same proof sizes as FRI as an univariate PCS, without the the overhead of generic univariate-to-multivariate transformations such as the ones from [BCHO22, CBBZ22] or [PH23].

In a concurrent work [ACFY24b] the authors elaborate WHIR, a multilinear variant of STIR [ACFY24a]. Their soundness analysis in the list decoding regime relies on a conjecture called *mutual correlated agreement*, which is stronger than the above mentioned non-linear generalization of the subcode agreement. We are confident that our generalization is sufficient for proving soundness of their protocol, but this will be addressed in a separate document.

The write-up is organized as follows. In Section 2 we introduce notation and basic facts on multilinear polynomials. Section 3 recaps the Basefold protocol for the specific case of Reed-Solomon codes, over finite fields with a smooth multiplicFRIBinius:Galoisative subgroup. We sketch the underlying mechanics of our soundness proof, and explain the need of strengthened correlated agreement. In the following Section 4, we then go over the corner pillars of the Guruswami-Sudan list decoder analysis from [BCI+20] and derive the above mentioned strengthenings of the correlated agreement theorem. While this part, considered in full depth, is the most challenging, we stress the fact that it can be easily understood on high-level, without going into the algebraic knits and grits of the decoder analysis. Section 5 is the core of the document: We prove soundness of Reed-Solomon Basefold in the plain oracle model, for proximity parameters up to the Johnson bound. As aforementioned, the proof is in the spirit of that of FRI in [BCI+20] and makes use of the correlated agreement

strengthening from Section 4. The remainder of the document, Section 6, is devoted to generalizations of Basefold: Its adaption to other FFTs is outlined in Section 6.1, Basefold for more expressive inner products is addressed in Section 6.2), and FRI-Binius, in its optimized form [Dia24b], is discussed in Section 6.3.

Acknowledgements. The author would like to thank: Ben E. Diamond and Jim Posen for posing the question addressed in this write-up; Swastik Kopparty for his feedback on Section 4; Giacomo Fenzi for his feedback and for sharing an early version of [ACFY24b]; and Al Kindi for a thorough proof read.

2 Preliminaries

An *interactive proof* [BS84, BM88, GMR89] between two parties is an interactive protocol, in which one party (the *prover*) wishes to convince another party (the *verifier*) upon the validity of a claimed statement. An *interactive oracle proof* (IOP) [BCS16, RRR16], or an interactive oracle proof in the *plain oracle model*, is an information-theoretic model of an interactive proof in which the prover is able to commit data strings via *oracles*. Oracles serve the ideal functionality of a binding and hiding commitment, and allow partial revelations (local *openings*) at arbitrary positions of the string, the *queries* from the verifier. The oracles we use will be of the form $f \in F^D$, where D is some well-defined subset of a finite field F, and the queries will be points $x \in D$ selected by the verifier.

The interactive oracle proofs discussed in this document will be as follows. Given some functions $g_0, \ldots, g_M \in F^D$ committed by the prover, the interactive oracle proof of $(g_0, \ldots, g_M)$ belonging to a claimed relation $\mathcal{R}$, is comprised of a fixed number of rounds, and has a *public coin verifier*: In each round the verifier sends a random challenge, and the prover answers with (the commitments to) some other set of functions, besides some plain message data. We shall use the common terms that the prover *sends* (or, *shares*) the functions to (resp., with) the verifier, and the verifier *queries* the functions, knowing that these are secured by oracles and not accessible by the verifier in full length. After these rounds, the verifier queries the functions it has received in the course of protocol at a fixed number of random positions (from their respective domain of definition), an uses the local openings in order to decide whether to accept or reject. Although not touched by this document, we assume that the reader is aware of the fact that in the random oracle model, interactive oracle proofs with a public coin verifier can be compiled into non-interactive arguments of knowledge via the BCS-transform [BCS16], see also [CY24].

Let F be a finite field of arbitrary characteristic. We denote the set of

univariate polynomials of degree less than 2^n by

$$\mathscr{P}_n = F[X]^{<2^n},$$

for any integer $n \geq 0$. The n-dimensional *Boolean hypercube* over F is the set $H_n = \{0,1\}^n$, regarded as a subset of F^n, and a multivariate polynomial $P \in F[X_1, \ldots, X_n]$ in n variables $X_1, \ldots X_n$ is *multilinear*, if it is at most linear in each of its variables, $\deg_{X_i}(P) \leq 1$ for every i, and thus of the form

$$P(X_1, \ldots, X_n) = \sum_{i=(i_1,\ldots,i_n)\in\{0,1\}^n} c_i \cdot X_1^{i_1} \cdot \ldots \cdot X_n^{i_n},$$

with coefficients $c_i \in F$. For any $\vec{x} = (x_1, \ldots, x_n) \in F^n$ we write $P(\vec{x})$ for the value $P(x_1, \ldots, x_n)$. Multilinear polynomials are uniquely determined by their values over the Boolean hypercube, and their coefficients can be computed from the values by a multidimensional FFT, and vice-versa.

The n-dimensional *Lagrange kernel* (also called $eq()$ in several works) is the multilinear polynomial

$$L(X_1, \ldots, X_n, Y_1, \ldots, Y_n) = \prod_{i=1}^{n} (1 - (X_i + Y_i) + 2 \cdot X_i \cdot Y_i),$$

where we will use the same notation for different dimensions n, without causing confusion. In particular, for $\vec{y} = (y_1, \ldots, y_n) \in \{0,1\}^n$ its specialization $L(X_1, \ldots, X_n, \vec{y}) = L(X_1, \ldots, X_n, y_1, \ldots, y_n)$ is the unique multilinear from $F[X_1, \ldots, X_n]$ which evaluates to 1 at $\vec{y}$, and is equal to 0 elsewhere on H_n. This property induces that for arbitrary multilinear $P \in F[X_1, \ldots, X_n]$ and $\vec{y} \in F^n$, the inner product

$$\langle P, L(\,.\,, \vec{y}) \rangle_{H_n} = \sum_{\vec{x}\in H_n} P(\vec{x}) \cdot L(\vec{x}, \vec{y}) = P(\vec{y}),$$

and thus yields the value of P at $\vec{y}$. We shall call this inner product the *evaluation inner product*.

The multivariate sumcheck protocol [LFKN92] is an interactive proof for that a multivariate polynomial $G \in F[X_1, \ldots, X_n]$ satisfies

$$\sum_{\vec{x}\in H_n} G(\vec{x}) = s.$$

Typically, the multivariate G is a virtual polynomial, a composition of committed multilinears, and its maximum individual degree $d = \max_i \deg_{X_i}(G)$ is at least 2. Although we shall describe our protocols in a self-contained manner, we quickly sketch the protocol, as it is an ingredient of Basefold. Before the first round the prover claims the refinement polynomial

$$q_0(X) = \sum_{(x_2,\ldots,x_n)\in H_{n-1}} G(X, x_2, \ldots, x_n),$$

which is of degree at most d. The verifier checks the sanity condition that $s = q_0(0) + q_0(1)$, chooses a random λ_1, and asks the prover to prove the specialized claim

$$q_0(\lambda_1) = \sum_{(x_2,\ldots,x_n)\in H_{n-1}} G(\lambda_1, x_2, \ldots, x_n)$$

instead. The protocol continues in the same manner, letting the prover provide refinement polynomials

$$q_i(X) = \sum_{(x_2,\ldots,x_n)\in H_{n-(i+1)}} G(\lambda_1, \ldots, \lambda_i, X, x_{i+2}, \ldots, x_n),$$

for $i = 1, \ldots, n-2$ as response of the received challenges $\lambda_2, \ldots, \lambda_{n-1}$. In the last step the verifier samples $\lambda_n \leftarrow\!\!\$\ F$, which eventually reduces the sumcheck claim to that

$$q_{n-2}(\lambda_n) = G(\lambda_1, \ldots, \lambda_n),$$

a evaluation claim for the multivariate polynomial G at the random point $\vec{\lambda} = (\lambda_1, \ldots, \lambda_n)$, comprised of the random challenges received in the course of the protocol. For a detailed treatment of the sumcheck protocol, and its importance to multivariate proofs in general, we refer to [Tha23] and the references therein.

3 Basefold for Reed-Solomon codes

In this section we describe Basefold [ZCF23] for the simplest case of Reed-Solomon codes, i.e. over a finite field F of characteristic $p > 2$, with a smooth multiplicative group, meaning that

$$G_n = \{x \in F \,:\, x^{2^n} = 1\}$$

is a multiplicative subgroup of order $|G_n| = 2^n$, for a given fixed integer $n \geq 1$. For such a smooth group (as well as any subgroup and its cosets) interpolation and evaluation of polynomials can be done via the Fast Fourier Transform and its inverse. Likewise, there is the FRI low degree test [BBHR18] for the Reed-Solomon code

$$\mathscr{C}_0 = \mathsf{RS}_{2^n}[F, D] = \big\{q(x)|_{x\in D} \,:\, q(X) \in \mathscr{P}_n\big\},$$

generated by the space of polynomials $\mathscr{P}_n = F[X]^{<2^n}$ of degree at most 2^n, over any evaluation domain D which is a union of cosets of G_n. We fix the evaluation domain D, and with it the rate $\rho = 2^n/|D|$ of the code. The FRI protocol goes along the chain of projected domains

$$D = D_0 \xrightarrow{\pi} D_1 \xrightarrow{\pi} \ldots \xrightarrow{\pi} D_n,$$

where the domain size halves under each application of $\pi(x) = x^2$, and their codes $\mathscr{C}_k = \mathsf{RS}_{2^{n-k}}[F, D_k]$ of the same rate, generated by the space of polynomials $\mathscr{P}_{n-k}$ of degree less than 2^{n-k}, for $k = 0, \ldots, n$. We assume that the reader is familiar with the mechanics of FRI, the round-wise folding of odd and even parts.

In Reed-Solomon Basefold, a multilinear polynomial $P \in F[X_1, \ldots, X_n]$ is committed through its *univariate representation* $p(X) \in \mathscr{P}_n$, which determined from its values over the Boolean hypercube $H_n = \{0, 1\}^n$,

$$p(X) = \sum_{i=0}^{2^n-1} P(i_1, \ldots, i_n) \cdot X^i, \tag{1}$$

where $(i_1, \ldots, i_n) \in \{0, 1\}^n$ are the bits of $i = i_1 + i_2 \cdot 2 + \ldots + i_n \cdot 2^{n-1}$. Concretely, it is committed as word $f \in \mathscr{C}_0$, by evaluating $p(X)$ on the evaluation domain D. (Note that (1) establishes a one-to-one correspondence between the space of multilinear polynomials from $F[X_1, \ldots, X_n]$ and the space $\mathscr{P}_n$ of univariate polynomials of degree less than 2^n.)

With the identification (1) in mind, the even and odd part $p_0(X)$, $p_1(X)$ of the univariate representation $p(X)$, subject to

$$p(X) = p_0(X^2) + X \cdot p_1(X^2),$$

correspond to the prefixed polynomials $P(0, X_2, \ldots, X_n)$ and $P(1, X_2, \ldots, X_n)$, and more generally, any linear combination of the form

$$p_{\lambda_1}(X) = (1 - \lambda_1) \cdot p_0(X) + \lambda_1 \cdot p_1(X)$$

corresponds to the specialization $P_{\lambda_1}(X_2, \ldots, X_n) = P(\lambda_1, X_2, \ldots, X_n)$ of the multilinear polynomial. In other words, *FRI-like folding corresponds to partial evaluation of the multilinear representation*[2], and repeating the argument we conclude that, eventually the last oracle of the folding cascade corresponds to the value

$$v = P(\lambda_1, \ldots, \lambda_n),$$

where $\lambda_1, \ldots, \lambda_n$ are the randomnesses drawn in the course of the protocol. To leverage this observation for custom, non-random, queries $\vec{\omega} = (\omega_1, \ldots, \omega_n) \in F^n$, Basefold considers the evaluation inner product

$$v = \big\langle L(\vec{\omega}, \,.\,), P(\,.\,) \big\rangle_{H_n} = \sum_{\vec{x} \in H_n} L(\vec{\omega}, \vec{x}) \cdot P(\vec{x}),$$

and proves it via the multivariate sumcheck protocol *intertwined* with the folding rounds of the FRI proximity test: In each round, both reductions, the one of the sumcheck protocol and the one of FRI, use the *same* randomness.

[2]This and similar connections have been used in previous constructions, see [BCHO22, CBBZ22, Ham22, KT23] and [PH23].

3.1 The protocol

We slightly deviate from the presentation in [ZCF23] in regards to how the prover provides the sumcheck polynomials

$$q_i(X) = \sum_{\vec{x}=(x_{i+2},\ldots,x_n)\in H_{n-(i+1)}} L((\lambda_1,\ldots,\lambda_i,X,\vec{x}),\vec{\omega})\cdot P(\lambda_1,\ldots,\lambda_i,X,\vec{x}), \quad (2)$$

determined by the round challenges $\lambda_1,\ldots,\lambda_i$, where $i=1,\ldots,n-1$. Instead of (2) we let the prover send the *linear* polynomial

$$\Lambda_i(X) = \sum_{\vec{x}=(x_{i+2},\ldots,x_n)\in H_{n-(i+1)}} L(\vec{x},(\omega_{i+2},\ldots,\omega_n))\cdot P(\lambda_1,\ldots,\lambda_i,X,\vec{x}), \quad (3)$$

from which the quadratic sumcheck polynomial are derived via

$$q_i(X) = L(\lambda_1,\ldots,\lambda_i,\omega_1,\ldots,\omega_i)\cdot L(X,\omega_{i+1})\cdot \Lambda_i(X). \quad (4)$$

(The equality is an immediate consequence of the tensor product structure of the Lagrangian.) Although this can be seen as prover optimization (see also [Gru24, DT24]), the reason for this change is a different one: It allows to relate the folding specialization on the FRI-side of the protocol with the refinement on the sumcheck side, by means of the correlated agreement theorem for *linear subcodes* of Reed-Solomon codes. (See the sketch of soundness in the following section.)

The concrete protocol, formulated as interactive oracle proof is as follows.

Protocol 1 (Reed-Solomon Basefold [ZCF23])**.** *Let $P\in F[X_1,\ldots,X_n]$ be a multilinear polynomial, $\vec{\omega}=(\omega_1,\ldots,\omega_n)$ be any query from F^n, and $v=P(\vec{\omega})$. The prover takes the Reed-Solomon code word $f_0\in\mathscr{C}_0$ corresponding to the univariate representation of P, computes the linear polynomial $\Lambda_0(X)$ as in* (3) *and shares*

$$f_0,\Lambda_0(X),$$

and $s_0=v$ with the verifier. Then both engage in the following protocol, consisting of a commit phase and a subsequent query phase.

1. *Commit phase. This is the multivariate sumcheck protocol intertwined with the folding cascade of FRI, using n reduction steps.*
   (a) *In round i, $1\le i\le n-1$, the verifier previously received $f_{i-1}\in F^{D_{i-1}}$, $\Lambda_{i-1}(X)$ from the prover, and knows the value s_{i-1}. It checks that the sumcheck polynomial $q_{i-1}(X)$ determined from $\Lambda_{i-1}(X)$ using* (4) *satisfies*
   $$q_{i-1} = q_{i-1}(0)+q_{i-1}(1),$$
   and returns a random scalar $\lambda_i\leftarrow\!\!\$\, F$ to the prover.

(In the first round $i = 1$, Formula (4) collapses to $q_0(X) = L(X, \omega_1) \cdot \Lambda_0(X)$.)

The prover computes the Reed-Solomon codeword $f_i \in \mathscr{C}_i$ corresponding to the univariate representation of the specialization $P_{\lambda_1,\ldots,\lambda_i} \in F[X_{i+1}, \ldots X_n]$, and its linear polynomial $\Lambda_i(X)$ as in (3), and answers the verifier with

$$f_i, \Lambda_i(X).$$

Both take $s_i = q_{i-1}(\lambda_i)$ for the next round.

(In round $i = n - 1$, Definition (3) reduces to the singleton sum $\Lambda_i(X) = P(\lambda_1, \ldots, \lambda_{n-1}, X)$.)

(b) *The last round $i = n$ is as the previous rounds, except that the prover replies on $\lambda_n \leftarrow\!\!\$\ F$ with the constant $c = P(\lambda_1, \ldots, \lambda_n)$ for the constant code word f_n (no Λ_n either). Both take $s_n = q_{n-1}(\lambda_n)$.*

The verifier does the final check whether $s_n = L(\vec{\omega}, \lambda_1, \ldots, \lambda_n) \cdot c$, and proceeds with the query phase, which consists of $s \geq 1$ query rounds:

2. *Query Phase. In each of the rounds, the verifier samples a random $x_0 \leftarrow\!\!\$\ D_0$, and queries the oracles $f_0, \ldots, f_{n-1}$ for the values that are needed to check the consistency relations*

$$f_{i+1}(x_{i+1}) = \frac{f_0(x_i) + f_0(-x_i)}{2} + \lambda_i \cdot \frac{f_0(x_i) + f_0(-x_i)}{2 \cdot x_i},$$

for every $i = 0, \ldots, n - 1$, along the projection trace of x_0, given by $x_1 = \pi_1(x_0)$, $x_2 = \pi_2(x_1)$, $\ldots$, $x_n = \pi_n(x_{n-1})$. (f_n is taken as the constant function c.)

If all verifier checks of the protocol, including those of the sumcheck, pass then the verifier accepts. (Otherwise, it rejects.)

Of particular importance is the generalization of Protocol 1 to lists (often called "batches") of multilinears, as correlated agreement across all the committed words plays a crucial role in proving soundness of any IOP on top of the scheme. We keep with a single common query $\vec{\omega} \in F^n$ for the entire batch of multilinears. Multi-query proofs are discussed in Section 6.2.

Protocol 2 (Batch Reed-Solomon code Basefold)**.** *The prover shares the Reed-Solomon codewords $g_0, \ldots, g_M \in \mathscr{C}_0$ of the multilinears $G_0, \ldots, G_M$, together with their evaluation claims $v_0, \ldots v_M$ at $\vec{\omega} \in F^n$ with the verifier. Then they engage in the following extension of Protocol 1:*

1. *In a preceding round $i = 0$, the verifier sends a random $\lambda_0 \leftarrow\!\!\$\ F$, and the*

prover answers with the oracle for

$$f_0 = \sum_{k=0}^{M} \lambda_0^k \cdot g_k. \tag{5}$$

Then both prover and verifier engage in Protocol 1 on f_0 and the claim $v_0 = \sum_{k=0}^{M} \lambda_0^k \cdot v_k$. In addition to the checks in Protocol 1, the verifier also checks that equation (5) *holds at every sample x from D_0.*

Without formal definition, we state the soundness error of Protocol 2, as an interactive oracle proof of proximity for the evaluation claims, with a proximity parameter in the list decoding regime,

$$\theta \in \left(\frac{1-\rho}{2}, 1-\sqrt{\rho}\right).$$

That is, if a (possibly unbounded) algorithm P^* succeeds the verifier with a probability greater than that soundness error, then there exist polynomials

$$p_0(X), \ldots, p_M(X) \in \mathscr{P}_n$$

which agree with the committed words $g_0, \ldots, g_M$ on a *joint set* of density at least $1-\theta$, and the multilinear representations $P_0, \ldots, P_M$ of which, satisfy the evaluation claims $P_k(\vec{\omega}) = v_k$, for each $k = 0, \ldots, M$.

Theorem 1 (Basefold soundness)**.** *Let F and D and $\mathscr{C} = \mathsf{RS}_{2^n}[F, D]$ as above, and choose a proximity parameter $\theta = (1 + \frac{1}{2m}) \cdot \sqrt{\rho}$, where ρ is the rate of $\mathscr{C}$ and $m \geq 3$. The soundness error ε of the batch evaluation proof, Protocol 2, is bounded by*

$$\varepsilon < \varepsilon(\mathscr{C}_0, M, 1, \theta) + \sum_{i=1}^{n} \left(\frac{1}{|F|} + \varepsilon\big(\mathscr{C}_i, 1, B_i, \theta\big)\right) + (1-\theta)^s. \tag{6}$$

where $\varepsilon(\mathscr{C}_i, M_i, B_i, \theta)$ is the soundness error of the weighted correlated agreement theorem, Theorem 4, for subcodes of $\mathscr{C}_i = \mathsf{RS}_{2^{n-i}}[F, D_i]$, on a batch of $M_i + 1$ words, weight denominator bound $B_i = |D|/|D_i| = 2^i$, and proximity parameter θ.

A formal treatment, including a proof of the theorem, is given in Section 5. Nevertheless, we quickly explain the components of the soundness error in Theorem 1. The overall sum in the first line of equation (6) corresponds to the soundness error of the commit phase, expressing the maximum probability for an adversary, on a set of words $(g_1, \ldots g_M)$ which does not satisfy the claim of the proof (which is, there exists a correlated θ-proximate $(p_1, \ldots, p_M) \in \mathscr{P}_n^M$ satisfying the evaluation claims) is able to provide oracles $f_0, f_1, \ldots, f_n$ with only few folding inconsistencies, assuming admissible sumcheck refinement claims $\Lambda_0(X), \ldots, \Lambda_{n-1}(X)$.

- The first term $\varepsilon(\mathscr{C}_0, M, 1, \theta)$ is the error for the batching step of the $M+1$ given words over $D_0 = D$. This is the same error as given by the regular (i.e. non-weighted) correlated agreement theorem for linear subcodes, Theorem 3, and we chose the weighted error only for notational convenience.
- The terms $\varepsilon(\mathscr{C}_i, 1, B_i, \theta)$, $i = 1, \ldots, n$, are the errors for the further, FRI-like reduction steps, when folding $M_i = 2$ words over the domain D_i, where the additional $1/|F|$ term is a remainder contribution to the sumcheck. One would expect a double as large term, since the sumcheck is quadratic. However, half of it is already covered by the subcode correlated agreement theorem, and that half can be even dropped when using the improved non-linear variant from Section 4.2, see Lemma 2 in Section 6.2.

Finally, the $(1-\theta)^s$-term is the soundness error of the query phase: The probability that a given set of round oracles $f_0, f_1, \ldots, f_n$ with a folding inconsistency set (on which one of the folding checks does not hold) of density at least θ is not detected by s samples.

Remark 2. As in regular FRI, one can stop the Protocol 1 at any step k, $0 \leq k < n$, and let the prover provide the last oracle $f_k \in F^{D_k}$ in plain, without the refinement polynomial $\Lambda_k(X)$, e.g. via the coefficients of the univariate representation of $P_{\lambda_1,\ldots,\lambda_k}$. The verifier then checks the remaining inner product

$$L((\omega_1, \ldots, \omega_k), (\lambda_1, \ldots, \lambda_k)) \cdot \sum_{\vec{x} \in H_{n-k}} L((\omega_{k+1}, \ldots, \omega_n), \vec{x}) \cdot P_{\lambda_1,\ldots,\lambda_k}(\vec{x})$$

against its expected value $s_k = q_{k-1}(\lambda_k)$, and accepts if equal. The soundness error in Theorem 1 is then reduced to the sum of folding errors ranging only over the corresponding steps.

3.2 Sketch of soundness

The crucial ingredient for proving the soundness of Protocol 2 is the following strengthening of the correlated agreement theorem for Reed-Solomon codes [BCI+20]. It allows to relate the specialization on the FRI-side of the protocol with the refinement of the sumcheck claim. A proof of the theorem, as well as its weighted variant Theorem 4, is given in Section 4.1.

Theorem 3 (Correlated agreement for subcodes)**.** *Let F be a finite field of arbitrary characteristic, and $\mathscr{C} = \mathsf{RS}_k[F, D]$ the Reed-Solomon code over F with evaluation domain $D \subseteq K$ and rate $\rho = k/|D|$. Let $\mathscr{C}'$ be a linear subcode of $\mathscr{C}$, generated by a subspace $\mathscr{P}'$ of polynomials from $F[X]^{<k}$. Given a proximity parameter $\theta = 1 - \sqrt{\rho} \cdot \left(1 + \frac{1}{2m}\right)$, with $m \geq 3$, and words $f_0, f_1, \ldots, f_M \in F^D$*

for which

$$\frac{\left|\left\{z \in F : d\big(f_0 + z \cdot f_1 + \ldots + z^M \cdot f_M, \mathscr{C}'\big) < \theta\right\}\right|}{|F|} > \varepsilon,$$

where

$$\varepsilon = M \cdot \frac{\left(m + \frac{1}{2}\right)^7}{3 \cdot \rho^{\frac{3}{2}}} \cdot \frac{|D|^2}{|F|}, \tag{7}$$

Then there exist polynomials $p_0, p_2, \ldots, p_M \in \mathscr{P}'$, and a set $A \subseteq D$ of density $|A|/|D| \geq 1 - \theta$ on which $f_0, f_1, \ldots, f_M$ jointly coincide with the values of $p_0, p_1, \ldots, p_M$, respectively.

Let us illustrate the role of Theorem 3 in proving soundness in a round-by-round manner. We do this by means of the first round of Protocol 1, for a given proximity parameter θ. (The importance of the theorem in the batching round of Protocol 2 is even easier to explain.)

Suppose that a (possibly malicious) prover is able to provide f_0 and Λ_0, the latter of which is sumcheck compliant (i.e. $s_0 = q_0(0) + q_0(1)$), so that with noticeable probability the linear combination

$$f_{\lambda_1} = (1 - \lambda_1) \cdot f_{0,0} + \lambda_1 \cdot f_{0,1}$$

is θ-close to a polynomial $p_{\lambda_1}(X) \in \mathscr{P}_{n-1}$ with a multilinear representation $P_{\lambda_1} \in F[X_2, \ldots, X_n]$ satisfying the refinement on the sumcheck side,

$$\big\langle L((\lambda_1, \,.\,), \omega), P_{\lambda_1} \big\rangle_{H_{n-1}} = q_0(\lambda_1).$$

We wish to conclude that then both $f_{0,0}, f_{0,1}$ agree with polynomials $p_0(X), p_1(X)$ from $\mathscr{P}_{n-1}$ on a joint set $A \subseteq D_1$ of density at least $1 - \theta$, the multilinear representations P_0 and P_1 of which are sumcheck compliant,

$$\langle L((0, \,.\,), \vec{\omega}), P_0 \rangle_{H_{n-1}} = q_0(0), \tag{8}$$
$$\langle L((1, \,.\,), \vec{\omega}), P_1 \rangle_{H_{n-1}} = q_0(1). \tag{9}$$

It is then easy to see that, over the preimage $\pi^{-1}(A) \subseteq D_0$, the word f_0 agrees with the combined polynomial $p(X) = p_0(X^2) + X \cdot p_1(X^2)$ from $\mathscr{P}_n$, the multilinear representation of which satisfies the sumcheck claim, since $q_0(0) + q_1(0) = s_0$.

To show (8) and (9) we reduce the quadratic sumcheck compliance to consistency with the linear polynomial Λ_0. By Equation (4), sumcheck compliance translates to

$$L(\lambda_1, \omega_1) \cdot \big\langle L(\,.\,, \omega_2, \ldots, \omega_n), P_{\lambda_1} \big\rangle_{H_{n-1}} = L(\lambda_1, \omega_1) \cdot \Lambda_0(\lambda_1),$$

and hence, except for a set of probability $1/|F|$ (for the zero of $L(\,.\,,\omega_1)$), we get that

$$\big\langle L(\,.\,,\omega_2,\ldots,\omega_m), P_{\lambda_1}\big\rangle_{H_{n-1}} = \Lambda_0(\lambda_1),$$

which, after centering, reads as

$$\big\langle L(\,.\,,\omega_2,\ldots,\omega_m), P_{\lambda_1} - \Lambda_0(\lambda_1)\big\rangle_{H_{n-1}} = 0.$$

By linearity, $\Lambda_0(\lambda_1) = (1-\lambda_1)\cdot\Lambda_0(0) + \lambda_1\cdot\Lambda_0(1)$, and we conclude that, with noticable probability the random linear combination of the centered parts,

$$f'_{\lambda_1} = (1-\lambda_1)\cdot\underbrace{(f_{0,0} - \Lambda_0(0))}_{=:f'_{0,0}} + \lambda_1\cdot\underbrace{(f_{0,1} - \Lambda_0(1))}_{=:f'_{0,1}},$$

is θ-close to the centered polynomial $p'_{\lambda_1} = p_{\lambda_1} - \Lambda_0(\lambda_1)$, which belongs to the space

$$\mathscr{P}'_{n-1} = \big\{u(X)\in\mathscr{P}_{n-1} \,:\, \langle L(\,.\,,\omega_2,\ldots,\omega_n), U\rangle_{H_{n-1}} = 0\big\}.$$

This subspace defines a linear subcode $\mathscr{C}'_1$ of the Reed-Solomon code $\mathscr{C}_1$, and by Theorem 3 we eventually conclude the following: Both centered words $f'_{0,0}, f'_{0,1}$ agree with polynomials $p'_0(X), p'_1(X)$ from $\mathscr{P}'_{n-1}$ on a joint set $A\subseteq D_1$ of density $\geq 1-\theta$. Back in terms of non-centered functions, both even and odd parts $f_{0,0}$ and $f_{0,1}$ agree over A with the respective polynomials

$$\begin{aligned} p_0(X) &= p'_0(X) + \Lambda_0(0),\\ p_1(X) &= p'_1(X) + \Lambda_0(1),\end{aligned}$$

from $\mathscr{P}_{n-1}$, which moreover are consistent with the sumcheck side of the protocol: Their multilinear representations $P_0, P_1\in F[X_2,\ldots,X_n]$ satisfy

$$\begin{aligned}\langle L(\,.\,,\omega_2,\ldots,\omega_n), P_0\rangle_{H_{n-1}} &= \Lambda_0(0),\\ \langle L(\,.\,,\omega_2,\ldots,\omega_n), P_1\rangle_{H_{n-1}} &= \Lambda_0(1).\end{aligned}$$

Again by (4) the latter two constraints imply the sumcheck compliance, equation (8) and (9).

The complete argument in Section 5 faces the same technicalities as the soundness proof for FRI in [BCI+20]: It takes care of the FRI consistency sets, by means of weights (for the conditional probability of that the folding checks on $f_1,\ldots,f_{i-1}$ hold "above" a point $x\in D_i$), and the weighted variant of the correlated agreement theorem, Theorem 4.

4 Correlated agreement with constraints

In this section we discuss how the correlated agreement theorem for linear subcodes, Theorem 3 as well as its non-linear generalization, Section 4.2, are obtained from the Guruswami-Sudan list decoder analysis [BCI+20]. To this end, we give a sufficiently detailed overview of [BCI+20, full version, Chapter 5].

Let $F = \mathbb{F}_q$ be a finite field (of arbitrary characteristic), and $\mathscr{C} = \mathsf{RS}_k[F, D]$ be the Reed-Solomon code of rate $\rho = k/|D|$ generated by $\mathscr{P} = F[X]^{<k}$, the set of all polynomials of degree less than $k \geq 1$, using an arbitrary evaluation set $D \subseteq F$ of size $|D|$. We fix a proximity parameter

$$\theta = 1 - \left(1 + \frac{1}{2 \cdot m}\right) \cdot \sqrt{\rho},$$

with integer $m \geq 3$ being the multiplicity parameter for the Guruswami-Sudan list decoder.

Assume that we have given words $f_0, f_1 \in F^D$ so that $S = \{z \in F : d(f_0 + z \cdot f_1, \mathscr{C}) < \theta\}$ is of size

$$|S| > \frac{\left(m + \frac{1}{2}\right)^7}{3 \cdot \rho^{\frac{3}{2}}} \cdot |D|^2, \tag{10}$$

and let

$$P_z \in \mathscr{P}, z \in S, \tag{11}$$

be an *arbitrary* selection of θ-proximates, satisfying $d(f_0 + z \cdot f_1, P_z) < \theta$ for every $z \in S$. (Here d is the fractional Hammming distance over D. Weighted agreement is discussed later on.) In order to understand list decoding across different z's, [BCI+20] analyze the decoder over the rational function field $K = F(Z)$ in the indeterminate Z.

The proof starts with determining

$$Q(X, Y, Z) \in F[X, Y, Z],$$

a (carefully selected) polynomial variant of the Guruswami-Sudan interpolant for the K-valued word

$$f_0 + Z \cdot f_1,$$

meaning that $Q(x, f_0(x) + Z \cdot f_1(x), Z) = 0$ in K, with multiplicity at least the given parameter m, for each $x \in D$. (This is Step 1 in [BCI+20, Chapter 5].) Let

$$Q(X, Y, Z) = C(X, Z) \cdot \prod_i R_i(X, Y, Z)^{e_i}$$

be its decomposition into irreducible polynomials.

The main part of the proof, Step 2 to Step 7 in [BCI+20, Chapter 5], is then devoted to showing that one[3] of the irreducible factors $R_i(X, Y, Z)$ is in fact of the form $Y - P(X, Z)$, with $\deg_X(P) < k$ and $\deg_Z(P) \leq 1$, and hence leads to a solution

$$Y = P(X, Z) = p_0(X) + Z \cdot p_1(X), \tag{12}$$

[3] Namely, any factor which covers a sufficiently large fraction of the claimed proximates.

with $p_0(X), p_1(X) \in \mathscr{P} = F[X]^{<k}$. The proof uses techniques used in factorizing bivariates over finite fields (i.e. computing a power series solution from a simple root) carried over to an algebraic extension of K; a step that requires certain familiarity with algebraic function fields.

Proposition 1 ([BCI+20], Proposition 5.5)**.** *Under the above assumptions, there exists a polynomial $P(X, Z) \in F[X, Z]$ of degree $\deg_X P < k$, $\deg_Z P \leq 1$, and so that*

$$|\{z \in S \,:\, P(X, z) = P_z(X)\}| > \frac{|S|}{2 \cdot \ell_m},$$

where $\ell_m = \left(m + \frac{1}{2}\right) / \sqrt{\rho}$ is the Guruswami-Sudan list size bound for multiplicity parameter $m \geq 3$. *In a nutshell, the factor ℓ_m reflects the pigeon-hole principle for at most list size many irreducible factors, the additional factor 2 is for excluding poles of the sufficiently many coefficients of the power series solution.*

The analysis generalizes to more expressive linear combinations, in particular to the case $f_0 + Z \cdot f_1 + \ldots + Z^M \cdot f_M$, with $f_0, \ldots, f_M \in F^D$ for any $M \geq 1$, while the larger degree in Z demands the size of the set

$$S = \{z \in F \,:\, d(f_0 + z \cdot f_1 + \ldots + z^M \cdot f_M, \mathscr{P}) < \theta\}$$

being scaled by the degree M,

$$|S| > M \cdot \frac{\left(m + \frac{1}{2}\right)^7}{3 \cdot \rho^{\frac{3}{2}}} \cdot |D|^2. \tag{13}$$

This generalization is proven in [BCI+20, Section 6.2], and we cite it as a separate proposition.

Proposition 2 ([BCI+20], Section 6.2)**.** *Under the above assumptions, there exists a polynomial $P(X, Z) \in F[X, Z]$ of degree $\deg_X P < k$, $\deg_Z P \leq M$, and so that*

$$|\{z \in S \,:\, P(X, z) = P_z(X)\}| > \frac{|S|}{2 \cdot \ell_m},$$

where $\ell_m = \left(m + \frac{1}{2}\right) / \sqrt{\rho}$ is the Guruswami-Sudan list size bound for multiplicity parameter $m \geq 3$.

Proposition 1 and respectively Proposition 2 are the core results from which correlated agreement, as well as a weighted variant of it, are derived. For details, see Step 8 in [BCI+20, Chapter 5] and more generally [BCI+20, Chapter 6] for regular correlated agreement, and [BCI+20, Section 7] for the weighted variant.

4.1 For subcodes ...

Let us now extend the list decoder analysis to an arbitrary linear subcode $\mathscr{C}'$ of $\mathscr{C} = \mathsf{RS}_k[F, D]$. We directly do this for the general case $f_0, \ldots, f_M \in F^D$

covered by Proposition 2. Assume that for every $z \in S$ as claimed we even have proximity to the subcode,

$$d(f_0 + z \cdot f_1 + \ldots + z^M \cdot f_M, \mathscr{C}') < \theta,$$

and take any selection of proximates $P_z(X)$, $z \in S$, from the linear subspace $\mathscr{P}' \subseteq \mathscr{P}$ behind $\mathscr{C}'$. We claim that then, the polynomial

$$P(X, Z) = p_0(X) + Z \cdot p_1(X) + \ldots + Z^M \cdot p_M(X)$$

from Proposition 2 *additionally satisfies that*

$$p_0(X), p_1(X), \ldots, p_M(X) \in \mathscr{P}',$$

and thus belong to the subcode $\mathscr{C}'$.

To see this, assume that $\mathscr{C}'$ is generated by a single linear constraint, and thus its space of polynomials is $\mathscr{P}' = \{p \in \mathscr{P} \,:\, \Lambda p = 0\}$, where Λ is a linear functional on $\mathscr{P} = F[X]^{<k}$. (The general case bears no additional difficulties.) Since

$$\frac{|S|}{2 \cdot \ell_m} \geq M \cdot \frac{\left(m + \frac{1}{2}\right)^6}{6 \cdot \rho} \cdot |D|^2 \geq 3.7 \cdot M \cdot |D|^2 \tag{14}$$

even for the smallest choices of $|D|$ and m, regardless of $\rho < 1$, there exist at least $M + 1$ different points $z_0, z_1, \ldots, z_M \in S$ for which $P(X, z_i) = P_{z_i}(X) \in \mathscr{P}'$. By linearity,

$$0 = \Lambda P(X, z_i) = \Lambda p_0(X) + z_i \cdot \Lambda p_1(X) + \ldots + z_i^M \cdot \Lambda p_M(X),$$

for each z_i, and we conclude that

$$\Lambda p_0(X) = \Lambda p_1(X) = \ldots = \Lambda p_M(X) = 0,$$

showing that all $p_i(X)$ belong to the subcode $\in \mathscr{C}'$, as claimed. (If $\mathscr{C}'$ is defined by several linear functionals, then $0 = \Lambda p_i(X)$ for each of the functionals Λ, yielding the same conclusion.)

From this sharpening of Proposition 2, both regular as well as weighted correlated agreement of $P(X, Z)$ are proven as in [BCI+20, Section 6 and 7], without any changes. We only cite the weighted variant; the regular case is already mentioned in Theorem 3. Given a sub-probability measure μ on D, and $f \in F^D$, we write

$$\mathsf{agree}_\mu(f, \mathscr{C}') \geq 1 - \theta$$

if there exists a polynomial $p(X) \in \mathscr{P}'$ such that $\mu(\{x \in D \,:\, f(x) = p(x)\}) \geq 1 - \theta$.

Theorem 4. *(Weighted correlated agreement for subcodes) Let* $\mathscr{C}'$ *be a linear subcode of* $\mathsf{RS}_k[F, D]$*, and choose* $\theta = 1 - \sqrt{\rho} \cdot \left(1 + \frac{1}{2 \cdot m}\right)$*, for some integer* $m \geq 3$*,*

where $\rho = k/|D|$. Assume a density function $\delta : D \longrightarrow [0,1] \cap \mathbb{Q}$ with common denominator $B \geq 1$, i.e. for all x in D

$$\delta(x) = \frac{m_x}{B},$$

for an integer value $m_x \in [0, B]$, and let μ be the sub-probability measure with density δ, defined by $\mu(\{x\}) = \delta(x)/|D|$. If for $f_0, f_1, \ldots, f_M \in F^D$,

$$\frac{\left|\left\{z \in F : \mathsf{agree}_\mu(f_0 + z \cdot f_1 + \ldots + z^M \cdot f_M, \mathscr{C}') \geq 1 - \theta\right\}\right|}{|F|} > \varepsilon(\mathscr{C}, M, B, \theta),$$

where

$$\varepsilon(\mathscr{C}, M, B, \theta) = \frac{M}{|F|} \cdot \frac{\left(m + \frac{1}{2}\right)}{\sqrt{\rho}} \cdot \max\left(\frac{\left(m + \frac{1}{2}\right)^6}{3 \cdot \rho} \cdot |D|^2, \quad 2 \cdot (B \cdot |D| + 1)\right),$$

then there exist polynomials $p_0(X), p_1(X), \ldots, p_M(X)$ belonging to the subcode $\mathscr{C}'$, and a set A with $\mu(A) \geq 1 - \theta$ on which $f_0, f_1, \ldots, f_M$ coincide with $p_0(X), p_1(X), \ldots, p_M(X)$, respectively.

4.2 . . . and beyond

We emphasize the fact, that Proposition 2 applies also to a more general setting, in which the subspace of polynomials is characterized by a challenge-dependent function

$$\Gamma : \mathscr{P} \times F \longrightarrow F,$$

a polynomial (in the coordinates of an arbitrary basis) of typically small total degree $d = \deg \Gamma$, and

$$S = \Big\{ z \in F \ : \exists P_z(X) \in \mathscr{P} \text{ s.t.} \\ d(f_0 + z \cdot f_1 + \ldots + z^M \cdot f_M, P_z) < \theta \ \wedge \ \Gamma(P_z, z) = 0 \Big\},$$

where as before $\mathscr{P} = F[X]^{<k}$. In many applications, such as a Basefold proof for inner products with less-structured multilinears (see Section 6.2), we have $d = 2$.

As before, given the usual size bound (13) for S, and choosing the proximates accordingly (i.e., satisfying Γ), the polynomial $P(X, Z) = p_0(X) + Z \cdot p_1(X) + \ldots + Z^M \cdot p_M(X)$ from Proposition 2 satisfies that

$$\Gamma(p_0 + z \cdot p_1 + \ldots + z^M \cdot p_M, z) = 0$$

for z from a fraction of S, of size larger than

$$\frac{|S|}{2 \cdot \ell_m} \geq M \cdot \frac{\left(m + \frac{1}{2}\right)^6}{6 \cdot \rho} \cdot |D|^2 > 3.7 \cdot M \cdot |D|^2$$

regardless of the choice of m and $\rho < 1$. Thus if $d \cdot M + 1 \leq 3.7 \cdot M \cdot |D|^2$ (which in the case $d = 2$ holds even for smallest domains), we obtain that

$$\Gamma(p_0 + Z \cdot p_1 + \ldots + Z^M \cdot p_M, Z) = 0$$

as a formal identity. In other words, the θ-proximate $P(X, Z)$ satisfies the Γ-constraint over K, and in particular at every $Z = z$. Regular and weighted correlated agreement of $P(X, Z)$ is proven as before. We again cite only the weighted variant.

Theorem 5. *(Non-linear generalization of Theorem 4) Let $\mathscr{C} = \mathsf{RS}_k[F, D]$ be the Reed-Solomon code generated by the space of polynomials $\mathscr{P} = F[X]^{<k}$, and let $\Gamma : \mathscr{P} \times F \longrightarrow F$ be of total degree $d = 2$. As in Theorem 4, we take $\theta = 1 - \sqrt{\rho} \cdot \left(1 + \frac{1}{2 \cdot m}\right)$ for some integer $m \geq 3$, where $\rho = k/|D|$, and let μ be the sub-probability measure with density $\delta : D \longrightarrow [0, 1] \cap \mathbb{Q}$. If for $f_0, f_1, \ldots, f_M \in F^D$,*

$$\frac{\left|\left\{z \in F : \begin{array}{c} \exists P_z(X) \in \mathscr{P} \text{ s.t. } \Gamma(P_z, z) = 0 \\ \wedge\, \mathsf{agree}_\mu(f_0 + z \cdot f_1 + \ldots + z^M \cdot f_M, P_z) \geq 1 - \theta \end{array}\right\}\right|}{|F|} > \varepsilon(\mathscr{C}, M, B, \theta),$$

where

$$\varepsilon(\mathscr{C}, M, B, \theta) = \frac{M}{|F|} \cdot \frac{\left(m + \frac{1}{2}\right)}{\sqrt{\rho}} \cdot \max\left(\frac{\left(m + \frac{1}{2}\right)^6}{3 \cdot \rho} \cdot |D|^2, \quad 2 \cdot (B \cdot |D| + 1)\right),$$

with $B \geq 1$ being the common denominator of δ, then there exist polynomials $p_0(X), \ldots, p_M(X) \in \mathscr{P}$ such that

$$\Gamma(p_0 + Z \cdot p_1 + \ldots + Z^M \cdot p_m, Z) = 0 \in F[Z],$$

and which coincide with $f_0, f_1, \ldots, f_M$, respectively, on a joint set A of weight $\mu(A) \geq 1 - \theta$.

Remark 6. The soundness error of the non-weighted variant of Theorem 5 is equal to $\varepsilon(\mathscr{C}, M, B, \theta)$ with $B = 1$.

5 Soundness in the oracle model

In this section we prove Theorem 1, i.e. the soundness error of the batch variant of Basefold for Reed-Solomon codes, Protocol 2, as an interactive oracle proof of the relation

$$\mathcal{R} = \left\{ (g_0, \ldots, g_M) \; : \; \begin{array}{c} \exists p_0, \ldots, p_M \in \mathscr{F}[X]^{<2^n} \text{ s.t.} \\ d((g_0, \ldots, g_M), (p_0, \ldots, p_M)) < \theta \\ \wedge \bigwedge_{k=0}^{M} P_k(\omega_1, \ldots, \omega_M) = v_k \end{array} \right\}, \tag{15}$$

for given words $g_0, \ldots, g_M \in F^{D_0}$, query $\vec{\omega} = (\omega_1, \ldots, \omega_n)$, and evaluation claims $v_0, \ldots, v_M$. (In words, a tuple of words $(g_0, \ldots, g_M)$ belong to $\mathcal{R}$ if it has a θ-proximate $(p_0, \ldots, p_M)$, in the correlated agreement sense, the multilinear representations $P_0, \ldots, P_M$ of which satisfies the evaluation claims.) The proximity parameter θ is taken from the list decoding regime, that is $\theta = 1 - \alpha$ with

$$\alpha = \left(1 + \frac{1}{2m}\right) \cdot \sqrt{\rho},$$

where $\rho = 2^n/|D|$ is the rate of the Reed-Solomon code $\mathscr{C} = \mathsf{RS}_{2^n}[F, D]$, and $m \geq 3$ is the multiplicity parameter of the Guruswami-Sudan list decoder. As the soundness proof of FRI [BCI$^+$20, Chapter 8.2], the analysis of Basefold is done in a round-by-round manner, considering each round of the protocol as a probabilistic reduction from one relation to another, simpler relation, thereby defining its own soundness error. These errors are runtime independent and depend only on protocol parameters, making a translation into the formal terms of round-by-round soundness straight-forward. We postpone that step to an extended version of the writeup.

Although we confine ourselves to the list decoding regime, we stress the fact that this is merely for brevity. The entire analysis can be carried over verbatim to the unique decoding regime, using the respective adaptions of Theorem 3 and Theorem 4.

Let us denote $\mathscr{F}_i$ the space of polynomials of the Reed-Solomon code $\mathscr{C}_i = \mathsf{RS}_{2^{n-i}}[F, D_i]$ over the projected domain $D_i = \pi^i(D)$, $i = 0, \ldots, n$. As sketched in Section 3.2, the linear subcodes $\mathscr{C}'_i < \mathscr{C}_i$, defined by the subspaces

$$\mathscr{F}'_i = \{p(X) \in \mathscr{F}_i \,:\, P(\omega_{i+1}, \ldots, \omega_n) = 0\}, \tag{16}$$

will play a crucial role in the soundness analysis of Protocol 2. This is the "punctured" space of all polynomials from $\mathscr{F}_i$, the multilinear representation of which has a zero at the given query. (In the edge case $i = n$, definition (16) is understood as $\mathscr{F}'_n = \{0\}$.)

For an agreement parameter $\alpha \in (0, 1)$, and integer r, $0 \leq r \leq n$, we say that a prover P^* *succeeds the commitment phase* with r rounds, if in interaction with the verifier it is able to provide

$$f_0, \Lambda_0, f_1, \Lambda_1, f_2, \Lambda_2, \ldots, \Lambda_{r-1}, \text{ and } f_r,$$

such that with $q_{-1}(X) := \sum X^k \cdot v_k$ and the other $q_i(X)$ as in (4) we have

$$q_{i-1}(\lambda_i) = q_i(0) + q_i(1),$$

for all $i = 0, \ldots, r-1$, and $f_0, f_1, \ldots, f_r$ define a sufficiently consistent fold-down of $g_0, \ldots, g_M$ to a sumcheck compliant codeword from $\mathscr{C}_r$. That is, there exists

$p_r(X) \in \mathscr{F}_r$ with its multilinear representation P_r satisfying

$$\begin{aligned} L((\omega_1, \ldots, \omega_r), (\lambda_1, \ldots, \lambda_r)) \cdot P_r(\omega_{r+1}, \ldots, \omega_n) &= q_{r-1}(\lambda_r) \\ &= L((\omega_1, \ldots, \omega_r), (\lambda_1, \ldots, \lambda_r)) \cdot \Lambda_{r-1}(\lambda_r) \end{aligned} \quad (17)$$

and for which

$$\left| \left\{ x \in D_0 \ : \ \begin{array}{c} (f_0, \ldots, f_r) \text{ satisfy all folding checks along } x \\ \wedge\ f_r(\pi^r(x)) = p_r(\pi^r(x)) \end{array} \right\} \right| \geq \alpha \cdot |D_0|. \quad (18)$$

(Note that in the edge case $r = 0$, Equation (17) collapses to $P_r(\omega_1, \ldots, \omega_n) = q_{-1}(\lambda_0) = \sum_k \lambda_0^k \cdot v_k$, and for $r = n$, $P_r(\omega_{r+1}, \ldots, \omega_n)$ in Equation 17 collapses to the constant $P_r \in F$.) Any such $(f_0, \Lambda_0, f_1, \Lambda_1, \ldots, f_r)$ will be called *α-good for* $(\lambda_0, \ldots, \lambda_r)$.

The soundness proof of Protocol 2 goes along the following lines. Starting with the relation $\mathcal{R}_{-1} = \mathcal{R}$ as defined above, each round of the commit phase, $0 \leq r \leq n$, is a randomized reduction from a transcript

$$\mathsf{tr}_{r-1} = (\lambda_0, f_0, \Lambda_0, \lambda_1, f_1, \Lambda_1, \ldots, \lambda_{i-1}, f_{i-1}, \Lambda_{r-1})$$

belonging $\mathcal{R}_{i-1}$ (where for $r = 0$ we take $\mathsf{tr}_{-1} = (g_0, \ldots, g_M)$) so that its continuation

$$\mathsf{tr}_i = \mathsf{tr}_{i-1} \| (\lambda_i, f_i, \Lambda_i)$$

is a member of

$$\mathcal{R}_i = \left\{ (\lambda_0, f_0, \Lambda_0, \ldots, \lambda_i, f_i, \Lambda_i) \ : \ \begin{array}{c} (f_0, \Lambda_0, \ldots, f_i) \\ \text{is } \alpha\text{-good for} (\lambda_0, \ldots, \lambda_i) \end{array} \right\},$$

with α-goodness as defined above. (In the edge case $i = r$, there is no Λ_r in the definition of $\mathcal{R}_r$.) The error of such a reduction step will be given from the subcode correlated agreement theorem (Theorem 3) in the first step, and its weighted variant, Theorem 4, in the FFT-like folding steps. The overall error is then dominated by the sum of the round-wise errors.

Lemma 1 (Soundness commit phase)**.** *Take a proximity parameter $\theta = 1 - \left(1 + \frac{1}{2 \cdot m}\right) \cdot \sqrt{\rho}$, with $m \geq 3$. Suppose that a (possibly computationally unbounded) algorithm P^* succeeds the commitment phase with $r \geq 0$ rounds with probability larger than*

$$\varepsilon_C = \varepsilon_0 + \varepsilon_1 + \ldots + \varepsilon_r,$$

where $\varepsilon_0 = \varepsilon(\mathscr{C}_i, M, \theta)$ is the soundness error from Theorem 3, and

$$\varepsilon_i := \varepsilon(\mathscr{C}_i, 1, B_i, \theta) + \frac{1}{|F|},$$

with $\varepsilon(\mathscr{C}_i, 1, B_i, \theta)$ being the soundness error from Theorem 4, where $B_i = |D|/|D_i| = 2^i$. Then $(g_0, \ldots, g_M)$ belongs to $\mathcal{R}$.

Proof. We prove the Lemma by induction on r, $0 \leq r \leq n$. We start by proving the base case $r = 0$. If with probability greater than $\varepsilon_0 = \varepsilon(\mathscr{C}_0, M, \theta)$, the prover is able to answer with an α-good $f_0 \in F^{D_0}$, where $\alpha = 1 - \theta$, then in particular the folding of the centered functions $g'_k = g_k - v_k$ are θ-proximate to the subcode $\mathscr{C}'_0$ of $\mathscr{C}_0$, with probability

$$\Pr\left[\lambda_0 \ : \ \exists p'_0 \in \mathscr{F}'_0 \text{ s.t. } \mathsf{agree}\left(\sum_{k=0}^{M} g'_k \cdot \lambda_0^k,\ p'_0(X)\right) \geq \alpha\right] > \varepsilon(\mathscr{C}_0, M, \theta),$$

where $\mathscr{F}'_0$ is as defined above. By the correlated agreement theorem for subcodes, Theorem 3, we conclude that there exists polynomials

$$p'_0(X), \ldots, p'_M(X) \in \mathscr{F}'_0$$

which agree with $g_0, \ldots, g_M$ on a joint set of density $\geq \alpha$. Over the same set the non-centered polynomials

$$p'_0(X) + v_0, \ldots, p'_M(X) + v_M \in \mathscr{F}_0$$

agree with $g_0, \ldots, g_M$. Their multilinear representations $P_k \in F[X_1, \ldots, X_n]$ satisfy $P_k(\vec{\omega}) = v_k$, showing that $(g_1, \ldots, g_M) \in \mathcal{R}$.

Assume that the Lemma holds for r, where $0 \leq r < n$, and that a prover P^* succeeds the commitment phase for $(r+1)$ rounds with probability greater than $(\varepsilon_1 + \ldots + \varepsilon_r) + \varepsilon_{r+1}$. Then the set $\mathfrak{T}$ of transcripts $\mathsf{tr}_r = (\lambda_0, f_0, \Lambda_0, \ldots, \lambda_r, f_r, \Lambda_r)$ for which the conditional success probability of P^* is greater than ε_{r+1}, and thus

$$\Pr\left[\lambda_{r+1} \ : \ \begin{array}{c}\exists f_{r+1} \text{ s.t. } \big(f_0, \Lambda_0, \ldots, f_r, \Lambda_r, f_{r+1})\big) \\ \text{is } \alpha\text{-good for } (\lambda_0, \ldots, \lambda_{r+1})\end{array}\right] > \varepsilon_{r+1},$$

has probability $\Pr[\mathfrak{T}] > \varepsilon_0 + \ldots + \varepsilon_r$. By the definition of α-goodness, for each of these λ_{r+1} there exists a sumcheck compliant polynomial $p_{r+1} \in \mathscr{F}_{r+1}$ so that

$$\mathsf{agree}_{\nu_r}((1 - \lambda_{r+1}) \cdot f_{r,0} + \lambda_{r+1} \cdot f_{r,1},\ p_{r+1}) \geq \alpha,$$

where ν_r is the sub-probability measure with density function

$$\delta_r(y) := \frac{\left|\left\{x \in \pi^{-(r+1)}(y) \ : \ (f_0, \ldots, f_r) \text{ satisfies all folding checks along } x\right\}\right|}{\left|\pi^{-(r+1)}(y)\right|},$$

for $y \in D_{r+1}$. (The claimed weighted agreement is an application of the law of total probability, using the conditional probabilities given by $\delta_r(y)$.) Using sumcheck compliance as in our proof sketch from Section 3.2, we conclude that except for a set of λ_{r+1} of probability $1/|F|$ (for the possible zero of the Lagrangian in formula (4)), and thus still of probability greater than

$$\varepsilon_{r+1} - \frac{1}{|F|} = \varepsilon(\mathscr{C}_{r+1}, 1, B_{r+1}, \theta),$$

the polynomial $p'_{r+1} = p_{r+1} - \Lambda_r(\lambda_{r+1})$ is from the subcode $\mathscr{F}'_{r+1}$ as defined above, and the centered parts $f'_{r,0} = f_{r,0} - \Lambda_r(0)$ and $f'_{r,1} = f_{r,1} - \Lambda_r(1)$ satisfy

$$\mathsf{agree}_{\nu_r}((1-\lambda_{r+1}) \cdot f'_{r,0} + \lambda_{r+1} \cdot f'_{r,1},\ p'_{r+1}) \geq \alpha,$$

altogether

$$\Pr\Big[\lambda_{r+1} \ : \ \begin{array}{c}\exists p'_{r+1} \in \mathscr{F}'_{r+1} \text{ s.t.} \\ \mathrm{agree}_{\nu_r}\left((1-\lambda_{r+1}) \cdot f'_{r,0} + \lambda_{r+1} \cdot f'_{r,1}, p'_{r+1}\right) \geq \alpha,\end{array}\Big] > \varepsilon(\mathscr{C}_{r+1}, 1, B_{r+1}, \theta).$$

By Theorem 4, we conclude that $f'_{r,0}$ and $f'_{r,1}$ agree with some

$$p'_{r,0}(X), p'_{r,1}(X) \in \mathscr{F}'_{r+1}$$

on a set $A_{r+1} \subseteq D_{r+1}$ of weight $\nu_r(A_{r+1}) \geq 1 - \theta$. Over the same set, the non-centered parts $f_{r,0}$ and $f_{r,1}$ agree with

$$p_{r,0}(X) = p'_{r,0}(X) + \Lambda_r(0), \quad p_{r,1}(X) = p'_{r,1}(X) + \Lambda_r(1) \in \mathscr{F}_{r+1},$$

the multilinear representations $P_{r,0}$, $P_{r,1}$ of which satisfy

$$\begin{aligned} P_{r,0}(\omega_{r+2}, \ldots, \omega_n) &= \Lambda_r(0), \\ P_{r,1}(\omega_{r+2}, \ldots, \omega_n) &= \Lambda_r(1). \end{aligned}$$

(In the edge case $r+1 = n$ the two equations collapse to $P_{n-1,0} = \Lambda_{n-1}(0)$ and $P_{n-1,1} = \Lambda_{n-1}(1)$.) Over the preimage $A_r = \pi^{-1}(A_{r+1})$, the word f_r coincides with the values of

$$p_r(X) = p_{r,0}(X^2) + X \cdot p_{r,1}(X^2) \in \mathscr{F}_r,$$

and by construction its multilinear representation P_r satisfies

$$\begin{aligned} P_r(\omega_{r+1}, \omega_{r+2}, \ldots, \omega_n) &= (1-\omega_{r+1}) \cdot \Lambda_r(0) + \omega_{r+1} \cdot \Lambda_r(1) \\ &= L(\omega_{r+1}, 0) \cdot \Lambda_r(0) + L(\omega_{r+1}, 1) \cdot \Lambda_r(1), \end{aligned}$$

and hence is sumcheck compliant,

$$L(\omega_1, \ldots, \omega_r, \lambda_1, \ldots, \lambda_r) \cdot P_r(\omega_{r+1}, \omega_{r+2}, \ldots, \omega_n) = q_r(0) + q_r(1) = q_{r-1}(\lambda_r).$$

Furthermore, the set of $x \in \pi^{-r}(A_r)$ on which all folding checks against hold against $f_0, \ldots, f_r$, is of density

$$\begin{aligned} &\frac{|\{x \in \pi^{-r}(A_r) \ : \ \text{all folding checks hold for } f_0, \ldots, f_r\}|}{|D_0|} \\ &\quad = \frac{1}{|D_0|} \cdot \sum_{y \in A_{r+1}} \delta(y) \cdot \left|\pi^{-(r+1)}(y)\right| = \frac{1}{|D_{r+1}|} \cdot \sum_{y \in A_{r+1}} \delta(y) = \nu_r(A_{r+1}), \end{aligned}$$

which is at least α. Putting everything together, the prover message $(f_0, \Lambda_0, \ldots, f_r)$ is α-good for $(\lambda_0, \ldots, \lambda_r)$. Since the probability of P^* producing such a trace is greater than $\varepsilon_1 + \ldots + \varepsilon_r$, we conclude from the induction hypothesis that $(g_0, \ldots, g_M)$ belongs to $\mathcal{R}$, completing the proof of the Lemma. $\square$

With Lemma 1 the soundness error of the entire Protocol 2 is obtained within a minor additional step: Considering the query phase as probabilistic reduction from $\mathcal{R}_n$ to the verifier relation

$$\mathcal{R}_{n+1} = \left\{ \begin{matrix} (\lambda_0, f_0, \Lambda_0, \ldots, \lambda_n, f_n), \\ (x_1, \ldots, x_s) \end{matrix} \; : \; \begin{matrix} f_n \in \mathscr{C}_n \wedge \text{ all verifier checks hold for} \\ (\lambda_0, \ldots, \lambda_n) \text{ and samples } x_1, \ldots x_s \end{matrix} \right\},$$

its soundness error corresponds to the probability that a prover message $t = (f_0, \Lambda_0, \ldots, f_n)$ which is not α-good for $(\lambda_0, \ldots, \lambda_n)$ being not detected by s random samples. A message t is not α-good, if either one of the sumcheck equations fail, or the folding consistency set is of density $< \alpha$. The former is always detected, and the latter is detected except with probability

$$\varepsilon_Q < \alpha^s.$$

Eventually, the soundness error of the entire oracle proof is dominated by sum of the roundwise errors

$$\varepsilon < \varepsilon_C + (1 - \theta)^s,$$

completing the proof of Theorem 1.

6 Generalizations

In this section we discuss several generalizations of Basefold, and how the soundness analysis from Section 5 extends to these cases. First of all, in Section 6.1, we quickly discuss the adaption of Basefold to different FFT environments, with emphasis on the additive FFT over binary fields. Then in Section 6.2, we generalize the type of sumcheck expressions to be proven by Basefold, covering multi-query evaluation proofs as well as the recent FRI-Binius optimization [Dia24b], which we shortly discuss in Section 6.3.

6.1 Other FFTs

Adapting Basefold to other FFT-encodable algebraic geometry codes is merely a technical step. For simplicity, and in particular in view of Section 6.3, we restrict ourselves to Reed-Solomon codes over binary fields, with the *additive FFT* [LCH14] used for encoding. Other codes such as EC-FFT codes [BCKL21] or Circle Codes [HLP24] are treated likewise.

Let F be a finite binary field, i.e. a finite field of characteristic 2. Instead of subgroups of the multiplicative group, the additive FFT takes subgroups of the additive group, i.e. $\mathbb{F}_2$-linear subspaces of F as the domains. Similar to the regular case, the FFT works along a chain of projected subspaces

$$U = U_0 \xrightarrow{\pi_1} U_1 \xrightarrow{\pi_2} \ldots \xrightarrow{\pi_n} U_n, \tag{19}$$

where

$$\pi_i(x) = x \cdot (x - b_i),$$

with $b_i \in F$, are suitably chosen quadratic maps[4]which halve the size of the subspaces in each of the steps, until ending up with a singleton domain U_n. (Recall that quadratic maps are $\mathbb{F}_2$-linear.) As "twiddle functions" $t_i : U_i \longrightarrow F$ one may take again $t_i(X) = X$, for $i = 0, \ldots, n-1$, or a properly normalized[5]variant if one desires an optimized butterfly network.

Given a function $f \in F^U$ over $U = U_0$, with values in F (or more generally, an extension of F), the additive FFT computes the coefficients of the interpolant

$$p(X) = \sum_{i=0}^{2^n-1} c_i \cdot b_{n,i}(X)$$

with respect to a polynomial basis $\mathscr{B}_n = \{b_{n,i}(X)\}$, which is different to the monomial one, defined by the projection chain and the twiddle functions,

$$b_{n,i}(X) = t_0(X)^{i_0} \cdot (t_1 \circ \pi_1)(X)^{i_1} \cdot \ldots \cdot (t_{n-1} \circ \pi_{n-1} \circ \ldots \circ \pi_1)(X)^{i_{n-1}},$$

where $(i_0, \ldots, i_{n-1})$ are the bits of $i = \sum_{k=0}^{n-1} i_k \cdot 2^k$. By the degrees of t_i and π_i, these polynomials are of degree less than 2^n and hence form a basis of $\mathscr{P}_n$. As in the multiplicative case, the algorithm is based on the decomposition of a function $f_i \in K^{U_i}$ into "even" and "odd" parts,

$$f_i(x) = f_{i,0}(\pi_{i+1}(x)) + t_i(x) \cdot f_{i,1}(\pi_{i+1}(x)), \tag{20}$$

which are computed by value, using a "butterfly" along the fibers of the projection, which are of the form $\{x, x + b_{i+1}\}$, see formula (21) below.

Likewise *additive FRI* [BBHR18], which proves proximity to the Reed-Solomon code $\mathscr{C}_0 = \mathsf{RS}_{2^n}[F, D]$ over an evaluation domain D, a disjoint coset union of U_0, goes along the projected domains

$$D = D_0 \xrightarrow{\pi_1} D_1 \xrightarrow{\pi_2} \ldots \xrightarrow{\pi_n} D_n,$$

which are also halved in each step. Starting with $f_0 \in F^D$, FRI recursively takes random linear combinations of the "even" and "odd" parts,

$$f_{i+1}(\pi_{i+1}(x)) = f_{i,0}(\pi_{i+1}(x)) + \lambda_i \cdot f_{i,1}(\pi_{i+1}(x)),$$

thereby reducing the initial proximity claim to gradually simpler claims, which are with respect to the codes $\mathscr{C}_i = \mathsf{RS}_{2^{n-i}}[F, D_i]$ over the projected domains, for $i = 1, \ldots, n$.

[4]In fact, if $U_i = u_i + V_i$ with V_i being a non-affine subspace, then any quadratic map of the above form, which annihilates a basis vector of V_i, is fine.

[5]The proper normalization is so that $t_i(x + b_{i+1}) = t_i(x) + 1$. Equivalently, one may integrate normalization into the projection, see [GM10, LCH14, DP24].

In regards to Basefold, a multilinear $P \in F[X_1, \ldots, X_n]$ is again represented as a univariate polynomial $p(X) \in \mathscr{P}_n$ via

$$p(X) = \sum_{i=0}^{2^n-1} P(i_0, \ldots, i_{n-1}) \cdot b_{n,i}(X),$$

and committed as word from $\mathscr{C}_0$. By the recursive structure of the polynomial basis, this identification preserves the connection between partial substitution and FRI-like folding of even and odd parts. With this different FFT environment, Protocol 1 and 2 remain essentially unchanged, except that "even" and "odd" parts are computed (and verified) via

$$\begin{pmatrix} f_{i,0}(\pi_{i+1}(x)) \\ f_{i,1}(\pi_{i+1}(x)) \end{pmatrix} = \begin{pmatrix} 1 & t_i(x) \\ 1 & t_i(x + b_{i+1}) \end{pmatrix}^{-1} \cdot \begin{pmatrix} f_i(x) \\ f_i(x + b_{i+1}) \end{pmatrix}. \tag{21}$$

The subcode correlated agreement theorem from Section 4.1 is valid for fields over arbitrary characteristic (as does its non-linear generalization), and hence Theorem 1 with its proof from Section 5 holds verbatim.

6.2 More expressive inner products

Protocol 1 and its batched variant Protocol 2 are easily extended to more expressive hypercube sums than the evaluation inner product for a single query on the multilinears $G_0, \ldots, G_M$. For simplicity, we restrict to assertions of the form

$$\langle G_k, R \rangle_{H_n} = v_k, \tag{22}$$

sharing the same $R \in F[X_1, \ldots, X_n]$ for $k = 0, \ldots, M$, any succinct evaluable multilinear. This simple case covers linear combinations of Lagrangians, as used in multi-point evaluation proofs, or the expression used by the row-batched optimization of FRI-Binius, which we address in Section 6.3.

For assertions of the form (22), the sumcheck is still quadratic, but since R is in general not decomposable as a Lagrangian $L(\,.\,, \vec{\omega})$, we cannot work with linear refinement polynomials as we did in Section 3. This rules out a reduction of the soundness analysis to correlated agreement for linear subcodes, Theorem 4. Instead, we return to the regular sumcheck convention and let the prover directly provide the quadratic polynomials

$$q_i(X) = \big\langle P(\lambda_1, \ldots, \lambda_i, X, \,.\,), R(\lambda_1, \ldots, \lambda_i, X, \,.\,) \big\rangle_{H_{n-i-1}},$$

in each of the rounds $i = 0, \ldots, n-1$, and correlated agreement with the needed additional algebraic properties will be guaranteed by the non-linear generalization Theorem 5.

The differences to Section 5 are again merely technical. The relation to be proven is

$$\mathcal{R} = \left\{ \begin{array}{c} \exists p_0, \ldots, p_M \in \mathscr{F}[X]^{<2^n} \text{ s.t.} \\ (g_0, \ldots, g_M) \ : \ d((g_0, \ldots, g_M), (p_0, \ldots, p_M)) < \theta \\ \wedge \bigwedge_{k=0}^{M} \langle P_k, R \rangle_{H_n} = v_k \end{array} \right\}, \tag{23}$$

and the further relations $\mathcal{R}_1, \ldots, \mathcal{R}_n$, including the notion of prover *success* and α-goodness of prover messages are as in Section 5, except for replacing $\Lambda_i(X)$ by $q_i(X)$, and adapting the sumcheck compliance formula accordingly. Soundness of the commit phase, is as done for Lemma 1, with the main difference that we work with non-linear constraints imposed by the sumcheck refinements, and no centering is needed. Notably, the use of Theorem 5 slightly tightens the soundness error from Lemma 1, allowing the sumcheck error being completely covered by the soundness error of the correlated agreement theorem.

Lemma 2 (Soundness commit phase)**.** *Take a proximity parameter* $\theta = 1 - \left(1 + \frac{1}{2 \cdot m}\right) \cdot \sqrt{\rho}$*, with* $m \geq 3$*. Suppose that a (possibly computationally unbounded) algorithm* P^* *succeeds the commitment phase with* $r \geq 0$ *rounds with probability larger than*

$$\varepsilon_C = \varepsilon_0 + \varepsilon_1 + \ldots + \varepsilon_r,$$

where

$$\varepsilon_i := \varepsilon(\mathscr{C}_i, M_i, B_i, \theta)$$

is the soundness error from Theorem 5 for $M_0 = M$*,* $B_0 = 1$*, and* $M_i = 1$*,* $B_i = |D|/|D_i| = 2^i$ *otherwise. Then* $(g_0, \ldots, g_M)$ *belongs to* $\mathcal{R}$*.*

Proof. We only point out the differences to the proof of Lemma 1, which essentially are due to the usage of Theorem 5 and its non-linear constraint Γ.

In the base case $r = 0$ for the batching round, we choose the constraint $\Gamma_0 : \mathscr{F}_0 \times F \longrightarrow F$ defined by

$$\Gamma_0(p, z) = \langle P, R \rangle_{H_n} - \sum_{k=0}^{M} v_k \cdot z^k,$$

which is of degree M in z, and linear in the coordinates of $\mathscr{P}$. In particular, $\Gamma(p_0 + Z \cdot p_1 + \ldots + Z^M \cdot p_M, Z)$ is still of degree $\leq M$, and by the discussion preceding Theorem 5 shows the existence of

$$p_0(X), \ldots, p_M(X) \in \mathscr{F}_0$$

which agree with $g_0, \ldots, g_M$ on a joint set of density $\geq 1 - \theta$, and moreover $\Gamma_0(p_0 + Z \cdot p_1 + \ldots + Z^M \cdot p_M, Z) = 0 \in F[Z]$. Thus,

$$\sum_{k=0}^{M} (\langle P_k, R \rangle_{H_n} - v_k) \cdot Z^k = 0$$

as a formal identity, showing that each P_k satisfies its sumcheck claim.

In the induction step from r to $r+1$, where $1 \leq r \leq n-1$, the non-linear constraint $\Gamma_r : \mathscr{F}_{r+1} \times F \longrightarrow F$ is

$$\Gamma_r(p, z) = \langle P, R(\lambda_1, \ldots, \lambda_r, z, \ .\ )\rangle_{H_{n-r-1}} - q_r(z),$$

where $q_r(z)$ is the claimed sumcheck polynomial. Here, the total degree of Γ_r is $d = 2$. (In the edge case $r = n-1$, $\mathscr{F}_{r+1}$ is the space of constants $\mathscr{F}_n$, and the inner product reduces to the singleton sum $\Gamma_{n-1}(p, z) = p \cdot R(\lambda_1, \ldots, \lambda_{n-1}, z)$.) By the induction assumption, the set $\mathfrak{T}$ of up-to-round-r transcripts for which

$$\Pr\left[\lambda_{r+1} \ : \ \begin{array}{c} \exists p_{r+1} \in \mathscr{F}_{r+1} \text{ s.t. } \Gamma(p_{r+1}, \lambda_{r+1}) = 0 \\ \wedge\, \mathrm{agree}_{\nu_r}\left((1-\lambda_{r+1}) \cdot f_{r,0} + \lambda_{r+1} \cdot f_{r,1}, p_{r+1}\right) \geq \alpha \end{array}\right] > \varepsilon_{r+1} = \varepsilon(\mathscr{C}_{r+1}, 1, B_{r+1}, \theta),$$

is of probability $\Pr[\mathfrak{T}] > \varepsilon_0 + \ldots + \varepsilon_r$. (We use the same sub-probability measure ν_r as in Section 5). We apply Theorem 5 to conclude that $f_{r,0}$ and $f_{r,1}$ agree over a joint set A_{r+1} of density $\nu_r(A_{r+1}) \geq 1 - \theta$ with polynomials

$$p_{r,0}(X), p_{r,1}(X) \in \mathscr{F}_{r+1},$$

satisfying $\Gamma_r((1-Z) \cdot p_{r,0} + Z \cdot p_{r,1}, Z) = 0 \in F[Z]$. In particular $\Gamma(p_{r,0}, 0) = \Gamma(p_{r,1}, 1) = 0$, meaning that their multilinear representations $P_{r,0}$ and $P_{r,1}$ satisfy

$$\begin{aligned} \langle P_{r,0}, R(\lambda_1, \ldots, \lambda_r, 0)\rangle_{H_{n-r-1}} &= q_r(0), \\ \langle P_{r,1}, R(\lambda_1, \ldots, \lambda_r, 1)\rangle_{H_{n-r-1}} &= q_r(1). \end{aligned}$$

Over the preimage $A_r = \pi^{-1}(A_{r+1})$, the word f_r coincides with the values of

$$p_r(X) = p_{r,0}(X^2) + X \cdot p_{r,1}(X^2) \in \mathscr{F}_r,$$

and by construction its multilinear representation P_r satisfies

$$\begin{aligned} \langle P_r, R(\lambda_1, \ldots, \lambda_r, \ .\ )\rangle_{H_{n-r}} &= \langle P_{r,0}, R(\lambda_1, \ldots, \lambda_r, 0)\rangle_{H_{n-r-1}} \\ &\quad + \langle P_{r,1}, R(\lambda_1, \ldots, \lambda_r, 1)\rangle_{H_{n-r-1}} = q_r(0) + q_r(1) = q_{r-1}(\lambda_r), \end{aligned}$$

and hence is sumcheck compliant. The density of the folding consistency set for the oracles $f_0, \ldots, f_r$ is shown as in Lemma 1, yielding that $(f_0, q_0(X), \ldots, f_r)$ is α-good for $(\lambda_1, \ldots, \lambda_r)$. This completes the proof of the lemma. □

6.3 FRI-Binius

FRI-Binius [DP24] is a *small-field polynomial commitment scheme*, in which multilinears over some small field F (for example, $F = \mathbb{F}_2$) are committed via a "packed" representation over some larger field

$$E > F.$$

We restrict ourselves to the case of binary fields and Reed-Solomon codes, and for brevity we omit certain details which are not essential for the core of the construction.

Let F be a field of characteristic two, and E be an extension field of dimension $d = \dim E/F$, with an (additive) FFT domain D of size $|D| > n$. Fix a basis $\{\beta_1, \ldots, \beta_d\}$ of E/F, then any collection[6]of small-field multilinears

$$P_1, \ldots P_d \in F[X_1, \ldots, X_n]$$

is committed via the "packed" multilinear

$$P = \sum_{i=1}^{d} \beta_i \cdot P_i \in E[X_1, \ldots, X_n], \tag{24}$$

over the larger field E, as code word f from $\mathscr{C} = \mathsf{RS}_{2^n}[E, D]$, using the same identification of multilinears and univariates as before. Conversely, given any $P \in E[X_1, \ldots, X_n]$ it can be uniquely decomposed into component polynomials $P_1, \ldots, P_d \in F[X_1, \ldots, X_n]$ satisfying (24).

Now, for a given a query $\vec{\omega} \in Q^n$ from any other field $Q \geq F$ (the *query field*, typically a cryptographically large extension), we wish to prove the values

$$v_i = P_i(\vec{\omega}) \in Q, \quad i = 1, \ldots, d, \tag{25}$$

by means of the packed polynomial P. While in [DP24], these claims are expressed as an evaluation inner product over the tensor algebra $Q \otimes E$, we largely avoid the notion of tensor algebras, and directly describe the row-batched optimization from [Dia24b]:

Let $d' = \dim Q/F$ and $\{\gamma_1, \ldots, \gamma_{d'}\}$ be a basis of Q/F, and write

$$v_i = \sum_{j=1}^{d'} v_{i,j} \cdot \gamma_j \quad \text{and} \quad L(\vec{\omega}, \,.\,) = \sum_{j=1}^{d'} \gamma_j \cdot L_{\vec{\omega},j}(\,.\,) \tag{26}$$

with $v_{i,j} \in F$ and component multilinears $L_{\vec{\omega},j} \in F[X_1, \ldots, X_n]$. Then the claims (25) are equivalent to that

$$\langle P, L_{\vec{\omega},j} \rangle_{H_n} = \sum_{i=1}^{d} \beta_i \cdot \langle P_i, L_{\vec{\omega},j} \rangle_{H_n} = \sum_{i=1}^{d} \beta_i \cdot v_{i,j},$$

for each $j = 1, \ldots, d'$, which is then proven via the random linear combination

$$\left\langle P, \sum_{j=1}^{d'} \lambda^{j-1} \cdot L_{\vec{\omega},j} \right\rangle_{H_n} = \sum_{j=1}^{d'} \lambda^{j-1} \cdot \sum_{i=1}^{d} \beta_i \cdot v_{i,j}, \tag{27}$$

[6]In FRI-Binius these polynomials are a Lagrange decomposition of a larger multilinear over F.

where λ is taken from a cryptographically large extension K of E. This is done using Basefold from Section 6.2, adapted to the additive FFT as described in Section 6.1. Note that the polynomials $L_{\vec{\omega},j}$ are multilinear, and more importantly, they can be evaluated simultaneously, *still in a succinct manner*, at any point $\vec{\lambda} \in K^n$. (This can be done by means of tensor algebra operations[7], see [DP24].) The protocol for a batch of packed multilinears is as follows.

Protocol 3 (Optimized FRI-Binius [Dia24b])**.** *Let F, E and Q as above. Given multilinears $G_0, \ldots, G_M \in E[X_1, \ldots, X_n]$ over E, committed as Reed-Solomon code words $g_0, \ldots, g_M \in \mathscr{C} = \mathsf{RS}_{2^n}[E, D]$, and evaluation claims $v_i^{(k)} \in Q$, $i = 1, \ldots, d$, $k = 0, \ldots, M$ for their component polynomials $G_{k,i}$, at some query $\vec{\omega} \in Q^n$.*

1. *The verifier sends a random $\lambda \leftarrow\!\!\$\ K$, from an extension field K of E, to the prover.*
2. *Both prover and the verifier now run Basefold from Section 6.2, with proximity parameter $\theta = (1 + \frac{1}{2m}) \cdot \sqrt{\rho}$ and $s \geq 1$ samples, on $G_0, \ldots, G_M$ for their inner products $\langle G_k, L_{\vec{\omega},\lambda} \rangle_{H_n} = v_\lambda^{(k)}$, $0 \leq k \leq M$, where*

$$L_{\vec{\omega},\lambda} = \sum_{j=1}^{d'} \lambda^{j-1} \cdot L_{\vec{\omega},j}, \qquad v_\lambda^{(k)} = \sum_{j=1}^{d'} \sum_{i=1}^{d} \lambda^{j-1} \cdot \beta_i \cdot v_{i,j}^{(k)}. \tag{28}$$

   using the decomposition of $L_{\vec{\omega}}$ and $v_i^{(k)}$ as in (26)*.*

Protocol 3 is an interactive oracle proof for the relation

$$\mathcal{R} = \left\{ (g_0, \ldots, g_M) \ : \ \begin{array}{c} \exists p_0, \ldots, p_M \in E[X]^{<2^n} \text{ s.t.} \\ d((g_0, \ldots, g_M), (p_0, \ldots, p_M)) < \theta \\ \wedge \bigwedge_{k=0}^{M} \bigwedge_{i=1}^{[E:F]} P_{k,i}(\omega_1, \ldots, \omega_n) = v_i^{(k)} \end{array} \right\}, \tag{29}$$

and its soundness error is the soundness error of Basefold, plus the error the first round of the protocol, which is random reduction from $\mathcal{R}$ to

$$\mathcal{R}' = \left\{ (g_0, \ldots, g_M) \ : \ \begin{array}{c} \exists p_0, \ldots, p_M \in E[X]^{<2^n} \text{ s.t.} \\ d((g_0, \ldots, g_M), (p_0, \ldots, p_M)) < \theta \\ \wedge \bigwedge_{k=0}^{M} \langle P_k, L_{\vec{\omega},\lambda} \rangle_{H_n} = v_\lambda^{(k)} \end{array} \right\}. \tag{30}$$

The error of this round is proven as for any other interactive oracle proof on top of Basefold, taking into account the size of

$$\mathcal{L} = \left\{ (p_0, \ldots, p_M) \in \left(E[X]^{<2^n} \right)^M \ : \ d((p_0, \ldots, p_M), (g_0, \ldots, g_M)) < \theta \right\},$$

[7] The values $L_{\vec{\omega},j}(\vec{\lambda})$ are the rows of the Lagrangian $L(\vec{\omega}^t, \vec{\lambda})$ evaluated in the tensor algebra $Q \otimes K$, where $\vec{\omega}^t$ is an element in the vertical embedding of Q, and $\vec{\lambda}$ is an element in the horizontal embedding of K.

the list of all correlated θ-proximates, which is bounded by

$$|\mathcal{L}| \leq \frac{m+\frac{1}{2}}{\sqrt{\rho}},$$

the Guruswami-Sudan bound for $\mathsf{RS}_{2^n}[E(Z), D]$, over $E(Z)$ the field of rational functions over E.[8]

Theorem 7 (FRI-Binius soundness)**.** *The soundness error ε of Protocol 3 as an interactive oracle proof for the relation $\mathcal{R}$ defined in* (29)*, is bounded by*

$$\varepsilon < \ell_m \cdot \frac{d'-1}{|K|} + \varepsilon_{BF}(\mathscr{C}, M, \theta, s),$$

where $\varepsilon_{BF}(\mathscr{C}, M, \theta, s)$ is the soundness error of Basefold for inner products (Section 6.2) for the code $\mathscr{C} = \mathsf{RS}_{2^n}[K, D]$ over the extension field K, with batch size $M+1$, proximity parameter $\theta = \left(1+\frac{1}{2m}\right)\cdot\sqrt{\rho}$, where $m \geq 3$, and $s \geq 1$ samples, and $\ell_m = \frac{m+\frac{1}{2}}{\sqrt{\rho}}$ is the Guruswami-Sudan list size bound.

Proof. Write $\varepsilon_1 = \ell_m \cdot \frac{d'-1}{|K|}$ and $\varepsilon_2 = \varepsilon_{BF}(\mathscr{C}, M, \theta, s)$, and let $\mathcal{L}$ be the (possibly empty) list of all correlated θ-proximates of $(g_0, \ldots, g_M)$, bounded by ℓ_m as above.

Assume that a (possibly computationally unbounded) algorithm P^* passes the verifier on given words $g_0, \ldots, g_M \in E^D$ and claims $v_i^{(k)} \in Q$, $1 \leq i \leq d$, $0 \leq k \leq M$, with a probability larger than $\varepsilon_1 + \varepsilon_2$. Then the number of "good" λ, on which P^* is able to succeed the Basefold verifier with a (conditional) probability larger than ε_2, is at bounded from below by

$$|\{\lambda \in K \,:\, \Pr[P^* \text{ succeeds}|\lambda] > \varepsilon_2\}| > \ell_m \cdot (d'-1)$$

For each such "good" λ, the soundness of Basefold from Section 6.2 enforces that $(g_0, \ldots, g_M)$ belongs to the relation $\mathcal{R}'$ defined in (30), meaning there exists some $(p_0, \ldots, p_M) \in \mathcal{L}$ for which $\langle P_k, L_{\vec{\omega},\lambda}\rangle = v_\lambda^{(k)}$ for every $k = 0, \ldots, M$. By the size bound on $\mathcal{L}$ and the pigeon-hole principle, at least one of the proximates from $\mathcal{L}$, which we again denote by $(p_0, \ldots, p_M)$, has $(d'-1)$ different $\lambda_1, \ldots, \lambda_{d'-1}$ for which

$$\langle P_k, L_{\vec{\omega},\lambda_i}\rangle = v_{\lambda_i}^{(k)},$$

for every $i = 1, \ldots, d'-1$ and $k = 0, \ldots, M$. Both $L_{\vec{\omega},\lambda}$ and $v_\lambda^{(k)}$ are polynomials in λ of degree at most $d'-1$, and we conclude from their definitions in (28) that

$$\langle P_k, L_{\vec{\omega},j}\rangle = \sum_{i=1}^{d} \beta_i \cdot v_{i,j}^{(k)},$$

for every $j = 1, \ldots, d'$ and $k = 0, \ldots, M$. In other words, $\langle P_{k,i}, L_{\vec{\omega},j}\rangle = v_i^{(k)}$, for all i and k. This shows that $(g_0, \ldots, g_M)$ belongs to $\mathcal{R}$, proving the theorem. □

[8] The Guruswami-Sudan list size bound applies to infinite fields as well. See the discussion in the appendix of [Hab22].

References

[ACFY24a] Gal Arnon, Alessandro Chiesa, Giacomo Fenzi, and Eylon Yogev. STIR: Reed–Solomon proximity testing with fewer queries. In *CRYPTO 2024*, volume 14929 of *LNCS*, 2024. Full paper: `https://eprint.iacr.org/2024/390`.

[ACFY24b] Gal Arnon, Alessandro Chiesa, Giacomo Fenzi, and Eylon Yogev. WHIR: Reed–solomon proximity testing with super-fast verification. 2024. in preparation.

[AFLN23] Martin R. Albrecht, Giacomo Fenzi, Oleksandra Lapiha, and Ngoc Khanh Nguyen. SLAP: Succinct lattice-based polynomial commitments from standard assumptions. 2023. `https://eprint.iacr.org/2023/1469`.

[AHIV17] Scott Ames, Carmit Hazay, Yuval Ishai, and Muthuramakrishnan Venkitasubramaniam. Ligero: Lightweight sublinear arguments without a trusted setup. In *CCS 2017*, 2017. Full paper: `https://eprint.iacr.org/2022/1608`.

[BBB+18] Benedikt Bünz, Jonathan Bootle, Dan Boneh, Andrew Poelstra, Pieter Wuille, and Greg Maxwell. Bulletproofs: Short proofs for confidential transactions and more. In *in IEEE Symposium on Security and Privacy*, pages 315–334, 2018. Full paper: `https://eprint.iacr.org/2017/1066`.

[BBHR18] Eli Ben-Sasson, Iddo Bentov, Yinon Horesh, and Michael Riabzev. Fast Reed-Solomon interactive oracle proofs of proximity. In *ICALP 2018*, 2018. Full paper: `https://eccc.weizmann.ac.il/report/2017/134/`.

[BCC+16] Jonathan Bootle, Andrea Cerulli, Pyrros Chaidos, Jens Groth, and Christophe Petit. Efficient zero-knowledge arguments for arithmetic circuits in the discrete log setting. In M. Fischlin and J.S. Coron, editors, *EUROCRYPT 2016*, volume 9666 of *LNCS*. Springer, 2016. Full paper: `https://eprint.iacr.org/2016/263`.

[BCHO22] Jonathan Bootle, Alessandro Chiesa, Yuncong Hu, and Michele Orr'u. Gemini: Elastic SNARKs for diverse environments. In *EUROCRYPT 2022*, 2022. Full paper: `https://eprint.iacr.org/2022/420`.

[BCI+20] Eli Ben-Sasson, Dan Carmon, Yuval Ishai, Swastik Kopparty, and Shubhangi Saraf. Proximity gaps for Reed-Solomon codes. In *FOCS 2020*, 2020. Full paper: `https://eprint.iacr.org/2020/654`.

[BCKL21] Eli Ben-Sasson, Dan Carmon, Swastik Kopparty, and David Levit. Elliptic Curve Fast Fourier Transform (ECFFT) Part I: Fast polynomial algorithms over all finite fields. In *Electronic Colloquium on Compputational Complexity*, volume TR21-103, 2021. https://eccc.weizmann.ac.il/report/2021/103/.

[BCKL22] Eli Ben-Sasson, Dan Carmon, Swastik Kopparty, and David Levit. Scalable and transparent proofs over all large fields, via elliptic curves (ECFFT part II). In *IACR preprint archive*, 2022. https://eprint.iacr.org/2022/1542.

[BCS16] Eli Ben-Sasson, Alessandro Chiesa, and Nicholas Spooner. Interactive oracle proofs. In *TCC 2016*, pages 31–60, 2016.

[BFS20] Benedikt Bünz, Ben Fisch, and Alan Szepieniec. Transparent SNARKs from DARK compilers. In *EUROCRYPT 2020*, 2020. Full paper: https://eprint.iacr.org/2019/1229.

[BGKS20] Eli Ben-Sasson, Lior Goldberg, Swastik Kopparty, and Shubhangi Saraf. DEEP-FRI: Sampling outside the box improves soundness. In *ITCS 2020*, 2020. Full paper: https://eprint.iacr.org/2019/336.

[BLNR22] Sarah Bordage, Mathieu Lhotel, Jade Nardi, and Hugues Randriam. Interactive oracle proofs of proximity to algebraic geometry codes. In *CCS'22*, 2022. Full paper: https://arxiv.org/abs/2011.04295.

[BM88] Lázló Babai and Shlomer Moran. Arthur-Merlin games: A randomized proof system and a hierarchy of complexity classes. In *Journal of Computer Sciences*, 1988.

[BS84] Lázló Babai and Endre Szemarédi. On the complexity classes of matrix group problems. In *Proc. of the 25th Symposium on Foundation of Computer Sciences*, 1984. https://doi.org/10.1109/SFCS.1984.71591.

[CBBZ22] Binyi Chen, Benedikt Bünz, Dan Boneh, and Zhenfei Zhang. Hyperplonk: PLONK with a linear-time prover and high-degree costum gates. In *IACR ePrint Archive 2020/1355*, 2022. https://eprint.iacr.org/2022/1355.

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

[CY24] Alessandro Chiesa and Eylon Yogev. Building cryptographic proofs from hash functions. 2024. https://snargsbook.org/.

[Dia24a] Benjamin E. Diamond. Galois FRI-Binius: Joseph Jonston's Idea. hackmd.io, 2024. https://hackmd.io/@benediamond/BkLYpf__0.

[Dia24b] Benjamin E. Diamond. Overhead-Free FRI-Binius Without Galois Theory. hackmd.io, 2024. https://hackmd.io/@benediamond/BJgKxUau0.

[DP23] Benjamin E. Diamond and Jim Posen. Succinct arguments over towers of binary fields. In *IACR preprint archive 2023/1784*, 2023. https://eprint.iacr.org/2023/1784.

[DP24] Benjamin E. Diamond and Jim Posen. Polylogarithmic proofs for multilinears over binary towers. In *IACR preprint archive 2024/504*, 2024. https://eprint.iacr.org/2024/504.

[DT24] Quand Dao and Justin Thaler. More optimizations to sum-check proving. In *IACR ePrint Archive 2024/1210*, 2024. https://eprint.iacr.org/2024/1210.

[GM10] Shuhong Gao and Todd Mateer. Additive Fast Fourier Transforms over finite fields. In *IEEE Transactions on Information Theory*, volume 56 (12), 2010.

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof systems. In *SIAM J. Comput.*, 1989.

[Gru24] Angus Gruen. Some improvements for the PIOP for ZeroCheck. In *IACR ePrint Archive 2024/108*, 2024. https://eprint.iacr.org/2024/108.

[Hab22] Ulrich Haböck. A summary on the FRI low-degree test. In *IACR ePrint Archive 2022/1216*, 2022. https://eprint.iacr.org/2022/1216.

[Ham22] Adrian Hamelink. Gemini. hackmd.io, 2022. https://hackmd.io/@adrian-aztec/BJxoyeCqj.

[HLP24] Ulrich Haböck, David Levit, and Shahar Papini. Circle STARKs. In *IACR preprint archive*, 2024. https://eprint.iacr.org/2024/278.

[KT23] Tohru Kohrita and Patrick Towa. Zeromorph: Zero-knowledge multilinear-evaluation proofs from homomorphic univariate commitments. In *IACR preprint archive 2023/917*, 2023. https://eprint.iacr.org/2023/917.

[KZG10] Aniket Kate, Gregory M. Zaverucha, and Ian Goldberg. Constant-size commimtments to polynomials and their applications. In Abe M., editor, *ASIACRYPT 2010*, volume 6477 of *LNCS*. Springer, 2010.

[LCH14] Sian-Jheng Lin, Wei-Ho Chung, and Yunghsiang S. Han. Novel polynomial basis and its application to Reed-Solomon erasure codes. In *FOCS 2014*, 2014.

[LFKN92] Carsten Lund, Lance Fortnow, Howard Karloff, and Noam Nisan. Algebraic methods for interactive proof systems. In *Journal of the Association for Computing Machinery*, volume 39, pages 859–868, 1992. Full paper: `https://eprint.iacr.org/2017/1066`.

[NS24] Ngoc Khanh Nguyen and Gregor Seiler. Greyhound: Fast polynomial commitments from lattices. 2024. `https://eprint.iacr.org/2024/1293`.

[PH23] Shahar Papini and Ulrich Haböck. Improving logarithmic derivative lookups using GKR. In *IACR ePrint Archive 2023/1284*, 2023. `https://eprint.iacr.org/2023/1284`.

[PST13] Charalampos Papamanthou, Elaine Shi, and Roberto Tamassia. Signatures of correct computation. In *TCC 3*, volume 7785 of *LNCS*. Springer, 2013.

[RRR16] Omer Reingold, Ron Rothblum, and Guy Rothblum. Constant-round interactive proofs for delegating computation. In *STOC'16*, 2016. Full paper: `https://eccc.weizmann.ac.il/report/2016/061/`.

[SB23] István András Seres and Péter Burcsi. Behemoth: transparent polynomial commitment scheme with constant opening proof size and verifier time. 2023. `https://eprint.iacr.org/2023/670`.

[Tha23] Justin Thaler. Proofs, arguments, and zero-knowledge. 2023. `https://people.cs.georgetown.edu/jthaler/ProofsArgsAndZK.html`.

[ZCF23] Hadas Zeilberger, Binyi Chen, and Ben Fisch. BaseFold: efficient field-agnostic polynomial commitment schemes from foldable codes. In *IACR preprint archive 2023/1705*, 2023. `https://eprint.iacr.org/2023/1705`.

[ZXZS20] Jiaheng Zhang, Tianchen Xie, Yupeng Zhang, and Dawn Song. Transparent polynomial delegation and its applications to zero knowledge proof. In *IEEE S&P 2020*, 2020. Full paper: `https://eprint.iacr.org/2019/1482`.