Electronic Colloquium on Computational Complexity, Report No. 149 (2016)

A security analysis of Probabilistically Checkable Proofs *

Eli Ben-Sasson†
eli@cs.technion.ac.il
Technion

Iddo Ben-Tov
iddo333@gmail.com
Cornell

Ariel Gabizon
arielga@cs.technion.ac.il
Technion

Michael Riabzev
mriabzev@cs.technion.ac.il
Technion

September 23, 2016

Abstract

Probabilistically Checkable Proofs (PCPs) [Babai et al. FOCS 90; Arora et al. JACM 98] can be used to construct asymptotically efficient cryptographic zero knowledge arguments of membership in any language $L \in \mathbf{NEXP}$, with minimal communication complexity and computational effort on behalf of both prover and verifier [Babai et al. STOC 91; Kilian STOC '92; Micali SICOMP '00]. Though PCP constructions are *asymptotically* efficient, it is still far from clear how well they perform in practice on concrete input sizes. One of the most important parameters to study in this context is PCP *soundness*, defined as the probability of rejecting false statements; this parameter determines to a large extent the communication complexity of PCP based protocols, as well as prover and verifier running time. Of particular importance is studying soundness for *concrete*, non-asymptotic, input sizes. This underlies the definition of the *concrete efficiency threshold* of PCP (and related systems) [Ben-Sasson et. al. STOC '13], defined informally as the smallest instance size for which using a PCP verifier is more efficient than naive verification via re-execution.

To further advance the study of concrete PCP efficiency we initiate a *security analysis* of PCPs and two related systems: PCPs of Proximity (PCPP) [Dinur and Reingold, FOCS 2004; Ben-Sasson et al., SICOMP 2006] and Interactive Oracle Proofs of Proximity (IOPP) [Reingold et al., STOC 2016; Ben-Sasson et al., TCC-A+B 2016]. Security analysis means measuring soundness only with respect to the set of known *attacks*; an attack is a randomized efficient algorithm that produces a pseudo-proof for false statements. In the context of proximity testing (PCPP/IOPP), an attack receives as input a specification of an error correcting code C and a word w that is very far from C and outputs a pseudo-proof of proximity for the (false) statement "$w \in C$".

To jumpstart security analysis one needs a set of attacks. We define one non-trivial attack, called the *rational* attack, on the quasilinear PCP of [Ben-Sasson and Sudan, SICOMP 2008], and two basic attacks — *row-* and *column-compliant* attacks — on the specific PCPP system for Reed-Solomon codes (RS-PCPP) that underlies that PCP as well as mixed strategies combining the two attacks. Our main results are:

- Rational attacks force the attacker to prove proximity of a rational function (like $1/x$) to low-degree polynomials. Rational functions are maximally far from low-degree so this suggests that PCP security may be significantly higher than soundness;
- row- and column-compliant attacks on rational functions have very large security (i.e., large probability of being rejected by the verifier); the same holds for most (random) functions. This gives additional (preliminary) support to the view that PCPP security may be higher than predicted by soundness.

We also give an improved unconditional soundness analysis for the aforementioned RS-PCPP, reducing its concrete efficiency threshold from the previous state of the art of 2^{43} [Ben-Sasson et al., STOC '13] to 2^{23}. Using this measure of concrete efficiency threshold as our gauge, we reduce it further to 2^{19} by applying security analysis; and finally reduce it to 2^{14} for IOPP systems.

*An earlier version of this report appeared as Technical Report TR16-073 on the Electronic Colloquium on Computational Complexity [BBGR16]; this report subsumes and replaces that version.

†Research supported by the Israel Science Foundation (grant 1501/14) and the US-Israel Binational Science Foundation (grant 2021036).

ISSN 1433-8092

Contents

1 Introduction

The study of the *soundness* of PCP systems — the minimal probability of rejecting a pseudo-proof of false statements — is a cornerstone of modern computational complexity. This study is typically conducted by fixing soundness to a constant (say, $1/2$) and constructing systems that optimize other parameters. Applications to inapproximability [FGL$^+$96] require minimizing *query complexity* and *alphabet size*, as studied, e.g., in [Hås01, Din07, MR08]. On the other hand, cryptographic applications such as polylogarithmic verification of computation [BFL90, BFLS91] and succinct Computationally Sound (CS) zero-knowledge argument systems [Kil92, KPT97, Mic00] require minimizing *prover* and *verifier* running time and overall *communication complexity*, as studied, e.g., in [PS94, HS00, GS06, BSVW03, BGH$^+$06, BS08, Mie08].

One of the goals of this research, at least implicitly, has been to facilitate the practical implementation of PCP based systems for the purpose of enforcing and verifying computations that are delegated to external untrusted parties. This goal is part of a larger recent effort, spanning theory and practice, to implement efficient interactive protocols to solve the problem of succinct verification of delegated computation, with or without zero-knowledge [GOS06, IKO07, KR08, GKR08, Gro09, GS09, Gro10, SBW11, Lip12, CMT12, BM88, SMBW12, SVP$^+$12, BCI$^+$13, GGPR13, BCG$^+$13, VSBW13, PGHR13] (see [WB15] for a recent survey). These implementations have used numerous ideas arising in PCP constructions, like arithmetization and the encoding of computations by (evaluations of) low-degree polynomials. However, with the exception of the recent [BBC$^+$16], they have avoided using the "heavy PCP machinery", e.g., probabilistically checkable proofs of proximity (PCPPs) [BGH$^+$06, DR04] and proof composition [AS98], partially because it is not sufficiently clear whether these tools are efficient enough in practice.

In spite of the remarkable recent progress towards implementing cryptographic proof system for general computations, PCP-based interactive proof systems (and noninteractive computationally sound proofs) offer a number of advantages over the systems implemented thus far: (i) they do not require a "trusted setup" preprocessing phase like some of the above mentioned systems (e.g., [GGPR13, PGHR13, BCG$^+$13]) and (ii) apply to any language in $\mathbf{NEXP}$, as opposed to more restricted complexity classes to which some implemented approaches (like [GKR08, CMT12]) are limited to. It is therefore of significance to understand how well PCPs perform when they are "scaled down" to "small" input lengths as may arise in practice.

1.1 Definition of PCP and PCPP security

PCP security The starting point of the current work is an empirically observed gap between soundness and security of proof systems. Indeed, soundness is established by *proving lower bounds* on the probability of rejecting pseudo-proofs of false statements, even when we do not know to construct a pseudo-proof that matches that bound, let alone do this in an efficient manner. Therefore, it may be the case that PCP systems are far more secure in practice that currently believed, which means that to obtain a specified security level we may use far more efficient versions of them. To define *security* let us first recall the standard definition of a PCP system.

Definition 1.1 (PCP system)**.** *A probabilistically checkable proof (PCP) system for a language $L \in \mathbf{NTIME}(T(n))$ with soundness (function) $s(\cdot)$ is a pair of randomized algorithms (P, V) satisfying all of the following requirements:*

- *P is a deterministic machine that on input $(x, y) \in R_L$, $|x| = n$ outputs a proof $\pi = \pi_{(x,y)}$ of length $|\pi| = \ell(n)$ in time $\mathrm{poly}(T(n))$. The proof is viewed as a function $\pi : [\ell(n)] \to \{0, 1\}$*
- *V is a randomized machine that on input $x, |x| = n$ generates in time $\mathrm{poly}(n)$, and using randomness string R, a pair (I_R, D_R) where $I_R \subset [\ell(n)]$ and D_R is a decision circuit that computes a decision predicate $D_R : \{0, 1\}^I \to \{0, 1\}$ (an output of 0 is identified with* reject *and 1 is identified with* accept*). We denote by*

  $$V^{\pi}[x, R] = D_R(\pi|_{I_R})$$

  the output of the verifier on randomness string R, where $\pi|_{I_R} : I_R \to \{0, 1\}$ is the restriction of the function π to inputs I_R.
- **Completeness:** *For every $(x, y) \in R_L$, V accepts x when given oracle access to the proof $\pi_P(x, y) = P(x, y)$,*

  $$\Pr\left[V^{\pi_P(x,y)}[x] = \mathsf{accept}\right] = 1$$

- **Soundness:** *For every $x \notin L$ and pseudo-proof π^*,*

$$\Pr_R\left[V^{\pi^*}[x,R] = \mathsf{reject}\right] \geq s(n) \tag{1.1}$$

where $s(\cdot)$ is called the soundness (function) *of V and the* soundness error (function) *is $1 - s(\cdot)$.*

Security is similar to soundness but measured only with respect to a limited set of *efficient* pseudo-provers which are called *attacks*.

Definition 1.2 (PCP security)**.** *An* attack *on a PCP system $S = (P, V)$ is a randomized polynomial time algorithm P^*. We say S has* security $\mathbf{s}(\cdot)$ against the set of attacks $\mathcal{P}$ *if for all $x \notin L, |x| = n$ and $P^* \in \mathcal{P}$ we have*

$$\Pr_R\left[V^{P^*(x)}[x,R] = \mathsf{reject}\right] \geq \mathbf{s}(n) \tag{1.2}$$

PCPP security Proof methods involving arithmetization, starting from Lund et. al [LFKN92] (see also [BFL90, AS98, ALM+98, PS94]) reduce the problem of verifying membership in a language L to a constant number of *proximity testing* problems (see Definition 1.7). In this problem one is given a specification of an error correcting code C and oracle access to a purported codeword w, and needs to distinguish with high probability and small query complexity between the case that $w \in C$ and the case that w is far in relative hamming distance from C. This problem can be solved efficiently if C is locally testable. But some high-rate "PCP-friendly" codes, like Reed-Solomon (RS) codes, are not locally testable while some high-rate locally testable codes (in terms of rate), like [KRS15, KMRS15], are not known to "PCP-friendly", i.e., it is not clear how to build PCPs in a black-box manner using these codes.

To enable using a simple code of high rate like RS in a PCP, the prover adds an auxiliary proof y that assert $w \in C$. Miraculously, even if C has no local structure, V can verify that w is at least close to C by querying a few locations in both w and y. This auxiliary proof is what is commonly called a PCPP[1] for the code C [BGH+06, DR04]. The particular PCP that we study involves such a PCPP for RS codes and the "best" attacks we are aware of on that PCP system will employ attacks on the underlying PCPP system, because PCPP soundness is the "weakest link" in terms of soundness.

Before dealing with the specific system that interests us we give a general definition of PCPP security for general linear codes. We use the standard notation that an $[n,k,d]_{\mathbb{F}}$-code C is a k-dimensional subspace of $\mathbb{F}^n$ in which no two distinct members have hamming distance smaller than d.

Definition 1.3 (PCPP for a code C)**.** *Fix integers $A, \mathsf{q} \in \mathbb{N}$. Let C be an $[n = n(C), k = k(C), d = d(C)]_{\mathbb{F}}$-code. An (A, q)-*PCPP system *$\mathcal{S}_C$ for C with* soundness function *$s(\cdot)$ is a pair $\mathcal{S} = (P, V)$, where*

- *P is a* systematic *mapping $P : C \to \mathbb{F}^A$.*

  That is, for any $x \in C$, $P(x) = (x, y)$ for some $y \in \mathbb{F}^{A-n}$.

- *V is a* q-local *randomized mapping $V : \mathbb{F}^A \to \{\mathsf{accept}, \mathsf{reject}\}$. That is, after choosing its internal randomness, $V(z)$ always depends on at most q indices of $z \in \mathbb{F}^A$.*

Such that

- **Completeness:** *For any $x \in C$, $V(P(x)) = \mathsf{accept}$ with probability one.*
- **Soundness:** *For any $x \in \mathbb{F}^n$ such that $\Delta(x, C) = \delta n$, and any $y \in \mathbb{F}^{A-n}$,*

$$\Pr\left[V(x,y) = \mathsf{reject}\right] \geq s(\delta).$$

When the soundness function is not explicitly mentioned, it is fixed to be the default function

$$s(\delta) = \begin{cases} \frac{1}{2} & \delta \geq \frac{d}{3n} \\ 0 & \textit{otherwise} \end{cases}$$

*Fix an ensemble of linear codes $\mathcal{C} = \left\{C \subseteq \mathbb{F}^{n(C)}\right\}$, and functions $A, \mathsf{q} : \mathbb{N} \to \mathbb{N}$. An (A, q)-*PCPP system for *$\mathcal{C}$ is an ensemble of PCPP systems $\mathcal{S} = \{\mathcal{S}_C | C \in \mathcal{C}\}$ where $\mathcal{S}_C$ is an $(A(k(C)), \mathsf{q}(k(C)))$-PCPP system for C.*

[1] However, it is technically convenient for us to define the PCPP as a mapping that outputs *both* w and y.

Definition 1.4 (PCPP security)**.** *An* attack *on a PCPP system $S = (P, V)$ for a family of codes $\mathcal{C} = \{C_n\}_{n\in\mathbb{N}}$ is a randomized polynomial time algorithm P^*. We say S has* security $\mathbf{s}(\cdot,\cdot)$ against the set of attacks $\mathcal{P}$ *if for all $x \notin C_n$ of relative hamming distance δ from C_n and $P^* \in \mathcal{P}$ we have*

$$\Pr\left[V(x, P^*(x)) = \mathsf{reject}\right] \geq \mathbf{s}(\delta, n) \tag{1.3}$$

Remark 1.5 ($\mathbf{POLY}$-Security)**.** *One may generalize the security definitions above by defining $\mathcal{P}$ to be the set of polynomial-time (randomized) algorithms. To obtain a meaningful definition for this class, one should use standard asymptotic conventions and define the security function $\mathbf{s}()$ as such that for each $P \in \mathcal{P}$ there exists some n_P such that instances of size $n \geq n_P$ are rejected with probability $\geq \mathbf{s}(n)$.*

Since this work is motivated by concrete rather than asymptotic security, we leave this study to future work.

Discussion of PCP/PCPP security analysis A reasonable question to ask at this point is why should one engage in a security analysis of PCP systems that inevitably leads to conditional results, when one can apply soundness analysis that is valid unconditionally. We give three answers:

- As mentioned, analyzing security may lead to more efficient PCP systems (say, with smaller query complexity) than what soundness allows for.
- PCPs used in some practical cryptographic applications (e.g., [Kil92, Mic00]) use cryptographic primitives (like a collision resistant hash function) for which no known construction is provably sound; in other words, avoiding conditional security analysis of PCPs will not remove the need for conditional cryptographic assumptions (backed by security analysis) in PCP-based applications.
- Studying attacks on PCPs and PCPPs may lead to new insights and interesting theoretical questions. This already seems to be the case for the attacks in this paper, that quickly lead to algebraic geometry questions about distinguishing rational functions from low-degree polynomials (Section 3.3).

Related work Kalai and Raz introduced in [KR09] the notion of a probabilistically checkable argument (PCA). In the PCA model soundness is relaxed to hold only against computationally-bounded adversaries, and the verifier, which is designated, uses cryptographic primitives (namely, a computational private information retrieval scheme) to make the interactive process more efficient. The emphasis of our security analysis is different. Rather than proving security against all computationally bounded adversaries under cryptographic assumptions, we focus on analyzing a few natural computationally bounded attacks without cryptographic assumptions.

1.2 Attacks

To initiate PCP security analysis one needs a specific PCP system and a set of attacks. We focus on the system of quasi-linear PCPs of [BS08] and describe a family of attacks on it, called *rational attacks*. Different members of this family correspond to different variants of *axis-compliant* attacks on the PCPP that underlies that PCP system. We start by describing the PCP attacks, followed by PCPP attacks.

The PCP system of [BS08] uses a reduction to a certain *linear algebraic Constraint Satisfaction Problem (CSP)* [BCGV16], which reduces membership in an $\mathbf{NP}$-complete language to problems of proximity-testing to a specific sub-family of RS codes. We describe this particular language and code-family next.

Definition 1.6 (Reed-Solomon codes on subspaces)**.** *Fix a positive integer η. Let $\mathbb{F}$ be a field of characteristic two. Fix an $\mathbb{F}_2$-subspace $L \subseteq \mathbb{F}$. We define $\mathsf{RS}_L[\eta]$ to be the set of functions $g : L \to \mathbb{F}$, each being the evaluation (on L) of a univariate polynomial $p \in \mathbb{F}[Z]$ of degree at most $2^{-\eta} \cdot |L| - 1$.*

We define $\mathrm{RS}_{\mathrm{a}}[\eta]$ to be the ensemble of Reed-Solomon codes over some subspace $L \subseteq \mathbb{F}$, of degree at most $2^{-\eta} \cdot |L| - 1$. That is,

$$\mathrm{RS}_{\mathrm{a}}[\eta] \triangleq \{\mathsf{RS}_L[\eta] | L \subseteq \mathbb{F} \text{ is a subspace}\}.$$

Note that in this notation the fixed field $\mathbb{F}$ is implicit.

The PCPs we study are for the following $\mathbf{NP}$-complete "PCP-friendly" language. The definition is similar to ones appearing in [PS94, HS00, BCGV16] and known to be $\mathbf{NP}$-complete even when q, c are fixed to some absolute constants (cf. [BCGV16, Section 7]). Later on, when we prove our main theorems, we shall use a more technically detailed version of the definition below (Definition A.6).

Definition 1.7 (Linear algebraic CSP (LACSP))**.** *An instance of the linear algebraic CSP (LACSP) corresponding to the [BS08]-PCP is given by a tuple $x = (C_0, C_1, \phi)$ where $C_0 \in \mathrm{RS}_{\mathrm{a}}[c], C_1 \in \mathrm{RS}_{\mathrm{a}}[3]$ are RS-codes of blocklength ℓ_0, ℓ_1 respectively and $\phi : \mathbb{F}^{\ell_0} \to \mathbb{F}^{\ell_1}$ is a q-local map of degree $d \leq \log c$, meaning that the value of the ith coordinate of $\phi(\cdot)$ is computed by a degree 2^c, q-variate polynomial, that, in particular, depends on at most q coordinates of the input. An instance is* satisfiable *iff there exists $f \in C_0, g \in C_1$ such that $g = \phi(f)$, and the language $L_{LACSP}[\mathsf{q}, c]$ is the language of satisfiable instances in which ϕ is q-local and of degree $\leq 2^c$.*

Rational attacks on [BS08]-PCP Using the notation of Definition 1.7, a PCP for membership in L_{LACSP} is given by a pair (f, g) of degree-d (univariate) polynomials that are evaluated on larger domains, of size $\ell_0, \ell_1 \geq 8d$, and which must satisfy a system of algebraic constraints. Informally, a *rational attack* (Definition 2.1) picks f to be a low-degree polynomial and sets g to equal $\phi(f)$, so that the pair (f, g) maximizes the probability of satisfying a random q-local constraint encoded by ϕ. Rational attacks differ (only) in the way they construct a PCPP for g (see below); since f is low-degree all such attacks construct (the same) PCPP for f which, by construction, is accepted by the verifier with probability 1.

When the pair (f, g) is chosen in this manner, and furthermore the instance x is "almost satisfiable" (see Definition 2.3), then g turns out being an evaluation of a *rational* function of the form $g(x) = a/(x - b)$ for some fixed $a, b \in \mathbb{F}$ (see Lemma 2.4). In other words, when f is "maximally close" to satisfying x and g is "maximally close" to agreeing with f on ϕ, then g is *maximally far* from being a low-degree polynomial; its relative hamming distance from $\mathrm{RS}_{\mathrm{a}}[3]$ is $7/8$. We find this phenomenon intriguing because it suggests that there may be further dependencies between different parts of a PCP construction that can be exploited to improve soundness. Moreover, when attempting to complement a rational function with a PCPP, the attacks we are aware of will lead to pseudo-proofs that are rejected with very high probability, as described next.

Axis-compliant attacks on RS-PCPP As a first step in the study of PCPP security, we suggest two "pure" attacks on the RS-PCPP of [BS08], that we shall call the *row-compliant* and the *column-compliant* attacks; an attacker that "mixes" the two attacks will be called an *axis-compliant* attacker. Rational (PCP) attacks use some axis-compliant attack to construct a PCPP for g. To explain the attacks we first recall informally (and, for simplicity, inaccurately) the construction of the relevant RS-PCPP.

Fix a function $g : L \to \mathbb{F}$. The purpose of the PCPP is to convince V (the verifier) that g is (close to) an element of $\mathsf{RS}_L[\eta]$, i.e., that it is an evaluation of a polynomial of degree at most $d \triangleq 2^{-3} \cdot |L| - 1$. For this overview, let us assume that V is willing to read $O(\sqrt{d})$ entries in total from g and an auxiliary proof π_g. Observe first, that reading any $t < d$ entries of g gives V no information on whether $g \in \mathsf{RS}_L[\eta]$; as a degree d univariate polynomial could be interpolated to match any t values. The basic idea is to "embed" g into a *bivariate* polynomial Q of degree $\sqrt{d}$; for this overview we use the term degree of a bivariate polynomial to mean individual degree. The proof π_g consists of values of Q on a carefully chosen product set[2] $A \times B, |A| = 7\sqrt{d}, |B| = 8\sqrt{d}$. Furthermore, we partition g (viewed as a table of values of length $|L|$) into $8\sqrt{d}$ sets of size $\sqrt{d}$ each; each set g_b is labeled by a unique $b \in B$ and "appended" to the b'th *row* of Q which consists of (the values of Q) on the set $A \times \{b\}$. We thus end up with a bivariate function π_g on a product set $A' \times B, |A'| = |B| = 8\sqrt{d}$.

It turns out that when $g \in \mathsf{RS}_L[\eta]$ then there exists a function Q such that when appended to g (as described above) each row and column of π_g is the evaluation of a degree $\sqrt{d}$ polynomial on $8\sqrt{d}$ points, i.e., it is a member of some $\mathsf{RS}_{L'}[3]$ for some L' (that happens to be a linear space). More crucially, the bivariate low-degree testing Theorem of [PS94] is used by [BS08] to show a converse: if g is far from $\mathsf{RS}_L[3]$ then, on average, rows and columns of π_g are far from $\mathsf{RS}_{L'}[3]$. Thus, to verify that g is close to $\mathsf{RS}_L[3]$ the verifier V selects a uniformly random row/column of π_g and verifies that it is close to $\mathsf{RS}_{L'}[3]$ by recursively applying the same univariate-to-bivariate construction as described above.

[2] Actually, the carefully chosen set only *contains* a product set but this doesn't affect the high-level description of the construction and attack.

Returning to a description of our attacks, suppose now that the attacker P^* has in its possession a function $g : L \to \mathbb{F}$ that is far from any element of $\mathsf{RS}_L[3]$. He wishes to devise a pseudo-proof that will cause V to accept with high probability, i.e., V will conclude wrongly that $g \in \mathsf{RS}_L[3]$. One natural way to attack the verifier is to choose π_g such that all its rows are degree $\sqrt{d}$ polynomials. This requires interpolating g_b (a table of size $\sqrt{d}$) to fill the remaining entries of the b'th row, leading to the *row-compliant attack*, named so because all rows of π_g are of degree $\sqrt{d}$ as required. Similarly, the *column-compliant attack* results from ensuring that each column of π_g is of degree $\sqrt{d}$, while also maximizing the number of rows that are low-degree. Finally, an *axis-compliant attack* mixes the two attacks when building the recursively constructed PCPP.

In Section 3, we show that for rational functions g such as those that arise from the rational PCP attack described above, V rejects the corresponding PCPP π_g with much higher probability than what can be shown for a general pseudo-proof. Roughly speaking, we show that "while making all rows low degree, you will make many columns far from low degree"; we show the same holds vice-versa for depth-1 column-compliant proofs and conjecture it to hold for any depth. Furthermore, we show the same is true (with high probability) when axis compliant attacks (of any depth) are applied to random functions, which, like rational functions, are also maximally far from low-degree polynomials.

Combining our results about the two attacks (on PCP and PCPP systems) leads to our main result (see Theorem 4.1 for the formal statement).

Theorem 1.8 (Main — informal)**.** *There exists a verifier V for L_{LACSP} for which the following holds. Suppose ψ is a "nearly-satisfiable" instance (see Definition 2.3) of L_{LACSP}. Let π be a pseudo-proof for ψ given by the rational attack, using an axis-compliant PCPP attack of depth d. Assuming Conjecture 3.2, V rejects ψ with probability at least*

$$\frac{7}{15} \cdot \left(\frac{3}{7}\right)^{d-1} \tag{1.4}$$

Furthermore, the inequality above holds unconditionally (i.e., without assuming Conjecture 3.2) for $d = 1$ and/or for the case that the PCPP attack is a pure row-compliant one.

When applied to concrete input lengths, the query complexity that is implied by the (conditional) theorem above is far better than what our best (unconditional) soundness analysis give. For example, fixing security to $1/2$, a PCP statement involving codewords of length $N = 2^{35}$ with recursion depth $d = 4$ requires only 19 repetitions of the PCPP verifier for each of the two RS-codeword to reach the target soundness. Calculation[3] shows that for $d = 4$, each "test" of the verifier queries a set of at most $2^4 = 16$ field elements; this test is repeated 19 times to reach the target soundness of $1/2$, reaching a total of 304 queries for each of the two invocations of the RS-PCPP verifier, and a total of 608 queries for RS-PCPP verification[4] (see Figure 1). State of the art soundness analysis [BCGT13] requires far greater query complexity (as large as N when $N \leq 2^{40}$). This rather large gap between security and soundness at the concrete (non-asymptotic) range calls for further study, with a goal of matching provable soundness lower-bounds with (efficient) attacks that show these bounds to be tight.

[3]Starting with a space of dimension $n = \log N$, one computes the sequence $n_0 = n, n_1, \ldots, n_d$ given by $n_i = \lceil n_{i-1}/2 \rceil + 1$; in our case $n_0 = 35$ which gives $n_1 = 18$, $n_2 = 10$, $n_3 = 6$ and $n_4 = 4$ meaning the base-case test queries $2^{n_4} = 16$ field elements.

[4]The ACSP consistency verifier (see Definition A.8) typically requires fewer repetitions but also depends on the ACSP query complexity (denoted q in Definition 1.7), hence we omit it from our calculations.

Table 1: Optimal PCPP query complexity as a function of RS blocklength, according to Theorem 1.8 (cf. Theorem 4.1 and Lemma 5.1). Reported below for each range of blocklengths are optimal PCPP recursion depth, number of repetitions of base-test needed to reach soundness $1/2$, and upper bounds on query complexity for a single RS-PCPP test as well as total query complexity (each query is a single field element).

blocklength	recursion depth	# repetitions	query complexity	
			single test	total
$2^{10} - 2^{11}$	2	4	16	64
$2^{12} - 2^{15}$	2	4	32	128
$2^{16} - 2^{19}$	3	8	16	128
$2^{20} - 2^{27}$	3	8	32	256
$2^{28} - 2^{35}$	4	19	16	304
$2^{36} - 2^{50}$	4	19	32	608

1.3 Illustration of results for PCPPs and IOPPs for concrete input sizes

To compare our security results to prior works for concrete input lengths, the *concrete efficiency threshold* of [BCGT13] is a useful metric, as it captures informally the notion of the "smallest input length where PCP verification is useful".

Definition 1.9 (Cost and concrete efficiency threshold of a PCPP)**.** *Using the notation of Definition 1.3, we define the* cost *of an (A, q)-PCPP system for a code C to be $A \cdot \mathsf{q}$. We say the system is* efficient *if the cost $A \cdot \mathsf{q} \leq k(C)^2/2$. (The efficiency factor $1/2$ is an arbitrary choice in [BCGT13] that we keep for consistency.)*

Given an ensemble of PCPP systems $\mathcal{S}$ for an ensemble of linear codes $\mathcal{C}$, the concrete efficiency threshold of *$\mathcal{S}$ is the smallest integer $\mathbf{k}$ such that for any $C \in \mathcal{C}$ of dimension $k(C) \geq \mathbf{k}$, $\mathcal{S}_C$ is efficient.*

An Interactive Oracle Proof of Proximity (IOPP) [BSCS16, BCG+16] (introduced independently by [RRR16] under the name Probabilistically Checkable Interactive Proof of Proximity) is similar to a PCPP (Definition 1.3), but in an IOPP the prover P writes down the proof as part of an interactive process with V. However, contrary to a standard interactive protocol and similarly to a PCPP, V does not pay the cost of receiving and storing P's messages, but only the cost of accessing the locations it actually reads. This results in shorter proofs and hence a lower cost and concrete efficiency threshold, compared to PCPPs. The previous definition applies naturally also to IOPPs by redefining A to be the sum of lengths of interactive proofs supplied by P.

To gain some intuition for the above definition, note that the cost of the "trivial system" where P simply sends the message m encoded by the codeword $x \in C$, and V reads of all m and verifies its encoding is x, has cost $k(C)^2$.

The previous state of the art for the concrete efficiency threshold of $\mathrm{RS}_{\mathrm{a}}[3]$ was 2^{43} [BCGT13] (down from 2^{683} in [BS08]). By careful analysis and using a new PCPP verifier we reduce this to 2^{23} (see Theorem B.1). By replacing soundness with security against axis-compliant attacks, we show a concrete threshold of 2^{19} for random functions (see Section 5.2) and conjecture the same holds for rational functions (Conjecture 3.2). Finally, when replacing a PCPP with an Interactive Oracle Proof of Proximity [BSCS16] this number drops to 2^{14} (see Section 5.3).

Cost comparison Let us define an *(A, q)-PCPP system (or IOPP system) with soundness error $2^{-\lambda}$* identically to Definition 1.3 but with the rejection probability of V being at least $1 - 2^{-\lambda}$ for "large" λ, rather than $1/2$ ($\lambda = 1$); we call λ the *security parameter*. Figure 1.1 compares the cost (as per Definition 1.9) of the proof system for $\mathsf{RS}_L[3]$ as a function of the security parameter λ on messages of size 2^{35}. We note that the previous analysis of [BCGT13] does not enable improving upon the trivial system for such message size. The following four cases are displayed:

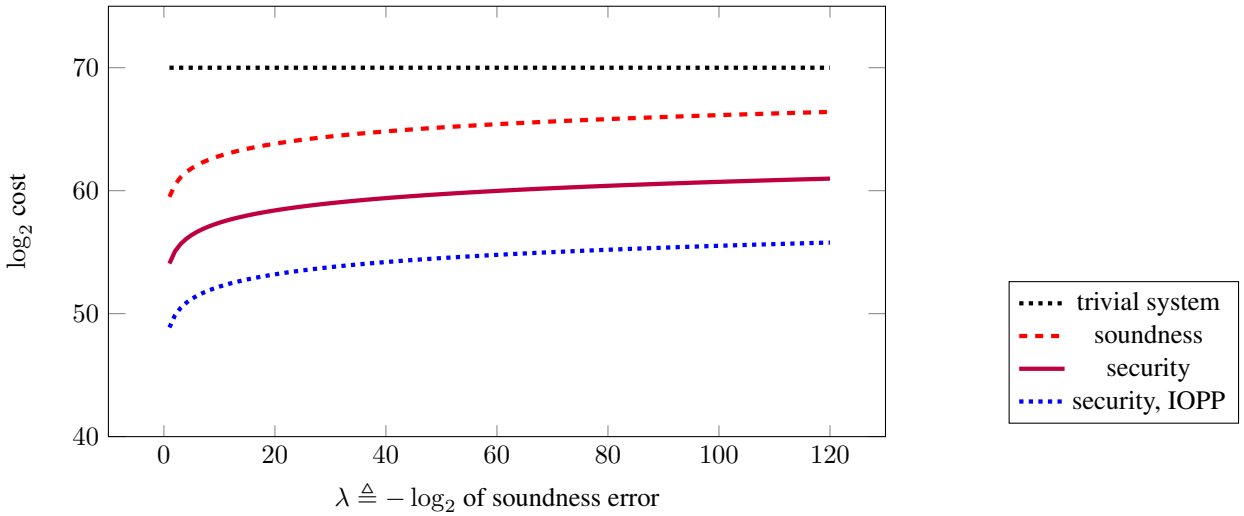
1. **black dotted line:** The "trivial" system, where prover simply sends the message being encoded, and the verifier reads all of it.

2. **red broken line:** Our improved (unconditional) soundness analysis, described in Section B;

3. **purple solid line:** The security analysis described in Section 5 applied to random functions, assuming the prover can only use the row and column-compliant attacks; see Section 5.2

4. **blue dotted line:** The security analysis in the IOPP model, applied to random functions, assuming the prover can only use the row and column-compliant attacks; see Section 5.3

Figure 1.1: A comparison of cost as a function of security parameter λ for messages of length 2^{35} with $\mathsf{RS}_L[3]$.



1.4 Organization of the paper

In the next section we analyze PCP security with respect to the rational attack. In Section 3.3 we study the security of the RS-PCPP against axis-compliant attacks, applied to rational functions. The combination of these two sections (on PCP and PCPP security) lead to the proof of the Main Theorem 1.8, appearing in Section 4. We continue the study of security applied to random functions in Section 5 because random functions are similar to rational functions in being maximally far from low-degree polynomials; they also informally are the closest approximation we currently have for a "typical" function, one that may arise from an attack on a PCP system. The technical definitions of PCP and PCPP systems are deferred to Section A because they are fairly close to standard ones appearing in previous works. Section B gives a better concrete soundness analysis than the current state of the art [BBGR16]. Finally, Section C presents visualizations of the RS-PCPP construction studied earlier.

2 PCP security against rational attacks

In this section we focus on security of the PCP system of [BS08]. First we formally define the "rational" attack on it. Then we study this attack on the "easiest-to-satisfy" instances, ones in which all but a single (algebraic) constraint can be satisfied. The main point in the section (Lemma 2.4) is that the rational attack forces the attacker to prove proximity to Reed-Solomon codes of a rational function. Later on (Section 3.3) we shall strengthen this argument by studying the security of attacks that are forced to start with a rational functions and showing that their rejection probability is also higher than what can be proved by soundness analysis.

2.1 The rational attack

The name "rational attack" below refers to the fact that when applied to minimally unsatisfiable instances (defined later) it forces the prover to prove "low-degreeness" of rational functions $1/(aX + b)$ which are maximally far from low-degree polynomials (see Lemma 2.4).

Definition 2.1 (Rational attack)**.** *Given an unsatisfiable* $\mathsf{UACSP}[2^{-\eta}]$ *instance* $\psi = (L, H, Q, \mathcal{N})$*, the attacker* P^* *operates as follows:*

- *Choose* $f \in \mathsf{RS}_L[\eta + \log \deg(\psi)]$ *as to minimize (see Remark 2.2)*

$$\Pr_{x \in H} \left[(Q \circ f \circ \mathcal{N})(x) \neq 0 \right]. \tag{2.1}$$

  Compute the depth-r *BSS extension* $\Omega_{\mathsf{r}}(f)$ *(as in the third step of the prover in Definition A.8).*

- *Let* $g : \mathbb{F} \to \mathbb{F}$ *be the function defined by*

$$g(x) = \begin{cases} \frac{(Q \circ f \circ \mathcal{N})(x)}{\mathsf{Zero}_H(x)} & \mathsf{Zero}_H(x) \neq 0 \\ 0 & \textit{otherwise} \end{cases} \tag{2.2}$$

- *Compute a depth-*r *axis-compliant extension* h *for* g *as in Definition A.9.*

- *Publish the concatenation of both* $\Omega_{\mathsf{r}}(f)$ *and* h *as the proof* π^* *for* ψ*.*

Remark 2.2. *The problem of minimizing the number of unsatisfied constraints of a constraint satisfaction problem is often known to be* $\mathbf{NP}$*-hard (cf. [Hås01]), though we are not aware of an inapproximability result that applies specifically to the (specially structured) language* $\mathsf{UACSP}[2^{-\eta}]$*.*

Choosing f *to minimize (2.1) leads to* stronger *attacks (i.e., ones that are harder to detect). Indeed, letting* ϵ *equal the probability in (2.1) and* ρ *be the rate of the relevant RS-code for* f*, the ACSP consistency test from Definition A.8 rejects* (f, g) *with rather large probability, at least* $1-(1-\rho\epsilon)^\rho \geq 1-e^\epsilon$*. Therefore, intuitively, an attacker has higher overall probability of evading rejection when minimizing* ϵ *and attempting to "fool" other sub-verifiers, notably the RS-PCPP one.*

2.2 Security of rational attack on minimal unsatisfiable instances

An unsatisfiable instance of a constraint satisfaction problem (like 3SAT) is said to be *minimally unsatisfiable* (MU) if there exists an assignment that satisfies all but one constraint, so removing that constraint would make the residual instance satisfiable). Minimally unsatisfiable CSPs are interesting in their own right and are also useful in the study of proof complexity (cf. [NR11] for a recent example and references therein). The following definition is the natural analog of minimal unsatisfiability for ACSPs.

Definition 2.3 (Minimally unsatisfiable ACSP)**.** *An unsatisfiable* $\mathsf{UACSP}[\rho]$*-instance* $\psi = (\mathbb{F}, H, Q, \mathcal{N})$ *is called* minimally unsatisfiable *(MU) if there exists* $A \in \mathbb{F}[X], \deg(A) < \rho|\mathbb{F}|$ *and* $\sigma \in H$ *such that* $Q \circ A \circ \mathcal{N}$ *vanishes on* $H \setminus \{\sigma\}$*. In this case we say* A *is a* minimally unsatisfying assignment *for* ψ*.*

The following algebraic characterization of MU instances and assignments informally shows that to construct pseudo-proofs for evaluations of MU assignments, one should construct pseudo-proofs for rational functions of the form $f(X) := a/(X-b)$. PCPP-attacks for such functions are discussed in Section 3.3 where it is argued that they are rejected with rather high probability.

Lemma 2.4 (Rational attacks lead to rational functions)**.** *An unsatisfiable* $\mathsf{UACSP}[\rho]$*-instance* $\psi = (\mathbb{F}, H, Q, \mathcal{N})$ *is minimally unsatisfiable if and only if there exist* $A, B \in \mathbb{F}[X], \deg(A) < \rho|\mathbb{F}|, \deg(B) < \rho \deg(\psi)|\mathbb{F}|$ *and* $\sigma \in H$ *such that*

$$\frac{Q \circ A \circ \mathcal{N}(X)}{\mathsf{Zero}_H(X)} = B(X) + \frac{c}{X - \sigma} \tag{2.3}$$

for some non-zero $c \in \mathbb{F}$*.*

Remark 2.5. *Definition 2.3 can be extended to a* t*-unsatisfiable instances, in which* $Q \circ A \circ \mathcal{N}$ *vanishes on a* $H \setminus H', |H'| \leq t$*. In such a case the right hand side of (2.3) is replaced by a rational function in which the denominator is of degree* t*.*

Notice that A and B are of sufficiently low-degree to ensure that their evaluations, denoted f and g respectively henceforth, have RS-PCPP proofs that are accepted with probability 1 by the PCP verifier from Definition A.8. Moreover, the linearity of the RS-PCPP prover, along with the right hand side of (2.3) imply that a pseudo-proof has soundness ϵ if and only if the function $c/(X-\sigma)$ has a pseudo-PCPP with soundness error ϵ. We study the security of this function in Section 3.3.

Proof of Lemma 2.4. ($\Rightarrow$) Suppose ψ is minimally unsatisfiable and let A be a minimally non-satisfying assignment, so $\deg(A) < \rho|\mathbb{F}|$. Then $Q \circ A \circ \mathcal{N}(X)$ vanishes on $H \setminus \{\sigma\}$, hence it is divisible by $\mathsf{Zero}_{H\setminus\{\sigma\}} = \frac{\mathsf{Zero}_H(X)}{X-\sigma}$. Let

$$\tilde{B}(X) = \frac{Q \circ A \circ \mathcal{N}(X)}{\mathsf{Zero}_{H\setminus\{\sigma\}}} = \frac{(X-\sigma)Q \circ A \circ \mathcal{N}(X)}{\mathsf{Zero}_H(X)} \tag{2.4}$$

with $\deg(\tilde{B}) < \rho \deg(\psi)|\mathbb{F}|$. The unsatisfiability of ψ implies that $\tilde{B}$ is not divisible by $X-\sigma$. So it can be written as $\tilde{B}(X) = c + (X-\sigma)B(X), c \neq 0$. Dividing both sides of (2.4) gives (2.3).

In the other direction, multiply both sides of (2.3) by $\mathsf{Zero}_H(X)$. This gives

$$Q \circ A \circ \mathcal{N}(X) = \mathsf{Zero}_H(X)B(X) + c\mathsf{Zero}_{H\setminus\{\sigma\}}(X) = \mathsf{Zero}_{H\setminus\{\sigma\}}(X)\left(B(X)(X-\sigma)+c\right)$$

and the right hand side proves that $Q \circ A \circ \mathcal{N}(X)$ vanishes on $H \setminus \{\sigma\}$, i.e., ψ is minimally unsatisfiable. □

3 PCPP security against rational functions and axis-compliant attacks

In this section we analyze the security of the PCPP system for Reed-Solomon codes against two "pure" attacks (see Definition 1.4) and mixtures of them.

In Section 3.1 we define the attacks and then propose an optimized PCPP verifier for them (Section 3.2). We end by analyzing security on rational functions, as needed to prove our main theorem in the next section. Definitions of the [BS08] PCP (and PCPP) appear in Appendix A.

3.1 Two attacks

Our starting point is a function $g : L \to \mathbb{F}$ where L is a subspace, and g is possibly far from $\mathsf{RS}_L[\eta]$ (see Definition 1.6). Recalling the recursive construction of the BSS proof (Definition A.4), and the fact that the tests applied by the standard BSS verifier V to (the first level of) the BSS proof are either "row"-tests or "column"-tests, two natural attacks come to mind. The row-compliant attack constructs a pseudo-proof in which all rows are low-degree and the column-compliant attack constructs a pseudo-proof in which all columns and a nontrivial fraction of rows are low-degree. Axis-compliant attacks combine both row- and compliant-attacks at different points in the recursive PCPP construction. All axis-compliant attacks are a generalization of the BSS extension (Definition A.3) to the case that g does not belong to $\mathsf{RS}_L[\eta]$, meaning that all such attacks compute the (same) BSS extension when $g \in \mathsf{RS}_L[\eta]$. The formal definition appears in Appendix A.3 because it relies on the (somewhat technical) definition of the RS-PCPP.

3.2 An optimized verifier against combined row and column-compliant attacks

The purpose of this section is to construct a verifier that maximizes its rejection probability against axis-compliant attacks. By construction of the BSS verifier, if attacked by a row-compliant attack then all row-tests pass with probability 1. Similarly, if attacked by a column-compliant attack, all column-tests and an additional $2^{-\eta}$-fraction of row-tests are accepted with probability 1. This attack is better answered with a *biased* verifier, one that performs a row-test with higher probability than a column-test.

Definition 3.1 (Biased depth d RS-PCPP verifier)**.** *Fix integer η. Define $\alpha_\eta \triangleq \frac{1-2^{-\eta}}{2-2^{-\eta}}$. Given a function $f : \Omega_d(L) \to \mathbb{F}$, where $\Omega_d(L)$ is defined as in Definition A.4, the verifier V_η operates as follows:*

- *If $d = 0$, then read f entirely (on all of L) and accept iff $f \in \mathsf{RS}_L[\eta]$.*

- *For $d > 0$,*
  - *with probability α_η pick a uniformly random $\alpha \in L'_0$ and return $\mathrm{V}_\eta^{\Omega_{d-1}(f_\alpha^{\mathrm{col}})}$*
  - *with probability $1 - \alpha_\eta$ pick a uniformly random $\beta \in L_1$ and return $\mathrm{V}_2^{\Omega_{d-1}(f_\beta^{\mathrm{ext}})}$,*

3.3 Security on rational functions

We now study the security of axis-compliant attacks in the case when applied to a rational function that has a linear term in its denominator. More precisely, we use the terminology that $f : S \to \mathbb{F}$ is a *rational function* if

$$f = \frac{1}{x + \sigma}|_S,$$

for some $\sigma \in \mathbb{F} \setminus S$. (In this section we use the convention that $S \subset \mathbb{F}$ always denotes an affine subspace. That is, $S \triangleq H + \beta$ for some linear subspace $H \subset \mathbb{F}$ and $\beta \in \mathbb{F}$.)

The purpose of this section is to give evidence that axis-compliant attacks fail with high probability, when applied to a rational function. We prove this is the case for row-compliant attacks of arbitrary depth (Lemma 3.3) and for depth-1 column-compliant attacks (Lemma 3.5). We conjecture it to be true for arbitrary depth column-compliant attacks, and hence for all axis-compliant attacks (Conjecture 3.2). Another way to state the conjecture below is to say that axis-parallel attacks applied to rational functions are rejected with the same probability as axis-parallel attacks on *random* functions, which are also (with high probability) maximally far from low-degree polynomials (cf. Lemma 5.1).

Conjecture 3.2. *Let $L \subseteq \mathbb{F}$ be a linear subspace, and $\sigma \in \mathbb{F} \setminus L$. Let $g \triangleq f_{L,\sigma}$ be a rational function. Let $f : \Omega_d(L) \to \mathbb{F}$ be any depth d axis-compliant extension of g. Then*

$$\Pr[\mathrm{V}_\eta^f = \mathsf{reject}] \geq \alpha_\eta \cdot (3/7)^{d-1}$$

where V_η is the biased verifier from Definition 3.1.

Row compliant attack on rational functions We proceed to proving the conjecture for arbitrary depth row-compliant attacks by showing that in the row-compliant extension of a rational function $g : L \to \mathbb{F}$, all columns turn out to be rational functions as well (!). Inductively, therefore, recursive row-compliant attacks (to arbitrary depth d) result in columns that are far from low-degree, hence rejected by the PCPP verifier.

Lemma 3.3 (Columns of the row-compliant attack on a rational function, are rational functions)**.** *Let $L \subseteq \mathbb{F}$ be a linear subspace, and $\sigma \in \mathbb{F} \setminus L$. Let $g \triangleq f_{L,\sigma}$ be a rational function. Let $f \triangleq \Omega_-(g) : \Omega(L) \to \mathbb{F}$ be the row-compliant extension of g. Then for every $\alpha \in L'_0$,*

$$f_\alpha^{\mathrm{col}} = \frac{c_\alpha}{Y + \sigma'}|_{L'_1},$$

for some $0 \neq c_\alpha \in \mathbb{F}$ and $\sigma' \in \mathbb{F} \setminus L'_1$.

Consequently, for any depth-d row-compliant extension $f' : \Omega_d(L) \to \mathbb{F}$ of g,

$$\Pr[\mathrm{V}_\eta^{f'} = \mathsf{reject}] \geq \alpha_\eta^d \geq \alpha_\eta \cdot (3/7)^{d-1}$$

The key observation in our proof is given by the following simple expression for the algebraic representation of the (evaluation of a) rational function $f : S \to \mathbb{F}$ as a low-degree polynomial. From now on, let us denote by $f_{S,\sigma} \in \mathbb{F}[X]$ the unique polynomial of degree smaller than $|S|$, with $f_{S,\sigma}|_S = \frac{1}{X-\sigma}|_S$.

Claim 3.4. *Let $S \triangleq H + \beta$ for a linear subspace $H \subset \mathbb{F}$ and $\beta \in \mathbb{F}$. Let $q(X) \triangleq q_H(X)$. Fix $\sigma \in \mathbb{F} \setminus S$.*

Then,

$$f_{S,\sigma}(X) = \frac{q(X - \sigma)}{q(\beta - \sigma) \cdot (X - \sigma)}.$$

Proof. First note that the expression on the right-hand side is indeed a polynomial: σ is a root of $q(X-\sigma)$ and therefore, $(X-\sigma)|q(X-\sigma)$. In particular, it is a polynomial of degree less than $|H| = |S|$. Now, note that for any $b = a + \beta \in S$, where $a \in H$,

$$q(b-\sigma) = q(a) + q(\beta-\sigma) = q(\beta-\sigma).$$

Therefore, for any $b \in S$,

$$\frac{q(b-\sigma)}{q(\beta-\sigma)\cdot(b-\sigma)} = \frac{1}{b-\sigma},$$

which implies the claim. □

Proof of Lemma 3.3. Denote $q \triangleq q_{L_0}$. Fix $\beta \in L_1$. We know from Claim 3.4 that for any $x \in L_\beta$,

$$f_\beta^{\text{ext}}(x) = f_{L_0+\beta,\sigma}(x) = \frac{q(x-\sigma)}{q(\beta-\sigma)\cdot(x-\sigma)}.$$

Now fix any $\alpha \in L_0'$. Recall Definition A.1 that f_α^{col} is a function with domain $L_1' = q(L_1)$. Define $c_\alpha \triangleq \frac{q(\alpha-\sigma)}{\alpha-\sigma}$. Note that as $\sigma \notin L$ and $\alpha \in L$, so $\alpha - \sigma \notin L$ (and in particular $\alpha - \sigma \notin L_0$), and this implies $q(\alpha-\sigma) \neq 0$; therefore, c_α is non-zero. For any $\beta \in L_1$ we have

$$f_\alpha^{\text{col}}(q(\beta)) = \frac{c_\alpha}{q(\beta)-q(\sigma)},$$

where $q(\sigma) \notin L_1'$ as $\sigma \notin L_1$. Defining $\sigma' \triangleq q(\sigma)$, the first part of the lemma is proved. The second part follows by induction, recalling that (i) a column-test is selected with probability α_η. □

Column compliant attack on rational function The following claim shows that in a column-compliant extension of a rational function a $(1-2^{-\eta})$-fraction of the rows, can be written as a sum of a function of the allowed degree and a high degree function. In particular, these rows will have high degree.

Lemma 3.5 (Column-compliant extension of rational functions)**.** *Let $L \subseteq \mathbb{F}$ be a linear subspace, and $\sigma \in \mathbb{F}\backslash(L \cup L_1')$. Let $g \triangleq f_{L,\sigma}$ be a rational function. Let $f \triangleq \Omega_|(g) : \Omega(L) \to \mathbb{F}$ be the column-compliant extension of g. Then for a $(1-2^{-\mu})$-fraction $\beta \in L_1$, we can write*

$$f_\beta^{\text{ext}} = f_1 + f_2,$$

for functions $f_1, f_2 : L_\beta \to \mathbb{F}$ where $f_1 \in \mathsf{RS}_{L_\beta}[\mu+1]$ and $f_2(x) = \frac{c_\beta}{(x+\sigma)}$ for some $0 \neq c_\beta \in \mathbb{F}$ when $x \in L_0 + \beta$, and $f_2(x) = 0$ otherwise.

This lemma implies that the verifier V_η, given a depth one column-compliant extension $f = \Omega_|(g)$ of a rational function $g : L \to \mathbb{F}$, will reject when deciding to query a row, and then choosing a row outside of a set $S \subset L_1$ of density $2^{-\eta}$. This event (that leads the verifier to reject) occurs with probability

$$(1-\alpha_\eta)\cdot(1-2^{-\eta}) = \left(1 - \frac{1-2^{-\eta}}{2-2^{-\eta}}\right)\cdot(1-2^{-\eta}) = \frac{1-2^{-\eta}}{2-2^{-\eta}} = \alpha_\eta.$$

Proof of Lemma 3.5. Denote $q \triangleq q_{L_0}$. Recall the column-compliant extension $f = \Omega_|(g)$ is defined by first choosing a certain subspace $S \subset L_1$ of size $|S| = 2^{-\eta}\cdot|L_1|$, and defining $f_\beta^{\text{ext}} \triangleq \mathrm{LDE}_{L_\beta}(g|_{L_0+\beta})$. Using Claim 3.4 this implies for each $\beta \in S$ and $x \in L_\beta$

$$f_\beta^{\text{ext}}(x) = \frac{q(x-\sigma)}{q(\beta-\sigma)\cdot(x-\sigma)}.$$

In particular, for any $(\alpha, \beta) \in L_0' \times S$,

$$f(\alpha, q(\beta)) = \frac{q(\alpha-\sigma)}{(\alpha-\sigma)\cdot(q(\beta)-\sigma)}.$$

Fix any $\alpha \in L'_0$, and denote $c_\alpha \triangleq \frac{q(\alpha-\sigma)}{\alpha-\sigma}$. Denote $S' \triangleq q(S)$, and $q'(Y) \triangleq q_{S'}(Y)$. We currently know that $f^{\mathrm{col}}_\alpha|_{S'} = \frac{c_\alpha}{Y+\sigma}|_{S'}$, and thus $f^{\mathrm{col}}_\alpha(Y) = c_\alpha \cdot f_{S',\sigma}$. Using Claim 3.4 again with $H = S'$, we have that

$$f^{\mathrm{col}}_\alpha(Y) = \frac{c_\alpha \cdot q'(Y-\sigma)}{q'(\sigma)\cdot(Y-\sigma)}.$$

Hence, for any $\beta \in L_1 \setminus S$ and $x \in L'_0$,

$$f^{\mathrm{ext}}_\beta(x) = \frac{q(x-\sigma)\cdot q'(q(\beta)-\sigma)}{(x-\sigma)\cdot q'(\sigma)\cdot(q(\beta)-\sigma)}$$

Recall that the column-compliant extension now defines f^{ext}_β on $x \in L_\beta \setminus (L'_0 \cup (L_0+\beta))$ according to the unique polynomial of degree less than $|L_0|$ coinciding with f^{row}_β. Fix $\beta \in L_1 \setminus S$. For fixed β, the expression in the above equation already coincides with a polynomial $f_1(X)$ of degree less than $|L_0|$ in X, namely,

$$f_1(X) \triangleq \frac{q(X-\sigma)\cdot c'_\beta}{(X-\sigma)},$$

where $c'_\beta \triangleq \frac{q'(q(\beta)-\sigma)}{q'(\sigma)\cdot(q(\beta)-\sigma)}$. Thus, for all $x \in L_\beta \setminus (L_0+\beta)$, $f^{\mathrm{ext}}_\beta(x) = f_1(x)$. It follows from Claim 3.4 that $f_1|_{L_0+\beta} = \frac{c''_\beta}{X-\sigma}|_{L_0+\beta}$ for some $c''_\beta \in \mathbb{F}$.

Now define $f_2(x) : L_\beta \to \mathbb{F}$ by $f_2(x) \triangleq f^{\mathrm{ext}}_\beta(x) - f_1(x)$; thus, $f_2(x) = 0$ for $x \in L_\beta \setminus (L_0+\beta)$, and $f_2(x) = \frac{1+c''_\beta}{x-\sigma}$ for $x \in L_0+\beta$. It is now left to show that $c_\beta \triangleq 1 + c''_\beta \neq 0$: If $c_\beta = 0$, then $f^{\mathrm{ext}}_\beta = f_2$. We show this cannot be the case.

Let $g' = \mathrm{LDE}_L(g|_{L_0+S})$. As we used a column-compliant extension, we have $f_2 = \mathrm{LDE}_{L_\beta}(g'|_{L_0+\beta})$. On the other hand $f^{\mathrm{ext}}_\beta|_{L_0+\beta} = g|_{L_0+\beta}$ Note that $g'(x) \neq g(x)$ for any $x \notin L_0+S$, as the function $\frac{1}{X+\sigma}$ can agree with a polynomial of degree $2^{k-\eta}-1$ on at most $2^{k-\eta}$ elements, and $|L_0+S| = 2^{k-\eta}$. Thus, f_2 and f^{ext}_β must disagree on all of $L_0+\beta$. □

4 Proof of Main Theorem

We now state and prove our main Theorem, which appeared informally as Theorem 1.8.

Theorem 4.1 (Main)**.** *Let ϕ be a minimally unsatisfiable* $\mathsf{UACSP}[\frac{1}{8}]$ *instance. Let π be the pseudo-proof for ϕ supplied by a rational attack (Definition 2.1). Then*

1. *Assuming Conjecture 3.2, the verifier V_η from Definition 3.1 rejects π with probability at least*

$$\frac{7}{15}\cdot\left(\frac{3}{7}\right)^{d-1} \tag{4.1}$$

2. *The same probability of rejection (4.1) holds unconditionally when either $d = 1$ or the rational attack uses only a row-compliant PCPP attack.*

Proof. Let A be the minimal unsatisfying assignment for ϕ given in Lemma 2.4. Let f be the RS-codeword that is the evaluation of A and let g be its corresponding function as specified by Definition 2.1. Let π be the full PCP for ϕ specified by that attack.

By definition of the rational attack, the probability of rejecting π is at least the probability that the PCPP-verifier rejects g (and its PCPP). By (2.3) and the discussion preceding the proof of Lemma 2.4, this latter rejection probability is equal to the probability of rejecting the rational function $c/(x-\sigma)$ (along with its axis-compliant attack). Conjecture 3.2 completes the proof of the first statement and Lemmas 3.5, 3.3 prove the second statement. □

5 PCPP security on random functions against axis-compliant attacks

We proceed to analyze the security of the BSS PCPP on random functions. Random functions are a good starting point for analyzing security of the PCP system of [BS08] for two reasons:

- We are not aware of "natural" unsatisfiable instances for which a pseudo-proof leads to evaluations of functions that have lower (i.e., worse) soundness than what can be obtained for a uniformly random function.
- Random functions "resemble" rational functions as both are maximally far from low-degree polynomials.

The main result of this section is the following Lemma, which gives similar security against axis-compliant attacks for random functions, as conjectured for rational functions (cf. Conjecture 3.2).

Lemma 5.1. *Assume that* $|\mathbb{F}| > 10 \cdot 2^{k+2+2.6(d-1)}$, $k = \dim(L) \geq 2^d \cdot 2d$ *and* $\eta \geq 2$. *With probability* 0.9 *over random* $g : L \to \mathbb{F}$, *for all depth* d *axis-compliant extensions* $f : \Omega_L(d) \to \mathbb{F}$ *of* g,

$$\Pr\left[\mathrm{V}_\eta^f = \mathsf{reject}\right] \geq \alpha_\eta \cdot \alpha_2^{d-1} = \alpha_\eta \cdot (3/7)^{d-1},$$

We begin by introducing notation that will make it more convenient to discuss recursive queries on a function $f : \Omega_d(L) \to \mathbb{F}$. We remind again that when the term row is used, it always refers to extended rows in the terminology of Section A.

First, we need the following notation for a function $f : \Omega_d(L) \to \mathbb{F}$ (that is not necessary a BSS extension of some $g \in \mathsf{RS}_L[\eta(f)]$.

1. For $\alpha \in L'_0$ we denote by $\Omega_{d-1}(f^{\mathrm{col}}_\alpha)$ the "depth $d-1$-extension of f^{col}_α in f". Formally, in the notation of Definition A.4,
$$\Omega_{d-1}(f^{\mathrm{col}}_\alpha) \triangleq f|_{\Omega^\alpha_{d-1}(L'_1)}.$$
2. Similarly, for $\beta \in L_1$ we denote by $\Omega_{d-1}(f^{\mathrm{ext}}_\beta)$, the "depth $d-1$-extension of f^{ext}_β in f". Formally,
$$\Omega_{d-1}(f^{\mathrm{ext}}_\beta) \triangleq f|_{\Omega^{\beta^{\mathrm{ext}}}_{d-1}(L_\beta)}$$

Notation for recursive queries We define $q_{L,1}$ to be the set

$$q_{L,1} \triangleq (\mathrm{col}, L'_0) \cup (\mathrm{row}, L_1).$$

For $d > 1$, we define

$$q_{L,d} \triangleq (\mathrm{col}, L'_0) \times q_{L'_0,d-1} \cup_{\beta \in L_1} (\mathrm{row}, \beta) \times q_{L_\beta,d-1}.$$

For example, an element of $q_{L,3}$ can look like $((\mathrm{row}, \beta_1), (\mathrm{col}, \alpha_2), (\mathrm{row}, \beta_3))$ where $\beta_1 \in L_1, \alpha_2 \in (L_\beta)'_0$, and $\beta_3 \in ((L_\beta)'_0)_1$.

For a function $f : \Omega(L) \to \mathbb{F}$, and $\gamma \in q_{L,1}$, we define $f(\gamma)$ to be

- f^{col}_α when $\gamma = (\mathrm{col}, \alpha)$.
- f^{ext}_β when $\gamma = (\mathrm{row}, \beta)$.

More generally, given $d > 1$, a function $f : \Omega_d(L) \to \mathbb{F}$, and $\gamma \in q_{L,1}$ we define $f(\gamma)$ to be

- $\Omega_{d-1}(f^{\mathrm{col}}_\alpha)$ when $\gamma = (\mathrm{col}, \alpha)$.
- $\Omega_{d-1}(f^{\mathrm{ext}}_\beta)$ when $\gamma = (\mathrm{row}, \beta)$.

Finally, for $\gamma = (\gamma_1, \ldots, \gamma_d) \in q_{L,d}$ and such f, we define $f(\gamma)$ to be $f(\gamma_1)(\gamma_2, \ldots, \gamma_d)$. We also denote by L_γ the domain of f_γ.

In this section, we often denote by G a random variable taking values as a function $G : L \to \mathbb{F}$.

In our analysis we will end up with functions that have "at least one random element per row". This motivates the following definition.

Segment-wise independence We say a random variable $G : L \to \mathbb{F}$ is *segment-wise independent*, if for any $\beta \in L_1$, there exists an element $a \in L_0 + \beta$ such that $G(a)$ is uniform and independent of the values $\{G(a')\}_{a' \in L \setminus \{a\}}$.

A useful and immediate property, is that if G is segment-wise independent, it is unlikely it will be a low-degree polynomial. Specifically,

$$\Pr(G \in \mathsf{RS}_L[\eta]) \leq 1/|\mathbb{F}|,$$

for any $\eta \geq 1$.

The following claim will be useful for analyzing row-compliant extensions.

Claim 5.2 (Row-compliant extensions of random functions)**.** *Suppose that $G : L \to \mathbb{F}$ is segment-wise independent, and let F be the random variable $F \triangleq \Omega_{-}(G) : \Omega(L) \to \mathbb{F}$. Then, for any $\alpha \in L'_0$ we have that $F^{\mathrm{col}}_\alpha : L'_0 \to \mathbb{F}$ is a uniformly random function.*

Proof. Fix any $\beta \in L_1$, and $\alpha \in L'_0$. Recall that $F^{\mathrm{ext}}_\beta = \mathrm{LDE}_{L_\beta}(G|_{L_0+\beta})$. Thus, $F^{\mathrm{ext}}_\beta(\alpha)$ is a linear combination of the elements of the set $T_\beta \triangleq \{G(x)\}_{x \in L_0+\beta}$, where all elements have non-zero coefficient. As G is segment-wise independent there is an element $t_\beta \in T_\beta$ that is uniformly random and independent of other elements of T_β, and thus $f^{\mathrm{ext}}_\beta(\alpha)$ is uniformly random. Furthermore, the values $\left\{F^{\mathrm{ext}}_\beta(\alpha)\right\}_{\beta \in L_1}$ are independent, as the elements $\{t_\beta\}_{\beta \in L_1}$ are independent. As this is exactly the set of values of F^{col}_α, it follows that F^{col}_α is a uniformly random function. □

We proceed to discuss the column-compliant extension. We denote by $q^{\mathrm{row}}_{L,d}$ the subset of elements $\gamma \in q_{L,d}$ such that γ_i has the form (row, β_i) for all $i \in [d]$.

Claim 5.3 (Column-compliant extensions of random functions)**.** *Assume that $k = \dim(L) \geq 2^d \cdot 2d$. Let $G : L \to \mathbb{F}$ be uniformly distributed. Let $F \triangleq \Omega_{d,|}(G) : \Omega_d(L) \to \mathbb{F}$. Let γ be a uniform element of $q^{\mathrm{row}}_{L,d}$. Then, F_γ is segment-wise independent with probability at least $(1 - 2^{-\eta}) \cdot (3/4)^{d-1}$ over the choice of γ.*

Proof. Consider first the case $d = 1$ for simplicity. Recall that $F = \Omega_{|}(G)$ is defined outside of $T(L)$, only as a function of the values of G on $\hat{L} = \{L_0 + \beta | \beta \in S\}$, for a set $R_{\hat{L}} \subset L_1$ with $|R_{\hat{L}}| = 2^{-\eta} \cdot |L_1|$. On the other hand, for any $\beta \in L_1$, F^{ext}_β contains the $|L_\beta|/4$ values $T_\beta = \{G(x)\}_{x \in L_0+\beta}$. Thus, for $\beta \notin R_{\hat{L}}$, F^{ext}_β contains a set of values T_β that are each random and independent of all other values of F^{ext}_β. It follows from Remark A.5 that for any $\beta' \in (L_\beta)_1$, $F^{\mathrm{ext}}_\beta|_{L_0+\beta'}$ contains one of these values. Summing up, we have that for uniform $\gamma = (\mathrm{row}, \beta) \in q^{\mathrm{row}}_{L,1}$, F_γ is segment-wise independent with probability $(1 - 2^{-\eta})$.

Similarly, given $\gamma = ((\mathrm{row}, \beta_1), \ldots, (\mathrm{row}, \beta_d))$, such that β_i always "evades" the set of rows S_i on according to which $\Omega_{|}(F_{\gamma_{<i}})$ is constructed, we have, using Remark A.5, that F_γ is segment-wise independent. The set $S_1 = R_{\hat{L}}$ has density $2^{-\eta}$, while the sets $S_2, \ldots, S_d$ have density $1/4$. □

5.1 Proof of Lemma 5.1

Recall the biased RS verifier from Definition 3.1 We proceed to analyze its security against axis-compliant attacks (Definition A.10).

Using the bounds $\dim L_\beta, \dim L_1 \leq k/2 + 1.5$, calculation shows that given $f : \Omega_d(L) \to \mathbb{F}$, for L of dimension k, V^g_η reads at most $q_0(k, d) \triangleq 2^{k/2^d + 3(1-1/2^d)}$ locations of g.

The next notation, will formalize the notion of "the answer to the query that catches the prover when using an axis-compliant extension of a non-codeword".

For $\gamma \in q_{L,1}$, and $g : L \to \mathbb{F}$ we define the "evading answer of g on γ", denoted g^*_γ, as follows.

- $g^*_\gamma \triangleq f^{\mathrm{col}}_\alpha$ for $f = \Omega_{-}(g)$ when $\gamma = (\mathrm{col}, \alpha)$.
- $g^*_\gamma \triangleq f^{\mathrm{ext}}_\beta$ for $f = \Omega_{|}(g)$ when $\gamma = (\mathrm{row}, \beta)$ for $\beta \notin R_{\hat{L}}$ and undefined otherwise.

For $\gamma \in q_{L,d}$ when $d > 1$ we define

- $g^*_\gamma \triangleq f^{\mathrm{col}}_\alpha$ for $f = \Omega_{-}(g^*_{\gamma_{<d}})$ when $\gamma_d = (\mathrm{col}, \alpha)$.

- $g^*_\gamma \triangleq f^{\text{ext}}_\beta$ for $f = \Omega_|(g^*_{\gamma_{<d}})$ when $\gamma_d = (\text{row}, \beta)$, f is defined and $\beta \notin R_{\hat{L}_{\gamma_{<d}}}$; and undefined otherwise.

Before proceeding to the proof of Lemma 5.1 we define, for $\gamma \in q_{L,d}$, by $\eta(\gamma)$ "the rate we are testing f_γ for". Formally, for $\gamma \in q_{L,d}$ and assuming we have a parameter η implicit in the context,

- $\eta(\gamma) \triangleq \eta$ when γ is a pure column query, i.e., $\gamma = ((\text{col}, \alpha_1), (\text{col}, \alpha_2), \ldots, (\text{col}, \alpha_d))$
- $\eta(\gamma) \triangleq 2$ otherwise.

Proof of Lemma 5.1. Consider the following sampling procedure for a random variable $\gamma \in q_{L,d}$: For each $i \in [d]$, with probability $\alpha_{\eta(\gamma_{<i})}$ γ_i is a uniform column query, i.e., $\gamma_i = (\text{col}, \alpha)$ for uniform $\alpha \in (L_{\gamma_{<i}})'_0$; and with probability $1 - \alpha_{\eta(\gamma_{<i})}$, $\gamma_i = (\text{row}, \beta)$ for uniform $\beta \in (L_{\gamma_{<i}})_1$.

Let $G : L \to \mathbb{F}$ be uniformly distributed. Let $\gamma \in q_{L,d}$ be such that G^*_γ is defined. Then the same arguments as in Claims 5.2 and 5.3 can be used to show G^*_γ is segment-wise independent; and thus $G^*_\gamma \in \mathsf{RS}_{L_\gamma}[\eta(\gamma)]$ with probability at most $1/|\mathbb{F}|$. A union bound now shows that with probability 0.9 over uniform $g : L \to \mathbb{F}$, $g^*_\gamma \notin \mathsf{RS}_{L_\gamma}[\eta(\gamma)]$ for all $\gamma \in q_{L,d}$ for which it is defined.

On the other hand, for such $g : L \to \mathbb{F}$, and any depth d axis-compliant extension f of g, we claim that $f_\gamma = g^*_\gamma$ with probability at least $\alpha_\eta \cdot (3/7)^{d-1}$, when γ is sampled as above: Using the abbreviation $\eta_i \triangleq \eta(\gamma_{<i})$ for $i \in \{2, \ldots, d\}$ and $\eta_1 \triangleq \eta$, the event $f_\gamma = g^*_\gamma$ corresponds to the event that for each $i \in [d]$, either

1. $\Omega(f_{\gamma_{<i}}) \subset \Omega_d(f)$ is a row-compliant extension, and γ_i is a column query - which happens with probability α_{η_i}, or
2. $\Omega(f_{\gamma<i})$ is a column-compliant extension, and $\gamma_i = (\text{row}, \beta)$ is a row query for $\beta \notin S_i$; where S_i is the set of density $2^{-\eta_i}$ according to which the extension was constructed. This happens with probability

$$(1 - \alpha_{\eta_i}) \cdot (1 - 2^{-\eta_i}) = \alpha_{\eta_i},$$

where the equality can be verified from the definition of α_{η_i} (see Definition 3.1).

Using $\eta_i \geq 2$ for each $i \in [d]$, we have that indeed $f_\gamma = g^*_\gamma$ with probability at least $\alpha_\eta \cdot (3/7)^{d-1}$.

Noticing that V_η ends up reading f_γ for γ sampled as above, we are done. □

5.2 Concrete security threshold of depth-2PCPPs on random functions

To phrase our results on random functions, we give a formal definition of a PCPP system that is secure against *most* inputs x, when the prover is limited to a certain set of strategies for generating the auxiliary proof y.

Definition 5.4 (ϵ-PCPP for a code C against prespecified attacks)**.** *Fix integers $A, Q \in \mathbb{N}$ and $0 < \epsilon, \delta < 1$. Let $C \subseteq \mathbb{F}^n$ be an $[n = n(C), k = k(C), d = d(C)]$-code. Let $\mathcal{H}$ be a set of functions $h : \mathbb{F}^n \to \mathbb{F}^{A-n}$*

An (A, Q, ϵ)-$\mathcal{H}$-resistant PCPP system *$\mathcal{S}$ for C with soundness error δ is a pair $\mathcal{S} = (P, V)$, where*

- *P is a* systematic *mapping $P : C \to \mathbb{F}^A$.*

  That is, for any $x \in C$, $P(x) = (x, y)$ for some $y \in \mathbb{F}^{A-n}$.

- *V is a Q-*local *randomized mapping $V : \mathbb{F}^A \to \{\mathsf{accept}, \mathsf{reject}\}$. That is, after choosing its internal randomness, $V(z)$ always depends on at most Q indices of $z \in \mathbb{F}^A$.*

Such that

- *(Completeness) For any $x \in C$, $V(P(x)) = \mathsf{accept}$ with probability one.*
- *(Soundness) For a $(1 - \epsilon)$-fraction of $x \in \mathbb{F}^n$, and any $y \in \mathbb{F}^{A-n}$ of the form $y = h(x)$ for some $h \in \mathcal{H}$, $V((x, y)) = \mathsf{accept}$ with probability at most δ.*

We define a concrete efficiency threshold in this setting analogously to Definition 1.9.

Definition 5.5 (ϵ-Concrete efficiency threshold of a PCPP against prespecified attacks)**.** *We say an* (A, Q, ϵ)-$\mathcal{H}$-resistant PCPP system $\mathcal{S}$ *for* C *is* efficient *if the cost* $A \cdot Q \leq k(C)^2/2$.

Fix an ensemble of linear codes $\mathcal{C} = \left\{C \subseteq \mathbb{F}^{n(C)}\right\}$*, and functions* $A, Q : \mathbb{N} \to \mathbb{N}$*. An* (A, Q, ϵ)-$\mathcal{H}$-resistant PCPP system for $\mathcal{C}$ *is an ensemble of PCPP systems* $\mathcal{S} = \{\mathcal{S}_C | C \in \mathcal{C}\}$ *where* $\mathcal{S}_C$ *is an* $(A(k(C)), Q(k(C)), \epsilon)$-$\mathcal{H}$*-resistant PCPP system for* C*. The* ϵ-concrete efficiency threshold of $\mathcal{S}$ *is the smallest integer* $\mathbf{k}$ *such that for any* $C \in \mathcal{C}$ *of dimension* $k(C) \geq \mathbf{k}$*,* $\mathcal{S}_C$ *is efficient.*

For the sake of comparison with the improved concrete soundness studied in Section B and summarized in Table 4, we present here the concrete soundness threshold where soundness is measured on random functions using only the pair of attacks studied previously; for the sake of comparison, we fix the recursion depth to 2, as studied there. The bottom line is quite encouraging, showing a far better concrete threshold security, and stressing the importance of suggesting and analyzing other attacks on PCPP systems.

Corollary 5.6. *Fix positive integer* d*. Assume* $|\mathbb{F}| > 10 \cdot 2^{k+2+2.6(d-1)}$*. There is a* $(2^{\ell+5+2.6\cdot(d-1)}, q_0(k,d) \cdot m, 0.1)$-*PCPP system for* $\mathrm{RS}_{\mathrm{a}}[3]$ *resistant to axis-compliant attacks; where* $m \triangleq \lceil \frac{\log(0.5)}{\log(1-\frac{7}{15}\cdot(3/7)^{d-1})} \rceil$*, for* $k = \ell + 3$*.*

Proof. Fix a subspace $L \subseteq \mathbb{F}$ of dimension k. Fix $g : L \to F$. Let f be a depth d axis-compliant extension of g. Similarly to Lemma B.12 we simply run the verifier from Definition 3.1 on f m times and reject if one of the runs rejected. Using Lemma 5.1, with probability 0.9 over the choice of g, each run rejects with probability at least $(7/15) \cdot (3/7)^{d-1}$. As each run requires reading $q_0(k, d)$ entries of f, the claim follows. □

In particular,we can compare the improved concrete efficiency threshold to the concrete threshold with respect to axis-compliant "attacks".

Corollary 5.7 (Security on random functions)**.** *Assume* $|\mathbb{F}| \geq 10 \cdot 2^{k+2+2.6(d-1)}$*. There is a PCPP system for* $\mathrm{RS}_{\mathrm{a}}[3]$ *resistant to axis-compliant attacks with* 0.1*-concrete efficiency threshold* 2^{19}*.*

message length	codeword length	proof length	# of queries
23	26	30.6	10.75
24	27	31.6	11
25	28	32.6	11.25
26	29	33.6	11.5
27	30	34.6	11.75
28	31	35.6	12
29	32	36.6	12.25
30	33	37.6	12.5
31	34	38.6	12.75
32	35	39.6	13
33	36	40.6	13.25
34	37	41.6	13.5
35	38	42.6	13.75

Table 2: Instantiations of the system in Corollary 5.6 with $d = 2$. All numbers are logs in base two of the described quantity. The first column describes the length (in field elements) of the message w to be encoded into a word $x \in \mathsf{RS}_L[3]$. The second column is the length of x. The third column is the length of x together with the proof y that $x \in \mathsf{RS}_L[3]$. The fourth is the number of field elements a verifier needs to read to be reject with probability $1/2$ for a 0.9 fraction of $x : L \to \mathbb{F}$.

We end by asking whether, for random $f : L \to \mathbb{F}$, the minimal rejection probability taken over all pseudo-proofs, is significantly smaller than the rejection probability with respect to row- and column-compliant pseudo-proofs. Currently, we cannot rule out the possibility that no better pseudo-proof exists!

5.3 Reducing proof length with interactive oracle proofs of proximity

In this section we show that, when allowing a few rounds of interaction between the prover and verifier, the efficiency of the PCPP can be significantly improved. This interaction uses the IOPP model as presented in [BCG$^+$16], based on the IOP model defined in [BSCS16, RRR16]. We give a tailored definition of IOPPs convenient for our purposes.

Definition 5.8. *Fix integers $A, Q \in \mathbb{N}$ Let $C \subseteq \mathbb{F}^n$ be an $[n = n(C), k = k(C), d = d(C)]$-code.An (A, Q)-IOPP system $\mathcal{S}$ for C is a pair $\mathcal{S} = (P, V)$, of players that run an interactive protocol, where*

- *The first message is a systematic mapping $P : C \to \mathbb{F}^{A_1}$*
- *The total size of messages sent by P is at most A.*
- *The total number of locations read by V from P's answers is $\leq Q$.*

Such that

- *(Completeness) For any $x \in C$, $V(P(x)) =$ accept with probability one.*
- *(Soundness) For any $z = (x, y) \in \mathbb{F}^A$ such that $\Delta(x, C) \geq d/3$, $V(z) =$ reject with probability at least $1/2$.*

Given $\epsilon > 0$, we also define an (A, Q, ϵ)-IOPP system in a similar way to Definition 5.4; that is, the soundness condition needs to hold for a $(1 - \epsilon)$-fraction of $x \in \mathbb{F}^n$, rather than x that is $d/3$-far from C.

Lemma 5.9. *Assume $|\mathbb{F}| > 10 \cdot 2^{k+2+2.6(d-1)}$. There is a $(4 \cdot 2^k + 8 \cdot 4 \cdot (2^{k/2+1.5} + 2^{k/4+2.25}), 8 \cdot 2^{k/8+2.25}, 0.1)$-IOPP system for $\mathrm{RS}_{\mathrm{a}}[3]$ with resistant to axis-compliant attacks, where $k = \ell + 3$*

Proof. The proof is similar to that used in the results of [BCG$^+$16] on IOPPs for RS codes (see Theorem 1.2 and Section 5 there), and we do not give a fully formal argument. Fix $g : L \to \mathbb{F}$. As in the proof of Corollary 5.6 we use the verifier from Definition 3.1; here we fix depth $d = 3$. The difference is that know we are constructing an IOPP system so P only has to write down the depth one BSS-extension f of g at the start. Afterwards he will send the depth one extension f' of f^{col}_α or f^{ext}_β according to V's decision to query an extended row or column. As V is using depth 3, he will choose once more a depth one extension of a column or extended row of f' for P to send.

Thus, for each repetition, P will need to write at most $4 \cdot (2^{k/2+1.5} + 2^{k/4+2.25})$ elements in addition to writing f at the start, which has length 2^{k+2}. (The 4 factor is because a depth one extension of a function is four times as long as the function itself).

□

message length	codeword length	proof length	# of queries
23	26	28.1	8.9
24	27	29.1	9
25	28	30.1	9.2
26	29	31.1	9.3
27	30	32.1	9.4
28	31	33.1	9.5
29	32	34.1	9.7
30	33	35.1	9.8
31	34	36.1	9.9
32	35	37.1	10
33	36	38.1	10.2
34	37	39.1	10.3
35	38	40.1	10.4

Table 3: Instantiations of the system in Lemma 5.9 for $d = 3$. All numbers are logs in base two of the described quantity. The first column describes the length (in field elements) of the message w to be encoded into a word $g \in \mathsf{RS}_L[3]$. The second column is the length of g. The third column is the total length of P's messages while proving $g \in \mathsf{RS}_L[3]$. The fourth is the number of field elements a verifier needs to read to be reject with probability $1/2$ for a 0.9 fraction of $g : L \to \mathbb{F}$.

References

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, 1998. Preliminary version in FOCS '92.

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: a new characterization of NP. *Journal of the ACM*, 45(1):70–122, 1998. Preliminary version in FOCS '92.

[BBC+16] Eli Ben-Sasson, Iddo Bentov, Alessandro Chiesa, Ariel Gabizon, Daniel Genkin, Matan Hamilis, Evgenya Pergament, Michael Riabzev, Mark Silberstein, Eran Tromer, and Madars Virza. Computational integrity with a public random string from quasi-linear PCPs. *IACR Cryptology ePrint Archive*, 2016:646, 2016.

[BBGR16] Eli Ben-Sasson, Iddo Bentov, Ariel Gabizon, and Michael Riabzev. Improved concrete efficiency and security analysis of reed-solomon pcpps. *Electronic Colloquium on Computational Complexity (ECCC)*, 23:73, 2016.

[BCG+13] Eli Ben-Sasson, Alessandro Chiesa, Daniel Genkin, Eran Tromer, and Madars Virza. Snarks for C: verifying program executions succinctly and in zero knowledge. In *Advances in Cryptology - CRYPTO 2013 - 33rd Annual Cryptology Conference, Santa Barbara, CA, USA, August 18-22, 2013. Proceedings, Part II*, pages 90–108, 2013.

[BCG+16] Eli Ben-Sasson, Alessandro Chiesa, Ariel Gabizon, Michael Riabzev, and Nicholas Spooner. Short interactive oracle proofs with constant query complexity, via composition and sumcheck, 2016. Crypto ePrint 2016/324.

[BCGT13] Eli Ben-Sasson, Alessandro Chiesa, Daniel Genkin, and Eran Tromer. On the concrete efficiency of probabilistically-checkable proofs. In *Proceedings of the 45th ACM Symposium on the Theory of Computing*, STOC '13, 2013.

[BCGV16] Eli Ben-Sasson, Alessandro Chiesa, Ariel Gabizon, and Madars Virza. Quasi-linear size zero knowledge from linear-algebraic PCPs. In *13th Theory of Cryptography Conference, TCC*, 2016.

[BCI+13] Nir Bitansky, Alessandro Chiesa, Yuval Ishai, Rafail Ostrovsky, and Omer Paneth. Succinct non-interactive arguments via linear interactive proofs. In *TCC*, pages 315–333, 2013.

[BFL90] László Babai, Lance Fortnow, and Carsten Lund. Non-deterministic exponential time has two-prover interactive protocols. In *31st Annual Symposium on Foundations of Computer Science, St. Louis, Missouri, USA, October 22-24, 1990, Volume I*, pages 16–25, 1990.

[BFLS91] László Babai, Lance Fortnow, Leonid A. Levin, and Mario Szegedy. Checking computations in polylogarithmic time. In *Proceedings of the 23rd Annual ACM Symposium on Theory of Computing*, STOC '91, pages 21–32, 1991.

[BGH+06] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil Vadhan. Robust PCPs of proximity, shorter PCPs, and applications to coding. *SIAM Journal on Computing*, 36(4):889–974, 2006. Preliminary versions of this paper have appeared in Proceedings of the 36th ACM Symposium on Theory of Computing and in Electronic Colloquium on Computational Complexity.

[BM88] Mihir Bellare and Silvio Micali. How to sign given any trapdoor function. In *STOC '88: Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, pages 32–42, New York, NY, USA, 1988. ACM.

[BS08] Eli Ben-Sasson and Madhu Sudan. Short PCPs with polylog query complexity. *SIAM Journal on Computing*, 38(2):551–607, 2008. Preliminary version appeared in STOC '05.

[BSCS16] Eli Ben-Sasson, Alessandro Chiesa, and Nicholas Spooner. Interactive oracle proofs. Cryptology ePrint Archive, Report 2016/116, 2016. `http://eprint.iacr.org/`.

[BSVW03] Eli Ben-Sasson, Madhu Sudan, Salil Vadhan, and Avi Wigderson. Randomness-efficient low degree tests and short pcps via epsilon-biased sets. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing*, STOC '03, pages 612–621, 2003.

[CMT12] Graham Cormode, Michael Mitzenmacher, and Justin Thaler. Practical verified computation with streaming interactive proofs. In *Innovations in Theoretical Computer Science 2012, Cambridge, MA, USA, January 8-10, 2012*, pages 90–112, 2012.

[Din07] Irit Dinur. The PCP theorem by gap amplification. *Journal of the ACM*, 54(3):12, 2007.

[DR04] Irit Dinur and Omer Reingold. Assignment testers: Towards a combinatorial proof of the PCP theorem. In *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '04, pages 155–164, 2004.

[FGL+96] Uriel Feige, Shafi Goldwasser, Laszlo Lovász, Shmuel Safra, and Mario Szegedy. Interactive proofs and the hardness of approximating cliques. *Journal of the ACM*, 43(2):268–292, 1996. Preliminary version in FOCS '91.

[GGPR13] Rosario Gennaro, Craig Gentry, Bryan Parno, and Mariana Raykova. Quadratic span programs and succinct nizks without pcps. In *Advances in Cryptology - EUROCRYPT 2013, 32nd Annual International Conference on the Theory and Applications of Cryptographic Techniques, Athens, Greece, May 26-30, 2013. Proceedings*, pages 626–645, 2013.

[GKR08] Shafi Goldwasser, Yael Tauman Kalai, and Guy N. Rothblum. Delegating computation: interactive proofs for Muggles. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, STOC '08, pages 113–122, 2008.

[GOS06] Jens Groth, Rafail Ostrovsky, and Amit Sahai. Non-interactive zaps and new techniques for NIZK. In *Proceedings of the 26th Annual International Conference on Advances in Cryptology*, CRYPTO '06, pages 97–111, 2006.

[Gro09] Jens Groth. Linear algebra with sub-linear zero-knowledge arguments. In *Proceedings of the 29th Annual International Cryptology Conference on Advances in Cryptology*, CRYPTO '09, pages 192–208, 2009.

[Gro10] Jens Groth. Short non-interactive zero-knowledge proofs. In *Proceedings of the 16th International Conference on the Theory and Application of Cryptology and Information Security*, ASIACRYPT '10, pages 341–358, 2010.

[GS06] Oded Goldreich and Madhu Sudan. Locally testable codes and pcps of almost-linear length. *J. ACM*, 53(4):558–655, 2006.

[GS09] Vipul Goyal and Amit Sahai. Resettably secure computation. In *EUROCRYPT '09: Proceedings of the 28th Annual International Conference on Advances in Cryptology*, pages 54–71, Berlin, Heidelberg, 2009. Springer-Verlag.

[Hås01] Johan Håstad. Some optimal inapproximability results. *Journal of the ACM*, 48:798–859, July 2001. Preliminary version in STOC '97.

[HS00] Prahladh Harsha and Madhu Sudan. Small PCPs with low query complexity. *Computational Complexity*, 9(3–4):157–201, Dec 2000. Preliminary version in STACS '91.

[IKO07] Yuval Ishai, Eyal Kushilevitz, and Rafail Ostrovsky. Efficient arguments without short pcps. In *22nd Annual IEEE Conference on Computational Complexity (CCC 2007), 13-16 June 2007, San Diego, California, USA*, pages 278–291, 2007.

[Kil92] Joe Kilian. A note on efficient zero-knowledge proofs and arguments. In *Proceedings of the 24th Annual ACM Symposium on Theory of Computing*, STOC '92, pages 723–732, 1992.

[KMRS15] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally-testable codes with quasi-polylogarithmic query complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:110, 2015.

[KPT97] Joe Kilian, Erez Petrank, and Gábor Tardos. Probabilistically checkable proofs with zero knowledge. In *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing*, STOC '97, pages 496–505, New York, NY, USA, 1997. ACM.

[KR08] Yael Tauman Kalai and Ran Raz. Interactive PCP. In *ICALP '08: Proceedings of the 35th International Colloquium on Automata, Languages and Programming, Part II*, pages 536–547, Berlin, Heidelberg, 2008. Springer-Verlag.

[KR09] Yael Tauman Kalai and Ran Raz. Probabilistically checkable arguments. In *Advances in Cryptology-CRYPTO 2009*, pages 143–159. Springer, 2009.

[KRS15] Swastik Kopparty, Noga Ron-Zewi, and Shubhangi Saraf. High rate locally-correctable and locally-testable codes with sub-polynomial query complexity. *Electronic Colloquium on Computational Complexity (ECCC)*, 22:68, 2015.

[LFKN92] Carsten Lund, Lance Fortnow, Howard Karloff, and Nisan Noam. Algebraic methods for interactive proof systems. *Journal of the ACM*, 39(4):859–868, 1992.

[Lip12] Helger Lipmaa. Progression-free sets and sublinear pairing-based non-interactive zero-knowledge arguments. In *Proceedings of the 9th Theory of Cryptography Conference on Theory of Cryptography*, TCC '12, pages 169–189, 2012.

[Mic00] Silvio Micali. Computationally sound proofs. *SIAM Journal on Computing*, 30(4):1253–1298, 2000. Preliminary version appeared in FOCS '94.

[Mie08] Thilo Mie. Polylogarithmic two-round argument systems. *Journal of Mathematical Cryptology*, 2(4):343–363, 2008.

[MR08] Dana Moshkovitz and Ran Raz. Two-query PCP with subconstant error. *Journal of the ACM*, 57:1–29, June 2008. Preliminary version appeared in FOCS '08.

[NR11] Jakob Nordström and Alexander Razborov. On minimal unsatisfiability and time-space trade-offs for k-dnf resolution. In Luca Aceto, Monika Henzinger, and JiVrí Sgall, editors, *Automata, Languages and Programming*, volume 6755 of *Lecture Notes in Computer Science*, pages 642–653. Springer Berlin Heidelberg, 2011.

[PGHR13] Bryan Parno, Craig Gentry, Jon Howell, and Mariana Raykova. Pinocchio: Nearly practical verifiable computation. *IACR Cryptology ePrint Archive*, 2013:279, 2013.

[PS94] Alexander Polishchuk and Daniel A. Spielman. Nearly-linear size holographic proofs. In *Proceedings of the 26th Annual ACM Symposium on Theory of Computing*, STOC '94, pages 194–203, 1994.

[RRR16] Omer Reingold, Ron Rothblum, and Guy Rothblum. Constant-round interactive proofs for delegating computation. In *Proceedings of the 48th ACM Symposium on the Theory of Computing*, STOC '16, 2016.

[SBW11] Srinath Setty, Andrew J. Blumberg, and Michael Walfish. Toward practical and unconditional verification of remote computations. In *Proceedings of the 13th USENIX Conference on Hot Topics in Operating Systems*, HotOS '13, pages 29–29, 2011.

[SMBW12] Srinath Setty, Michael McPherson, Andrew J. Blumberg, and Michael Walfish. Making argument systems for outsourced computation practical (sometimes). In *Proceedings of the 2012 Network and Distributed System Security Symposium*, NDSS '12, 2012.

[SVP+12] Srinath Setty, Victor Vu, Nikhil Panpalia, Benjamin Braun, Andrew J. Blumberg, and Michael Walfish. Taking proof-based verified computation a few steps closer to practicality. In *Proceedings of the 21st USENIX Security Symposium*, Security '12, page ???, 2012.

[VSBW13] Victor Vu, Srujay Setty, Andrew J Blumberg, and Michael Walfish. A hybrid architecture for interactive verifiable computation. In *Security and Privacy (SP), 2013 IEEE Symposium on*, pages 223–237. IEEE, 2013.

[WB15] Michael Walfish and Andrew J. Blumberg. Verifying computations without reexecuting them. *Commun. ACM*, 58(2):74–84, 2015.

A Definitions

We present some notation and basic facts that will be used throughout the paper. $\mathbb{F}$ always denotes a finite field of characteristic two. Recall that $\mathbb{F}$ is a vector space over $\mathbb{F}_2$. When we refer to a *subspace* $L \subseteq \mathbb{F}$ we always mean *an $\mathbb{F}_2$-subspace of $\mathbb{F}$ when viewed as such a vector space*.

Product Sets Fix sets A, B, an element $a \in A$ and subset $S \subseteq B$. We denote by (a, S) or $a \times S$ the subset $\{(a, b) | b \in S\}$ of $A \times B$.

Functions and Polynomials For polynomials $F, G \in \mathbb{F}[Z]$ we will write $F = G$ or $F(Z) = G(Z)$ to denote equality of formal polynomials. For a subset $S \subseteq \mathbb{F}$ and polynomial $F \in \mathbb{F}[Z]$ we denote by $F|_S$ the function defined by $F|_S(x) \triangleq F(x)$. For functions $f, g : S \to \mathbb{F}$ we write $f = g$ to mean f and g are identical on S. For a domain $S' \subseteq S$ we use the notation $\Delta_{S'}(f, g)$ to denote the fractional disagreement of f and g on S'. That is,

$$\Delta_{S'}(f, g) \triangleq \Pr_{x \leftarrow S'}(f(x) \neq g(x)).$$

We also use the notation $\Delta(f, g) \triangleq \Delta_S(f, g)$. For $F \in \mathbb{F}[Z]$ and a function $g : S \to \mathbb{F}$ we use the abbreviated notations $F = g$ to mean $F|_S = g$, and $\Delta_S(F, g)$ to mean $\Delta_S(F|_S, g)$. We sometimes slightly abuse notation and denote by $g(Z)$ the unique polynomial of degree smaller than $|S|$ with $g|_S = g$. For a function $g : S \to \mathbb{F}$ and $S' \subset \mathbb{F}$, we denote by $\mathrm{LDE}_{S'}(g) : S' \to \mathbb{F}$ the function $g|_{S'}$ (where g is the unique polynomial of degree smaller than S with $g|_S = g$). When discussing bivariate polynomials $Q(X, Y) \in \mathbb{F}[X, Y]$ (in formal variables X, Y) we say Q *has degree* (d, e) if $\deg_X(Q) \leq d$ and $\deg_Y(Q) \leq e$. For two sets $A, B \subseteq \mathbb{F}$ and a bivariate function $f : A \times B \to \mathbb{F}, \alpha \in A$ and $\beta \in B$; we denote $f(X, \beta) \triangleq f|_{X \times \beta}$ and $f(\alpha, Y) \triangleq f|_{\alpha \times Y}$ We use similar notation for $Q(X, Y) \in \mathbb{F}[X, Y]$. For example, $Q(X, \beta)$ denotes the univariate polynomial obtained by substituting $Y = \beta$ in Q. Note that $Q(X, \beta) = f(X, \beta)$ means f and Q identify as functions on $A \times \beta$.

Subspace Polynomials For a subspace $L \subseteq \mathbb{F}$, we denote by q_L the *subspace polynomial of* L, defined as,

$$q_L(Z) \triangleq \prod_{v \in L} (Z - v). \tag{A.1}$$

These polynomials are well-studied. We mention the essential relevant properties for our results. Suppose that L is the direct sum of the linear spaces L_0 and L_1. Then

- $q_{L_0}|_L$ is $|L_0|$-regular.
- $q_{L_0}(L) = q_{L_0}(L_1) = L'_1$ for a subspace $L'_1 \subseteq \mathbb{F}$ with $|L'_1| = |L_1|$. In particular, q_{L_0} is injective on L_1.

A.1 The Reed-Solomon PCP of proximity

Notation For $a, b \in \mathbb{R}$, it will be convenient to use the notation $a \vee b \triangleq \min\{a, b\}$.

Fix a positive integer k. All fields $\mathbb{F}$ mentioned here are finite and of characteristic 2. We make the convention that an $\mathbb{F}_2$-subspace $L \subset \mathbb{F}$ of dimension k is always associated with some default basis $\{b_1, \ldots, b_k\}$ of L. Using this convention, for $c \in [k]$, we define $L_{\leq c} \subseteq L$ to be the subspace spanned by the first c vectors of this default basis. That is, $L_{\leq c} \triangleq \mathrm{span}\{b_1, \ldots, b_c\}$. We define the *midpoint of* k to be $\left\lfloor \frac{k-1}{2} \right\rfloor$. Note that for any k, $k/2 - 1 \leq \left\lfloor \frac{k-1}{2} \right\rfloor \leq k/2 - 0.5$.

The Ben-Sasson-Sudan PCPP [BS08] is based on a special, somewhat complex, subset of $\mathbb{F}^2$ that we describe next. We recommend looking at Appendix C for a helpful visualization of this set.

Definition A.1 (BSS sets)**.** *Let $L \subseteq \mathbb{F}$ be an $\mathbb{F}_2$-subspace of dimension k. The* Ben-Sasson-Sudan (BSS) Set of L *, denoted $\Omega(L)$, is defined as follows. First denote*

- $L_0 \triangleq L_{\leq \left\lfloor \frac{k-1}{2} \right\rfloor}$.

- $L'_0 \triangleq L_{\leq \lfloor \frac{k-1}{2} \rfloor + 1}$.
- $L_1 \triangleq \text{span}\left\{b_{\lfloor \frac{k-1}{2} \rfloor + 1}, \ldots, b_k\right\}$, *and* $L'_1 \triangleq q_{L_0}(L_1)$. *From the properties of subspace polynomials discussed in Section A we have* $|L'_1| = |L_1|$ *and* $L'_1 = q_{L_0}(L)$.
- *For each* $\beta \in L_1$, *let* $L_\beta \triangleq \text{span}\{L'_0, \beta\}$ *if* $\beta \notin L'_0$ *and* $L_\beta = \text{span}\left\{L'_0, b_{\lfloor \frac{k-1}{2} \rfloor + 2}\right\}$ *otherwise. (For simplicity, always think of* L_β *as "*L'_0 *with* β *added" , i.e.,* $\text{span}\{L'_0, \beta\}$*).*
- *For each* $\beta \in L_1$, *now define the '*β*'th extended row' as* $\text{extrow}_\beta \triangleq (L_\beta, q_{L_0}(\beta)) \subset \mathbb{F}^2$.

Finally, we define

$$\Omega(L) \triangleq \bigcup_{\beta \in L_1} \text{extrow}_\beta.$$

Note that $|\Omega(L)| = 4 \cdot |L| = 2^{k+2}$.

For $\alpha \in L'_0$, *define the '*α*'th column' as* $\text{col}_\alpha \triangleq (\alpha, L'_1) \subset \Omega(L)$. *A useful property of* $\Omega(L)$, *that can be easily verified, is that the following two sets are contained in it:*

- *The 'product set of* $\Omega(L)$*':* $S(L) \triangleq L'_0 \times L'_1 \subset \Omega(L)$. [5]
- *The 'curve of* $\Omega(L)$*':* $T(L) \triangleq \{(v, q_{L_0}(v)) | v \in L\} \subset \Omega(L)$.

(We think of $T(L)$ *as an embedding of* L *into* $\Omega(L)$, *and for this reason think of* $\Omega(L)$ *as an extension of* L. *We note again that a helpful visualization of the above sets is found in appendix C.)*

Now fix a function $f : \Omega(L) \to \mathbb{F}$. *Define the* univariate *function* $P[f] : L \to \mathbb{F}$ *by* $P[f](Z) \triangleq f(Z, q_{L_0}(Z))$. *We assume a function* f *on a BSS-Set* $\Omega(L)$ *will always be associated with an integer parameter* $\eta = \eta(f)$ *- which informally 'represents the degree the prover is claiming* $P[f]$ *has'. Thus, we can use the parameter* η *in definitions relating to* f.

We define various measures describing distances of restrictions of f *to low degree polynomials. We will be most interested in the distance of* $P[f]$ *from* $\mathsf{RS}_L[\eta]$.

- *Define* $\delta_{\text{uni}}(f) \triangleq \Delta(P[f], \mathsf{RS}_L[\eta])$.
- *For* $\alpha \in L'_0$, *define* $f^{\text{col}}_\alpha : L'_1 \to \mathbb{F}$ *by* $f^{\text{col}}_\alpha(Z) \triangleq f(\alpha, Z)$. *Define* $\delta_{c,\alpha}(f)$ *to be the distance of* f^{col}_α *from polynomials of degree* $2^{-\eta} \cdot |L_1| - 1$. *That is,*

$$\delta_{c,\alpha}(f) \triangleq \Delta(f^{\text{col}}_\alpha, \mathsf{RS}_{L'_1}[\eta]).$$

Finally, define $\delta_c(f) \triangleq \mathop{\mathrm{E}}_{\alpha \in L'_0}[\delta_{c,\alpha}(f)]$.

- *For* $\beta \in L_1$, *define* $f^{\text{row}}_\beta : L'_0 \to \mathbb{F}$ *by* $f^{\text{row}}_\beta(Z) \triangleq f(Z, q_{L_0}(\beta))$. *Let* $\delta_r(f) \triangleq \mathop{\mathrm{E}}_{\beta \in L_1}[\delta_{r,\beta}(f)]$ *where*

$$\delta_{r,\beta}(f) \triangleq \Delta(f^{\text{row}}_\beta, \mathsf{RS}_{L'_0}[1]).$$

- *For* $\beta \in L_1$ *define* $f^{\text{ext}}_\beta : L_\beta \to \mathbb{F}$ *by* $f^{\text{ext}}_\beta(Z) \triangleq f(Z, q_L(\beta))$. *Let* $\delta^{\text{ext}}_r(f) \triangleq \mathop{\mathrm{E}}_{\beta \in L_1}\left[\delta^{\text{ext}}_{r,\beta}(f)\right]$ *where*

$$\delta^{\text{ext}}_{r,\beta}(f) \triangleq \Delta(f^{\text{ext}}_\beta, \mathsf{RS}_{L_\beta}[2]).$$

Finally, define $\delta_{\text{rect}}(f)$ *to be the distance of* $f|_S$ *from the set of bi-variate polynomials of degree* $(|L_0|-1, 2^{-\eta} \cdot |L_1|-1)$.

Similarly to extending a subspace L to $\Omega(L)$ we want to have a canonical way of extending a function $g \in \mathsf{RS}_L[\eta]$ to a function on $\Omega(L)$. For this purpose the following claim from [BS08] will be useful.

[5] In [BS08, BCGT13] the set S was defined differently as a larger set.

Claim A.2. *Let $L \subseteq \mathbb{F}$ be a subspace, and fix $g \in \mathsf{RS}_L[\eta]$. There exists a bivariate polynomial $Q_g \in \mathbb{F}[X,Y]$ of degree $(|L_0|-1, 2^{-\eta}\cdot|L_1|-1)$ such that $Q_g(X,Y) \equiv g(X) \mod (Y - q_{L_0}(X))$. In particular, for any $z \in L$*

$$Q_g(z, q_{L_0}(z)) = g(z).$$

Definition A.3 (BSS Extension)**.** *Fix a positive integer η. Let $L \subseteq \mathbb{F}$ be a subspace, and fix $g \in \mathsf{RS}_L[\eta]$. We define the* Ben-Sasson-Sudan extension of g, $\Omega(g) : \Omega(L) \to \mathbb{F}$, *to be the evaluation of the polynomial Q_g from Claim A.2 on $\Omega(L)$.*

A.1.1 BSS sets and extensions of depth greater than one

Informally speaking, we wish to design tests where a verifier, examining a function on a BSS-Set, can recursively focus on a column, row, or extended row, and ask for the BSS-extension of the function restricted to that part. To formalize this we need to define BSS sets and BSS extensions of depth greater than one. We mention that in this section we only use extensions of depth one or two.

Definition A.4 (Ben-Sasson-Sudan sets and extensions of arbitrary depth)**.** *Fix a subspace $L \subseteq \mathbb{F}$ and integer $d > 0$. For each $\beta \in L_1$ we denote by β^{ext} a distinct symbol disjoint from $\mathbb{F}$.*

The depth d BSS-Set of L, *denoted $\Omega_d(L)$, is defined inductively as follows.*

- $\Omega_1(L) \triangleq \Omega(L)$. $\overline{\Omega}_1(L) \triangleq \Omega_1(L) \setminus T(L)$.
- $\Omega_d(L)$ *is defined as the disjoint union of*
  1. $\Omega(L)$.
  2. $\cup_{\alpha \in L_0'}(\alpha, \overline{\Omega}_{d-1}(L_1'))$.
  3. $\cup_{\beta \in L_1}(\beta, \overline{\Omega}_{d-1}(L_0'))$.
  4. $\cup_{\beta \in L_1}(\beta^{\text{ext}}, \overline{\Omega}_{d-1}(L_\beta))$.
- $\overline{\Omega}_d(L) \triangleq \Omega_d(L) \setminus T(L)$.

We add that

1. *As $\Omega_d(L)$ contains the depth one extension $\Omega(L)$, we can define for a function $f : \Omega_d(L) \to \mathbb{F}$ all measures that were defined for a function $f' : \Omega(L) \to \mathbb{F}$ - $\delta_c(f')$,$\delta_r(f')$,etc. as the corresponding value defined for the restriction $\Omega_d(f)|_{\Omega(L)}$. Similarly, "sub-functions" that were defined for a function $f' : \Omega(L) \to \mathbb{F}$ are defined for f according to $\Omega_d(f)|_{\Omega(L)}$. For example $f_\alpha^{\text{col}} \triangleq \left(\Omega_d(f)|_{\Omega(L)}\right)_\alpha^{\text{col}}$.*
2. *For $\alpha \in L_0'$, we define the 'α'th copy of $\Omega_{d-1}(L_1') \subset \Omega_d(L)$', denoted $\Omega_{d-1}^\alpha(L_1')$, as the union of (α, L_1') and $(\alpha, \overline{\Omega}_{d-1}(L_1'))$. We identify $\Omega_{d-1}^\alpha(L_1')$ with $\Omega_{d-1}(L_1')$ by mapping $(\alpha, z) \in (\alpha, \overline{\Omega}_{d-1}(L_1'))$ to $z \in \overline{\Omega}_{d-1}(L_1')$ and mapping $(\alpha, z) \in (\alpha, L_1')$ to $(z, q_{(L_1')_0}(z))$. We define $\Omega_{d-1}^\beta(L_0')$ and $\Omega_{d-1}^{\beta^{\text{ext}}}(L_\beta)$ similarly.*

Fix a function $g \in \mathsf{RS}_L[\eta]$. The depth d BSS-extension of g, *denoted $\Omega_d(g)$, is a function $\Omega_d(L) \to \mathbb{F}$ defined inductively as follows.*

- $\Omega_1(g) \triangleq \Omega(g)$. $\overline{\Omega}_1(g) \triangleq \Omega_1(g)|_{\overline{\Omega}_1(L)}$.
- $\Omega_d(g)$ *is defined as the union*[6] *of*
  1. $f : \Omega(L) \to \mathbb{F}$ *defined as the (depth one) BSS-extension of g.*
  2. $\cup_{\alpha \in L_0'} \overline{\Omega}_{d-1}(f_\alpha^{\text{col}})$ *viewed as a function* $(\alpha, \overline{\Omega}_{d-1}(L_1')) \to \mathbb{F}$.
  3. $\cup_{\beta \in L_1} \overline{\Omega}_{d-1}(f_\beta^{\text{row}})$ *viewed as a function* $(\beta, \overline{\Omega}_{d-1}(L_0')) \to \mathbb{F}$.

[6] We define a single function by a union of several functions on disjoint domains.

4. $\cup_{\beta \in L_1} \overline{\Omega}_{d-1}(f^{\mathrm{ext}}_{\beta})$ *viewed as a function* $(\beta^{\mathrm{ext}}, \overline{\Omega}_{d-1}(L_\beta)) \to \mathbb{F}$.

- $\overline{\Omega}_d(g) \triangleq \Omega_d(g)|_{\overline{\Omega}_d(L)}$.

1. *We mention that the depth* d *extension of* g *contains the depth* $d-1$ *extensions of the restrictions* $\Omega(g)^{\mathrm{col}}_{\alpha}, \Omega(g)^{\mathrm{row}}_{\beta}, \Omega(g)^{\mathrm{ext}}_{\beta}$ *of* $\Omega(g)$. *For example, identifying* $\Omega^{\alpha}_{d-1}(L'_1)$ *with* $\Omega_{d-1}(L'_1)$ *as described above, the restriction* $\Omega_d(g)|_{\Omega^{\alpha}_{d-1}(L'_1)}$ *is precisely* $\Omega_{d-1}(g^{\mathrm{col}}_{\alpha})$.
2. *In similar spirit, for an arbitrary function* $f : \Omega_d(L) \to \mathbb{F}$, *and* $\alpha \in L'_1$, *we define* $\Omega_{d-1}(f^{\mathrm{col}}_{\alpha}) \triangleq f|_{\Omega^{\alpha}_{d-1}(L'_1)}$. $\Omega_{d-1}(f^{\mathrm{row}}_{\beta})$ *and* $\Omega_{d-1}(f^{\mathrm{ext}}_{\beta})$ *are defined similarly.*

Calculation (details omitted) shows that if L has dimension k, $|\Omega_d(L)| \leq 2^{k+2+3(d-1)}$

Remark A.5 (Intersection of recursive extended rows with the curve of L)**.** *In Definition A.1 we made no assumptions regarding the default basis* $\{b_1, \ldots, b_k\}$ *of the subspace* L. *When working with BSS sets of depth greater than one, it is important that bases are chosen so that "all recursive extended rows have intersection with the original univariate polynomial". To formalize this let* $\Omega_d(L)$ *be a BSS-Set of depth* d. *For* $\beta_1, \ldots, \beta_d$ *define* $L_{\beta_1 \ldots \beta_d}$ *as* $L_{\beta_1 \ldots \beta_d} \triangleq (((L_{\beta_1})_{\beta_2})..)_{\beta_d}$. *Note that* $L_{\beta_1 \ldots \beta_d}$ *is defined only when* $\beta_i \in (L_{\beta_1 \ldots \beta_{i-1}})_1$ *for all* $i \in [d]$. *In such a case, let us call* $\beta_1, \ldots, \beta_d$ compatible. *We would like to view the subspace* $L_{\beta_1 \ldots \beta_d}$ *as a subset of* $\Omega_d(L)$. *We define such an embedding using induction on* d: *For* $d = 1$, *we identify* L_{β_1} *with* $(\beta_1, L_{\beta_1}) \subset \Omega(L)$ *in the natural way. Assume we have now embedded* $L_{\beta_2 \ldots \beta_d}$ *as a subset* $L_{\beta_2 \ldots \beta_d} \subset \Omega_{d-1}(L_{\beta_1})$. *Now using the identification of* $\Omega_{d-1}(L_{\beta_1})$ *with* $\Omega^{\beta^{\mathrm{ext}}_1}_{d-1}(L_{\beta_1}) \subset \Omega_d(L)$, *we obtain the embedding* $L_{\beta_1 \ldots \beta_d} \subset \Omega_d(L)$.

Assume $k \geq 2^d \cdot 2d$. *We claim that the bases of* $L_{\beta_1 \ldots \beta_i}$ *can be chosen such that, for every compatible* $\beta_1, \ldots, \beta_d$ *and every* $i \in [d]$, $L_{\beta_1 \ldots \beta_i} \cap T(L)$ *is an affine subspace of co-dimension* $2i$ *in* $L_{\beta_1 \ldots \beta_i}$ *(when* $L_{\beta_1 \ldots \beta_i}$ *is embedded into* $\Omega_d(L)$ *as described above). In particular,* $|L_{\beta_1 \ldots \beta_d} \cap T(L)| \geq 4^{-d} \cdot |L_{\beta_1 \ldots \beta_d}|$. *We show this by induction on* i. *For* $i = 1$, *we know that* $L_{\beta_1} \cap T(L)$ *is the subspace* $L_0 + \beta_1$ *and* $\dim(L_{\beta_1}) = \dim(L_0) + 2$. *Assume the claim for* i. *Let* $t \triangleq \dim(L_{\beta_1 \ldots \beta_i})$. *Thus, we have a basis* $v_1, \ldots, v_t$ *for* $L_{\beta_1 \ldots \beta_i}$ *such that*

$$L_{\beta_1 \ldots \beta_i} \cap T(L) = \{x = (x_1, \ldots, x_t) \in L_{\beta_1 \ldots \beta_i} | x_1 = a_1, \ldots, x_{2i} = a_{2i}\},$$

for some $a_1, \ldots, a_{2i} \in \mathbb{F}_2$ *when* x *is written in the basis* $v_1, \ldots, v_t$. *This is the basis of* $L_{\beta_1 \ldots \beta_i}$ *we use. Calculation shows* $\dim(L_{\beta_1 \ldots \beta_{i+1}}) \geq k/2^d$, *and therefore*

$$\dim((L_{\beta_1 \ldots \beta_i})_0) \geq k/2^d - 2 \geq 2i,$$

where the second inequality follows from our assumption on k *and* $i < d$. *Thus, when using* $v_1, \ldots, v_t$ *as a basis for* $L_{\beta_1 \ldots \beta_i}$, $(L_{\beta_1 \ldots \beta_i})_0$ *will have basis* $\{v_1, \ldots, v_{t'}\}$ *for* $t' \geq 2i$; *and so* $(L_{\beta_1 \ldots \beta_i})_0 \cap T(L)$ *will have co-dimension* $2i$ *in* $L_{\beta_1 \ldots \beta_i}$. *As* $(L_{\beta_1 \ldots \beta_i})_0 \subset L_{\beta_1 \ldots \beta_{i+1}}$ *and* $\dim(L_{\beta_1 \ldots \beta_{i+1}}) = \dim((L_{\beta_1 \ldots \beta_i})_0) + 2$, $L_{\beta_1 \ldots \beta_{i+1}} \cap T(L)$ *will have co-dimension* $2(i+1)$ *in* $L_{\beta_1 \ldots \beta_{i+1}}$.

A.2 The quasilinear PCP system

Soundness and security of a PCP system depend on the verifiers specification (cf. (1.1)). We study here a simple variant of the quasilinear PCP verifier of [BS08], defined later on in this section (Definition A.8), after we recall the necessary preliminary definitions from [BS08].

The PCP system of [BS08] is constructed for a $\mathbf{NEXP}$-complete problem[7] defined below.

Definition A.6 (Univariate algebraic CSP (ACSP))**.** *Instances of the language* UACSP *are tuples* $\psi = (L, \mathcal{N} = \{N_1, \ldots, N_{\mathsf{q}}\}, H, Q)$ *where*

- L *is an* $\mathbb{F}_2$*-subspace contained in a finite field* $\mathbb{F}$ *of characteristic two,*

[7] [BS08] uses a version that is $\mathbf{NP}$-complete. The $\mathbf{NEXP}$-complete version is from [BGH+06], cf. [BCGT13].

- $\mathcal{N} = \{N_1, \ldots, N_{\mathsf{q}}\} \subset \mathbb{F}[X]$ *is a set of degree-*1 *polynomials called the* neighborhood functions*,*
- $H \subset \mathbb{F}$ *is disjoint from* L[8]*,*
- *and* $Q \in \mathbb{F}[X, Y_1, \ldots, Y_{\mathsf{q}}]$ *is a polynomial of degree less than* $|H|$ *in* X*.*

The degree *of* ψ*, denoted* $\deg(\psi)$*, is* $\deg_{Y_1,\ldots,Y_{\mathsf{q}}}(Q)$*, i.e., it is the total degree of* Q *viewed as a polynomial in variables* $Y_1, \ldots, Y_{\mathsf{q}}$ *with coefficients in the ring* $\mathbb{F}[X]$

For $A \in \mathbb{F}[X]$ *denote by*

$$(Q \circ A \circ \mathcal{N})(X) \triangleq Q(X, A(N_1(X)), \ldots, A(N_{\mathsf{q}}(X))) \tag{A.2}$$

the composition of Q, A *and the neighborhood functions* $\mathcal{N}$*. A polynomial* $A \in \mathbb{F}[X]$ *is said to* satisfy ψ *if and only if*

$$\forall x \in H \quad (Q \circ A \circ \mathcal{N})(x) = 0. \tag{A.3}$$

Equivalently, A *satisfies* ψ *if and only if there exists* $B \in \mathbb{F}[X], \deg(B) < \deg(\psi) \cdot \deg(A)$ *such that*

$$(Q \circ A \circ \mathcal{N})(X) = B(X) \cdot \mathsf{Zero}_H(X) \tag{A.4}$$

where equality is in the ring $\mathbb{F}[X]$*.*

Let $\mathsf{UACSP}[\rho]$ *be the language of instances* ψ *satisfiable by a polynomial of degree* $\leq \frac{\rho}{\deg \psi} \cdot |L|$*.*

The following theorem from [BGH+06] (cf. [BCGT13]) justifies focusing on PCPs for UACSP.

Theorem A.7 (Reduction to UACSP)**.** *For any language* $L \in \mathbf{NTIME}(T(n))$ *there exists a polynomial-time reduction* R_L *from* L *to* $\mathsf{UACSP}[1/8]$*. Moreover, there exists a constant* d, q *such that* R_L *reduces instances of* L *to instances of* $\mathsf{UACSP}[1/8]$ *of degree at most* d *and query complexity at most* q *over a finite field* $\mathbb{F}$ *of characteristic* 2 *and in which* H *is an* $\mathbb{F}_2$*-affine subspace of* $\mathbb{F}$*.*

We now provide the definition of the PCP system for which the security analysis is applied. Recall the definition of the midpoint of k from the beginning of this section. For a field of size $|\mathbb{F}|$ and rate parameter η we define the *BSS recursion depth* to be the smallest integer r such that applying the midpoint operation r times sequentially to $k \triangleq \dim L$ gives a number that is smaller than $2^{2\eta}$; Formally, letting $k_1 = \left\lfloor \frac{k-1}{2} \right\rfloor, k_2 = \left\lfloor \frac{k_1-1}{2} \right\rfloor, \ldots$, the recursion depth is the first index r such that $k_r < 2^{2\eta}$.

Definition A.8 ([BS08]-PCP system)**.** *Given an instance* $\psi = (\mathbb{F}, \mathcal{N} = \{N_1, \ldots, N_{\mathsf{q}}\}, H, Q)$ *of* $\mathsf{UACSP}[2^{-\eta}]$ *where* $\mathbb{F}$ *has recursion depth* r*, the BSS-PCP system has the following prover and verifier.*

- **Prover** *The prover receives* ψ *and a satisfying assignment* $A \in \mathbb{F}[X], \deg(A) < \rho|\mathbb{F}|/\deg(\psi)$*. It computes and publishes as its PCP the following functions:*
  - *The evaluation* $f = A|_L$ *of* A *on* L
  - *The evaluation* g *of* $(Q \circ A \circ \mathcal{N})/\mathsf{Zero}_H$ *on* L
  - *The depth-*r *BSS-extension* $\Omega_{\mathsf{r}}(f)$
  - *The depth-*r *BSS-extension* $\Omega_{\mathsf{r}}(g)$
- **Verifier** *On input* ψ *and oracle access to* $\Omega_{\mathsf{r}}(f)$ *and* $\Omega_{\mathsf{r}}(g)$*, verifier performs the following tests, accepting iff all of them accept:*
  - *Invoke the depth-*r *RS-PCPP verifier for* $\mathsf{RS}_{\mathbb{F}}[\eta + \log \deg(\psi)]$ *from Definition 3.1 on* $\Omega_{\mathsf{r}}(f)$
  - *Invoke the depth-*r *RS-PCPP verifier for* $\mathsf{RS}_{\mathbb{F}}[\eta]$ *from Definition 3.1 on* $\Omega_{\mathsf{r}}(g)$
  - *Invoke the following* ACSP consistency verifier V_Q *for* $(Q, \mathcal{N})$ *on* f, g*:*
    - *Sample* $\alpha_1, \ldots, \alpha_{2/\rho} \in L$ *uniformly and independently;*
    - *accept iff for all* i *we have*[9] $(Q \circ f \circ \mathcal{N})(\alpha_i) = g(\alpha_i) \cdot \mathsf{Zero}_H(\alpha_i)$*; otherwise reject*

[8] This requirement is not made in [BS08], and typically $H \subset \mathbb{F}$. The disjointness requirement is convenient for using the results of Section 3.3.

[9] Here we are implicitly using the following "neighbor closure" property: $\alpha \in L \to N_j(\alpha) \in L$ for all $j \in \mathsf{q}$. When this is not the case, it is necessary for the prover to given an evaluation of f on a different, possibly larger space L' containing the sets $N_j(L)$.

A.3 Axis parallel attacks on the RS-PCPP

The following definition uses the notion of a depth-d BSS-extension of a function $g \in \mathsf{RS}_L[\eta]$ (Section A.1.1).

Definition A.9 (Row- and Column-compliant BSS extension of arbitrary functions)**.** *Fix a positive integer η. Let $L \subseteq \mathbb{F}$ be a subspace, and fix $g : L \to \mathbb{F}$. We define two "BSS extensions" of g:*

- **Row-compliant:** *We define the* row-compliant Ben-Sasson-Sudan extension of g, *denoted $\Omega_-(g) : \Omega(L) \to \mathbb{F}$, as follows. Let $f \triangleq \Omega_-(g)$ for brevity.*

  First, using the identification of L and $T(L)$ define $f(z, q_{L_0}(z)) = g(z)$, for $z \in L$. Recall that $T_\beta(L) = T(L) \cap \mathrm{extrow}_\beta = (L_0 + \beta, q_{L_0}(\beta))$ has size $|L_\beta|/4$. Thus, there is a unique $P \in \mathsf{RS}_{L_\beta}[2]$, i.e. a unique polynomial of degree at most $|L_\beta|/4 - 1$, such that $P(x) = f(x, \beta)$ for each $(x, \beta) \in T_\beta(L)$. Formally, $f(x, q_{L_0}(\beta)) \triangleq P(x)$ for $x \in L_\beta \setminus L_0 + \beta$. This process is equivalent to defining for each $\beta \in L_1$

  $$f^{\mathrm{ext}}_\beta \triangleq \mathrm{LDE}_{L_\beta}(g|_{L_0+\beta}).$$

  In particular, $f = \Omega_-(g)^{\mathrm{ext}}_\beta \in \mathsf{RS}_{L_\beta}[2]$ for each $\beta \in L_1$. Another definition that can be seen to be equivalent, is to define f as the evaluation of the polynomial Q_g from Claim A.2 on $\Omega(L)$, where Q_g is computed there with respect to $g \in \mathsf{RS}_L[0]$.

- **Column-compliant:** *Let $\hat{L}$ be a subspace of L of co-dimension η that contains L_0. Let $\hat{g} = \mathrm{LDE}_L(g|_{\hat{L}})$ be the low-degree extension to L of $g|_{\hat{L}}$; that is, $\hat{g} : L \to \mathbb{F}$ is the unique element of $\mathsf{RS}_L[\eta]$ that identifies with f on $\hat{L}$. We define the* column compliant Ben-Sasson-Sudan extension of g, *denoted $\Omega_|(g) : \Omega(L) \to \mathbb{F}$, to be $(g, \overline{\Omega}(\hat{g}))$. It can be seen that this is equivalent to the following. Let $f \triangleq \Omega_|(g)$ for brevity. Denote by $R_{\hat{L}} \subset L_1$ a space of size $2^{-\eta} \cdot |L_1|$ such that $\hat{L} = L_0 \oplus R_{\hat{L}}$. As in the row-compliant extension, define $\Omega_|(f)$ on $T(L)$ according to f. Now define f on extended rows extrow_β for $\beta \in R_{\hat{L}}$ by*

  $$f^{\mathrm{ext}}_\beta \triangleq \mathrm{LDE}_{L_\beta}(g|_{L_0+\beta}).$$

  Now complete each column col_α by a low degree extension of the values $\{(\alpha, \beta)\}_{\beta \in R_{\hat{L}}}$. That is,

  $$f^{\mathrm{col}}_\alpha \triangleq \mathrm{LDE}_{L'_1}(f^{\mathrm{col}}_\alpha|_{q_{L_0}(R_{\hat{L}})}),$$

  for each $\alpha \in L'_0$. At this stage, for each $\beta \in L_1$, f^{ext}_β is defined on L'_0; and it must coincide with a polynomial of degree smaller than $|L_0|$. Now complete the values $\mathrm{extrow}_\beta \setminus T_\beta(L)$ for $\beta \notin R_{\hat{L}}$ as a low degree extension of the values in row_β. That is,

  $$f^{\mathrm{ext}}_\beta|_{L_\beta \setminus (L_0+\beta)} = \mathrm{LDE}_{L_\beta \setminus (L_0+\beta)}(f^{\mathrm{ext}}_\beta|_{L_0}).$$

The depth d row-compliant BSS extension of f, denoted $\Omega_{d,-}(f)$ is obtained by using the row-compliant BSS extension recursively. Formally, this is obtained by setting $\Omega_1(f) = \Omega_-(f)$ in Definition A.4 and then recursively (for recursion depth $d-1$) computing the row-compliant BSS extension of every extended row and column of $\Omega_-(f)$. The depth d column-compliant BSS extension of f is denoted $\Omega_{d,|}(f)$ and computed recursively analogously by computing the column-compliant BSS extension of f and repeating this process for every extended row and column of $\Omega_|(f)$.

The following definition formally defines an attack that "mixes" recursively column- and row-compliant attacks in an arbitrary manner.

Definition A.10 (Axis-compliant attack)**.** *Given a function $g : L \to \mathbb{F}$, we say a function $f : \Omega(L) \to \mathbb{F}$ is an* axis-compliant extension of g, *if $f = \Omega_|(g)$ or $f = \Omega_-(g)$. For integer $d > 1$, we say that $f : \Omega_d(L) \to \mathbb{F}$ is a* depth d axis-compliant extension of g *if*

1. *$f|_{\Omega(L)} = \Omega_|(g)$ or $f|_{\Omega(L)} = \Omega_-(g)$.*
2. *For each $\alpha \in L'_0$, $\Omega_{d-1}(f^{\mathrm{col}}_\alpha)$ is a depth $d-1$ axis-compliant extension of f^{col}_α.*
3. *For each $\beta \in L_1$, $\Omega_{d-1}(f^{\mathrm{ext}}_\beta)$ is a depth $d-1$ axis-compliant extension of f^{ext}_β.*

Notice that when $f \in \mathsf{RS}_L[\eta]$ then by Claim A.2 we have $\Omega_-(f) = \Omega(f)$ by construction. Furthermore, in this case $\hat{f} = f$ so we also have $\Omega_|(f) = \Omega(f)$. In other words, when $f \in \mathsf{RS}_L[\eta]$ all axis-compliant attacks (including "pure" row- and column-compliant attacks) produce the "standard" BSS extension from Definition A.3.

log(message length)	log(codeword length)	log(proof length)	log(# of queries)
23	26	31	16.5
24	27	32	16.75
25	28	33	17
26	29	34	17.25
27	30	35	17.5
28	31	36	17.75
29	32	37	18
30	33	38	18.25
31	34	39	18.5
32	35	40	18.75
33	36	41	19
34	37	42	19.25
35	38	43	19.5

Table 4: Instantiations of our analysis of a depth two test (see Subsection B.2.4). All numbers are logs in base two of the described quantity. The first column describes the length (in field elements) of the message w to be encoded into a word $x \in \mathsf{RS}_L[3]$. The second column is the length of x. The third column is the length of x together with the proof y that $x \in \mathsf{RS}_L[3]$. The fourth is the number of field elements a verifier needs to read to be convinced with probability $1/2$ that x is at least $\frac{7}{24}$-close to $\mathsf{RS}_L[3]$.

B Improved analysis of the concrete efficiency of the BSS PCPP

The main result of this section is the following.

Theorem B.1. *Fix any field $\mathbb{F}$ of characteristic two. There is a PCPP system for $\mathrm{RS}_{\mathrm{a}}[3]$ with concrete efficiency threshold at most 2^{23}.*

Theorem B.1 improves the bound of 2^{43} from [BCGT13]. It will follow from the first item of the next theorem; the proof of the two parts of this theorem appear as Lemma B.12 and Lemma B.14 respectively.

Theorem B.2. *Fix any field $\mathbb{F}$ of characteristic two. There is a*

1. $(2^{\ell+5}, 2^{\ell/2+5.5})$*-PCPP system for* $\mathrm{RS}_{\mathrm{a}}[3]$,
2. $(2^{\ell+8}, 2^{\ell/4+10.75})$*-PCPP system for* $\mathrm{RS}_{\mathrm{a}}[3]$,

where 2^ℓ denotes the message length $k(C)$.

Table 4 shows instantiations of the second item of the above theorem for input lengths up to 2^{35}. We note again that filling the same table with the last column being the number of queries needed according to the analysis of [BCGT13], would give a trivial result where the verifier needs to read more locations than the original message length.

B.1 An overview of the proof of Theorem B.2

Fix a function $g : L \to \mathbb{F}$. The purpose of the Ben-Sasson-Sudan PCPP is to convince V that g is (close to) an element of $\mathsf{RS}_L[\eta]$, i.e., that it is an evaluation of a polynomial of degree at most $d \triangleq 2^{-\eta} \cdot |L| - 1$. For this overview, let us assume that V is willing to read $O(\sqrt{d})$ entries of g and an auxiliary proof $\pi(g)$. Observe first, that reading any $t < d$ entries of g gives V no information on whether $g \in \mathsf{RS}_L[\eta]$; as a degree d univariate polynomial could be interpolated to match any t values. The basic idea is to "embed" g into a *bivariate* polynomial Q of degree $O(\sqrt{d})$. The proof $\pi(g)$ consists of values of Q on a carefully chosen set. For this overview we use the term degree of a bivariate polynomial to mean individual degree.

The embedding of g works by choosing a certain univariate polynomial $q(Z)$ of degree $O(\sqrt{d})$, and then constructing $Q \in \mathbb{F}[X, Y]$ of (individual) degree[10] $\ell = O(\sqrt{d})$, such that the values of g on L correspond to the values of

[10] In the actual construction it is important to differentiate between the degree of Q in X and Y. We avoid this here, for simplicity of presentation.

Q on the "curve" $T(L) \triangleq \{(z, q(z)|z \in L\}$ (See Claim A.2 and Definition A.3). q is chosen such that $T(L)$ will have the following convenient properties.

- There are only $O(\sqrt{d})$ different "y values" in $T(L)$; that is $|q(L)| = |\{q(z)|z \in L\}| = O(\sqrt{d})$. Let $L_1 \subset L$ be a set of size $|q(L)|$ such that $q(L_1) = q(L)$.
- All "row restrictions" have the same size in $T(L)$; that is, for any $\beta \in L_1$ the set $\{(z, q(\beta)|z \in L, q(z) = q(\beta)\} \subset T(L)$ is of size $\frac{|T(L)|}{|q(L)|}$.

Checking whether g has degree d is reduced by this embedding to checking whether a bivariate function f on $T(L)$ has degree ℓ; that is, whether $f|_{T(L)}$ identifies with a bivariate polynomial Q of degree ℓ. This seems more doable as bivariate polynomials have more local structure. For example, their restriction to any "row" or "column" should satisfy the same degree bound as the polynomial itself. That is, if Q has degree $\leq \ell$, the univariate polynomials $Q(X, \beta)$ and $Q(\alpha, Y)$ also have degree at most ℓ. Polishchuk and Spielman[PS94] proved a strong converse to this: *Fix a table of a function $f(X, Y)$ on a product set $S = A \times B \subseteq \mathbb{F}^2$, where $|A|, |B| \geq 4\ell$. If many rows $f(X, \gamma)$ or many columns $f(\alpha, Y)$ are far from degree ℓ (as univariate polynomials), then f is far from any bivariate polynomial of (individual) degree ℓ.*

This is not immediately helpful, as the set $T(L)$ on which g's values correspond to values of Q, is quite far from a product set - for any value $z \in L$ of the first coordinate, there is only one value $(z, q(z)) \in T(L)$.

The proof $\pi(g)$ consists of the values of a function f - that is supposed to be Q - on a product set $S(L) \triangleq A \times q(L)$ where $|A| = O(\ell)$. Note that $|\pi(g)| = O(d)$. We think of $(g, \pi(g))$ as jointly describing a bivariate function f on the domain[11] $\Omega(L) \triangleq S(L) \cup T(L)$.

The sets $S(L)$ - on which we *know* how to test closeness of f to degree ℓ, and $T(L)$- on which we *desire* to test closeness to degree ℓ have almost no intersection. What ties them together are the *extended rows*

$$\mathrm{extrow}_\beta \triangleq (L, q(\beta)) \cap \Omega(L),$$

for $\beta \in L_1$. $\Omega(L)$ is precisely the disjoint union of $|q(L)|$ such rows. Any extended row will have large, specifically size $\Theta(\ell)$, intersection with both $S(L)$ and $T(L)$. Roughly speaking, this is what enables to relate the distance of $f|_{S(L)}$ from degree ℓ to the distance of $f|_{T(L)}$ from degree ℓ. Ultimately, the crux of the PCPP proof is to relate the following five measures

1. $\delta_{\mathrm{rect}}(f)$ - the distance of $f|_{S(L)}$ from bivariate polynomials of degree ℓ.
2. $\delta_c(f)$ - the average over $\alpha \in A$ of the distance of a "column" $f|_{\alpha \times q(L)}$ from univariate polynomials of degree ℓ.
3. $\delta_r(f)$ -the average over $\beta \in L_1$ of the distance of a "row" $f|_{\mathrm{extrow}_\beta \cap S(L)}$ from univariate polynomials of degree ℓ.
4. $\delta_r^{\mathsf{ext}}(f)$ -the average over $\beta \in L_1$ of the distance of an "extended row" $f|_{\mathrm{extrow}_\beta}$ from univariate polynomials of degree ℓ.
5. $\delta_{\mathrm{uni}}(f)$ - the distance of $f|_{T(L)}$ from bivariate polynomials of degree ℓ. This turns out to be the same as the distance of g from univariate polynomials of degree d.

What is implicit in the proofs of [BS08] and [BCGT13] is a relation of the form

$$c_1 \cdot \delta_c(f) + c_2 \cdot \delta_r(f) + c_3 \cdot \delta_r^{\mathsf{ext}}(f) \geq c \cdot \delta_{\mathrm{uni}}(f)$$

for non-negative c_1, c_2, c_3 with $c_1 + c_2 + c_3 = 1$ and $c > 0$. This relation suggests a natural recursive test: Think of c_1, c_2, c_3 as probabilities according to which either a random column, row, or extended row is chosen; Now check if f has the required degree on this restricted domain. The equation implies that if we started with a function g that is δ-far from degree d, we recurse, on average, on a function on a much smaller domain that is still $c \cdot \delta$-far from the degree we expect it to have. Intuitively, the larger c we can get, the larger bound we can get on V rejecting a function that

[11] In the actual definition (Definition A.1), $\Omega(L)$ actually needs to be defined as a similar but larger set, as the analysis sometimes requires that the restriction of $\Omega(L)$ to any row be an $\mathbb{F}_2$-subspace.

is far from low degree; and the larger the bound, the less repetitions, and therefore verifier queries, we need to reject with probability $1/2$.

Making this relation more explicit is the starting point for several improvements over the analysis of [BCGT13]. For example

- The [BS08], [BCGT13] verifier recurses only on columns and *extended* rows. They analyze this by using the bound $\delta_r^{\mathsf{ext}}(f) \geq \delta_r(f)/2$ to move to a relation with $c_2 = 0$. This decreases the obtained c.
- The [BS08],[BCGT13] verifier simply chooses a column or extended row each with probability $1/2$; so they are implicitly using a relation with $c_1 = 1/2, c_2 = 0, c_3 = 1/2$. This also decreases the obtained c.
- [BCGT13] using the analysis of Polishchuk-Spielman [PS94], in fact first implicitly obtain a bound the form $\min\{c \cdot \delta_{\mathrm{uni}}(f), \gamma\}$ for some $0 < \gamma < 1/100$ on the right-hand side of the relation. Since they need to work with expectations over such expressions in their proof they move to the more convenient and much smaller quantity $\gamma \cdot c \cdot \delta_{\mathrm{uni}}(f)$. We show that, at least when V uses only two recursion levels, there is no need to "move from the minimum to the product" and lose this large factor. See Lemma B.15.

Finally, we observe that for only one recursion level, a simple direct analysis is possible that avoids heavy factors coming from [PS94]. (See Claim B.11 and Lemma B.12).

B.2 Improved Soundness analysis

The following bound on the sizes of the subspaces arising in Definition A.1 will be useful for analyzing the efficiency of our tests.

Claim B.3. *Fix a subspace $L \subseteq \mathbb{F}$ of dimension k. Let L_1 and L_β be as in Definition A.1. Then*

- $|L_1|, |L_\beta| \leq 2^{k/2+1.5}$
- $|L_1| + |L_\beta| \leq 2^{k/2+2.1}$

Proof. The first item is immediate from the definition using $\dim(L_\beta) = \left\lfloor \frac{k-1}{2} \right\rfloor + 2 \leq k/2 + 1.5$ and $\dim(L_1) = k - \left\lfloor \frac{k-1}{2} \right\rfloor \leq k/2 + 1$. Moving to the second item, when k is even we have

$$|L_1| + |L_\beta| = 2^{k/2+1} + 2^{k/2+1} = 2^{k/2+2}.$$

When k is odd we have

$$|L_1| + |L_\beta| = 2^{k/2+1.5} + 2^{k/2+0.5} = (2^{1.5} + 2^{0.5}) \cdot 2^{k/2} \leq 2^{k/2+2.1}.$$

□

We proceed to show relations between the different measures defined. Intuitively we reduce the problem of verifying a function $g : L \to \mathbb{F}$ is close to some polynomial of low degree, to the problem of verifying $f = \Omega(g) : \Omega(L) \to \mathbb{F}$ is close to some bivariate polynomial of low degree. For this purpose we need to bound the distance of g from $\mathsf{RS}_L[\eta]$ by some attributes of f that are easier for the verifier to approximate. Lemma B.4 shows the distance of g from $\mathsf{RS}_L[\eta]$ can not be much larger than $\max\{\delta_{\mathrm{rect}}(f), \delta_r^{\mathsf{ext}}(f)\}$.

Lemma B.4. *Fix a subspace $L \subset \mathbb{F}$, and function $f : \Omega(L) \to \mathbb{F}$. We have*

$$\delta_{\mathrm{uni}}(f) \leq 2 \cdot \delta_{\mathrm{rect}}(f) + 4 \cdot \delta_r^{\mathsf{ext}}(f).$$

This improves the bound $\delta_{\mathrm{uni}}(f) \leq 2 \cdot \delta_{\mathrm{rect}}(f) + 8 \cdot \delta_r^{\mathsf{ext}}(f)$ implicit in Section 11 of [BCGT13] based on [BS08].

Proof. Fix $Q \in \mathbb{F}[X, Y]$ of degree $(|L_0| - 1, 2^{-\eta} \cdot |L_1| - 1)$ with $\Delta_{S(L)}(Q, f) = \delta_{\mathrm{rect}}(f)$. For $\beta \in L_1$, denote by

- P_β an element of $\mathsf{RS}_{L_\beta}[2]$ closest to f_β^{ext}; formally, P_β satisfies $\Delta_{L_\beta}(P_\beta, f_\beta^{\mathrm{ext}}) = \delta_{r,\beta}^{\mathrm{ext}}(f)$.

- Q_β the univariate polynomial $Q_\beta(X) \triangleq Q(X, q_L(\beta))$.

Denote

$$\gamma \triangleq \Pr_{\beta \in L_1}(Q_\beta \neq P_\beta).$$

For $\beta \in L_1$ such that $Q_\beta \neq P_\beta$, we have $\Delta_{L'_0}(Q_\beta, P_\beta) \geq 1/2$.

Thus

$$\frac{1}{2} \cdot \gamma \leq \mathop{\mathrm{E}}_{\beta \in L_1}\left[\Delta_{L'_0}(Q_\beta, P_\beta)\right]$$

Using the triangle inequality

$$\leq \mathop{\mathrm{E}}_{\beta \in L_1}\left[\Delta_{L'_0}(Q_\beta, f^{\mathrm{row}}_\beta)\right] + \mathop{\mathrm{E}}_{\beta \in L_1}\left[\Delta_{L'_0}(f^{\mathrm{row}}_\beta, P_\beta)\right].$$

Recall that $\delta_{\mathrm{rect}}(f)$ is the fractional distance between Q and f on $S(L)$, and note that $S(L) = L'_0 \times L'_1$ is the union over $\beta \in L_1$ of $\mathrm{row}_\beta \triangleq L'_0 \times \{q_L(\beta)\}$. Thus, $\delta_{\mathrm{rect}}(f)$ is equal to the average over $\beta \in L_1$ of the fractional distance between Q and f on row_β. Hence, we can replace the first term and get

$$= \delta_{\mathrm{rect}}(f) + \mathop{\mathrm{E}}_{\beta \in L_1}\left[\Delta_{L'_0}(f^{\mathrm{row}}_\beta, P_\beta)\right].$$

Let us call $\beta \in L_1$ *good* if $Q_\beta = P_\beta$. Calculation shows that $\deg(Q(Z, q_L(Z))) \leq 2^{k-\eta} - 1$ (see proof of Claim B.11); and so $P[Q] : L \to \mathbb{F}$ defined by $P[Q](z) \triangleq Q(Z, q_L(Z))$ is in $\mathsf{RS}_L[\eta]$. Also, $\Delta(P[Q], P[f]) = \Delta_{T(L)}(Q, f)$. Thus,

$$\delta_{\mathrm{uni}}(f) \leq \Delta_{T(L)}(Q, f) = \mathop{\mathrm{E}}_{\beta \in L_1}\left[\Delta_{T_\beta(L)}(Q_\beta, f^{\mathrm{ext}}_\beta)\right] \leq \gamma + (1-\gamma) \cdot \mathop{\mathrm{E}}_{\beta \ is \ good}\left[\Delta_{T_\beta(L)}(Q, f)\right]$$
$$\leq \gamma + \mathop{\mathrm{E}}_{\beta \in L_1}\left[\Delta_{T_\beta(L)}(P_\beta, f)\right]$$

Using our bound on γ we get

$$\leq 2 \cdot \delta_{\mathrm{rect}}(f) + \mathop{\mathrm{E}}_{\beta \in L_1}\left[2 \cdot \Delta_{L'_0}(f^{\mathrm{row}}_\beta, P_\beta) + \Delta_{T_\beta(L)}(P_\beta, f^{\mathrm{ext}}_\beta)\right]$$

Below we explain that this is

$$\leq 2 \cdot \delta_{\mathrm{rect}}(f) + 4 \cdot \mathop{\mathrm{E}}_{\beta \in L_1}\left[\delta^{\mathrm{ext}}_{r,\beta}(f)\right]$$
$$= 2 \cdot \delta_{\mathrm{rect}}(f) + 4 \cdot \delta^{\mathrm{ext}}_r(f).$$

We now explain the last inequality above. For $\beta \in L_1$, $\delta^{\mathrm{ext}}_{r,\beta}(f)$ is equal to the fractional number of disagreements between P_β and f^{ext}_β on L_β. Denote this set of locations of disagreements by $D \subseteq L_\beta$. Thus, $\delta^{\mathrm{ext}}_{r,\beta}(f) = \frac{|D|}{|L_\beta|}$.

L'_0 and $T_\beta(L)$ are disjoint subsets of L_β of density $1/2$ and $1/4$ respectively. Thus,

$$\Delta_{L'_0}(f^{\mathrm{row}}_\beta, P_\beta) = \frac{|D \cap L'_0|}{|L'_0|} = \frac{2 \cdot |D \cap L'_0|}{|L_\beta|}$$

and

$$\Delta_{T_\beta(L)}(P_\beta, f^{\mathrm{ext}}_\beta) = \frac{|D \cap T_\beta(L)|}{|T_\beta(L)|} = \frac{4 \cdot |D \cap T_\beta(L)|}{|L_\beta|}.$$

So

$$2 \cdot \Delta_{L'_0}(f^{\mathrm{row}}_\beta, P_\beta) + \Delta_{T_\beta(L)}(P_\beta, f^{\mathrm{ext}}_\beta) \leq \frac{4}{|L_\beta|} \cdot (|D \cap L'_0| + |D \cap T_\beta(L)|) \leq 4 \cdot \delta^{\mathrm{ext}}_{r,\beta}(f).$$

□

For integer $\eta > 1$ define the constant $\delta_\eta \triangleq (1/4 - 2^{-\eta-1})^2$. For example, $\delta_2 = 1/64$ and $\delta_3 = 9/256$. The following is a corollary of Lemma B.15 below.

Lemma B.5. *Fix a subspace $L \subset \mathbb{F}$, and function $f : \Omega(L) \to \mathbb{F}$ with $\eta(f) \geq \eta$ for integer $\eta > 1$. We have*

$$\delta_r(f) + \delta_c(f) \geq \delta_\eta \vee (2/3) \cdot \delta_{\mathrm{rect}}(f).$$

Corollary B.6. *Fix a subspace $L \subset \mathbb{F}$, and function $f : \Omega(L) \to \mathbb{F}$ with $\eta(f) \geq \eta$ for integer $\eta > 1$. Then*

$$3 \cdot \delta_c(f) + 3 \cdot \delta_r(f) + 4 \cdot \delta_r^{\mathsf{ext}}(f) \geq \delta_{\mathrm{uni}}(f) \vee 3 \cdot \delta_\eta.$$

Proof. Using Lemma B.5 we know that either

1. $\delta_r(f) + \delta_c(f) \geq \delta_\eta$ which implies $3 \cdot \delta_c(f) + 3 \cdot \delta_r(f) + 4 \cdot \delta_r^{\mathsf{ext}}(f) \geq 3 \cdot \delta_\eta$ or
2. $\delta_r(f) + \delta_c(f) \geq (2/3) \cdot \delta_{\mathrm{rect}}(f)$ which implies, using Lemma B.4,

$$\begin{aligned} &3 \cdot \delta_c(f) + 3 \cdot \delta_r(f) + 4 \cdot \delta_r^{\mathsf{ext}}(f) \\ &\geq 2 \cdot \delta_{\mathrm{rect}}(f) + 4 \cdot \delta_r^{\mathsf{ext}}(f) \geq \delta_{\mathrm{uni}}(f). \end{aligned}$$

□

For integer $\eta > 1$ let us define $c_\eta \triangleq \frac{3 \cdot \delta_\eta}{10}$. For example, $c_2 \geq \frac{1}{250}, c_3 \geq \frac{1}{100}$. Simple calculations now show

Corollary B.7. *Fix a subspace $L \subset \mathbb{F}$, and function $f : \Omega(L) \to \mathbb{F}$ with $\eta(f) \geq \eta$ for integer $\eta > 1$. Then*

$$\frac{3}{10} \cdot \delta_c(f) + \frac{3}{10} \cdot \delta_r(f) + \frac{4}{10} \cdot \delta_r^{\mathsf{ext}}(f) \geq \frac{\delta_{\mathrm{uni}}(f)}{10} \vee c_\eta.$$

B.2.1 Tightness of bounds

Note that the inequality of Lemma B.4 is obviously an equality if we take $f : \Omega(L) \to \mathbb{F}$ to be identically zero. One may wonder if for non-zero f the bound can be improved. That is, can we get a better lower bound on $\delta_{\mathrm{rect}}(f)$ and $\delta_r^{\mathsf{ext}}(f)$ in terms of $\delta_{\mathrm{uni}}(f)$? (The quality of such a lower bound directly relates to the soundness of our PCPPs). We show that the bound cannot be improved beyond a factor exponentially small in the dimension of L.

Claim B.8 (Tightness of bound)**.** *Fix any integers $\eta > 1$ and $i > 2$, and let $L^i \subset \mathbb{F}$ be a vector space of dimension i over $\mathbb{F}_2$. There is a function $f_i : \Omega(L^i) \to \mathbb{F}$ with $\eta(f_i) = \eta$ such that:*

$$\delta_{\mathrm{uni}}(f_i) \geq 2 \cdot \delta_{\mathrm{rect}}(f_i) + 4 \cdot \delta_r^{\mathsf{ext}}(f_i) - 2^{-(i-1)} > 0.$$

Proof. Fix integer $i > 2$, and let $L = L^i$ be an i dimensional subspace. Choose distinct $\beta_0 \neq \beta_1 \in L_1 \setminus L'_0$.

Let $h : L_0 \to \mathbb{F}$ be a mapping that is identically 0, except on $x = 0$. More precisely,

$$h(x) = \begin{cases} 1 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

We define $h_{\beta_0} : L_{\beta_0} \to \mathbb{F}$ to be the "low degree extension of h on L_{β_0}". That is, h_{β_0} is the unique polynomial of degree smaller than $|L_0|$, with $h_{\beta_0}|_{L_0} = h$. We define $f = f_i : \Omega(L) \to \mathbb{F}$ in the following way:

$$f(x,y) = \begin{cases} h_{\beta_0}(x) & \text{if } y = \beta_0 \\ 1 & \text{if } (x,y) \in T_{\beta_1}(L) \\ 0 & \text{otherwise} \end{cases}$$

Note that $P[f]$ differs from the zero polynomial Z, only on $T_{\beta_0}(L) \cup T_{\beta_1}(L)$, and thus

$$\delta_{\mathrm{uni}}(f) \leq \Delta_L(Z, P[f]) \leq \frac{2}{|L_1|},$$

and therefore, as $\frac{2}{|L_1|} \leq 2^{-(i-1)} < (1-2^{-\eta})/2$, which is less than half the unique distance of $\mathsf{RS}_L[\eta]$, $\delta_{\mathrm{uni}}(f) = \frac{2}{|L_1|}$.

The only non zero entries of f in S are $(0 \cup (L'_0 \setminus L_0), q_{L_0}(\beta_0))$, so the closest polynomial of degree $(|L_0| - 1, 2^{-\eta} \cdot |L_1| - 1)$ to f over S is the zero polynomial, and we get $\delta_{\mathrm{rect}}(f) = \frac{1}{2 \cdot |L_1|} + \frac{1}{|L|}$.

There is only one extended row which is not low degree, and it is $f^{\mathrm{ext}}_{\beta_0}$ which is not zero only on $T_{\beta_0}(L)$, so we get $\delta^{\mathsf{ext}}_r(f) = \frac{1}{4 \cdot |L_1|}$.

Putting everything together, we get the required result

$$\delta_{\mathrm{uni}}(f) = 2 \cdot \delta_{\mathrm{rect}}(f) + 4 \cdot \delta^{\mathsf{ext}}_r(f) - \frac{2}{|L|} = 2 \cdot \delta_{\mathrm{rect}}(f) + 4 \cdot \delta^{\mathsf{ext}}_r(f) - 2^{1-i}.$$

□

B.2.2 Viewing functions on BSS sets as PCPP proofs

When describing our PCPP systems, we explicitly describe only the verifiers, which we think of as tests that are given access to a function $f : \Omega_d(L) \to \mathbb{F}$, and try to determine whether $P[f] \in \mathsf{RS}_L[\eta]$. The PCPP system for $\mathsf{RS}_L[\eta]$ that is implicitly defined by the test works as follows. The prover P starts with a function $g : L \to \mathbb{F}$ that he wishes to convince the verifier V belongs to $\mathsf{RS}_L[\eta]$. He constructs a function $y : \overline{\Omega}_d(L) \to \mathbb{F}$ and sends (g, y) to V. V thinks of (g, y) as a single function $f : \Omega_d(L) \to \mathbb{F}$ by identifying L with $T(L) = \Omega_d(L) \setminus \overline{\Omega}_d(L)$ as described before, i.e., mapping z to $(z, q_{L_0}(z))$; now V applies the test on f. The honest prover, given $g \in \mathsf{RS}_L[\eta]$, will take y to be the BSS-extension of g restricted to $\overline{\Omega}_d(L)$, i.e., $y = \overline{\Omega}_d(g)$. In other words, the honest prover simply sends $\Omega_d(g)$ to V.

Before presenting our first verifier, we define the *soundness function* of a PCPP verifier V. Loosely speaking, on input δ the function equals the minimal probability that V rejects an input x that is δ-far from the code.

Definition B.9 (Soundness of test)**.** *Fix a randomized function $V : \mathbb{F}^* \to \{\mathsf{accept}, \mathsf{reject}\}$, and a code ensemble $\mathcal{C}$ over $\mathbb{F}$. Fix $0 < \delta < 1$. We define $S_{\mathcal{C}}[V](\delta)$ as the infimum over all $C \in \mathcal{C}$, all $x \in \mathbb{F}^{n(C)}$ that are δ-far from C, and all $y \in \mathbb{F}^*$ of*

$$\Pr(V(x, y) = \mathsf{reject}).$$

For integer $\eta > 0$, we use the shortened notation $S_\eta[V](\delta) \triangleq S_{\mathrm{RS_a}[\eta]}[V](\delta)$.

B.2.3 A test of depth one

We define the test $T_{1,\eta}$.

$\underline{T_{1,\eta}\textbf{(f)}}$ **input:** $f : \Omega(L) \to \mathbb{F}$.

1. Choose random $\alpha \in L'_0$. Check if f^{col}_α has degree at most $2^{-\eta} \cdot |L_1| - 1$.
2. Choose random $\beta \in L_1$. Check if f^{ext}_β has degree at most $|L_0| - 1$.
3. Return reject if one of the above checks failed. Return accept otherwise.

The following claim will be used to analyze the test.

Claim B.10. *Fix subsets $A = \{\alpha_1, \ldots, \alpha_n\}, B = \{\beta_1, \ldots, \beta_m\}$ of $\mathbb{F}$. Fix positive integers $d < n$ and $m' \leq m$. Suppose we have a function $f : A \times B \to \mathbb{F}$ such that*

- *for $1 \leq j \leq m'$ we have that $f(X, \beta_j)$ has degree at most d.*
- *For $1 \leq i \leq d + 1$ we have that $f(\alpha_i, Y)$ has degree at most e.*

Then there exists a bivariate polynomial $Q \in \mathbb{F}[X, Y]$ of degree (d, e) such that for every $1 \leq j \leq m'$ $f(X, \beta_j) = Q(X, \beta_j)$.

Proof. For $1 \le i \le d+1$, define a polynomial $\delta_i(X)$ of degree at most d such that $\delta_i(\alpha_i) = 1$, and $\delta_i(\alpha_\ell) = 0$ for $1 \le \ell \le d+1, \ell \ne i$. Let $g_i(Y)$ be the polynomial of degree at most e that identifies with $f(\alpha_i, Y)$ on B.

Now define

$$Q(X,Y) \triangleq \sum_{i=1}^{d+1} \delta_i(X) \cdot g_i(Y).$$

Clearly Q has degree (d,e). It is immediate from Q's formula that $Q(\alpha_i, Y) = g_i(Y)$ for $1 \le i \le d+1$. Thus, for any $1 \le i \le d+1$ and $1 \le j \le m'$, $Q(\alpha_i, \beta_j) = g_i(\beta_j) = f(\alpha_i, \beta_j)$. In other words, $Q(X, \beta_j)$ and $f(X, \beta_j)$ agree on $d+1$ points and therefore agree on all of A.

□

Lemma B.11 (Soundness of depth one test)**.** *For any* $0 < \delta < 1$

$$S_\eta[T_{1,\eta}](\delta) \ge \delta \vee 1/2 \ge \delta/2.$$

Proof. Fix $f : \Omega(L) \to \mathbb{F}$. Denote $d = |L_0| - 1$ and $e = 2^{-\eta} \cdot |L_1| - 1$. Suppose that $\Pr(T_{1,\eta}(f) = \mathsf{accept}) \ge \max\{1/2, 1-\delta\}$. In particular there is a set $C \subset L'_0$ with $|C| \ge |L'_0|/2$ such that $\deg(f^{\mathrm{col}}_\alpha) \le e$ for all $\alpha \in C$. It follows from Claim B.10 that there exists a bivariate polynomial $Q(X,Y)$ of degree (d,e) such that

$$Q(X, q_{L_0}(\beta)) = f(X, q_{L_0}(\beta))$$

for any $\beta \in R \subset L_1$ where R is a set of size at least $(1-\delta) \cdot |L_1|$. Now defining

$$P(Z) \triangleq Q(Z, q_{L_0}(Z))$$

We see that

1. P agrees with $P[f]$ on L with probability at least $1-\delta$: As

$$\Pr_{z \leftarrow L}(P(z) = P[f](z)) \ge \Pr_{\beta \leftarrow L_1}(Q(X,\beta) = f^{\mathrm{ext}}_\beta) \ge 1 - \delta.$$

2.

$$\begin{aligned} \deg P &\le |L_0| - 1 + |L_0| \cdot (2^{-\eta} \cdot |L_1| - 1) \\ &= 2^{k-\eta} - 1 \end{aligned}$$

Thus, as P is δ-close to $P[f]$ and $\deg(P) \le 2^{k-\eta} - 1$, $\delta_{\mathrm{uni}}(f) \le \delta$. □

Amplifying $T_{1,\eta}$ with repetitions we get the first part of Theorem B.2, stated as a separate lemma next.

Lemma B.12. *There is a* $(2^{\ell+5}, 2^{\ell/2+5.5})$*-PCPP system for* $\mathrm{RS_a}[3]$*, where* 2^ℓ *denotes the message length.*

> **Reminder**
>
> *A* $(2^{\ell+5}, 2^{\ell/2+5.5})$*-PCPP system for some code* C *will generate a proof of length* $A = 2^{\ell+5}$ *field elements, of which the verifier* V *reads only* $Q = 2^{\ell/2+5.5}$ *field elements, and is guaranteed to reject with probability at least* $\frac{1}{2}$ *any* x *that is* $d(C)/3$*-far from* C*. (Definition 1.3) In case* $C \in \mathrm{RS_a}[3]$ *we have* $d(C) = 7/8$*.*

Proof. Fix a subspace $L \subseteq \mathbb{F}$ of dimension $k = \ell + 3$. Fix $f : \Omega(L) \to \mathbb{F}$ with $\eta(f) = 3$ and $\delta_{\mathrm{uni}}(f) \ge \delta \triangleq \frac{7}{8} \cdot \frac{1}{3} \ge \frac{1}{4}$. Consider the verifier V that given such f runs $T_{1,3}(f)$ 3 times and accepts if and only if $T_{1,3}$ accepted every time. We have

$$\Pr(V(f) = \mathsf{accept}) \le (1 - 1/4)^3 \le 1/2.$$

At each repetition V queries $|L_\beta| + |L'_1|$ field elements (for some $\beta \in L'_0$) which is at most $2^{k/2+2.1}$ by Claim B.3. So $Q \le 3 \cdot 2^{k/2+2.1} \le 2^{k/2+4} = 2^{\ell/2+5.5}$. P needs to write down the table of $f : \Omega(L) \to \mathbb{F}$ which is of length $|\Omega(L)| = 4 \cdot |L| = 2^{k+2} = 2^{\ell+5}$.

□

We now complete the proof of Theorem B.1.

Proof of Theorem B.1. Using the system from Lemma B.12 we need to see for what ℓ we have

$$2^{\ell+5} \cdot 2^{\ell/2+5.5} \leq 2^{2\ell-1}.$$

Equivalently

$$\ell/2 \geq 11.5,$$

which is satisfied when $\ell \geq 2^{23}$. □

B.2.4 A test of depth two

We define the test $T_{2,\eta}$.

$\underline{T_{2,\eta}\textbf{(f)}}$ **input:** $f : \Omega_2(L) \to \mathbb{F}$.

1. With probability $\frac{3}{10}$, choose random $\alpha \in L'_0$; and return $T_{1,\eta}(\Omega(f^{\mathrm{col}}_\alpha))$.
2. With probability $\frac{3}{10}$, choose random $\beta \in L_1$; and return $T_{1,1}(\Omega(f^{\mathrm{row}}_\beta))$.
3. Otherwise, i.e. with probability $\frac{4}{10}$, choose random $\beta \in L_1$; and return $T_{1,2}(\Omega(f^{\mathrm{ext}}_\beta))$.

Lemma B.13. *[Soundness of depth 2 test] Fix any integer $\eta > 1$. For any $0 < \delta < 1$*

$$S_\eta[T_{2,\eta}](\delta) \geq \frac{\delta}{20} \vee \frac{c_\eta}{2}.$$

For instance,

$$S_3[T_{2,3}](\delta) \geq \frac{\delta}{20} \vee \frac{1}{200}$$

and

$$S_2[T_{2,2}](\delta) \geq \frac{\delta}{20} \vee \frac{1}{500}.$$

Proof. Fix any $f : \Omega(L) \to \mathbb{F}$ with $\delta_{\mathrm{uni}}(f) = \delta$.

$$\Pr[T_{2,\eta}(f) = \mathsf{reject}] =$$

$$\frac{3}{10} \cdot \mathop{\mathrm{E}}_{\alpha \in L'_0} \left[\Pr(T_{1,\eta}(\Omega(f^{\mathrm{col}}_\alpha)) = \mathsf{reject})\right] + \frac{3}{10} \cdot \mathop{\mathrm{E}}_{\beta \in L_1} \left[\Pr(T_{1,1}(\Omega(f^{\mathrm{row}}_\beta) = \mathsf{reject})\right] + \frac{4}{10} \cdot \mathop{\mathrm{E}}_{\beta \in L_1} \left[\Pr(T_{1,2}(\Omega(f^{\mathrm{ext}}_\beta) = \mathsf{reject})\right]$$

Using Claim B.11 this is

$$\geq \frac{3}{10} \cdot \mathop{\mathrm{E}}_{\alpha \in L'_0} [\delta_{c,\alpha}(f)/2] + \frac{3}{10} \cdot \mathop{\mathrm{E}}_{\beta \in L_1} [\delta_{r,\beta}(f)/2] + \frac{4}{10} \cdot \mathop{\mathrm{E}}_{\beta \in L_1} \left[\delta^{\mathrm{ext}}_{r,\beta}(f)/2\right]$$

$$\geq 1/2 \cdot \left(\frac{3}{10} \cdot \delta_c(f) + \frac{3}{10} \cdot \delta_r(f) + \frac{4}{10} \cdot \delta^{\mathsf{ext}}_r(f) \right)$$

Using Corollary B.7, this is

$$\geq 1/2 \cdot \left(\frac{\delta_{\mathrm{uni}}(f)}{10} \vee c_\eta \right) = \frac{\delta}{20} \vee \frac{c_\eta}{2}.$$

□

Amplifying the test by repetition to soundness half we can get the proof of the second part of Theorem B.2, stated as the lemma below.

Lemma B.14. *There is a* $(2^{\ell+8}, 2^{\ell/4+10.75})$*-PCPP system for* $\mathrm{RS}_{\mathrm{a}}[3]$.

Proof. Fix a subspace $L \subseteq \mathbb{F}$ of dimension $k = \ell+3$. Fix $f : \Omega(L) \to \mathbb{F}$ with $\eta(f) = 3$ and $\delta_{\mathrm{uni}}(f) \geq \delta \triangleq \frac{7}{8} \cdot \frac{1}{3} \geq \frac{1}{4}$. Consider the verifier V that given such f runs $T_{2,3}(f)$ 140 times and accepts if and only if $T_{2,3}$ accepted every time. We have

$$\Pr(V(f) = \mathsf{accept}) \leq (1 - \frac{1}{200})^{140} \leq 1/2.$$

Using Claim B.3, we can see that at each repetition V queries at most $2^{k/4+2.85}$ field elements. So $Q \leq 140 \cdot 2^{k/4+2.85} \leq 2^{k/4+10} = 2^{\ell/4+10.75}$. P needs to write down the table of $f : \Omega_2(L) \to \mathbb{F}$ of size $|\Omega_2(L)| \leq 2^{k+5} = 2^{\ell+8}$. □

B.3 Improving the bound obtained from the Polishchuk-Spielman analysis

Fix a bivariate function $f : A \times B \to \mathbb{F}$. As in [BCGT13] we denote in this section by

- $\delta^{(d,*)}_{A\times B}(f)$ - the distance of f from polynomials of degree $(d, |B|-1)$. This is the same as the average over $\gamma \in B$ of the distance between $f(X, \gamma)$ and polynomials of degree d.
- $\delta^{(*,e)}_{A\times B}(f)$ - the distance of f from polynomials of degree $(|A|-1, e)$. This is the same as the average over $\alpha \in A$ of the distance between $f(\alpha, Y)$ and polynomials of degree e.
- $\delta^{(d,e)}_{A\times B}(f)$ - the distance of f from polynomials of degree (d, e).

Using a more careful pass of the proof of Lemma 10.6 in [BCGT13], we prove

Lemma B.15. *Fix a field* $\mathbb{F}$*. Fix positive integers* d, m, e, n *such that* $d/m + e/n < 1$*. Fix* $\delta > 0$ *such that* $\delta < 1/2(1 - d/m - e/n)$*. Fix* $A, B \subseteq \mathbb{F}$ *with* $|A| = m$ *and* $|B| = n$*. Fix any function* $f : A \times B \to \mathbb{F}$*. Then*

$$\delta^{(d,*)}_{A\times B}(f) + \delta^{(*,e)}_{A\times B}(f) \geq \min\left\{\delta^2, \delta^{(d,e)}_{A\times B}(f)/1.5\right\}$$

We remark that the bound in Lemma 10.6 of [BCGT13] was $\delta^2 \cdot \delta^{(d,e)}_{A\times B}(f)$ which can be significantly smaller.

Proof. If $\delta^{(d,*)}_{A\times B}(f) + \delta^{(*,e)}_{A\times B}(f) \geq \delta^2$ we are done. Suppose from now on that $\delta^{(d,*)}_{A\times B}(f) + \delta^{(*,e)}_{A\times B}(f) < \delta^2$. Let R be the[12] polynomial of degree $(d, |B|-1)$ closest to f. We have $\Delta_{A\times B}(f, R) = \delta^{(d,*)}_{A\times B}(f)$. From the triangle inequality

$$\delta^{(d,e)}_{A\times B}(f) \leq \Delta_{A\times B}(f, R) + \delta^{(d,e)}_{A\times B}(R) = \delta^{(d,*)}_{A\times B}(f) + \delta^{(d,e)}_{A\times B}(R)$$

Similarly, let C be the polynomial of degree $(|A|-1, e)$ closet to f. We have $\Delta_{A\times B}(f, C) = \delta^{(*,e)}_{A\times B}(f)$. From the triangle inequality

$$\delta^{(d,e)}_{A\times B}(f) \leq \Delta_{A\times B}(f, C) + \delta^{(d,e)}_{A\times B}(C) = \delta^{(*,e)}_{A\times B}(f) + \delta^{(d,e)}_{A\times B}(C).$$

Now define $\delta' \triangleq \sqrt{\delta^{(d,*)}_{A\times B}(f) + \delta^{(*,e)}_{A\times B}(f)}$. We have

$$\Delta_{A\times B}(R, C) \leq \Delta_{A\times B}(R, f) + \Delta_{A\times B}(f, C) = \delta^{(d,*)}_{A\times B}(f) + \delta^{(*,e)}_{A\times B}(f) = \delta'^2.$$

It follows from the analysis of [BCGT13] of the Polishchuk-Spielman test (Theorem 10.5 of [BCGT13]) that there exists a polynomial Q of degree (d, e) that disagrees on a uniform point of $A \times B$ with either R or C with probability at most $2\delta'^2$. It follows that

$$\delta^{(d,e)}_{A\times B}(R) + \delta^{(d,e)}_{A\times B}(C) \leq 2\delta'^2 = 2 \cdot (\delta^{(d,*)}_{A\times B}(f) + \delta^{(*,e)}_{A\times B}(f))$$

[12] If there is not a unique polynomial closest to f, pick one arbitrarily.

and thus

$$\delta^{(d,e)}_{A\times B}(f) = \frac{\delta^{(d,e)}_{A\times B}(f)}{2} + \frac{\delta^{(d,e)}_{A\times B}(f)}{2} \leq \frac{\delta^{(d,*)}_{A\times B}(f) + \delta^{(d,e)}_{A\times B}(R)}{2} + \frac{\delta^{(*,e)}_{A\times B}(f) + \delta^{(d,e)}_{A\times B}(C)}{2}$$
$$< \frac{\delta^{(d,*)}_{A\times B}(f) + \delta^{(*,e)}_{A\times B}(f)}{2} + \delta'^2 = 1.5(\delta^{(d,*)}_{A\times B}(f) + \delta^{(*,e)}_{A\times B}(f)).$$

□

C Visualization of BSS Sets

The purpose of this appendix is to give a concrete example of a BSS set, with a visualization, in order to help clarify Definition A.1.

Let $L \subset \mathbb{F}$ be an $\mathbb{F}_2$-subspace of dimension $k = 5$, and let $\{b_1, \cdots, b_5\}$ be its basis. We visualize the set of elements L as an ordered vector of points. We think of the indices of the points as the boolean vectors in $\{0,1\}^5$. Each index $(a_1, \ldots, a_5)$ represents the element $\sum_{i=1}^{5} a_i \cdot b_i$ of L; and they are ordered by the standard alphanumeric order (see figure C.1).

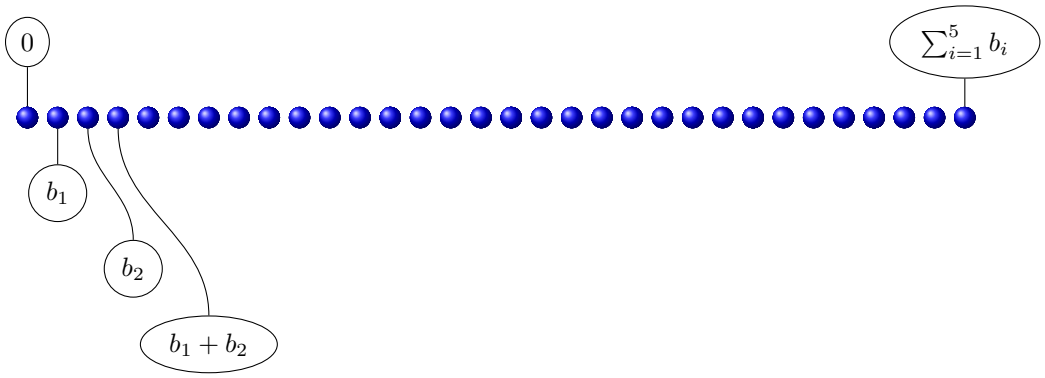


Figure C.1: Visualization of the vector space L

We view this space as a direct sum $L = L_0 \oplus L_1$ where $L_0 = \text{span}\,(b_1, b_2)$ and $L_1 = \text{span}\,(b_3, b_4, b_5)$.

A convenient way to visualize these spaces is thinking of L as the union of L_0 cosets shifted by L_1 elements as we see in figure C.2, where the beginning of a new coset is indicated by a change in color.

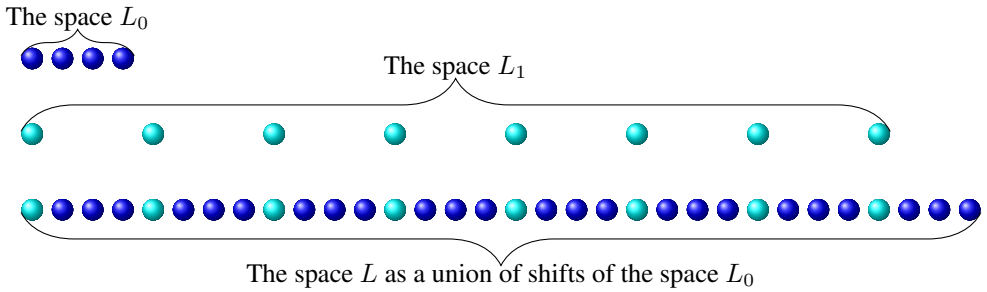


Figure C.2: Visualization of the spaces L_0 and L_1 and L as their direct sum. Each cyan node in L indicates the "beginning" of a new affine shift of L_0 by some element of L_1

The space L_0' is $\text{span}\,(b_1, b_2, b_3)$ and is the common subset of all the spaces L_β for all $\beta \in L_1$. The space L_β is defined as $L_\beta = \text{span}\,(b_1, b_2, b_3, \beta)$ for all $\beta \notin L_0'$, and otherwise $L_\beta = \text{span}\,(b_1, b_2, b_3, b_4)$.

An illustration of the construction of L_β for arbitrary value of β can be seen in figure C.3.

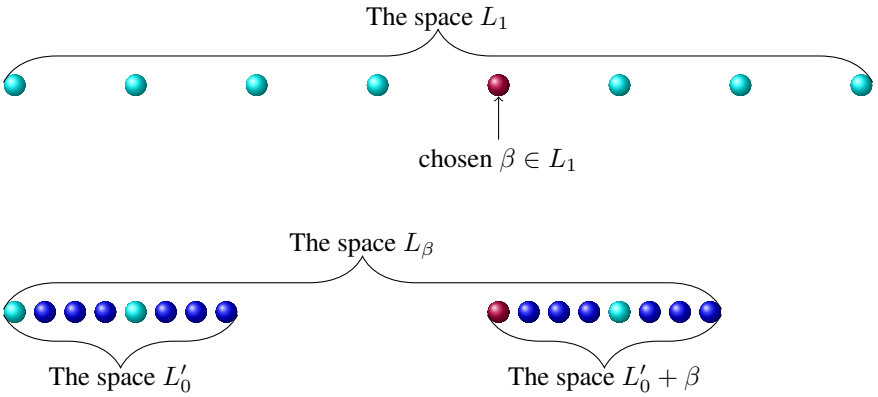


Figure C.3: Visualization of the spaces L'_0 and L_β

For all $\beta \in L_1$ we define the set $\mathrm{extrow}_\beta \subset \mathbb{F}^2$ to be the embedding of L_β into $\mathbb{F}^2$ using the mapping ϕ_β that is defined by $\forall \alpha \in L_\beta : \phi_\beta(\alpha) = (\alpha, q_{L_0}(\beta))$. The BSS set $\Omega(L)$ is defined as the union of extrow_β for all $\beta \in L_1$. We embed the space L into $\Omega(L)$ using the mapping ϕ_Ω that is defined by $\forall \alpha \in L_0, \beta \in L_1 : \phi_\Omega(\alpha + \beta) = \phi_\beta(\alpha)$ (see figure C.4).

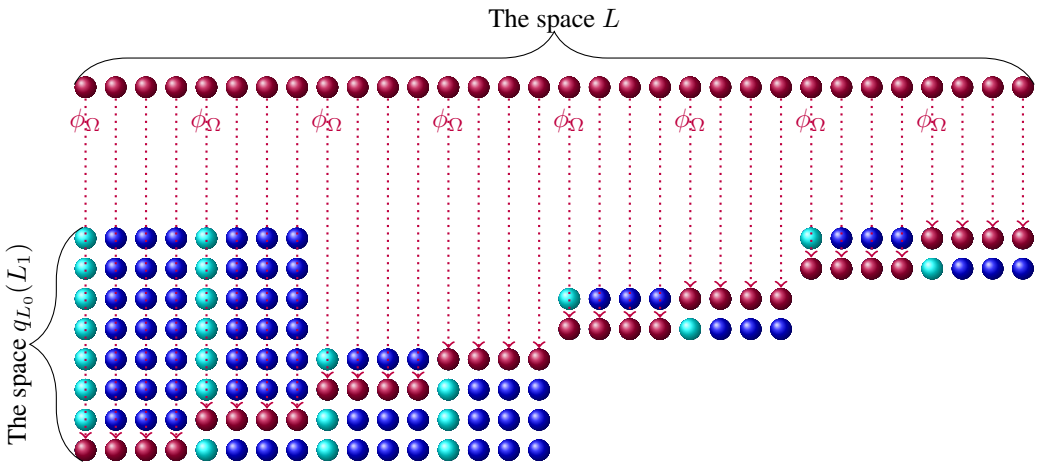


Figure C.4: Visualization of $\Omega(L)$ and the embedding of L into it. The subset of $q_{L_0}(L_1) \times L$ is $\Omega(L)$, and the purple arrows from the space L into $\Omega(L)$ illustrate the embedding ψ_Ω. The purple elements in $\Omega(L)$ is the image of this embedding, which is denoted $T(L)$ in Definition A.1. The product set $S(L) = q_{L_0}(L_1) \times L'_0$ can be easily seen as the block of full columns at the left part of $\Omega(L)$.

The set $\overline{\Omega}(L)$ is defined as $\Omega(L) \setminus T(L)$ and visually illustrated in figure C.5.

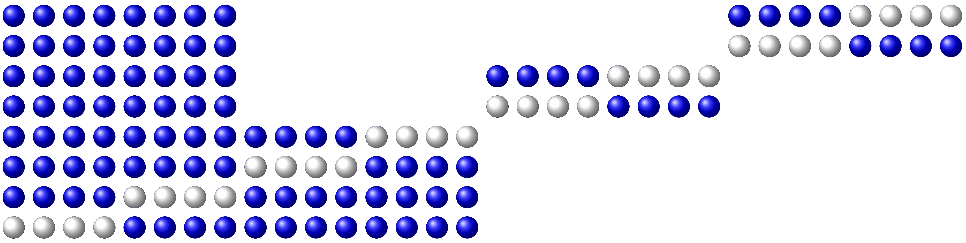


Figure C.5: Illustration of the set $\overline{\Omega}(L)$. The elements in the set are blue. The remaining elements are the set $T(L)$ from Definition A.1.

ECCC ISSN 1433-8092
http://eccc.hpi-web.de