Interactive Oracle Proofs with Constant Rate and Query Complexity

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

Alessandro Chiesa
alexch@berkeley.edu
UC Berkeley

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

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

Nicholas Spooner
spooner@cs.toronto.edu
University of Toronto

September 21, 2017

Abstract

We study *interactive oracle proofs* (IOPs) [BCS16, RRR16], which combine aspects of probabilistically checkable proofs (PCPs) and interactive proofs (IPs). We present IOP constructions and techniques that let us achieve tradeoffs in proof length versus query complexity that are not known to be achievable via PCPs or IPs alone. Our main results are:

1. *Circuit satisfiability has* 3*-round IOPs with linear proof length (counted in bits) and constant query complexity.*
2. *Reed–Solomon codes have 2-round IOPs of proximity with linear proof length and constant query complexity.*
3. *Tensor product codes have 1-round IOPs of proximity with sublinear proof length and constant query complexity.* (A familiar example of a tensor product code is the Reed–Muller code with a bound on individual degrees.)

For all the above, known PCP constructions give *quasilinear* proof length and constant query complexity [BS08, Din07]. Also, for circuit satisfiability, [BKK$^+$13] obtain PCPs with linear proof length but *sublinear* (and super-constant) query complexity. As in [BKK$^+$13], we rely on algebraic-geometry codes to obtain our first result; but, unlike that work, our use of such codes is much "lighter" because we do not rely on any automorphisms of the code.

We obtain our results by proving and combining "IOP-analogues" of tools underlying numerous IPs and PCPs:

- **Interactive proof composition.** Proof composition [AS98] is used to reduce the query complexity of PCP verifiers, at the cost of increasing proof length by an additive factor that is exponential in the verifier's randomness complexity. We prove a composition theorem for IOPs where this additive factor is linear.
- **Sublinear sumcheck.** The sumcheck protocol [LFKN92, Sha92] is an IP that enables the verifier to check the sum of values of a low-degree multi-variate polynomial on an exponentially-large hypercube, but the verifier's running time depends linearly on the bound on individual degrees. We prove a sumcheck protocol for IOPs where this dependence is sublinear (e.g., polylogarithmic).

Our work demonstrates that even constant-round IOPs are more efficient than known PCPs and IPs.

Keywords: probabilistically checkable proofs, interactive proofs, proof composition, sumcheck

Contents

1 Introduction

We study *Interactive Oracle Proofs* (also known as *Probabilistically Checkable Interactive Proofs*) [BCS16, RRR16], which combine aspects of probabilistically checkable proofs (PCPs) and interactive proofs (IPs). We present IOP constructions and general techniques that enable us to obtain tradeoffs in proof length versus query complexity that are *not known to be achievable by either PCPs or IPs alone*. In addition to being a natural notion to study, IOPs can substitute PCPs or IPs in some applications, offering a richer set of constructions. Perhaps most notably, IOPs can be used to construct non-interactive arguments in the random oracle model [BCS16], and techniques underlying our results have been used to achieve significant efficiency gains compared to PCP-based arguments [BBC$^+$17].

1.1 Proof length vs. query complexity

Probabilistically checkable proofs (PCPs) were introduced by [FRS88, BFLS91, FGL$^+$91, AS98, ALM$^+$98]: in a PCP, a probabilistic polynomial-time verifier has oracle access to the proof string. The complexity class $\mathbf{PCP}[\mathsf{r},\mathsf{q}]$ denotes those languages for which the verifier uses at most r random bits and queries at most q proof locations; the proof length is then at most 2^{r}. The PCP Theorem [AS98, ALM$^+$98] states that $\mathbf{NP} = \mathbf{PCP}[O(\log n), O(1)]$: every NP statement has a proof of polynomial length that can be verified via a constant number of queries (say, with soundness error $1/2$).

A fundamental question is how long a PCP needs to be, compared to the corresponding "standard" NP proof. Given $T\colon \mathbb{N} \to \mathbb{N}$, the PCP Theorem states that every language $\mathscr{L}$ in $\mathbf{NTIME}(T)$ has a proof of length $\mathrm{poly}(T(n))$ that can be verified with $O(1)$ queries. A sequence of works [PS94, HS00, GS06, BSVW03, BGH$^+$06, BS08, Din07] gradually reduced the proof length to quasilinear, i.e., $T(n) \cdot \mathrm{polylog}(T(n))$; much of this progress was accompanied by progress on efficient reductions from $\mathbf{NTIME}$ to "PCP-friendly" problems, as well as efficient constructions of PCPs of proximity (PCPPs) for key classes of linear codes. Despite much progress, the following question remains open: *are there PCPs with linear proof length and constant query complexity?*

Ben-Sasson et al. [BKK$^+$13] make progress in this direction by proving that there is $a > 0$ such that for every $\epsilon > 0$ there is a PCP for circuit satisfiability with proof length $2^{a/\epsilon}n$ and query complexity n^{ϵ}. Beyond the sublinear query complexity, [BKK$^+$13]'s result comes with other caveats not affecting most prior constructions: the verifier is *non-uniform*, namely it requires a polynomial-size advice string for every circuit size; and the verifier is not succinct, namely it does run in time that is sublinear in the circuit size even if the circuit comes from a uniform circuit family. (Recent constructions of constant-rate locally testable codes with sub-polynomial query complexity [KMRS16] are not yet known to be convertible to PCPs with similar parameters.)

In this paper, we continue the study of the tradeoff between proof length and query complexity, but we do so for a natural extension of the PCP model that can be thought of as a "multi-round PCP", described below. This extension of the PCP model suffices for notable applications, e.g., [BCS16].

From this point onwards, we switch to using relations instead of languages. We denote by $\mathscr{R}$ a relation consisting of pairs $(\mathbb{x}, \mathbb{w})$, where $\mathbb{x}$ is the *instance* and $\mathbb{w}$ is the *witness*; we think of $\mathscr{R}$ naturally induced by a non-deterministic language $\mathscr{L}$. We denote by $\mathscr{R}|_{\mathbb{x}}$ the (possibly empty) set of witnesses for a given instance $\mathbb{x}$, and by n the size of $\mathbb{x}$.

1.2 A more general model: interactive oracle proofs

Interactive Oracle Proofs (IOPs) are a type of proof system introduced in [BCS16, RRR16] that combines aspects of IPs [Bab85, GMR89] and PCPs [BFLS91, AS98, ALM$^+$98], and generalizes Interactive PCPs [KR08]. IOPs naturally extend the notion of a PCP to multiple rounds or, viewed from an another angle, they naturally extend the notion of an IP by allowing probabilistic checking. Prior work shows that IOPs can be used to construct non-interactive proofs in the random oracle model [BCS16], that IOPs efficiently achieve unconditional zero knowledge [BCGV16], and that IOPs can be used to obtain doubly-efficient constant-round IPs for polynomial-time bounded-space computations [RRR16].

Informally, an IOP extends an IP as follows: whenever the prover sends to the verifier a message, the verifier does not have to read the message in full but may probabilistically query it. In more detail, a k-round IOP comprises k rounds of interaction. In the i-th round of interaction: the verifier sends a message m_i to the prover; then the prover replies with a message f_i to the verifier, which the verifier can query in this and all later rounds (by having oracle access to it). After the k rounds of interaction, the verifier either accepts or rejects.

An *IOP system* for a relation $\mathscr{R}$ with round complexity k and soundness ε is a pair (P, V), where P, V are probabilistic algorithms, that satisfies natural notions of completeness and soundness: for every instance-witness pair $(\mathbb{x}, \mathbb{w})$ in $\mathscr{R}$, $V(\mathbb{x})$ always accepts after $\mathsf{k}(n)$ rounds of interaction with $P(\mathbb{x}, \mathbb{w})$; and, for every instance $\mathbb{x}$ with $\mathscr{R}|_{\mathbb{x}} = \emptyset$ and unbounded prover $\tilde{P}$, $V(\mathbb{x})$ accepts with probability at most $\varepsilon(n)$ after $\mathsf{k}(n)$ rounds of interaction with $\tilde{P}$.

Like the IP model, one efficiency measure is the round complexity k. Like the PCP model, two additional efficiency measures are the *proof length* l, which is the total number of alphabet symbols in all of the prover's messages, and the *query complexity* q, which is the total number of locations queried by the verifier across all of the prover's messages. Considering all of these parameters, we say that a relation $\mathscr{R}$ belongs to the complexity class $\mathbf{IOP}[\mathsf{k}, \mathsf{a}, \mathsf{l}, \mathsf{r}, \mathsf{q}, \varepsilon]$ if there is an IOP system for $\mathscr{R}$ in which on instances of size n: (1) the number of rounds is $\mathsf{k}(n)$; (2) the prover messages are over the alphabet $\mathsf{a}(n)$; (3) the proof length over this alphabet is $\mathsf{l}(n)$; (4) the verifier uses $\mathsf{r}(n)$ random bits; (5) the verifier queries the prover messages in $\mathsf{q}(n)$ locations; (6) the soundness error is $\varepsilon(n)$.

Many other definitions for IPs and PCPs carry over naturally. An IOP is *public coin* if each m_i is a random string and the verifier postpones any oracle queries until after receiving all the oracles from the prover (i.e., after the k-th round of interaction). An IOP is *non-adaptive* if the query locations do not depend on answers to any previous queries.

Prior work on IOPs. In prior work, [BCS16] prove that public-coin IOPs can be compiled into non-interactive proofs in the random oracle model; their compiler is as a generalization of the Fiat–Shamir paradigm for public-coin IPs [FS86, PS96], and of the "CS proof" constructions for PCPs of Micali [Mic00] and Valiant [Val08]. Also, [BCGV16] construct 2-round IOPs (called "duplex PCPs" there) with unconditional zero knowledge and quasilinear proof length; in comparison, short PCPs with unconditional zero knowledge are not known. Also, [RRR16] use IOPs to obtain doubly-efficient constant-round IPs for polynomial-time bounded-space computations. In this paper, we do not study compilers for cryptographic proofs, nor zero knowledge, nor applications to interactive proofs; instead, we focus on tradeoffs of proof length versus query complexity for IOPs.

Prior work on Interactive PCPs. An Interactive PCP [KR08] is a PCP followed by a standard IP; in particular, it is an IOP where the verifier sends an empty first message and may query only the first prover message (but must read any other prover messages in full). Prior work on Interactive PCPs obtains proof length that depends on the witness size rather than computation size [KR08, GKR08], as well as unconditional zero knowledge [GIMS10]. In this paper we also study proof length but our results to not seem to extend to the more restricted setting of Interactive PCPs.

1.3 Proximity and robustness

To facilitate upcoming technical discussions we briefly introduce two notions that strengthen a PCP.

- *PCPs of proximity* (PCPPs) [DR04, BGH$^+$06]. On the one hand, a PCP verifier has oracle access to a candidate proof π and only decides if $\mathscr{R}|_{\mathbb{x}} \neq \emptyset$ ($\mathbb{x} \in \mathscr{L}$) or $\mathscr{R}|_{\mathbb{x}} = \emptyset$ ($\mathbb{x} \notin \mathscr{L}$). On the other hand, a PCPP verifier has oracle access to a candidate witness $\mathbb{w}$ and proof π and decides if $\mathbb{w} \in \mathscr{R}|_{\mathbb{x}}$ or $\mathbb{w}$ is far from $\mathscr{R}|_{\mathbb{x}}$ (in particular, if $\mathscr{R}|_{\mathbb{x}} = \emptyset$, then $\mathbb{w}$ is far from $\mathscr{R}|_{\mathbb{x}}$). A quantity δ known as the *proximity parameter* specifies what "far" means: if $\mathbb{w}$ is δ-far from $\mathscr{R}|_{\mathbb{x}}$ then the PCPP verifier accepts with probability at most ε, where ε is the soundness error.

- *Robust PCPs* [BGH$^+$06]. When $\mathscr{R}|_{\mathbb{x}} = \emptyset$, the answers to the verifier's queries are, with high probability, far from any answers that make the verifier accept. A quantity ρ known as the *robustness parameter* specifies what "far" means: if $\mathscr{R}|_{\mathbb{x}} = \emptyset$ then, with probability at least $1 - \varepsilon$, the answers are ρ-far from accepting ones.

The two above notions can also be combined, yielding the definition of a *robust PCP of proximity*.

Extension to IOPs. The notions of proximity and robustness naturally extend to IOPs; see Section 2.3 for details. For example, we say that an IOP has *proximity parameter* δ if the analogous property for PCPs of proximity holds; we can then correspondingly define the complexity class $\mathbf{IOPP}[\mathsf{k}, \mathsf{a}, \mathsf{l}, \mathsf{r}, \mathsf{q}, \varepsilon, \delta]$.

1.4 Results

We obtain several IOP constructions with proof length and query complexity that are not known to be achievable either via PCPs or IPs alone (or even via Interactive PCPs [KR08]). First, we show that for circuit satisfiability we can obtain IOPs with linear proof length and constant query complexity; constant round complexity and public coins suffice.

Theorem 1.1 (informal)**.** *Let $\mathscr{R}$ be the relation consisting of instance-witness pairs (ϕ, w) where ϕ is a boolean circuit (of two-input NAND gates) and w is a binary input that satisfies ϕ; we use n to denote the number of gates in ϕ. There exists $a > 0$ and a public-coin IOP system that puts $\mathscr{R}$ in the complexity class*

$$\mathbf{IOP}\left[\begin{array}{llcl} \text{rounds} & \mathsf{k}(n) & = & 3 \\ \text{answer alphabet} & \mathsf{a}(n) & = & \mathbb{F}_2 \\ \text{proof length} & \mathsf{l}(n) & = & a \cdot n \\ \text{query complexity} & \mathsf{q}(n) & = & a \\ \text{soundness error} & \varepsilon(n) & = & 1/2 \end{array}\right] .$$

In particular, via [PF79]'s reduction from Turing machines to circuits, we deduce that

$$\mathbf{NTIME}(T) \subseteq \mathbf{IOP}\left[\begin{array}{llcl} \text{rounds} & \mathsf{k}(T) & = & 3 \\ \text{answer alphabet} & \mathsf{a}(T) & = & \mathbb{F}_2 \\ \text{proof length} & \mathsf{l}(T) & = & a \cdot T \log T \\ \text{query complexity} & \mathsf{q}(T) & = & a \\ \text{soundness error} & \varepsilon(T) & = & 1/2 \end{array}\right] .$$

The main points of comparison of the above theorem with prior work are the following.

- For PCPs with constant query complexity, prior work achieved only quasilinear proof length [BS08, Din07], with the "quasilinear" hiding several logarithmic factors. In comparison, we achieve linear proof length for circuit satisfiability, and $O(T \log T)$ proof length for nondeterministic T-time relations.

- Ben-Sasson et al. [BKK$^+$13] show that there is $a > 0$ such that for every $\epsilon > 0$ there is a non-uniform PCP for circuit satisfiability with proof length $2^{a/\epsilon} n$ and query complexity n^{ϵ}; the non-uniformity comes from the use of algebraic-geometry (AG) codes with transitive automorphism groups, for which uniform families are not known. In comparison, we simultaneously achieve linear proof length and constant query complexity; moreover, we make a much "lighter" use of AG codes, which also allows us to avoid non-uniformity. Namely, we rely only on the multiplication properties of AG codes [CC88, Mei13], and do not rely on any code automorphisms. Looking ahead, this is because we do not route circuits on Cayley graphs induced by the automorphisms of the underlying code, unlike [BKK$^+$13].

Second, we show that Reed–Solomon codes over binary fields (fields of characteristic 2) have 2-round IOPs of proximity with linear proof length and constant query complexity. Such codes are a key ingredient for constructing PCPs with quasilinear proof length [BS08]. Recall that a word $w \colon D \to \mathbb{F}$ is represented via $|w| = |D| \cdot \log |\mathbb{F}|$ bits.

Theorem 1.2 (informal)**.** *Given a "fractional degree" $\varrho \in (0, 1)$, define $\mathscr{R}$ to be the relation consisting of instance-witness pairs $((\mathbb{F}_{2^\lambda}, d), w)$ where $d \leq \varrho 2^{\lambda}$ and $w \colon \mathbb{F}_{2^\lambda} \to \mathbb{F}_{2^\lambda}$ is the evaluation of a polynomial of degree less than d; we define the instance size to be λ, and note that w has $|w| = 2^{\lambda} \cdot \lambda$ bits. For every $\delta \in (0, \frac{1}{2}(1 - \varrho))$ there exist $a > 0$ and a public-coin IOP of proximity (P, V) that puts $\mathscr{R}$ in the complexity class*

$$\mathbf{IOPP}\left[\begin{array}{llcl} \text{rounds} & \mathsf{k}(\lambda) & = & 2 \\ \text{answer alphabet} & \mathsf{a}(\lambda) & = & \mathbb{F}_2 \\ \text{proof length} & \mathsf{l}(\lambda) & = & a \cdot 2^{\lambda} \cdot \lambda \\ \text{query complexity} & \mathsf{q}(\lambda) & = & a \\ \text{soundness error} & \varepsilon(\lambda) & = & 1/2 \\ \text{proximity parameter} & \delta(\lambda) & = & \delta \end{array}\right] .$$

More generally, our result concerns *additive Reed–Solomon codes*, where the domain of a codeword is a λ-dimensional affine subspace S of a potentially larger binary field $\mathbb{F}$; in such cases the above statement involves more parameters but achieves the same asymptotics.[1] The main point of comparison of the above theorem with prior work is [BS08, Din07], who achieve PCPs of proximity with the same parameters but superlinear proof length: $a \cdot 2^{\lambda} \cdot \mathrm{poly}(\lambda)$.

[1] In fact, using the same proof technique, an analogous result holds also for Reed–Solomon codes evaluated over any multiplicative subgroup whose size is $O(1)$-smooth, as described in [BS08, Section 7]. Details omitted.

Third, we show that tensor product codes have 1-round IOPs of proximity with sublinear proof length and constant query complexity. Given a positive integer m and linear code C with domain D and alphabet $\mathbb{F}$, the tensor product code $C^{\otimes m}$ is the linear code that comprises all functions $w\colon D^m \to \mathbb{F}$ whose restriction to any axis-parallel line is in C; the message length, block length, and distance of $C^{\otimes m}$ are each the m-th power of the corresponding parameters of C. Tensor product codes are a large family, and they include Reed–Muller codes (at least when considering the definition that bounds the variables' individual degrees, which we do, as opposed to the one that bounds their sum).

Theorem 1.3 (informal)**.** *Let $m \geq 3$ and C be a linear code with domain D, alphabet $\mathbb{F}$, and relative distance τ; let $\ell := |D|$ be the block length. Define $\mathscr{R}$ to be the relation of instance-witness pairs $\big((C, m), w\big)$ such that $w \in C^{\otimes m}$; note that w has $|w| = \ell^m \cdot \log|\mathbb{F}|$ bits. For every $\delta \in (0, \frac{1}{2}\tau^m)$ there exist $a > 0$ and a public-coin IOPP system (P, V) that puts $\mathscr{R}$ in the complexity class*

$$\mathbf{IOPP}\left[\begin{array}{lcl} \text{rounds} & \mathsf{k}(\ell^m) & = & 1 \\ \text{answer alphabet} & \mathsf{a}(\ell^m) & = & \mathbb{F}_2 \\ \text{proof length} & \mathsf{l}(\ell^m) & = & o(\ell^m \cdot \log|\mathbb{F}|) \\ \text{query complexity} & \mathsf{q}(\ell^m) & = & a \\ \text{soundness error} & \varepsilon(\ell^m) & = & 1/2 \\ \text{proximity parameter} & \delta(\ell^m) & = & \delta \end{array}\right].$$

The main points of comparison of the above theorem with prior work are the following.

- Several works [BS06, Vid15, CMS17] give local testers with query complexity $\mathsf{q}(\ell^m) = \ell^2$; for all tensor product codes Dinur et al. [DSW06] give local testers with $\mathsf{q}(\ell^m) = \ell$ for certain tensor product codes. In contrast, we achieve constant query complexity, with only sublinear proof length, for all tensor product codes. Moreover, given additional mild conditions, we obtain constant soundness error *even for non-constant m*.

- The work of [BS08, Din07] implies PCPs of proximity for tensor product codes with superlinear proof length and constant query complexity. In contrast, we obtain sublinear proof length, with a single round of interaction.

Analogously to [Vid15], we can invoke Theorem 1.3 on different choices of linear codes so to derive different code families that have good properties and an IOP tester (instead of a local tester as in [Vid15]). For example, we can choose a family of linear codes with arbitrarily high rate, constant relative distance, linear-time encoding, and linear-time decoding from a constant fraction of errors [Spi96, GI05, RS06]; our theorem implies a code with the same properties that *also* has a 1-round IOP of proximity with sublinear proof length and constant query complexity (cf. [Vid15, Section 3.1]). Similar statements hold for list-decodable codes with good parameters [GGR11] (cf. [Vid15, Section 3.2]); and also for locally correctable and, more generally, locally decodable codes with good parameters [Yek08, Vid10, Efr12, KSY14, KMRS16] (cf. [Vid15, Section 3.3]). In each of these cases, the tensor product operation preserves the "key" properties of the choice of underlying code C, while endowing the resulting code with an IOP of proximity.

We obtain the above results via techniques of independent interest: we prove that, in the IOP model, there are more efficient analogues of tools that are fundamental to constructing PCPs and IPs. We now discuss these techniques.

1.5 Techniques

Recall that IOPs generalize both IPs, by treating the prover's messages as oracle strings, and PCPs, by allowing for multiple rounds of interaction; they also generalize Interactive PCPs [KR08]. We prove that IOPs can express two fundamental techniques in a more efficient way than in these prior models: (i) in *interactive proof composition*, the prover is more efficient than in PCP proof composition; and (ii) in *sublinear sumchecks*, the verifier is more efficient than in IP sumcheck protocols. We now discuss both of our new tools, and then how we use them.

Interactive proof composition. Proof composition [AS98] is used to reduce PCP query complexity, cf. [ALM$^+$98, HS00, BGH$^+$06]; it involves two PCPs: an outer one and an inner one. One should think of the outer proof system as having short proofs but large query complexity, while the inner proof system has long proofs but small query complexity.

The composed prover uses the outer prover to send a PCP to the composed verifier, who does not run the outer verifier but, instead, uses the inner verifier to check that the outer verifier would have accepted had it made its queries to the PCP. The composed verifier also needs an auxiliary sub-PCP for the claim that the outer verifier would have accepted; in fact, he needs one sub-PCP for each possible random string of the outer verifier. Hence, the composed prover also sends all of these sub-PCPs along with the first PCP. The benefit is that the query complexity of the composed verifier equals that of the inner verifier, which is typically verifying a much smaller statement than the outer verifier.

Beyond query complexity, most other parameters of the composed proof system are simply the sum of corresponding parameters of the outer and inner proof systems. An exception is the proof length l: it does not simply equal the sum $\mathsf{l}_{\mathsf{out}} + \mathsf{l}_{\mathsf{in}}$, but instead equals $\mathsf{l}_{\mathsf{out}} + 2^{\mathsf{r}_{\mathsf{out}}} \cdot \mathsf{l}_{\mathsf{in}}$, because the composed prover uses the inner proof system to generate a proof *for each choice of randomness of the outer proof system.* (The same is true for prover running time.)

We prove an **Interactive Proof Composition Theorem** that avoids the above limitations. The outer proof system is a robust PCP $(P_{\mathsf{out}}, V_{\mathsf{out}})$ for a relation $\mathscr{R}$, while the inner one is a k-round IOP $(P_{\mathsf{in}}, V_{\mathsf{in}})$ for V_{out}'s relation; the composed proof system is a $(\mathsf{k}+1)$-round IOP (P, V) for $\mathscr{R}$. The parameters of the composed proof system are exactly as before, except that now the new proof length is *much smaller*: $\mathsf{l}_{\mathsf{out}} + \mathsf{l}_{\mathsf{in}}$. (Ditto for the prover running time.) The crucial observation is that, after the prover sends the outer proof to the verifier, *soundness is not harmed if the verifier tells the prover his choice of outer randomness*; hence, the prover does not have to invest work for all randomness choices but can simply invest work only for the outer randomness that was chosen, which he now knows.

Theorem 1.4 (Interactive Proof Composition — informal)**.** *Suppose that the relation $\mathscr{R}$ satisfies the following:*

***(1)** there exists a robust PCPP system $(P_{\mathsf{out}}, V_{\mathsf{out}})$ that puts $\mathscr{R}$ in the complexity class*

$$\textbf{PCPP}\begin{bmatrix} \text{proof length} & \mathsf{l}_{\mathsf{out}} \\ \text{randomness} & \mathsf{r}_{\mathsf{out}} \\ \text{query complexity} & \mathsf{q}_{\mathsf{out}} \\ \text{soundness error} & \varepsilon_{\mathsf{out}} \\ \text{proximity parameter} & \delta_{\mathsf{out}} \\ \text{robustness parameter} & \rho_{\mathsf{out}} \end{bmatrix}$$

and

***(2)** for every $\mathbb{x}$ there exists an IOPP system $(P_{\mathsf{in}}, V_{\mathsf{in}})$ that puts V_{out}'s relation in the complexity class*

$$\textbf{IOPP}\begin{bmatrix} \text{rounds} & \mathsf{k}_{\mathsf{in}} \\ \text{proof length} & \mathsf{l}_{\mathsf{in}} \\ \text{randomness} & \mathsf{r}_{\mathsf{in}} \\ \text{query complexity} & \mathsf{q}_{\mathsf{in}} \\ \text{soundness error} & \varepsilon_{\mathsf{in}} \\ \text{proximity parameter} & \delta_{\mathsf{in}} \end{bmatrix}$$

If $\delta_{\mathsf{in}} \leq \rho_{\mathsf{out}}$ then there exists an IOPP system (P, V) that puts $\mathscr{R}$ in the complexity class

$$\textbf{IOPP}\begin{bmatrix} \text{rounds} & \mathsf{k} & = & 1 + \mathsf{k}_{\mathsf{in}} \\ \text{proof length} & \mathsf{l} & = & \mathsf{l}_{\mathsf{out}} + \mathsf{l}_{\mathsf{in}} \\ \text{randomness} & \mathsf{r} & = & \mathsf{r}_{\mathsf{out}} + \mathsf{r}_{\mathsf{in}} \\ \text{query complexity} & \mathsf{q} & = & \mathsf{q}_{\mathsf{in}} \\ \text{soundness error} & \varepsilon & = & \varepsilon_{\mathsf{out}} + \varepsilon_{\mathsf{in}} \\ \text{proximity parameter} & \delta & = & \delta_{\mathsf{out}} \end{bmatrix} .$$

The above discussion and informal theorem statement omit many technical details that already arise in non-interactive proof composition (e.g., see lengthy discussions in [BGH$^+$06, BGH$^+$05]), and we also need to deal with. For instance, one has to clarify the size of the sub-claim on which the the inner proof system is invoked; also, one has to carefully define the notion of a verifier to allow for the composed verifier's running time to be *smaller* than the outer verifier's query complexity. For more details, see Section 3.

Sublinear sumcheck. The *sumcheck protocol* [LFKN92, Sha92] is an interactive proof for the claim "$\sum_{\vec{\alpha} \in H^m} w(\vec{\alpha}) = 0$", where w is the evaluation on $\mathbb{F}^m$ of an m-variate polynomial of individual degree d and H is a subset of $\mathbb{F}$. More generally, w may be a codeword in the tensor product code $C^{\otimes m}$, for a given linear code C with domain D and alphabet $\mathbb{F}$, and H is a subset of D [Mei13]. The prover receives H and w as input, while the verifier receives H as input and w as an oracle. The protocol has m rounds and, if C has relative distance τ, the protocol has soundness error $1 - \tau^m$; also, the prover runs in time $\mathrm{poly}(\ell^m)$, and the verifier in time $\mathrm{poly}(\ell + m)$, where $\ell := |D|$ is C's block length.

In each round, the verifier receives a codeword w_i in C and checks that $\sum_{\alpha \in H} w_i(\alpha)$ equals a certain value γ_{i-1} determined in the previous round; in particular, the verifier reads $\Omega(\ell)$ bits. We show that the verifier complexity can be *sublinear* in ℓ, if the prover and verifier engage in an IOP instead of an IP. The intuition to "go sublinear" is simple: instead of doing these checks explicitly, the verifier uses proximity testers for doing so. Thus, in each round, the prover sends to the verifier two oracles: the codeword in w_i, and a proximity proof attesting that $w_i \in C$ and that $\sum_{\alpha \in H} w_i(\alpha) = \gamma_{i-1}$. The use of proximity proofs complicates the soundness analysis because the verifier only sees noisy codewords, but the backbone of the proof follows that of the standard sumcheck protocol. Overall, we obtain a sumcheck IOP protocol that enables a verifier to efficiently check sumchecks for codes of much larger blocklength than what he can afford in the standard sumcheck protocol.

We state our **Sublinear Sumcheck Theorem** below as a reduction: given a PCP of proximity $(P_{\mathrm{SC}}, V_{\mathrm{SC}})$ for subcodes of the form $C|_{H,\gamma} := \{w \in C \text{ s.t. } \sum_{\alpha \in H} w(\alpha) = \gamma\}$, we construct an IOP of proximity (P, V) for sumchecks over H^m for $C^{\otimes m}$. The complexity of the PCPP verifier V_{SC} determines the complexity of the resulting IOPP verifier V; e.g., if the former is sublinear in C's block length ℓ, so is the latter.

Theorem 1.5 (Sublinear Sumcheck — informal). *Let m be a positive integer, and C a linear code with relative distance τ and block length ℓ. Suppose that there is a PCP of proximity for subcodes of the form $C|_{H,\gamma} := \{w \in C \text{ s.t. } \sum_{\alpha \in H} w(\alpha) = \gamma\}$ with proof length l_{SC}, query complexity q_{SC}, soundness error $\varepsilon_{\mathrm{SC}}$, proximity parameter δ_{SC}, prover running time $\mathsf{tp}_{\mathrm{SC}}$, and verifier running time $\mathsf{tv}_{\mathrm{SC}}$. Then there is a public-coin IOP for sumchecks over H^m for $C^{\otimes m}$ with the following parameters:*

$$\textbf{IOP} \left[\begin{array}{lccl} \text{rounds} & \mathsf{k} & = & m \\ \text{proof length} & \mathsf{l} & = & m \cdot \mathsf{l}_{\mathrm{SC}} + m \cdot \ell \\ \text{query complexity} & \mathsf{q} & = & m \cdot \mathsf{q}_{\mathrm{SC}} + m + 1 \\ \text{soundness error} & \varepsilon & = & 1 - \tau^m + \big(\varepsilon_{\mathrm{SC}} + m \cdot \delta_{\mathrm{SC}}\big) \\ \text{prover time} & \mathsf{tp} & = & m \cdot \mathsf{tp}_{\mathrm{SC}} + m \cdot \ell^m \\ \text{verifier time} & \mathsf{tv} & = & m \cdot \mathsf{tv}_{\mathrm{SC}} + O(m) \end{array} \right] .$$

In later sections, it is more natural to state the theorem without assuming that w is a codeword in $C^{\otimes m}$, so the reduction also takes as input a PCP of proximity $(P_{\otimes}, V_{\otimes})$ for $C^{\otimes m}$ that is invoked on w; this introduces additional terms in the parameters. More generally, both of the PCPs of proximity $(P_{\mathrm{SC}}, V_{\mathrm{SC}})$ and $(P_{\otimes}, V_{\otimes})$ can in fact be IOPs of proximity, and we state our theorem for this more general case, which we need. For more details, see Section 4.

Applying the new tools. We now sketch how we use the above new tools to derive the results of Section 1.4. We begin by discussing our results on proximity testing to codes (stated later); we then turn to circuit satisfiability (stated earlier) because its proof requires one of these results on proximity testing.

Intuition behind Theorem 1.2. The construction of linear-size IOPs of proximity for Reed–Solomon codes over binary fields follows from one invocation of our Interactive Proof Composition Theorem with [BS08]'s robust PCPs of proximity for Reed–Solomon codes as the outer proof system, and [Mie09]'s PCPs of proximity for nondeterministic languages as the inner proof system. Informally, in the first round, the prover sends to the verifier a [BS08] PCP of proximity, which reduces proximity testing of codewords over $\mathbb{F}_{2^\lambda}$ to proximity testing of sub-codewords over $\mathbb{F}_{2^{\lambda/2}+O(1)}$ with only constant overheads; in the second round, the verifier sends his choice of outer randomness, and the prover replies with a [Mie09] PCP of proximity for the sub-codeword. The proof length of this latter component is quasilinear, but is applied to a claim of "square-root size" only, so we obtain linear proof length.

Intuition behind Theorem 1.3. The construction of sublinear-size IOPs of proximity for tensor product codes follows from one invocation of our Interactive Proof Composition Theorem with [BS06, Vid15, CMS17]'s robust local tester for tensor product codes as the outer proof system, and [Mie09]'s PCPs of proximity for nondeterministic languages as the inner proof system. Unlike before, we now use one round, because the outer proof system only relies on a local tester rather than a PCP of proximity. The verifier thus simply sends his choice of outer randomness, and the prover replies with a [Mie09] PCP of proximity for a suitable sublinear-size sub-codeword. Since the proof length of this latter component is quasilinear but is applied to a sublinear-size claim, we obtain sublinear proof length.

A summary: overall, we can summarize the above sketches via the following diagram of implications.

Theorem 1.2 linear-size IOPP for Reed–Solomon codes	⟵	Theorem 1.4 interactive proof composition	+	[BS08] robust PCPs of proximity for Reed–Solomon codes	+	[Mie09] PCP of proximity for $\mathbf{NTIME}$
Theorem 1.3 sublinear-size IOPP for tensor product codes	⟵	Theorem 1.4 interactive proof composition	+	[BS06, Vid15] robust local testing for tensor product codes	+	[Mie09] PCP of proximity for $\mathbf{NTIME}$

Intuition behind Theorem 1.1. We now turn to how to construct 3-round IOPs for circuit satisfiability with linear proof length and constant query complexity.

The first step of many PCP constructions is to arithmetize the NP statement at hand (in our case, the satisfiability of a boolean circuit) by reducing it to a "PCP-friendly" problem that looks like a constraint satisfaction problem over a well-chosen graph and whose assignments involve codewords in a well-chosen linear code C. Meir observes in [Mei12, Mei13] that key features of C are good relative distance and, moreover, a *multiplication property*: coordinate-wise multiplication of codewords yields codewords in a code whose relative distance is still good [CC88, Mei13]. Moreover, to obtain short PCPs, the aforementioned graph is typically chosen so to behave like a routing network [PS94]; for example, [BS08] use De Bruijn graphs, while [BKK$^+$13] use hypercubes. To support such graphs, the automorphism group of C has to be rich enough. This typically holds for Reed–Solomon codes [BS08] which have a doubly-transitive automorphism group, but is a significantly harder condition to fulfill for AG codes [BKK$^+$13], for which obtaining a transitive automorphism group is quite involved and, currently, can only be achieved non-uniformly.

The aforementioned first step would be problematic in our setting, because known routing techniques introduce either logarithmic overheads (as in [BS08]) or large query complexity (as in [BKK$^+$13]), so it is not clear how we could use them. Departing from these prior works, we do not rely on any routing, and instead immediately leverage one round of interaction to directly reduce circuit satisfiability to a sumcheck instance over a given linear code C. Also, we only assume that C has good relative distance and a multiplication property [CC88], *but we do not rely on any automorphisms*.

Informally, the prover first sends three codewords w_1, w_2, w_3 over a field $\mathbb{F}$; the first codeword encodes values of the left wires of all gates, the second encodes values for the right wires of all gates, and the third encodes values for the output wires of all gates. (When a gate has fan-out greater than 1 we still consider 1 output wire.) The verifier now must check several things. First, that wire values are boolean and the output gate wire equals 0. Second, that the wire values are "locally consistent" with each gate: for every $i \in [n]$, $w_3(i)$ is the NAND of $w_1(i)$ and $w_2(i)$. Third, that the three encodings of wire values are consistent with the circuit topology: namely, if $\ell(i)$ represents the left wire used to compute i, and $r(i)$ represents the right wire used to compute i, the topology requires that $w_3(\ell(i)) = w_1(i)$ and $w_3(r(i)) = w_2(i)$ for every i. The verifier cannot directly conduct these checks (as doing so would incur linear query complexity); instead, the verifier sends some randomness to the prover so to "bundle" the checks into one sumcheck.

But how should the verifier sample randomness to achieve this bundling? One option is to sample a random element in $\mathbb{F}$ per check so to construct a random subset sum, which can be viewed as an n-variate polynomial of total degree 1, whose coefficients are the checks, evaluated at a random point. If not all checks are satisfied, the polynomial is non-zero, and its random evaluation cannot attain any value with too large probability. However, constructing a random subset sum is inefficient because the verifier samples and sends to the prover $\Omega(n)$ random bits, in order to describe the random point. Nevertheless, the verifier may hope to do better by using a *different* low-degree polynomial for the bundling. In general, if the polynomial has m variables each of degree at most d, the verifier must sample and send m field elements; this preserves soundness provided that $|\mathbb{F}| = \Omega(md)$ (for a constant probability of avoiding any particular output value by the Schwartz–Zippel Lemma [Sch80, Zip79, DL78]) and $d^m = \Omega(n)$ (to bundle all checks). For example, the univariate case of $m = 1$ was considered in [BFLS91] when reducing to a sumcheck problem; the multivariate case of $m = \log n$ or $m = \frac{\log n}{\log \log n}$ was considered in later works. Unfortunately, either setting does not work for *constant-size* fields, which we ultimately use to obtain linear proof length.

Taking a step back from polynomials, we see that all we need is an *evading set* S for $\mathbb{F}^n$, which is a small set such that for any non-zero $v \in \mathbb{F}^n$ the inner product $\langle r, v \rangle$, for random $r \in S$, does not attain any particular value $a \in \mathbb{F}$ with too high probability. Good constructions of evading sets are known: they relax a well-studied notion called ϵ-biased sets [NN90]. In particular, results of [AGHP92] imply that, for any ϵ, $\mathbb{F}^n$ has an evading set S of size $\mathrm{poly}(\frac{n}{\epsilon})$ and the aforementioned probability is $\gamma := \epsilon + \frac{1}{|\mathbb{F}|}$; in particular, such a construction is suitable for constant-size fields.

Below we informally state the reduction (see Section 6 for details), using the following notion: we say that a linear code C' is a *degree d-closure* of C if, for every $w_1, \dots, w_m \in C$ and m-variate polynomial P of total degree at most d, it holds that $w' \in C'$ where the i-th entry of w' is the evaluation of P on the i-th coordinates of $w_1, \dots, w_m$.

Lemma 1.6 (Circuit SAT to Sumcheck — informal)**.** *Let n be a positive integer, $C \subseteq \mathbb{F}^D$ an n-systematic linear code, ϕ an n-gate boolean circuit (of two-input NAND gates), and S an evading set for $\mathbb{F}^n$. There is a 1-round IOP that reduces satisfiability of ϕ to proximity testing to C and a sumcheck over any degree-3 closure of C. Moreover, the IOP introduces only constant overheads in all relevant parameters, including proof length and query complexity.*

After reducing circuit satisfiability to sumcheck over the given code C, we are left to choose C so to ensure that the sumcheck can be carried out with 2 additional rounds, linear proof length, and constant query complexity.

For this, our starting point is [GS96, SAK$^+$01]'s efficient construction of a code family with constant rate, relative distance, and alphabet size. Note that since these codes are AG codes, they have a naturally-defined degree-3 closure. Also, their construction is uniform, and thus represents a much "lighter" use of AG codes as compared to in [BKK$^+$13].

If we simply choose C to be a code from this AG code family, then it is not clear how to efficiently conduct the sumcheck. However, what does work is to take C to be the tensor product of $O(1)$ copies of this AG code. Informally, in this way, we can invoke our Sublinear Sumcheck Theorem (Theorem 1.5) on the tensor product code C and we can test proximity to it by Theorem 1.3. See Section 7 for details.

Overall, we can summarize the above sketch via the following diagram of implications.

$$\underset{\substack{\text{linear-size IOP}\\ \text{for circuit SAT}}}{\textbf{Theorem 1.1}} \longleftarrow \underset{\substack{\text{from circuit SAT}\\ \text{to sumcheck}}}{\text{Lemma 1.6}} + \underset{\substack{\text{sublinear}\\ \text{sumcheck}}}{\text{Theorem 1.5}} + \underset{\substack{\text{sublinear-size IOP}\\ \text{for tensor product codes}}}{\text{Theorem 1.3}} + \underset{\substack{\text{efficient construction}\\ \text{of AG codes}}}{\text{[GS96, SAK}^{+}\text{01]}}$$

1.6 Open questions

The question of whether there exist PCPs with linear proof length and constant query complexity remains open. Nevertheless, our work suggests additional questions that may be stepping stones in this and other intriguing directions: (1) Is there a *one*-round IOP for circuit satisfiability with linear proof length and query complexity? (Our IOP for circuit satisfiability requires 3 rounds.) (2) Is there an IOP for $\mathbf{NTIME}(T)$ with linear proof length and query complexity, for *some* number of rounds? (Our results, like [BKK$^+$13], only imply proof length $O(T \log T)$.) (3) Is there an IOP for *succinct* circuit satisfiability with linear proof length and query complexity? (Our results, like [BKK$^+$13], "stop" at $\mathbf{NP}$ but do not extend to $\mathbf{NEXP}$.) Finally, while "positive" applications of IOPs are known (e.g., non-interactive proofs in the random oracle model [BCS16]), "negative" ones are not: do IOP constructions with good parameters imply inapproximability results that are not known to be implied by known PCP constructions?

1.7 Roadmap

The rest of this paper is organized as follows. In Section 2, we provide basic notations and definitions, including for PCPs and IOPs, and state prior results that we rely on. In Section 3, we state and prove the interactive composition theorem. In Section 4, we state and prove the sublinear sumcheck theorem. In Section 5, we state and prove our results for additive Reed–Solomon codes and tensor product codes. In Section 6, we state and prove the reduction from circuit satisfiability to sumcheck. In Section 7, we state and prove our result for circuit satisfiability.

2 Preliminaries

2.1 Basic notations

Functions, distributions, fields. We use $f\colon D \to R$ to denote a function with domain D and range R; given a subset $\tilde{D}$ of D, we use $f|_{\tilde{D}}$ to denote the restriction of f to $\tilde{D}$. Given a distribution $\mathcal{D}$, we write $x \leftarrow \mathcal{D}$ to denote that x is sampled according to $\mathcal{D}$. We denote by $\mathbb{F}$ a finite field and by $\mathbb{F}_q$ the field of size q; we say $\mathbb{F}$ is a *binary field* if its characteristic is 2. We typically use fields of polynomial size, and take field operations to have constant cost each (and inspection shows that accounting for their actual polylogarithmic cost does not change any of the stated results).

Distances. A distance measure is a function $\Delta\colon \Sigma^n \times \Sigma^n \to [0,1]$ such that for all $x, y, z \in \Sigma^n$: (i) $\Delta(x,x) = 0$, (ii) $\Delta(x,y) = \Delta(y,x)$, and (iii) $\Delta(x,y) \leq \Delta(x,z) + \Delta(z,y)$. We extend Δ to distances to sets: given $x \in \Sigma^n$ and $S \subseteq \Sigma^n$, we define $\Delta(x,S) := \min_{y \in S} \Delta(x,y)$ (or 1 if S is empty). We say that a string x is ϵ-close to another string y if $\Delta(x,y) \leq \epsilon$, and ϵ-far from y if $\Delta(x,y) > \epsilon$; similar terminology applies for a string x and a set S. Unless noted otherwise, we use the *relative Hamming distance* over alphabet Σ (typically implicit): $\Delta(x,y) := |\{i \,|\, x_i \neq y_i\}|/n$.

Languages and relations. We denote by $\mathscr{R}$ a (binary ordered) relation consisting of pairs $(\mathbb{x}, \mathbb{w})$, where $\mathbb{x}$ is the *instance* and $\mathbb{w}$ is the *witness*. We denote by $\mathrm{Lan}(\mathscr{R})$ the language corresponding to $\mathscr{R}$, and by $\mathscr{R}|_{\mathbb{x}}$ the set of witnesses in $\mathscr{R}$ for $\mathbb{x}$. As always, we assume that $|\mathbb{w}|$ is bounded by some computable function of $n := |\mathbb{x}|$; in fact, we are mainly interested in relations arising from nondeterministic languages: $\mathscr{R} \in \mathbf{NTIME}(T)$ if there exists a $T(n)$-time machine M such that $M(\mathbb{x}, \mathbb{w})$ outputs 1 if and only if $(\mathbb{x}, \mathbb{w}) \in \mathscr{R}$. Throughout, we assume that $T(n) \geq n$. We say that $\mathscr{R}$ has relative distance $\delta_{\mathscr{R}}\colon \mathbb{N} \to [0,1]$ if $\delta_{\mathscr{R}}(n)$ is the minimum relative distance among witnesses in $\mathscr{R}|_{\mathbb{x}}$ for all $\mathbb{x}$ of size n. Throughout, we assume that $\delta_{\mathscr{R}}$ is a constant.

Polynomials. We denote by $\mathbb{F}[X_1, \dots, X_m]$ the ring of polynomials in m variables over $\mathbb{F}$. Given a polynomial P in $\mathbb{F}[X_1, \dots, X_m]$, $\deg_{X_i}(P)$ is the degree of P in the variable X_i. We denote by $\mathbb{F}^{<d}[X_1, \dots, X_m]$ the subspace consisting of $P \in \mathbb{F}[X_1, \dots, X_m]$ with $\deg_{X_i}(P) < d$ for every $i \in \{1, \dots, m\}$.

2.2 Probabilistically checkable proofs

We define non-adaptive *PCPs* [BFLS91, AS98, ALM$^+$98], *PCPs of proximity* [DR04, BGH$^+$06], and *robust PCPs of proximity* [BGH$^+$06]; each notion strengthens the former. We follow [BGH$^+$06], where more details can be found.

Informally, given a relation $\mathscr{R}$ and an instance $\mathbb{x}$: a PCP verifier is given oracle access to a candidate proof π and decides whether $\mathscr{R}|_{\mathbb{x}} \neq \emptyset$ ($\mathbb{x}$ is in $\mathscr{R}$'s language) or $\mathscr{R}|_{\mathbb{x}} = \emptyset$ ($\mathbb{x}$ is not in $\mathscr{R}$'s language); in contrast, a PCPP verifier is given oracle access to a candidate witness $\mathbb{w}$ and proof π and decides whether $\mathbb{w} \in \mathscr{R}|_{\mathbb{x}}$ or $\mathbb{w}$ is far from $\mathscr{R}|_{\mathbb{x}}$ (in particular, if $\mathscr{R}|_{\mathbb{x}} = \emptyset$, then $\mathbb{w}$ is far from $\mathscr{R}|_{\mathbb{x}}$). A robust PCPP strengthens a (standard) PCPP when $\mathscr{R}|_{\mathbb{x}} = \emptyset$: in such a case, the answers to the verifier's queries are, with high probability, far from any answers that make the verifier accept.

More formally, in each of the above cases, the proof system is specified by a pair (P, V) that works as follows.

- The *prover* P is a probabilistic algorithm that, given as input an instance-witness pair $(\mathbb{x}, \mathbb{w})$ with $n := |\mathbb{x}|$, outputs a proof $\pi\colon D(n) \to \Sigma(n)$, where the domain $D(n)$ and alphabet $\Sigma(n)$ are finite sets.
- The *verifier* V is a pair $(V^{\mathrm{Q}}, V^{\mathrm{D}})$, where V^{Q} is the *query algorithm* and V^{D} is the *decision algorithm*, where:
  - V^{Q} is probabilistic and, given as input an instance $\mathbb{x}$, outputs a state string σ and a query set I; and
  - V^{D} is deterministic and, given as input the state string σ and the $|I|$ query answers, outputs a bit.

  The verifier V induces the relation $\mathrm{Rel}(V)$ comprising pairs (σ, ω) such that, for some choice of $\mathbb{x}$ and I, (σ, I) is in the support of $V^{\mathrm{Q}}(\mathbb{x})$ and $V^{\mathrm{D}}(\sigma, \omega) = 1$.

We sometimes treat I as an algorithm that implicitly defines the query set (e.g., $I(i)$ is the i-th query) to allow for verifier-efficient interactive proof composition (see Section 3); this follows [BGH$^+$05]'s notion of *verifier specification*. Below, we explicitly denote the prover's and verifier's randomness as r_P and r_V.

The proof system (P, V) is for a relation $\mathscr{R}$ if it satisfies certain completeness and soundness properties. The completeness property is essentially the same in all three types of proof system; the only difference is that in a PCP the verifier only queries the proof π while in a (standard or robust) PCPP the verifier also queries the candidate witness.

Completeness

- **for PCPs:** (P, V) has perfect completeness if for every instance-witness pair $(\mathbb{x}, \mathbb{w})$ in the relation $\mathscr{R}$,

$$\Pr_{r_P, r_V}\left[V^{\mathrm{D}}(\sigma, \pi|_I) = 1 \;\middle|\; \begin{array}{c} \pi \leftarrow P(\mathbb{x}, \mathbb{w}; r_P) \\ (\sigma, I) \leftarrow V^{\mathrm{Q}}(\mathbb{x}; r_V) \end{array} \right] = 1 \ .$$

- **for PCPPs:** (P, V) has perfect completeness if for every instance-witness pair $(\mathbb{x}, \mathbb{w})$ in the relation $\mathscr{R}$,

$$\Pr_{r_P, r_V}\left[V^{\mathrm{D}}(\sigma, (\mathbb{w}\|\pi)|_I) = 1 \;\middle|\; \begin{array}{c} \pi \leftarrow P(\mathbb{x}, \mathbb{w}; r_P) \\ (\sigma, I) \leftarrow V^{\mathrm{Q}}(\mathbb{x}; r_V) \end{array} \right] = 1 \ .$$

The soundness property differs across the three types, with each condition strengthening the previous one:

Soundness

- **for PCPs:** (P, V) has soundness error ε if for every instance $\mathbb{x}$ with $\mathscr{R}|_{\mathbb{x}} = \emptyset$ and proof $\pi \colon D(n) \to \Sigma(n)$,

$$\Pr_{r_V}\left[\ V^{\mathrm{D}}(\sigma, \pi|_I) = 1 \ \middle| \ (\sigma, I) \leftarrow V^{\mathrm{Q}}(\mathbb{x}; r_V) \ \right] \le \varepsilon(n) \ .$$

- **for (standard) PCPPs:** (P, V) has soundness error ε with proximity parameter δ if for every instance-witness pair $(\mathbb{x}, \mathbb{w})$ with $\Delta(\mathbb{w}, \mathscr{R}|_{\mathbb{x}}) \ge \delta(n)$ and proof $\pi \colon D(n) \to \Sigma(n)$,

$$\Pr_{r_V}\left[\ V^{\mathrm{D}}(\sigma, (\mathbb{w}\|\pi)|_I) = 1 \ \middle| \ (\sigma, I) \leftarrow V^{\mathrm{Q}}(\mathbb{x}; r_V) \ \right] \le \varepsilon(n) \ .$$

- **for robust PCPPs:** (P, V) has soundness error ε with proximity parameter δ and robustness parameter ρ if for every instance-witness pair $(\mathbb{x}, \mathbb{w})$ with $\Delta(\mathbb{w}, \mathscr{R}|_{\mathbb{x}}) \ge \delta(n)$ and proof $\pi \colon D(n) \to \Sigma(n)$,

$$\Pr_{r_V}\left[\ (\mathbb{w}\|\pi)|_I \text{ is } \rho(n)\text{-close to } \mathrm{Rel}(V)|_\sigma \ \middle| \ (\sigma, I) \leftarrow V^{\mathrm{Q}}(\mathbb{x}; r_V) \ \right] \le \varepsilon(n) \ .$$

(We use the familiar "Markov-type" definition for robust soundness, but researchers have also used a definition that imposes a lower bound on the expected distance from $(\mathbb{w}\|\pi)|_I$ to $\mathrm{Rel}(V)|_\sigma$. The two notions are related [BGH+06].)

We now introduce complexity classes for each of the three types of proof systems. A relation $\mathscr{R}$ belongs to the complexity class $\mathbf{PCP}[\mathsf{a}, \mathsf{l}, \mathsf{r}, \mathsf{q}, \mathsf{s}, \varepsilon, \mathsf{tp}, \mathsf{tvq}, \mathsf{tvd}]$ if there is a PCP system for $\mathscr{R}$ in which: (1) the proof alphabet is $\mathsf{a}(n)$ (i.e., $\Sigma(n) = \mathsf{a}(n)$); (2) the proof length over that alphabet is at most $\mathsf{l}(n)$ (i.e., $|D(n)| \le \mathsf{l}(n)$); (3) the verifier uses at most $\mathsf{r}(n)$ random bits (i.e., $|r_V| \le \mathsf{r}(n)$); (4) the verifier queries the witness and proof in at most $\mathsf{q}(n)$ locations (i.e., $|I| \le \mathsf{q}(n)$); (5) the verifier state size is at most $\mathsf{s}(n)$ (i.e., $|\sigma| \le \mathsf{s}(n)$); (6) the soundness error is $\varepsilon(n)$; (7) the prover algorithm runs in time $\mathsf{tp}(n)$; (8) the verifier query algorithm runs in time $\mathsf{tvq}(n)$; (9) the verifier decision algorithm runs in time $\mathsf{tvd}(n)$. Similarly, we say that $\mathscr{R}$ belongs to the complexity class $\mathbf{PCPP}[\mathsf{a}, \mathsf{l}, \mathsf{r}, \mathsf{q}, \mathsf{s}, \varepsilon, \delta, \mathsf{tp}, \mathsf{tvq}, \mathsf{tvd}]$ if there is a PCPP system for $\mathscr{R}$ with the above parameters and having proximity parameter δ; also, we say that $\mathscr{R}$ belongs to the complexity class $\mathbf{PCPP}[\mathsf{a}, \mathsf{l}, \mathsf{r}, \mathsf{q}, \mathsf{s}, \varepsilon, \delta, \rho, \mathsf{tp}, \mathsf{tvq}, \mathsf{tvd}]$ if there is a robust PCPP system for $\mathscr{R}$ with the above parameters and having proximity parameter δ and robustness parameter ρ. Sometimes we write tv to denote the sum of tvq and tvd, i.e., the total running time of the verifier.

Throughout, we assume that the proximity parameter is less than the "unique-decoding radius" of the relation $\mathscr{R}$, i.e., that $\delta < \frac{1}{2}\delta_{\mathscr{R}}$ where $\delta_{\mathscr{R}}$ is the relative distance of $\mathscr{R}$ (see Section 2.1).

2.3 Interactive oracle proofs

We define *interactive oracle proofs* (IOPs) [BCS16, RRR16], which combine aspects of IPs [Bab85, GMR89] and PCPs [BFLS91, AS98, ALM+98]; also, they generalize Interactive PCPs [KR08].

Informally, an IOP extends an interactive proof as follows: whenever the prover sends to the verifier a message, the verifier does not have to read the message in full but may probabilistically query it. In more detail, a k-round IOP comprises k rounds of interaction. In the i-th round of interaction: the verifier sends a message m_i to the prover; then the prover replies with a message f_i to the verifier, which the verifier can query in this and all later rounds (by having oracle access to it). After the k rounds of interaction, the verifier either accepts or rejects.

An *IOP system* for a relation $\mathscr{R}$ with round complexity k and soundness error ε is a pair (P, V), where P, V are probabilistic algorithms, that satisfies the following properties. (See [BCS16, RRR16] for more details.)

Completeness: For every instance-witness pair $(\mathbb{x}, \mathbb{w})$ in the relation $\mathscr{R}$, $\langle P(\mathbb{x}, \mathbb{w}), V(\mathbb{x})\rangle$ is a $\mathsf{k}(n)$-round interactive oracle protocol with accepting probability 1.

Soundness: For every instance $\mathbb{x}$ not in $\mathscr{R}$'s language and unbounded malicious prover $\tilde{P}$, $\langle \tilde{P}, V(\mathbb{x})\rangle$ is a $\mathsf{k}(n)$-round interactive oracle protocol with accepting probability at most $\varepsilon(n)$.

Like the IP model, a fundamental measure of efficiency is the round complexity k. Like the PCP model, two additional fundamental measures of efficiency are the *proof length* l, which is the total number of alphabet symbols in all of the prover's messages, and the *query complexity* q, which is the total number of locations queried by the verifier across all of the prover's messages. Considering all of these parameters, we say that a relation $\mathscr{R}$ belongs to the complexity class $\mathbf{IOP}[\mathsf{k}, \mathsf{a}, \mathsf{l}, \mathsf{r}, \mathsf{q}, \varepsilon, \mathsf{tp}, \mathsf{tv}]$ if there is an IOP system for $\mathscr{R}$ in which: (1) the number of rounds is at most $\mathsf{k}(n)$; (2) the prover messages are over the alphabet $\mathsf{a}(n)$; (3) the proof length over this alphabet is at most $\mathsf{l}(n)$; (4) the verifier uses at most $\mathsf{r}(n)$ random bits; (5) the verifier queries the prover messages in at most $\mathsf{q}(n)$ locations; (6) the soundness error is $\varepsilon(n)$; (7) the prover algorithm runs in time $\mathsf{tp}(n)$; (8) the verifier algorithm runs in time $\mathsf{tv}(n)$.

Finally, we say that an IOP system is *non-adaptive* if the verifier queries are non-adaptive (i.e., the queried locations depend only on the verifier's inputs); also, we say that an IOP system is *public coin* if m_i is a random string and queries to $f_1, \ldots, f_{i-1}$ before the i-th round depend only on $m_1, \ldots, m_{i-1}$.

Proximity. We also study IOPs of proximity, which extend IOPs the same way that PCPs of proximity extend PCPs. An *IOPP system* for a relation $\mathscr{R}$ with round complexity k, soundness error ε, and proximity parameter δ is a pair (P, V) that satisfies the following properties.

Completeness: For every instance-witness pair $(\mathbb{x}, \mathbb{w})$ in the relation $\mathscr{R}$, $\langle P(\mathbb{x}, \mathbb{w}), V^{\mathbb{w}}(\mathbb{x})\rangle$ is a $\mathsf{k}(n)$-round interactive oracle protocol with accepting probability 1.

Soundness: For every instance-witness pair $(\mathbb{x}, \mathbb{w})$ with $\Delta(\mathbb{w}, \mathscr{R}|_{\mathbb{x}}) \geq \delta(n)$ and unbounded malicious prover $\tilde{P}$, $\langle \tilde{P}, V^{\mathbb{w}}(\mathbb{x})\rangle$ is a $\mathsf{k}(n)$-round interactive oracle protocol with accepting probability at most $\varepsilon(n)$.

Similarly to above, a relation $\mathscr{R}$ belongs to the complexity class $\mathbf{IOPP}[\mathsf{k}, \mathsf{a}, \mathsf{l}, \mathsf{r}, \mathsf{q}, \varepsilon, \delta, \mathsf{tp}, \mathsf{tv}]$ if there is an IOPP system for $\mathscr{R}$ with the corresponding parameters. As in Section 2.2, we always assume that δ is less than $\frac{1}{2}\delta_{\mathscr{R}}$.

Robustness. Analogously to non-adaptive PCPs, a robustness parameter ρ can be defined for non-adaptive IOPs. For such IOPs, we can think of the verifier algorithm as two algorithms: a query algorithm that, given as input an instance $\mathbb{x}$, interacts with the prover and then outputs a state string σ and a query set I; then, a decision algorithm takes as input the state string σ and the $|I|$ query answers (across all prover messages), and outputs a bit. Then one requires that, with probability at most $\varepsilon(n)$, the answers to the verifier's queries are $\rho(n)$-close to any answers that make the verifier accept. We do not spell out all the details of this definition because robustness of IOPs will arise in informal discussions only.

Remark 2.1 (comparison with PCPs and IPCPs)**.** An IOP can be viewed as a "multi-round PCP": a PCP is the special case of an IOP where $\mathsf{k} = 1$ (and the first verifier message is empty). Similarly, a PCP of proximity is the special case of an IOP of proximity where $\mathsf{k} = 1$ (and the first verifier message is empty). Also, IOPs generalize Interactive PCPs [KR08], which are IOPs where the verifier sends an empty first message and may query only the first prover message (but must read any other prover messages in full).

2.4 Codes

An error correcting code C is a set of functions $w \colon D \to \Sigma$, where D, Σ are finite sets known as the domain and alphabet; we write $C \subseteq \Sigma^D$. The message length of C is $k := \log_{|\Sigma|} |C|$, its block length is $\ell := |D|$, its rate is $\rho := k/\ell$, its (minimum) distance is $d := \min\{\Delta(w, z) \,|\, w, z \in C\,,\, w \neq z\}$ when Δ is the (absolute) Hamming distance, and its (minimum) relative distance is $\tau := d/\ell$. At times we write $k(C), \ell(C), \rho(C), d(C), \tau(C)$ to make the code under consideration explicit. All the codes we consider are linear codes, discussed next.

Linearity. A code C is *linear* if Σ is a finite field and C is a Σ-linear space in Σ^D. The dual code of C is the set $C^\perp$ of functions $z\colon D \to \Sigma$ such that, for all $w\colon D \to \Sigma$, $\langle z, w\rangle := \sum_{i\in D} z(i)w(i) = 0$. We denote by $\dim(C)$ the dimension of C (as a linear space); it holds that $\dim(C) + \dim(C^\perp) = \ell$ and $\dim(C) = k$ (dimension equals message length). The support of C, denoted $\mathrm{supp}(C)$, is the union of the support (non-zero entries) of functions in C.

Systematicity. Given $s \in \mathbb{N}$, a code $C \subseteq \Sigma^D$ is *s-systematic* if there exists a size-s subset of D, which for convenience we identify with $[s]$, such that for every $x \in \Sigma^{[s]}$ there exists $w \in C$ such that $x = w|_{[s]}$. Observe that the message length k is at least s, but may be larger; with this in mind, we define the *pseudorate* to be $\hat{\rho} := s/\ell \le \rho$.

Tensor products. Given a positive integer m, the *tensor product code* $C^{\otimes m}$ is the linear code with domain D^m, alphabet Σ, message length k^m, block length ℓ^m, and distance d^m that comprises all functions $w\colon D^m \to \Sigma$ whose restriction to any axis-parallel line is in C. Namely, for every $j \in \{1,\dots,m\}$ and $a_1,\dots,a_{j-1},a_{j+1},\dots,a_\ell \in D$, the function $w'\colon D \to \Sigma$ defined by $w'(i) := w(a_1,\dots,a_{j-1},i,a_{j+1},\dots,a_\ell)$ is in C.

Polynomial closures. Given $C, C' \subseteq \mathbb{F}^D$ and positive integer t, we say that C' is a *degree-t closure* of C if, for every $w_1,\dots,w_m \in C$ and $P \in \mathbb{F}[X_1,\dots,X_m]$ of total degree at most t, it holds that $w' := P(w_1,\dots,w_m)$ is in C', where $w'\colon D \to \Sigma$ is defined coordinate-wise by the equation $w'(i) := P(w_1(i),\dots,w_m(i))$ (i.e., P is applied coordinate-wise to $(w_1,\dots,w_m)$). We now state a simple lemma about polynomial closures.

Claim 2.2. *Let $C, C' \subseteq \mathbb{F}^D$, $t, m \in \mathbb{N}$, and $P \in \mathbb{F}[X_1,\dots,X_m]$. Suppose that C' is a degree-t closure of C, P has total degree at most t, and $f_1,\dots,f_m \in \mathbb{F}^D$ have relative distance less than $\frac{\tau(C')}{2m}$ to C'. Let w' be the codeword in C' closest to $P(f_1,\dots,f_m)$, and, for $i = 1,\dots,m$, let w_j be the codeword in C closest to f_j. Then $w' = P(w_1,\dots,w_m)$.*

Proof. Let T be the set of $i \in D$ for which there exist $j \in [m]$ for which $f_j(i) \neq w_j(i)$. By assumption, $|T|$ is less than $\frac{1}{2}\tau(C')|D|$. The codeword $P(w_1,\dots,w_m)$ is in C' and disagrees with $P(f_1,\dots,f_m)$ in less than $\frac{1}{2}\tau(C')|D|$ places, and so must be the unique codeword in C' closest to $P(f_1,\dots,f_m)$. Thus, $w' = P(w_1,\dots,w_m)$, as desired. □

Code families. We also consider families of codes, and the notation above typically extends in the natural way.

- *Parameters:* A code family $\mathscr{C} = \{C_n\}_{n\in\mathbb{N}}$ has domain $D(\cdot)$ and alphabet $\mathbb{F}(\cdot)$ if each code C_n has domain $D(n)$ and alphabet $\mathbb{F}(n)$. Similarly, $\mathscr{C}$ has message length $k(\cdot)$, block length $\ell(\cdot)$, rate $\rho(\cdot)$, distance $d(\cdot)$, and relative distance $\tau(\cdot)$ if each code C_n has message length $k(n)$, block length $\ell(n)$, rate $\rho(n)$, distance $d(n)$, and relative distance $\tau(n)$. We also define $\rho(\mathscr{C}) := \inf_{n\in\mathbb{N}} \rho(n)$ and $\tau(\mathscr{C}) := \inf_{n\in\mathbb{N}} \tau(n)$; note that $\mathscr{C}$ has constant rate iff $\rho(\mathscr{C}) > 0$ and has constant relative distance iff $\tau(\mathscr{C}) > 0$.

- *Systematicity:* A code family $\mathscr{C} = \{C_n\}_{n\in\mathbb{N}}$ is $s(\cdot)$-systematic if each code C_n is $s(n)$-systematic. We define $\hat{\rho}(\mathscr{C}) := \inf_{n\in\mathbb{N}} \hat{\rho}(C_n) \le \rho(\mathscr{C})$ and note that $\mathscr{C}$ has constant pseudorate iff $\hat{\rho}(\mathscr{C}) > 0$.

- *Tensor products:* Given a code family $\mathscr{C} = \{C_n\}_{n\in\mathbb{N}}$ and a positive integer m, $\mathscr{C}^{\otimes m}$ is the family $\{C_n^{\otimes m}\}_{n\in\mathbb{N}}$. Note that if $\mathscr{C}$ is $s(\cdot)$-systematic then $\mathscr{C}^{\otimes m}$ is $s(\cdot)^m$-systematic.

- *Polynomial closures:* Given two code families $\mathscr{C} = \{C_n\}_{n\in\mathbb{N}}$ and $\mathscr{D} = \{D_n\}_{n\in\mathbb{N}}$ over the same domain $D(n)$ and alphabet $\mathbb{F}(n)$ and a positive integer t, $\mathscr{D}$ is a degree-t closure of $\mathscr{C}$ if each D_n is a degree-t closure of C_n.

- *Efficiency:* We say that a n-systematic code family $\mathscr{C} = \{C_n\}_{n\in\mathbb{N}}$ is $T(\cdot)$-efficient if computing a codeword as well as checking that a codeword lies in the code can be done in time $T(\cdot)$. More precisely, there exist two deterministic algorithms Enc and Chk such that, for every $n \in \mathbb{N}$:
  - on input $x \in \mathbb{F}(n)^{[n]}$, Enc outputs $w \in C_n$ such that $w|_{[n]} = x$;
  - on input $w \in \mathbb{F}(n)^{\ell(n)}$, Chk outputs 1 if and only if $w \in C_n$.

  If a code family is not systematic, we still talk about its efficiency, but in such a case we only require the second condition (about checking if a codeword is in the code) and ignore the first one (about computing codewords).

  If $\mathscr{C}$ is $T(\cdot)$-efficient then $\mathscr{C}^{\otimes m}$ is $T'(\cdot)$-efficient with $T'(n) := m \cdot \ell(n)^{m-1} \cdot T(n)$; see [Vid15, Appendix C.2].

2.4.1 Reed–Solomon codes

The Reed–Solomon (RS) code is the code consisting of evaluations of *univariate* low-degree polynomials: given a field $\mathbb{F}$, subset S of $\mathbb{F}$, and positive integer d with $d \leq |S|$, we denote by $\mathrm{RS}[\mathbb{F}, S, d]$ the linear code consisting of evaluations $w \colon S \to \mathbb{F}$ over S of polynomials in $\mathbb{F}^{<d}[X]$. The code's message length is $k = d$, block length is $\ell = |S|$, rate is $\rho = \frac{d}{|S|}$, and relative distance is $\tau = 1 - \frac{d-1}{|S|}$.

2.4.2 Reed–Muller codes

The Reed–Muller (RM) code is the code consisting of evaluations of *multivariate* low-degree polynomials: given a field $\mathbb{F}$, subset S of $\mathbb{F}$, and positive integers m, d with $d \leq |S|$, we denote by $\mathrm{RM}[\mathbb{F}, S, m, d]$ the linear code consisting of evaluations $w \colon S^m \to \mathbb{F}$ over S^m of polynomials in $\mathbb{F}^{<d}[X_1, \dots, X_m]$ (i.e., we bound individual degrees rather than their sum). The code's message length is $k = d^m$, block length is $\ell = |S|^m$, rate is $\rho = (\frac{d}{|S|})^m$, and relative distance is $\tau = (1 - \frac{d-1}{|S|})^m$. Also, one can verify that (i) $\mathrm{RS}[\mathbb{F}, S, d] = \mathrm{RM}[\mathbb{F}, S, 1, d]$ and (ii) $\mathrm{RM}[\mathbb{F}, S, m, d]$ is the m-wise tensor product of $\mathrm{RS}[\mathbb{F}, S, d]$.

2.4.3 Algebraic-geometry codes

Codewords in an algebraic-geometry (AG) code [Gop81] consist of functions evaluated over domains that arise as the solutions of certain systems of rational equations. AG codes generalize RS codes and "behave" similarly but, unlike RS codes, can "work" over *constant-size* fields: AG codes can achieve constant rate and relative distance over a constant-size field and, moreover, they can be closed under coordinate-wise multiplication, a property first used by [CC88] in constructing error correcting codes. A formal definition of AG codes is not necessary to understand this paper, so we refer the reader to the excellent reference of Stichtenoth [Sti08]. Instead, below we provide a self-contained statement of the result we rely on; it follows from [SAK$^+$01, Theorem 7], which gives an efficient construction of AG codes based on [GS96]'s explicit towers of function fields.

Theorem 2.3. *For every $\tau_0 \in (0,1)$ and $t_0 \in \mathbb{N}$, there exist $\hat{\rho}_0 \in (0,1)$, a prime power q_0, and two code families $\mathscr{A} = \{A_n\}_{n \in \mathbb{N}}$ and $\mathscr{B} = \{B_n\}_{n \in \mathbb{N}}$ that satisfy the following properties:*
1. $\mathscr{A}$ and $\mathscr{B}$ are n-systematic codes with alphabet $\mathbb{F}_{q_0}$;
2. $\mathscr{A}$ and $\mathscr{B}$ have $\hat{\rho}(\mathscr{A}), \hat{\rho}(\mathscr{B}) \geq \hat{\rho}_0$;
3. $\mathscr{A}$ and $\mathscr{B}$ have $\tau(\mathscr{A}), \tau(\mathscr{B}) \geq \tau_0$;
4. $\mathscr{A}$ is a degree-t_0 closure of $\mathscr{B}$;
5. $\mathscr{A}$ and $\mathscr{B}$ are $T(n)$-efficient with $T(n) = O(n^3 \cdot \log n)$.

2.5 Evading sets

Given a field $\mathbb{F}$ and positive integer n, an evading set for $\mathbb{F}^n$ is a small set S such that for any non-zero $v \in \mathbb{F}^n$ the inner product $\langle r, v \rangle$, for a uniformly random $r \in S$, does not attain any particular value $a \in \mathbb{F}$ with too high probability. The notion is captured by the following definition.

Definition 2.4. *Let $n \in \mathbb{N}$, $\gamma \in (0,1)$, and $\mathbb{F}$ be a field. A subset S of $\mathbb{F}^n$ is* γ-evading *for $\mathbb{F}^n$ if $\Pr_{r \leftarrow S}[\langle r, v \rangle = a] \leq \gamma$ for every $v \in \mathbb{F}^n$ and $a \in \mathbb{F}$ with $v \neq 0^n$. We say that a family $\mathscr{S} = \{S_n\}_{n \in \mathbb{N}}$ is $T(\cdot)$-efficient $\gamma(\cdot)$-evading for $\mathbb{F}(\cdot)$ if, for every $n \in \mathbb{N}$: (i) S_n is $\gamma(n)$-evading for $\mathbb{F}(n)^n$, and (ii) one can sample a random vector in S_n in time $T(n)$.*

We state a construction of a small evading set, for any finite field. The construction is a straightforward generalization of the one for $\mathbb{F}_2$ in [AGHP92], where it is shown to have the stronger property of being an ϵ-biased set [NN90]; for completeness we also provide a proof sketch.

Lemma 2.5 (based on [AGHP92])**.** *Let $\epsilon \in (0,1)$, $n \in \mathbb{N}$, and $m := \lceil \log_q(n/\epsilon) \rceil$. Let q be a prime power and $\phi \colon \mathbb{F}_{q^m} \to \mathbb{F}_q^m$ an isomorphism of $\mathbb{F}_q$-vector spaces. Define $S := \{s_{x,y}\}_{x \in \mathbb{F}_{q^m}, y \in \mathbb{F}_q^m}$ where $s_{x,y} \in \mathbb{F}_q^n$ is the vector $\big(\langle \phi(1), y \rangle, \langle \phi(x), y \rangle, \dots, \langle \phi(x^{n-1}), y \rangle\big)$. Then S is $(\epsilon + 1/q)$-evading for $\mathbb{F}_q^n$.*

In particular, there is algorithm that, on input $n \in \mathbb{N}$ and $i \in [q^{2m}]$, runs in time $n \cdot \mathrm{polylog}(nq)$ and outputs the i-th element of a set S of size q^{2m} that is $(\epsilon + 1/q)$-evading for $\mathbb{F}_q^n$.

Proof. We only sketch the proof, because it is very similar to [AGHP92, Proposition 3]. Fix $v \in \mathbb{F}^n$ and $a \in \mathbb{F}$ with $v \neq 0^n$. The inner product $\langle s_{x,y}, v\rangle$ equals the inner product of y with $\phi(P_v(x))$, where P_v is the polynomial $P_v(X) \triangleq \sum_{i=0}^{n-1} v_i \cdot X^i$. Hence, $P_v(x) = 0$ with probability at most ϵ. Given that $\phi(P_v(x)) \neq 0$, the probability its inner product with a uniform y equals a is at most $1/q$. □

2.6 Constructions of proximity testers

We state results about constructions of proximity testers that we use in this paper.

2.6.1 Robust PCPs of proximity for additive Reed–Solomon codes

Ben-Sasson and Sudan [BS08] show that Reed–Solomon codes over binary fields have quasilinear-size PCPs of proximity with constant query complexity. A central ingredient of their construction is a robust PCP of proximity for these codes that reduces proximity testing for dimension λ to proximity testing for dimension $\lambda/2 + O(1)$; this also reduces the query complexity from 2^{λ} to $O(2^{\lambda/2})$.

Below, we state [BS08]'s robust PCP of proximity for these codes. The general statement involves the notion of an *additive* Reed–Solomon code $\mathrm{RS}^+[\mathbb{F}, S, d]$, which is a Reed–Solomon code $\mathrm{RS}[\mathbb{F}, S, d]$ where $\mathrm{char}(\mathbb{F}) = 2$ and S is an $\mathbb{F}_2$-linear affine subspace. In this case there exist $\alpha_0 \in \mathbb{F}$ and $\mathbb{F}_2$-linearly-independent $\alpha_1, \ldots, \alpha_\lambda \in \mathbb{F}$ such that S equals the $\mathbb{F}_2$-linear span of $(\alpha_1, \ldots, \alpha_\lambda)$ shifted by α_0; S can then be succinctly represented via the list $(\alpha_0, \alpha_1, \ldots, \alpha_\lambda)$. If $\mathrm{char}(\mathbb{F}) = 2$, we then denote by $\mathrm{Rel}(\mathbb{F}, \varrho)$ the relation of instance-witness pairs $(\mathbbm{x}, \mathbbm{w}) = \big((S, d), w\big)$ such that $w \in \mathrm{RS}^+[\mathbb{F}, S, d]$, S is λ-dimensional, and $d \leq \varrho|S|$; we define the size of $\mathbbm{x}$ to be λ.

Theorem 2.6. *For every $\varrho > 0$ there exist $\alpha \in (0, 1)$ and $a > 0$ such that for every binary field $\mathbb{F}$ and $\delta \in (0, \frac{1}{2}(1-\varrho))$ there exists a robust PCPP system (P, V) that puts $\mathrm{Rel}(\mathbb{F}, \varrho)$ in the complexity class*

$$\mathbf{PCPP}\left[\begin{array}{lrcl} \text{answer alphabet} & \mathsf{a}(\lambda) & = & \mathbb{F} \\ \text{proof length} & \mathsf{l}(\lambda) & = & a \cdot 2^{\lambda} \\ \text{randomness} & \mathsf{r}(\lambda) & = & \lambda + a \\ \text{query complexity} & \mathsf{q}(\lambda) & = & a \cdot 2^{\lambda/2} \\ \text{state size} & \mathsf{s}(\lambda) & = & \lambda/2 + a \\ \text{soundness error} & \varepsilon(\lambda) & = & 1 - \alpha \cdot \delta \\ \text{proximity parameter} & \delta(\lambda) & = & \delta \\ \text{robustness parameter} & \rho(\lambda) & = & \alpha \cdot \delta \\ \text{prover time} & \mathsf{tp}(\lambda) & = & \lambda^a 2^{\lambda} \\ \text{verifier query time} & \mathsf{tvq}(\lambda) & = & \lambda/2 + a \\ \text{verifier decision time} & \mathsf{tvd}(\lambda) & = & \lambda^a \cdot 2^{\lambda/2} \end{array}\right].$$

Moreover, $\mathrm{Rel}(V)$ *is a subset of* $\mathrm{Rel}(\mathbb{F}, \varrho)$.

Remark 2.7. All of the existential quantifiers in the theorem statement are "efficient" in the sense that there is a (uniform) polynomial-time procedure to construct the relevant objects. For example, (P, V) efficiently depends on the field $\mathbb{F}$ and proximity parameter δ.

Also, the running time of the verifier query algorithm in the above statement is $\lambda/2 + a$ while the number of queries is $a \cdot 2^{\lambda/2}$. This is possible because the verifier query algorithm outputs an algorithm that implicitly represents the query set rather than the query set itself; see Section 2.2.

2.6.2 Robust local testers for tensor product codes

Viderman [Vid15] shows that tensor product codes have robust local testers; his work builds on [BS06], who showed a similar result but only for codes with large enough distance; both of these can be viewed as extending [RS97]'s tester for low-degree polynomials to general linear codes. Suitable settings of parameters yield a generic construction of codes with rate $1/\mathrm{poly}(n)$ that are testable with query complexity $\mathrm{polylog}(n)$.

Given an integer $m > 1$ and a linear code $C \subseteq \mathbb{F}^D$ with relative distance τ, there is a natural $|D|^{m-1}$-query test for checking that a function $w\colon D^m \to \mathbb{F}$ is close to $C^{\otimes m}$: sample random $i \in D$ and $j \in \{1, \ldots, m\}$ and check that w with the j-th coordinate restricted to i is in $C^{\otimes m-1}$. Remarkably, this test is robust [BS06, Vid15, CMS17]: if $m \geq 3$ then the tester is $\frac{\tau^m}{12}$-robust. Query complexity can then be reduced via a composition theorem for Tanner product codes [Tan81], of which tensor product codes are a special case [BS06, Lemma 4.1].

For example, [Vid15, Theorem 3.1] states that: the $|D|^2$-query test "*pick a random axis-parallel 2-dimensional plane H in D^m and check that $w|_H \in C^{\otimes 2}$*" is $\frac{\tau^{2m}}{m^8}$-robust. (Recall that an n-dimensional axis-parallel plane H is a set of points in D^m obtained by restricting all but n coordinates to constants.)

Below, we state the aforementioned result in a more general form (which we need): for a given divisor μ of m with $3\mu \leq m$, the test picks a random axis-parallel 2μ-dimensional plane H and checks that $w|_H \in C^{\otimes 2\mu}$. This more general form follows by invoking the original result on the tensor product code $C_0^{\otimes m/\mu}$ with $C_0 := C^{\otimes \mu}$.

Lemma 2.8. *Let C be a linear code with alphabet $\mathbb{F}$ and relative distance τ, m a positive integer, and μ a divisor of m. If $m \geq 3\mu$ then the following two conditions hold:*

1. (completeness) *For every $w\colon D^m \to \mathbb{F}$ in the tensor product code $C^{\otimes m}$, it holds that* $\Pr_H[w|_H \in C^{\otimes 2\mu}] = 1$.

2. (robust soundness) *For every $w\colon D^m \to \mathbb{F}$, it holds that* $\mathbb{E}_H[\Delta(w|_H, C^{\otimes 2\mu})] \geq \frac{\tau^{2(m/\mu)}}{(m/\mu)^8} \cdot \Delta(w, C^{\otimes m})$.

2.6.3 PCPs of proximity for nondeterministic languages

Mie [Mie09] constructs PCPs of proximity for nondeterministic languages, where the prover runs in quasilinear time and the verifier in polylogarithmic time, with constant proximity parameter, soundness error, and query complexity.

Theorem 2.9. *For every relation $\mathscr{R} \in \mathbf{NTIME}(T(n))$ with distance $\delta_{\mathscr{R}}$, every $\delta_0 \in (0, \frac{1}{2}\delta_{\mathscr{R}})$, and every $\varepsilon_0 > 0$, there exist $q_0 > 0$ and a PCPP system (P, V) that puts $\mathscr{R}$ in the complexity class*

$$\mathbf{PCPP}\begin{bmatrix} \text{answer alphabet} & \mathsf{a}(n) & = & \mathbb{F}_2 \\ \text{proof length} & \mathsf{l}(n) & = & \tilde{O}(T(n)) \\ \text{randomness} & \mathsf{r}(n) & = & \log T(n) + O(\log\log T(n)) \\ \text{query complexity} & \mathsf{q}(n) & = & q_0 \\ \text{state size} & \mathsf{s}(n) & = & \operatorname{polylog} T(n) \\ \text{soundness error} & \varepsilon(n) & = & \varepsilon_0 \\ \text{proximity parameter} & \delta(n) & = & \delta_0 \\ \text{prover time} & \mathsf{tp}(n) & = & \operatorname{poly}(n) + \tilde{O}(T(n)) \\ \text{verifier query time} & \mathsf{tvq}(n) & = & \operatorname{poly}(n + \log T(n)) \\ \text{verifier decision time} & \mathsf{tvd}(n) & = & \operatorname{poly}(n + \log T(n)) \end{bmatrix}.$$

Remark 2.10. The statement in [Mie09] only shows that the prover runs in $\operatorname{poly}(T(n))$ time. However, that prover is composed of first running the "outer prover" of [BS08], which was shown to run in $\operatorname{poly}(n) + \tilde{O}(T(n))$ time [BCGT13], and then applying the "inner prover" of [Din07] (which runs in polynomial time) on instances of size $\operatorname{polylog} T(n)$. Combined, the resulting prover runs in $\operatorname{poly}(n) + \tilde{O}(T(n))$ time, as stated above.

3 Interactive proof composition

We prove an interactive proof composition theorem that, when compared to its non-interactive counterpart [AS98], provides significant savings in proof length, as well as prover running time, for the composed proof system. Later on, we leverage this result to obtain short interactive oracle proofs (see Section 5).

Proof composition. Proof composition is a fundamental technique for reducing the query complexity of PCP verifiers; it was introduced in [AS98] and later used and refined in numerous PCP constructions [ALM^{+}98, HS00, BGH^{+}06]. Proof composition involves two probabilistically-checkable proof systems: an outer one and an inner one. One should think of the outer proof system as having short proofs but large query complexity, while the inner proof system has long proofs but small query complexity. Proof composition combines these two so as to obtain a new proof system that inherits the good properties of each; one can think of this as a "proof analogue" of code concatenation [For65].

Informally, the composed prover uses the outer prover to send a PCP to the composed verifier; the composed verifier does not run the outer verifier but, instead, uses the inner verifier to check that the outer verifier would have accepted had it made its queries to the PCP. In order to do so, the composed verifier also needs an auxiliary sub-PCP for the claim that the outer verifier would have accepted; in fact, he needs one such sub-PCP for each possible random string of the outer verifier because each randomness choice induces a corresponding claim. Hence, the composed prover also sends all of these sub-PCPs along with the first PCP. The benefit is that the query complexity of the composed verifier equals that of the inner verifier, which is typically verifying a much smaller statement than the outer verifier (this statement's size is roughly the outer query complexity).

Turning the above sketch into a proof requires distinct properties from the outer PCP and inner PCP. A useful choice is from [BGH^{+}06]: the outer PCP should be a robust PCP while the inner PCP should be a PCP of proximity.[2]

Limitations of proof composition. Beyond query complexity, most other parameters of the composed proof system are simply the sum of the corresponding parameters of the outer and inner proof systems: roughly, the randomness complexity is $\mathsf{r} = \mathsf{r}_{\mathsf{out}} + \mathsf{r}_{\mathsf{in}}$, the soundness error is $\varepsilon = \varepsilon_{\mathsf{out}} + \varepsilon_{\mathsf{in}}$, and the verifier running time is $\mathsf{tv} = \mathsf{tv}_{\mathsf{out}} + \mathsf{tv}_{\mathsf{in}}$. There are two exceptions: the proof length l and prover running time tp are at least $\mathsf{l}_{\mathsf{out}} + 2^{\mathsf{r}_{\mathsf{out}}} \cdot \mathsf{l}_{\mathsf{in}}$ and $\mathsf{tp}_{\mathsf{out}} + 2^{\mathsf{r}_{\mathsf{out}}} \cdot \mathsf{tp}_{\mathsf{in}}$, because the composed prover uses the inner proof system to generate a proof of proximity *for each choice of randomness of the outer proof system*. Thus, the outer randomness complexity puts a significant efficiency limitation on proof composition.

Avoiding the limitations by interacting. We prove an interactive analogue of proof composition that avoids the above limitations. Our theorem involves two proof systems: the outer proof system is a robust PCP $(P_{\mathsf{out}}, V_{\mathsf{out}})$ for a relation $\mathscr{R}$ (as before), while the inner proof system is a k-round IOP $(P_{\mathsf{in}}, V_{\mathsf{in}})$ for V_{out}'s relation;[3] the composed proof system is a $(\mathsf{k}+1)$-round IOP (P, V) for $\mathscr{R}$. The parameters of the composed proof system are exactly as in the non-interactive case, except that now the new proof length and prover running time are *much smaller*: $\mathsf{l}_{\mathsf{out}} + \mathsf{l}_{\mathsf{in}}$ and $\mathsf{tp}_{\mathsf{out}} + \mathsf{tp}_{\mathsf{in}}$, i.e., there is no multiplicative factor of $2^{\mathsf{r}_{\mathsf{out}}}$ in front of l_{in} and $\mathsf{tp}_{\mathsf{in}}$.

The crucial observation is that, after the prover sends the outer proof to the verifier, *soundness is not harmed if the verifier tells the prover his choice of outer randomness*; in this way, the prover does not have to invest work for all randomness choices but can simply invest work for the outer randomness that was actually chosen because he now knows this choice. Thus, after receiving the outer randomness, the prover and verifier use the inner proof system for proving proximity, to a satisfying assignment, of the oracle locations chosen by this outer randomness.

From sketch to proof. The above sketch glosses over many technical details that, in a proof, need to be addressed.

First, we treated the outer verifier as a unit, while instead it consists of two algorithms: a query algorithm that, given the instance $\mathbb{x}$, outputs a state σ and the query set I; and a decision algorithm that, given the state σ and answers to these queries, outputs a bit. The composed prover and verifier explicitly invoke only the query algorithm; the dependence on the decision algorithm is only implicit, via the invocation of the inner proof system on its induced relation. This requires additional bookkeeping in parameters, e.g., the instance size for the inner proof system is not $|\mathbb{x}|$ but $|\sigma|$, which is denoted $\mathsf{s}_{\mathsf{out}}(|\mathbb{x}|)$. These details also arise in the non-interactive case [BGH^{+}06].

Second, most treatments of proof composition obtain a composed verifier that, despite obtaining savings in query complexity, still runs in time that is at least the outer query complexity, because the outer query algorithm outputs the

[2] Alternative useful choices, which we do not explore, are described in [MR08, DH13, DHK15].

[3] We could also make the outer proof system interactive by considering a robust IOP but we do not make use of this additional degree of freedom.

query set I; such uses of proof composition offer no savings in verifier running time [PS94, HS00, GS06, BSVW03, BGH$^+$06, BS08]. In contrast, [BGH$^+$05] introduce the notion of *verifier specification*, which allows the query algorithm to implicitly specify I (e.g., via an algorithm), potentially running in time that is polylogarithmic in the query complexity; then, they prove a (non-interactive) composition theorem that leverages this notion to obtain a composed verifier whose running time may be much smaller than the outer verifier's. We have adopted [BGH$^+$05]'s definitions (see Section 2.2), and our interactive proof composition theorem takes [BGH$^+$05] as a starting point so that we too may benefit from savings in verifier running time. (The main difference of our theorem from [BGH$^+$05]'s is that we do not consider verifier specifications for the composed proof system.)

Finally, analogously to the non-interactive case [BGH$^+$06], if the outer PCP has proximity parameter δ_{out}, then the composed proof system is an IOP with proximity parameter $\delta(n) = \delta_{\mathsf{out}}(n)$; moreover, if the inner IOP has robustness parameter ρ_{in} then the composed proof system is an IOP with robustness parameter $\rho(n) = \rho_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n))$. Our theorem below considers only the first case, ignoring any robustness of the inner proof system since we do not make use of it.

We are now ready to state and prove our interactive composition theorem; we have highlighted in blue the parameters for proof length and prover running time, since these are the main differences from the parameters of a non-interactive composition theorem (such as that in [BGH$^+$05]).

Theorem 3.1 (Interactive Proof Composition — formal statement of Theorem 1.4)**.** *Suppose that the relation $\mathscr{R}$ satisfies the following two conditions with $\delta_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n)) \le \rho_{\mathsf{out}}(n)$:*

***(1)** there exists a robust PCPP system $(P_{\mathsf{out}}, V_{\mathsf{out}})$ that puts $\mathscr{R}$ in the complexity class*

$$\textbf{PCPP}\left[\begin{array}{ll}
\text{answer alphabet} & \mathsf{a} \\
\text{proof length} & \mathsf{l}_{\mathsf{out}} \\
\text{randomness} & \mathsf{r}_{\mathsf{out}} \\
\text{query complexity} & \mathsf{q}_{\mathsf{out}} \\
\text{state size} & \mathsf{s}_{\mathsf{out}} \\
\text{soundness error} & \varepsilon_{\mathsf{out}} \\
\text{proximity parameter} & \delta_{\mathsf{out}} \\
\text{robustness parameter} & \rho_{\mathsf{out}} \\
\text{prover time} & \mathsf{tp}_{\mathsf{out}} \\
\text{verifier query time} & \mathsf{tvq}_{\mathsf{out}} \\
\text{verifier decision time} & \mathsf{tvd}_{\mathsf{out}}
\end{array}\right]$$

and

***(2)** there exists an IOPP system $(P_{\mathsf{in}}, V_{\mathsf{in}})$ that puts* $\mathrm{Rel}(V_{\mathsf{out}})$ *in the complexity class*

$$\textbf{IOPP}\left[\begin{array}{ll}
\text{rounds} & \mathsf{k}_{\mathsf{in}} \\
\text{answer alphabet} & \mathsf{a} \\
\text{proof length} & \mathsf{l}_{\mathsf{in}} \\
\text{randomness} & \mathsf{r}_{\mathsf{in}} \\
\text{query complexity} & \mathsf{q}_{\mathsf{in}} \\
\text{soundness error} & \varepsilon_{\mathsf{in}} \\
\text{proximity parameter} & \delta_{\mathsf{in}} \\
\text{prover time} & \mathsf{tp}_{\mathsf{in}} \\
\text{verifier time} & \mathsf{tv}_{\mathsf{in}}
\end{array}\right]$$

Then there exists an IOPP system (P, V) that puts $\mathscr{R}$ in the complexity class

$$\textbf{IOPP}\left[\begin{array}{llcl}
\text{rounds} & \mathsf{k}(n) & = & 1 + \mathsf{k}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n)) \\
\text{answer alphabet} & \mathsf{a}(n) & & \\
\text{proof length} & \mathsf{l}(n) & = & \mathsf{l}_{\mathsf{out}}(n) + \mathsf{l}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n)) \\
\text{randomness} & \mathsf{r}(n) & = & \mathsf{r}_{\mathsf{out}}(n) + \mathsf{r}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n)) \\
\text{query complexity} & \mathsf{q}(n) & = & \mathsf{q}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n)) \\
\text{soundness error} & \varepsilon(n) & = & \varepsilon_{\mathsf{out}}(n) + (1 - \varepsilon_{\mathsf{out}}(n)) \cdot \varepsilon_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n)) \\
\text{proximity parameter} & \delta(n) & = & \delta_{\mathsf{out}}(n) \\
\text{prover time} & \mathsf{tp}(n) & = & \mathsf{tp}_{\mathsf{out}}(n) + \mathsf{tvq}_{\mathsf{out}}(n) + \mathsf{tp}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n)) \\
\text{verifier time} & \mathsf{tv}(n) & = & \mathsf{tvq}_{\mathsf{out}}(n) + \mathsf{tv}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n))
\end{array}\right].$$

Moreover, if V_{in}'s queries are non-adaptive so are V's queries; also, if V_{in} is public coin so is V.

Proof. We construct the IOPP system (P, V), then analyze its completeness, its soundness, and efficiency parameters.

Construction. Construct the IOPP system (P, V) as follows. Let $(\mathbbm{x}, \mathbbm{w})$ be an instance-witness pair in the relation $\mathscr{R}$ with $n := |\mathbbm{x}|$; the prover P receives $(\mathbbm{x}, \mathbbm{w})$ as input, while the verifier V receives $\mathbbm{x}$ as input and $\mathbbm{w}$ as oracle. In the first round: V sends an empty message to P; then P computes the proximity proof $\pi_{\mathsf{out}} \leftarrow P_{\mathsf{out}}(\mathbbm{x}, \mathbbm{w})$ and sends π_{out} to V. Next, V samples randomness r_{out} for $V_{\mathsf{out}}^{\mathsf{Q}}(\mathbbm{x})$ and sends r_{out} to P; both P and V compute $(\sigma, I) \leftarrow V_{\mathsf{out}}^{\mathsf{Q}}(\mathbbm{x}; r_{\mathsf{out}})$; in parallel, P and V engage in an interactive oracle protocol attesting to the proximity of $(\mathbbm{w} \| \pi_{\mathsf{out}})|_I$ to $\mathrm{Rel}(V)|_\sigma$, by invoking $P_{\mathsf{in}}(\sigma, (\mathbbm{w} \| \pi_{\mathsf{out}})|_I)$ and $V_{\mathsf{in}}^{(\mathbbm{w} \| \pi_{\mathsf{out}})|_I}(\sigma)$, respectively. The verifier V accepts if and only if V_{in} does.

Completeness. Completeness of (P, V) follows from that of $(P_{\mathsf{out}}, V_{\mathsf{out}})$ and $(P_{\mathsf{in}}, V_{\mathsf{in}})$. Namely, for every instance-witness pair $(\mathbb{x}, \mathbb{w})$ in the relation $\mathscr{R}$, $(\sigma, (\mathbb{w}\|\pi_{\mathsf{out}})|_I) \in \mathrm{Rel}(V_{\mathsf{out}})$ with probability 1, where $\pi_{\mathsf{out}} \leftarrow P_{\mathsf{out}}(\mathbb{x}, \mathbb{w})$ and $(\sigma, I) \leftarrow V_{\mathsf{out}}^{\mathrm{Q}}(\mathbb{x})$. Hence, when interacting with $P_{\mathsf{in}}(\sigma, (\mathbb{w}\|\pi_{\mathsf{out}})|_I)$, $V_{\mathsf{in}}^{(\mathbb{w}\|\pi_{\mathsf{out}})|_I}(\sigma)$ accepts with probability 1.

Soundness. Soundness of (P, V) follows from that of $(P_{\mathsf{out}}, V_{\mathsf{out}})$ and $(P_{\mathsf{in}}, V_{\mathsf{in}})$, as we now explain. Consider any instance-witness pair $(\mathbb{x}, \mathbb{w})$ and with $\Delta(\mathbb{w}, \mathscr{R}|_{\mathbb{x}}) \geq \delta_{\mathsf{out}}(n)$, and an unbounded malicious prover $\tilde{P}$. Letting $\tilde{\pi}_{\mathsf{out}}$ be $\tilde{P}$'s first message, we know that $\Delta((\mathbb{w}\|\tilde{\pi}_{\mathsf{out}})|_I, \mathrm{Rel}(V_{\mathsf{out}})|_\sigma) \leq \rho_{\mathsf{out}}(n)$ with probability at most $\varepsilon_{\mathsf{out}}(n)$ over r_{out}, where $(\sigma, I) \leftarrow V^{\mathrm{Q}}(\mathbb{x}; r_{\mathsf{out}})$. Call r_{out} 'bad' if the distance in the previous sentence is at most $\rho_{\mathsf{out}}(n)$; else call r_{out} 'good'. For any bad r_{out}, we know only that $V_{\mathsf{in}}^{(\mathbb{w}\|\tilde{\pi}_{\mathsf{out}})|_I}(\sigma)$ accepts with probability at most 1; for any good r_{out}, we know that $V_{\mathsf{in}}^{(\mathbb{w}\|\tilde{\pi}_{\mathsf{out}})|_I}(\sigma)$ accepts with probability at most $\varepsilon_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n))$, because of the hypothesis that $\delta_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n)) \leq \rho_{\mathsf{out}}(n)$. Overall, we deduce that V accepts with probability at most $\varepsilon_{\mathsf{out}}(n) \cdot 1 + (1 - \varepsilon_{\mathsf{out}}(n)) \cdot \varepsilon_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n))$.

Efficiency parameters. The constructed IOPP system has $\mathsf{k}(n) := 1 + \mathsf{k}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n))$ rounds, because in the first round the verifier sends an empty message and the prover replies with a proximity proof; in the remaining $\mathsf{k}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n))$ rounds, the prover and verifier run $(P_{\mathsf{in}}, V_{\mathsf{in}})$ on σ (and note that the random string r_{out} can be sent in parallel to the first verifier message of $(P_{\mathsf{in}}, V_{\mathsf{in}})$). The proof length is $\mathsf{l}(n) := \mathsf{l}_{\mathsf{out}}(n) + \mathsf{l}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n))$ because the first term accounts for the proximity proof and the second for the prover messages in the subsequent interactive oracle protocol. The randomness complexity is $\mathsf{r}(n) := \mathsf{r}_{\mathsf{out}}(n) + \mathsf{r}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n))$ because the first term accounts for running $V_{\mathsf{out}}^{\mathrm{Q}}$ and the second term for running V_{in}. The query complexity is $\mathsf{q}(n) := \mathsf{q}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n))$ because verifier queries are due only to V_{in}. The prover running time is $\mathsf{tp}(n) := \mathsf{tp}_{\mathsf{out}}(n) + \mathsf{tvq}_{\mathsf{out}}(n) + \mathsf{tp}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n))$ because the prover runs P_{out} and $V_{\mathsf{out}}^{\mathrm{Q}}$ on $\mathbb{x}$ and P_{in} on σ. The verifier running time is $\mathsf{tv}(n) := \mathsf{tvq}_{\mathsf{out}}(n) + \mathsf{tv}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(n))$. Finally, the construction clearly preserves non-adaptivity of queries and public coins, if present. $\square$

Remark 3.2. The statement of Theorem 3.1 asserts that the outer and inner proof systems use the same alphabet for the prover messages. This is not essential: if the two alphabets differ, the same argument as above goes through, and the composed proof system will have some messages over one alphabet and other messages over the other alphabet.

4 Sublinear sumcheck

We show how to use IOPs to obtain a sumcheck protocol in which the verifier complexity is *sublinear* in the individual degree of the polynomial being verified. More generally, we phrase our result for tensor product codes [Wol65, WE63, Tan81], and the verifier complexity is then sublinear in the block length of one copy of the code. Later on, we leverage this result to obtain interactive oracle proofs for circuit satisfiability (see Section 7). We now review the sumcheck protocol, discuss intuition behind our result, and then formally state and prove it.

The sumcheck protocol. The sumcheck protocol [LFKN92, Sha92] is a fundamental building block of numerous results in complexity theory and cryptography. The protocol consists of an interactive proof for the claim "$\sum_{\vec{\alpha} \in H^m} w(\vec{\alpha}) = 0$", where w is the evaluation on $\mathbb{F}^m$ of an m-variate polynomial of individual degree d and H is a subset of $\mathbb{F}$. The prover receives H and w as input, while the verifier receives H as input and w as an oracle. The sumcheck protocol has soundness error $1 - (1 - \frac{d}{|\mathbb{F}|})^m$, the prover runs in time $\mathrm{poly}(|\mathbb{F}|^m)$, the verifier runs in time $\mathrm{poly}(|\mathbb{F}| + m)$, the communication complexity is $\mathrm{poly}(|\mathbb{F}| + m)$, and the number of rounds is m; moreover, the protocol is public coin and the verifier queries w only at one random index.[4]

Limitations, and how to avoid them. In each of the m rounds, the prover sends to the verifier the evaluation on $\mathbb{F}$ of a univariate polynomial of degree d or, alternatively, the prover sends the coefficients of this polynomial; then the verifier checks that the sum of this polynomial over H equals a certain value determined in the previous round. In particular, the verifier reads $\Omega(md)$ bits and its running time is also $\Omega(md)$. We show that the verifier complexity can be *sublinear* in d, if the prover and verifier engage in an interactive oracle proof (rather than an interactive proof).

Recall that, in an IOP, the verifier has oracle access to the prover's messages, so the verifier may read as many locations of these as are sufficient to perform the necessary checks. The intuition to "go sublinear" is simple: instead of performing these checks explicitly, the verifier relies on proximity testers for doing them. Thus, in each of the m rounds, the prover sends to the verifier two oracles: the evaluation on $\mathbb{F}$ of a univariate polynomial of degree d, and a proximity proof attesting that this evaluation has degree d and has the appropriate sum over H. The use of proximity proofs somewhat complicates the soundness analysis (e.g., the verifier only sees noisy codewords) but the backbone of the proof follows that of the standard sumcheck protocol; overall, this high level intuition can be turned into a proof.

More generally, instead of sending (non-interactive) proximity proofs, the prover may interact with the verifier in an interactive oracle sub-proof of proximity for the appropriate codewords.

Beyond Reed–Muller. The sumcheck protocol can be phrased in a more general setting [Mei13]: an interactive proof for the claim "$\sum_{\vec{\alpha} \in H^m} w(\vec{\alpha}) = 0$" where w is a codeword in the tensor product code $C^{\otimes m}$, for a given linear code C with domain D and alphabet $\mathbb{F}$, and H is a subset of D. Low-degree polynomials are a special case: the Reed–Muller code is a tensor of the Reed–Solomon code. Conveniently, the parameters in the more general case are analogous to those in the special case: the soundness error becomes $1 - \tau^m$ where τ is C's relative distance ($\tau = 1 - \frac{d}{|\mathbb{F}|}$ for the Reed–Solomon code), and C's block length ℓ replaces the field size $|\mathbb{F}|$ in the running times and communication complexity. In particular, the verifier reads $\Omega(m\ell)$ bits and its running time is also $\Omega(m\ell)$. Below, we phrase our sublinear sumcheck result in the language of tensor product codes not only because of the greater generality but also because we invoke this result on tensors of algebraic-geometry codes (see Section 7), which are not Reed–Muller codes.

We state our theorem as a reduction: given a PCP of proximity $(P_{\mathrm{SC}}, V_{\mathrm{SC}})$ for subcodes of the form $C|_{H,\gamma} := \{w \in C \text{ s.t. } \sum_{\alpha \in H} w(\alpha) = \gamma\}$, we construct an IOP of proximity (P, V) for sumchecks over H^m for $C^{\otimes m}$. The complexity of the PCPP verifier V_{SC} determines the complexity of the resulting IOPP verifier V; e.g., if the former is sublinear in C's block length ℓ, so is the latter. In fact, we find it more natural to state the theorem without assuming that w is promised to be a codeword in $C^{\otimes m}$, so the reduction also takes as input a PCP of proximity $(P_{\otimes}, V_{\otimes})$ for $C^{\otimes m}$ that is invoked on w.[5] More generally, both PCPPs can in fact be IOPPs, and we state our theorem for this more general case.

As an example of an instantiation, in the case of low-degree polynomials, one can invoke the theorem with a low-degree test [BFL90, BFLS91, ALM$^+$98, AS03] for the tensor product, and proximity proofs of [BS08] for the subcodes of the Reed–Solomon code; this yields sumcheck for low-degree polynomials where the verifier complexity is

[4] The sumcheck protocol's standard analysis yields a soundness error of $\frac{md}{|\mathbb{F}|}$, but a more careful analysis yields the smaller error of $1 - (1 - \frac{d}{|\mathbb{F}|})^m$.

[5] Indeed, since we think of ℓ as large, the setting in which the verifier knows all of w (as in [LFKN92, Sha92]) does not apply in general; that said, one can easily specialize the theorem's statement and proof to the cases where the promise "$w \in C^{\otimes m}$" holds.

polylogarithmic in the individual degree. At the other extreme, one can invoke the theorem with the "trivial" proximity testers that read a codeword in full; this collapses our construction to the standard non-sublinear sumcheck protocol.

Below, we use the following notation. Let $\mathscr{C} = \{C_n\}_{n\in\mathbb{N}}$ be a code family, $\mathscr{H} = \{H_n\}_{n\in\mathbb{N}}$ a family where each H_n is a subset of C_n's domain, and $m\colon \mathbb{N} \to \mathbb{N}$. We denote by:

- $\mathrm{Rel}(\mathscr{C}, m)$ the relation of instance-witness pairs (n, w) s.t. $n \in \mathbb{N}$ and $w \in C_n^{\otimes m(n)}$;
- $\mathrm{Rel}(\mathscr{C}, m, \mathscr{H})$ the relation of instance-witness pairs $((n,\gamma), w)$ s.t. $n \in \mathbb{N}$, $w \in C_n^{\otimes m(n)}$, and $\sum_{\vec{\alpha}\in H_n^{m(n)}} w(\vec{\alpha}) = \gamma$.

With this notation, our theorem can be viewed as a reduction from proximity testing to $\mathrm{Rel}(\mathscr{C}, m, \mathscr{H})$ to proximity testing to $\mathrm{Rel}(\mathscr{C}, m)$ and $\mathrm{Rel}(\mathscr{C}, 1, \mathscr{H})$. (Note that both a particular code C_n and subset H_n need not be represented explicitly, via a generator matrix and a set of indices; they may be represented in a more succinct way, when possible.) When reading the theorem statement, it is helpful to keep in mind that the constructed IOP relies on a single invocation of proximity testing to $\mathrm{Rel}(\mathscr{C}, m)$, and m invocations of proximity testing to $\mathrm{Rel}(\mathscr{C}, 1, \mathscr{H})$; it is also helpful to keep in mind that the term $1 - \tau^m$ in the soundness error (highlighted in blue) is inherited from the standard sumcheck protocol while the other terms are due to proximity testing.

Theorem 4.1 (Sublinear Sumcheck — formal statement of Theorem 1.5)**.** *Suppose that:*

- *$\mathscr{C} = \{C_n\}_{n\in\mathbb{N}}$ is a linear code family with relative distance $\tau(\cdot)$, block length $\ell(\cdot)$ and alphabet $\mathbb{F}(\cdot)$;*
- *$\mathscr{H} = \{H_n\}_{n\in\mathbb{N}}$ is a family where each H_n is a subset of C_n's domain;*
- *$m\colon \mathbb{N} \to \mathbb{N}$ is a (polynomial-time computable) function.*

Suppose further that the following two conditions hold:

***(1)** there exists an IOPP system $(P_\otimes, V_\otimes)$ that puts $\mathrm{Rel}(\mathscr{C}, m)$ in the complexity class*

$$\textbf{IOPP}\left[\begin{array}{ll} \text{rounds} & \mathsf{k}_\otimes \\ \text{answer alphabet} & \mathsf{a} \\ \text{proof length} & \mathsf{l}_\otimes \\ \text{randomness} & \mathsf{r}_\otimes \\ \text{query complexity} & \mathsf{q}_\otimes \\ \text{soundness error} & \varepsilon_\otimes \\ \text{proximity parameter} & \delta_\otimes \\ \text{prover time} & \mathsf{tp}_\otimes \\ \text{verifier time} & \mathsf{tv}_\otimes \end{array}\right]$$

and

***(2)** there exists an IOPP system $(P_{\mathrm{SC}}, V_{\mathrm{SC}})$ that puts $\mathrm{Rel}(\mathscr{C}, 1, \mathscr{H})$ in the complexity class*

$$\textbf{IOPP}\left[\begin{array}{ll} \text{rounds} & \mathsf{k}_{\mathrm{SC}} \\ \text{answer alphabet} & \mathsf{a} \\ \text{proof length} & \mathsf{l}_{\mathrm{SC}} \\ \text{randomness} & \mathsf{r}_{\mathrm{SC}} \\ \text{query complexity} & \mathsf{q}_{\mathrm{SC}} \\ \text{soundness error} & \varepsilon_{\mathrm{SC}} \\ \text{proximity parameter} & \delta_{\mathrm{SC}} \\ \text{prover time} & \mathsf{tp}_{\mathrm{SC}} \\ \text{verifier time} & \mathsf{tv}_{\mathrm{SC}} \end{array}\right]$$

Then there exists a public-coin IOPP system (P, V) that puts $\mathrm{Rel}(\mathscr{C}, m, \mathscr{H})$ in the complexity class

$$\textbf{IOPP}\left[\begin{array}{llcl} \text{rounds} & \mathsf{k}(n) & = & \max\{\mathsf{k}_\otimes(n), m(n)\cdot \mathsf{k}_{\mathrm{SC}}(n)\} \\ \text{answer alphabet} & \mathsf{a} & & \\ \text{proof length} & \mathsf{l}(n) & = & \mathsf{l}_\otimes(n) + m(n)\cdot \mathsf{l}_{\mathrm{SC}}(n) + m(n)\cdot \ell(n)\cdot \log|\mathbb{F}(n)| \\ \text{randomness} & \mathsf{r}(n) & = & \mathsf{r}_\otimes(n) + m(n)\cdot \mathsf{r}_{\mathrm{SC}}(n) + m(n)\cdot \log \ell(n) \\ \text{query complexity} & \mathsf{q}(n) & = & \mathsf{q}_\otimes(n) + m(n)\cdot \mathsf{q}_{\mathrm{SC}}(n) + m(n) + 1 \\ \text{soundness error} & \varepsilon(n) & & \\ \text{proximity parameter} & \delta(n) & = & \delta_\otimes(n) \\ \text{prover time} & \mathsf{tp}(n) & = & \mathsf{tp}_\otimes(n) + m(n)\cdot \mathsf{tp}_{\mathrm{SC}}(n) + m(n)\cdot \ell(n)^{m(n)} \\ \text{verifier time} & \mathsf{tv}(n) & = & \mathsf{tv}_\otimes(n) + m(n)\cdot \mathsf{tv}_{\mathrm{SC}}(n) + O(m(n)) \end{array}\right]$$

where the soundness error $\varepsilon(n)$ is

$$\varepsilon(n) = \max\Big\{\varepsilon_\otimes(n)\,,\ \varepsilon_{\mathrm{SC}}(n)\,,\ 1 - \tau(n)^{m(n)} + \Big(1 - (1-\delta_\otimes(n))\cdot(1-\delta_{\mathrm{SC}}(n))^{m(n)}\Big)\Big\}$$

Moreover, if $V_\otimes$'s and V_{SC}'s queries are non-adaptive, then so are V's.

Proof. We first prove the theorem in the case where both of the given IOPPs are in fact PCPPs (in particular, $\mathsf{k}_\otimes = \mathsf{k}_{\mathrm{SC}} = 1$); at the end of the proof we explain the straightforward extension to the general case. So now, for the case of PCPPs, we construct the IOPP system (P, V), then analyze its completeness, its soundness, and efficiency parameters.

Construction. Construct the IOPP system (P, V) as follows. Let $(\mathbb{x}, \mathbb{w}) = \big((n, \gamma_0), w\big)$ be an instance-witness pair in the relation $\mathrm{Rel}(\mathscr{C}, m, \mathscr{H})$; the prover P receives the instance and witness as input, while the verifier V receives the instance as input and the witness as an oracle.

- In the first round, the verifier V sends an empty message to P; next, the prover P proceeds as follows:
  - compute the proximity proof $\pi_0 \leftarrow P_{\otimes}(n, w)$, which attests that w is in the tensor product code $C_n^{\otimes m(n)}$;
  - compute the codeword $w_1 \colon D(n) \to \mathbb{F}(n)$ defined by $w_1(x) := \sum_{a_2,\dots,a_{m(n)} \in H_n} w(x, a_2, \dots, a_{m(n)})$;
  - compute the proximity proof $\pi_1 \leftarrow P_{\mathrm{SC}}\big((n, \gamma_0), w_1\big)$, which attests that w_1 is in the subcode $C_n|_{H_n, \gamma_0}$;
  - send the proof string (π_0, w_1, π_1) to the verifier V.
- For $i = 2, \dots, m(n)$, in the i-th round, the verifier V draws $r_{i-1} \in D(n)$ uniformly and independently at random, and sends r_{i-1} to P; next, the prover P proceeds as follows:
  - compute the codeword $w_i \colon D(n) \to \mathbb{F}(n)$ defined by $w_i(x) := \sum_{a_{i+1},\dots,a_{m(n)} \in H_n} w(r_1, \dots, r_{i-1}, x, a_{i+1}, \dots, a_{m(n)})$;
  - set $\gamma_{i-1} := w_{i-1}(r_{i-1})$;
  - compute the proximity proof $\pi_i \leftarrow P_{\mathrm{SC}}\big((n, \gamma_{i-1}), w_i\big)$, which attests that w_i is in the subcode $C_n|_{H_n, \gamma_{i-1}}$;
  - send (w_i, π_i) to V.
- After the $m(n)$-th round, the verifier V proceeds as follows:
  - set $\gamma_i := w_i(r_i)$ for every $i \in \{1, \dots, m(n)\}$;
  - check that $V_{\otimes}^{w, \pi_0}(n)$ accepts;
  - check that $V_{\mathrm{SC}}^{w_i, \pi_i}\big((n, \gamma_{i-1})\big)$ accepts for every $i \in \{1, \dots, m(n)\}$;
  - check that $w(r_1, \dots, r_{m(n)}) = \gamma_{m(n)}$.

Completeness. Completeness of (P, V) follows from that of $(P_{\otimes}, V_{\otimes})$ and $(P_{\mathrm{SC}}, V_{\mathrm{SC}})$, and the fact that the partial sums belong to the appropriate subcodes. Namely, for every instance-witness pair $(\mathbb{x}, \mathbb{w}) = \big((n, \gamma_0), w\big)$ in the relation $\mathrm{Rel}(\mathscr{C}, m, \mathscr{H})$ it holds that:

- w is in the tensor product code $C_n^{\otimes m(n)}$, so that $V_{\otimes}^{w, \pi_0}(n)$ accepts with probability 1, and its sum over $H_n^{m(n)}$ is γ_0;
- for every $i \in \{1, \dots, m(n)\}$ and $r_1, \dots, r_{i-1} \in D(n)$, the codeword w_i, which depends on $r_1, \dots, r_{i-1}$, is in the code C_n and its sum over H_n is $\gamma_{i-1} = w_{i-1}(r_{i-1})$ so that $V_{\mathrm{SC}}^{w_i, \pi_i}\big((n, \gamma_{i-1})\big)$ accepts with probability 1;
- $w(r_1, \dots, r_{m(n)}) = \gamma_{m(n)} = w_{m(n)}(r_{m(n)})$.

We conclude that P makes V accept with probability 1.

Soundness. Consider any instance-witness pair $(\mathbb{x}, \mathbb{w}) = \big((n, \gamma_0), w\big)$ and unbounded malicious prover $\tilde{P}$. Suppose for now that w does not sum to γ_0 on $H_n^{m(n)}$ but is in the tensor product code $C_n^{\otimes m(n)}$ and, moreover, each w_i sent by $\tilde{P}$ is in the subcode $C_n|_{H_n, \gamma_{i-1}}$. In this case, the standard soundness analysis of the sumcheck protocol (when extended to tensor product codes) shows that the probability that the verifier accepts is at most $1 - \tau(n)^{m(n)}$, where $\tau(n)$ is the relative distance of C_n. However, the verifier does not explicitly check if each w_i is in $C_n|_{H_n, \gamma_{i-1}}$ but, instead, relies on the PCPP verifier V_{SC} to test proximity to this code; also, the verifier is not guaranteed that w is in $C_n^{\otimes m(n)}$ but, instead, relies on the PCPP verifier $V_{\otimes}$ to test proximity to this code. Overall, this means that the verifier incurs additional soundness errors due to the proximity testing and, hence, accessing noisy codewords. We now describe how to account for these.

Suppose that w is $\delta_{\otimes}(n)$-far from any codeword in $C_n^{\otimes m(n)}$ that sums to γ_0 on $H_n^{m(n)}$. We distinguish among the following cases.

- *Case 1: w is $\delta_{\otimes}(n)$-far from the tensor product code $C^{\otimes m}$.* In this case, the PCPP verifier $V_{\otimes}^{w, \pi_0}(n)$ accepts with probability at most $\varepsilon_{\otimes}(n)$.
- *Case 2: there exists i such that w_i is $\delta_{\mathrm{SC}}(n)$-far from the subcode $C_n|_{H_n, \gamma_{i-1}}$.* In this case, the PCPP verifier $V_{\mathrm{SC}}^{w_i, \pi_i}\big((n, \gamma_{i-1})\big)$ accepts with probability at most $\varepsilon_{\mathrm{SC}}(n)$.
- *Case 3: the above two cases do not happen.* In this case, let $\hat{w}$ be the unique codeword in $C_n^{\otimes m(n)}$ closest to w and, for each i, let $\hat{w}_i$ be the unique codeword in $C_n|_{H_n, \gamma_{i-1}}$ that is closest to w_i; recall that proximity parameters

are less than the unique-decoding radius (see Section 2.2) so these unique codewords exist. Note that $\hat{w}$ cannot sum to γ_0 on $H_n^{m(n)}$ (because we have assumed that w is $\delta_\otimes(n)$-far from any codeword in $C_n^{\otimes m(n)}$ that sums to γ_0 on $H_n^{m(n)}$). At this point we apply the standard analysis of the sumcheck, but relative to the codewords $\hat{w}$ and $\hat{w}_1, \dots, \hat{w}_{m(n)}$: if the verifier has access to these codewords, then the verifier accepts with probability at most $1 - \tau(n)^{m(n)}$. However the verifier only has access to functions that are close to these, which incurs an additional soundness error of $1 - (1 - \delta_\otimes(n)) \cdot (1 - \delta_{\mathrm{SC}}(n))^{m(n)}$ because the verifier queries w and $w_1, \dots, w_{m(n)}$ each at at a uniformly and independently random location.

We deduce that the verifier accepts with probability that is at most the maximum of the acceptance probability across the three cases, namely,

$$\max\left\{\varepsilon_\otimes(n)\,,\ \varepsilon_{\mathrm{SC}}(n)\,,\ 1 - \tau(n)^{m(n)} + \Big(1 - (1 - \delta_\otimes(n)) \cdot (1 - \delta_{\mathrm{SC}}(n))^{m(n)}\Big)\right\} \ .$$

For any particular choice of code C and explicit bounds on the soundness errors and proximity parameters, the final soundness error can be further improved by "balancing" the above three cases; we do not do so for the general case.

Efficiency parameters. The constructed IOPP system has $\mathsf{k}(n) := m(n)$ rounds. (Recall that for now we are assuming that both of the given IOPPs are in fact PCPPs; see below for the general case.) The proof length is $\mathsf{l}(n) := \mathsf{l}_\otimes(n) + m(n) \cdot \mathsf{l}_{\mathrm{SC}}(n) + m(n) \cdot \ell(n)$ because the prover sends the proximity proof π_0 output by $P_\otimes$, $m(n)$ proximity proofs output by P_{SC}, and $m(n)$ codewords with block length $\ell(n)$. The randomness complexity is $\mathsf{r}(n) := \mathsf{r}_\otimes(n) + m(n) \cdot \mathsf{r}_{\mathrm{SC}}(n) + m(n) \cdot \log \ell(n)$ because the verifier runs $V_\otimes$, runs V_{SC} for $m(n)$ times, and samples $m(n)$ elements in $D(n)$. The query complexity is $\mathsf{q}(n) := \mathsf{q}_\otimes(n) + m(n) \cdot \mathsf{q}_\otimes(n) + m(n) + 1$ because the verifier runs $V_\otimes$, runs V_{SC} for $m(n)$ times, and makes $m(n) + 1$ additional queries (one to each of $w_1, \dots, w_{m(n)}, w$). The prover running time is $\mathsf{tp}(n) := \mathsf{tp}_\otimes(n) + m(n) \cdot \mathsf{tp}_{\mathrm{SC}}(n) + m(n) \cdot \ell(n)^{m(n)}$ because the prover runs $P_\otimes$, runs P_{SC} for $m(n)$ times, and computes $m(n)$ partial sums over domains of size at most $\ell(n)^{m(n)}$. The verifier running time is $\mathsf{tv}(n) := \mathsf{tv}_\otimes(n) + m(n) \cdot \mathsf{tv}_{\mathrm{SC}}(n) + O(m(n))$ because the verifier runs $V_\otimes$, runs V_{SC} for $m(n)$ times, and performs $O(m(n))$ additional work. Finally, the protocol is clearly public coins.

From PCPPs to IOPPs. The extension from PCPPs to IOPPs is straightforward: whenever the prover would have sent to the verifier a (non-interactive) proof of proximity, the prover now interacts with the verifier in an interactive oracle proof of proximity. Thus, testing proximity of w to $C_n^{\otimes m(n)}$ takes $\mathsf{k}_\otimes(n)$ rounds, while testing proximity of each of w_i to $C_n|_{H_n, \gamma_{i-1}}$ takes $\mathsf{k}_{\mathrm{SC}}(n)$ rounds. The first can be done in parallel to the second, so the overall number of rounds is now $\max\{\mathsf{k}_\otimes(n), m(n) \cdot \mathsf{k}_{\mathrm{SC}}(n)\}$. The rest of the proof, mutatis mutandis, is unaffected. □

5 Short IOPs of proximity with constant query complexity

We use interactive proof composition (see Section 3) to obtain results on proximity testing for notable classes of linear codes: we obtain IOPs of proximity with proof length and query complexity that are not known to be achievable by any PCP of proximity. We consider the following two classes of codes.

- **Additive Reed–Solomon codes** (Section 5.1)

  We show that additive Reed–Solomon codes over binary fields have *IOPs of proximity with linear proof length and constant query complexity*; moreover, 2 rounds of interaction and public coins suffice. In contrast, for these codes, we only know how to construct PCPs of proximity with *quasilinear* proof length and constant query complexity [BS08, Din07, Mie09].

- **Tensor product codes** (Section 5.2)

  We show that tensor product codes have *IOPs of proximity with sublinear proof length and constant query complexity*; moreover, 1 round of interaction and public coins suffice. In contrast, for these codes, we only know how to construct local testers with sublinear query complexity [BS06, Vid15], or PCPs of proximity with superlinear proof length and constant query complexity [Mie09].

Above, all statements are relative to a constant soundness error (with a necessary linear dependence on the proximity parameter), and involve a polylogarithmic-time verifier. We now describe, state, and prove the above results.

5.1 For additive Reed–Solomon codes

We show that additive Reed–Solomon codes over binary fields have *linear*-size IOPs of proximity with constant query complexity; moreover, 2 rounds of interaction and public coins suffice. The construction follows from one invocation of our interactive composition theorem with [BS08]'s robust PCPs of proximity for additive Reed–Solomon codes as the outer proof system, and [Mie09]'s PCPs of proximity for nondeterministic languages as the inner proof system. See Section 2.6.1 and Section 2.6.3 for these two components; also, see Section 2.6.1 for the definition of the relation $\mathrm{Rel}(\mathbb{F}, \varrho)$, corresponding to the class of additive Reed–Solomon codes, over a binary field $\mathbb{F}$, with fractional degree ϱ.

Informally, [BS08]'s robust PCPs of proximity reduce proximity testing for $\mathrm{Rel}(\mathbb{F}, \varrho)$ from dimension λ to dimension $\lambda/2 + O(1)$; this also reduces the query complexity from 2^{λ} to $O(2^{\lambda/2})$. Thus, in our 2-round IOP, in the first round the prover sends a [BS08]-type PCP of proximity for the function over a domain of dimension λ and, after receiving the randomness from the verifier, in the second round the prover sends a [Mie09]-type PCP of proximity for the statement that [BS08]'s verifier accepts. Since this statement lies in $\mathbf{NTIME}(\tilde{O}(2^{\lambda/2}))$ and this latter PCP of proximity is quasilinear in the decider running time, we obtain the desired result.

Below, we state the theorem. After the theorem, we also give a weaker theorem that forgoes the use of a "heavy" tool such as [Mie09], incurring a larger round complexity and soundness error but with better concrete constants.

Theorem 5.1 (formal statement of Theorem 1.2)**.** *For every $\varrho > 0$ there exists $c > 0$ such that the following holds for every binary field $\mathbb{F}$. Define $\mathscr{R} := \mathrm{Rel}(\mathbb{F}, \varrho)$ and note that $\delta_{\mathscr{R}} = 1 - \varrho$. For every $\delta \in (0, \frac{1}{2}\delta_{\mathscr{R}})$ there exists a public-coin IOPP system (P, V) that puts $\mathscr{R}$ in the complexity class*

$$\mathbf{IOPP}\left[\begin{array}{lrcl} \text{rounds} & \mathsf{k}(\lambda) & = & 2 \\ \text{answer alphabet} & \mathsf{a}(\lambda) & = & \mathbb{F}_2 \\ \text{proof length} & \mathsf{l}(\lambda) & = & c \cdot 2^{\lambda} \cdot \log|\mathbb{F}| \\ \text{randomness} & \mathsf{r}(\lambda) & = & c \cdot \lambda \\ \text{query complexity} & \mathsf{q}(\lambda) & = & c \\ \text{soundness error} & \varepsilon(\lambda) & = & 1/2 \\ \text{proximity parameter} & \delta(\lambda) & = & \delta \\ \text{prover time} & \mathsf{tp}(\lambda) & = & \lambda^{c} \cdot 2^{\lambda} \\ \text{verifier time} & \mathsf{tv}(\lambda) & = & \lambda^{c} \end{array}\right].$$

Proof. We invoke the interactive proof composition theorem (Theorem 3.1) as follows.

- The PCPP system for $\mathscr{R} = \mathrm{Rel}(\mathbb{F}, \varrho)$ of Theorem 2.6 as the "outer" proof system.

  We invoke the theorem with the same fractional degree ϱ as in this proof, which gives us $\alpha \in (0, 1)$ and $a > 0$ such that for every binary field $\mathbb{F}_{\mathsf{out}}$ and $\delta_{\mathsf{out}} \in (0, \frac{1}{2}\delta_{\mathscr{R}})$ there exists a robust PCPP system $(P_{\mathsf{out}}, V_{\mathsf{out}})$ for $\mathrm{Rel}(\mathbb{F}, \varrho)$ with the parameters described in the theorem statement. In this proof, we choose $\mathbb{F}_{\mathsf{out}} := \mathbb{F}$ and $\delta_{\mathsf{out}} := \delta$.

- The PCPP system for nondeterministic languages of Theorem 2.9 as the "inner" one.

  The relation that we choose is $\mathscr{R}_{\mathsf{in}} := \mathrm{Rel}(V_{\mathsf{out}})$; hence, $\delta_{\mathscr{R}_{\mathsf{in}}} = \delta_{\mathscr{R}}$. Because the state size of the outer proof system is $\mathsf{s}_{\mathsf{out}}(\lambda) = \lambda/2 + a$, we deduce that we can decide if $(\sigma, \omega) \in \mathscr{R}_{\mathsf{in}}$ in $T := \tilde{O}(2^{\lambda/2+a})$ time. We thus get that for every $\delta_{\mathsf{in}} \in (0, \frac{1}{2}\delta_{\mathscr{R}_{\mathsf{in}}})$ and $\varepsilon_{\mathsf{in}} > 0$ there exist $q_{\mathsf{in}} > 0$ and a PCPP system $(P_{\mathsf{in}}, V_{\mathsf{in}})$ for $\mathscr{R}_{\mathsf{in}}$ with the parameters described in the theorem statement. In this proof, we choose $\delta_{\mathsf{in}} := \rho_{\mathsf{out}}(\lambda) = \alpha \cdot \delta_{\mathsf{out}}$ and $\varepsilon_{\mathsf{in}} := \frac{1}{100}$.

This composition gives us an IOPP system (P, V) that puts $\mathscr{R}$ as a subset of the complexity class

$$
\mathbf{IOPP}\left[\begin{array}{llcl}
\text{rounds} & \mathsf{k}(\lambda) & = & 1 + \mathsf{k}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(\lambda)) \\
 & & = & 1 + \mathsf{k}_{\mathsf{in}}(\lambda/2 + a) \\
 & & = & 1 + 1 \\[1ex]
\text{answer alphabet} & \mathsf{a}(\lambda) & = & \mathbb{F}_2 \\[1ex]
\text{proof length} & \mathsf{l}(\lambda) & = & \mathsf{l}_{\mathsf{out}}(\lambda) + \mathsf{l}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(\lambda)) \\
 & & = & a2^{\lambda} \cdot \log|\mathbb{F}| + \mathsf{l}_{\mathsf{in}}(\lambda/2 + a) \\
 & & = & a2^{\lambda} \cdot \log|\mathbb{F}| + \tilde{O}(2^{(\lambda/2+a)} \cdot \log|\mathbb{F}|) \\[1ex]
\text{randomness} & \mathsf{r}(\lambda) & = & \mathsf{r}_{\mathsf{out}}(\lambda) + \mathsf{r}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(\lambda)) \\
 & & = & (\lambda + a) + \mathsf{r}_{\mathsf{in}}(\lambda/2 + a) \\
 & & = & (\lambda + a) + (\lambda/2 + a) + O(\log(\lambda/2 + a)) \\[1ex]
\text{query complexity} & \mathsf{q}(\lambda) & = & \mathsf{q}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(\lambda)) = \mathsf{q}_{\mathsf{in}}(\lambda/2 + a) = q_{\mathsf{in}} \\[1ex]
\text{soundness error} & \varepsilon(\lambda) & = & \varepsilon_{\mathsf{out}}(\lambda) + (1 - \varepsilon_{\mathsf{out}}(\lambda)) \cdot \varepsilon_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(\lambda)) \\
 & & = & 1 - \alpha \cdot \delta_{\mathsf{out}} + \alpha \cdot \delta_{\mathsf{out}} \cdot \varepsilon_{\mathsf{in}}(\lambda/2 + a) \\
 & & = & 1 - \alpha \cdot \delta_{\mathsf{out}} \cdot (1 - \varepsilon_{\mathsf{in}}) \\[1ex]
\text{proximity parameter} & \delta(\lambda) & = & \delta_{\mathsf{out}}(\lambda) = \delta_{\mathsf{out}} \\[1ex]
\text{prover time} & \mathsf{tp}(\lambda) & = & \mathsf{tp}_{\mathsf{out}}(\lambda) + \mathsf{tvq}_{\mathsf{out}}(\lambda) + \mathsf{tp}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(\lambda)) \\
 & & = & \lambda^{a}2^{\lambda} + (\lambda/2 + a) + \mathsf{tp}_{\mathsf{in}}(\lambda/2 + a) \\
 & & = & \lambda^{a}2^{\lambda+a} + (\lambda/2 + a) + \tilde{O}(2^{\lambda/2+a}) \\[1ex]
\text{verifier time} & \mathsf{tv}(\lambda) & = & \mathsf{tvq}_{\mathsf{out}}(\lambda) + \mathsf{tv}_{\mathsf{in}}(\mathsf{s}_{\mathsf{out}}(\lambda)) \\
 & & = & (\lambda/2 + a) + \mathsf{tv}_{\mathsf{in}}(\lambda/2 + a) \\
 & & = & (\lambda/2 + a) + \mathrm{poly}(\lambda/2 + a)
\end{array}\right] ,
$$

which implies the theorem statement, after a constant number of parallel repetitions, and for a large enough choice of a positive constant c that depends on the positive constant a and other constants hidden in Theorem 2.9. □

The alternative construction is an IOP of proximity with $O(\log\lambda)$ rounds, and follows from recursively invoking our interactive proof composition theorem $O(\log\lambda)$ times on [BS08]'s PCPs of proximity for additive Reed–Solomon codes. Here we exploit the fact that the relation of [BS08]'s verifier is itself a subrelation of $\mathrm{Rel}(\mathbb{F}, \varrho)$, so that we can again use the same PCP of proximity without going through a generic reduction. While this alternative construction has $O(\log\lambda)$ rounds rather than 2 and a weaker soundness guarantee, the construction is simpler and the underlying constants (other than soundness) are smaller because the "inner" proof system is much less complex.

Theorem 5.2. *For every $\varrho > 0$ there exists $c > 0$ such that the following holds for every binary field $\mathbb{F}$. Define $\mathscr{R} := \mathrm{Rel}(\mathbb{F}, \varrho)$ and note that $\delta_{\mathscr{R}} = 1 - \varrho$. For every $\delta \in (0, \frac{1}{2}\delta_{\mathscr{R}})$ there exists a public-coin IOPP system (P, V) that* *puts $\mathscr{R}$ in the complexity class*

$$\textbf{IOPP}\left[\begin{array}{llcl} \text{rounds} & \mathsf{k}(\lambda) & = & c \cdot \log \lambda \\ \text{answer alphabet} & \mathsf{a}(\lambda) & = & \mathbb{F}_2 \\ \text{proof length} & \mathsf{l}(\lambda) & = & c \cdot 2^{\lambda} \cdot \log |\mathbb{F}| \\ \text{randomness} & \mathsf{r}(\lambda) & = & c \cdot \lambda \\ \text{query complexity} & \mathsf{q}(\lambda) & = & c \\ \text{soundness error} & \varepsilon(\lambda) & = & 1 - \lambda^{c} \cdot \delta \\ \text{proximity parameter} & \delta(\lambda) & = & \delta \\ \text{prover time} & \mathsf{tp}(\lambda) & = & \lambda^{c} \cdot 2^{\lambda} \\ \text{verifier time} & \mathsf{tv}(\lambda) & = & \lambda^{c} \end{array}\right].$$

5.2 For tensor product codes

We show that tensor product codes have *sublinear*-size IOPs of proximity with constant query complexity; moreover, 1 round of interaction and public coins suffice. The construction follows from one invocation of our interactive composition theorem with [BS06, Vid15]'s robust local testers as the outer proof system, and [Mie09]'s PCPs of proximity for nondeterministic languages as the inner proof system. (See Section 2.6.2 and Section 2.6.3 for these.)

Let $\mathscr{C} = \{C_n\}_{n \in \mathbb{N}}$ be a $T(\cdot)$-efficient linear code family with block length $\ell(\cdot)$, m a positive integer, and μ a divisor of m. (See Section 2.4 for the definition of code families and their efficiency.) Informally, the robust local tester reduces proximity testing of $C_n^{\otimes m}$ to $C_n^{\otimes 2\mu}$ by restricting a function $w \colon D(n)^m \to \mathbb{F}(n)$ to a random axis-parallel 2μ-dimensional plane H, provided a mild condition holds ($m \geq 3\mu$). In our 1-round IOP of proximity, the verifier sends the random plane H to the prover, who replies with a PCP of proximity of [Mie09] for the claim "$w|_H \in C_n^{\otimes 2\mu}$". We then obtain the desired result, provided that the claim "$w|_H \in C_n^{\otimes 2\mu}$" lies in $\mathbf{NTIME}(T'(n))$ with $\tilde{O}(T'(n)) = o(\ell(n)^m)$, because [Mie09]'s proof length is quasilinear in the time complexity of the nondeterministic language. This latter condition is also mild, as we now explain. The code family $\mathscr{C}^{\otimes 2\mu}$ is $T'(\cdot)$-efficient with $T'(n) := 2\mu \cdot \ell(n)^{2\mu-1} \cdot T(n)$ (see discussion about efficiency of code families in Section 2.4), and the condition we need is $\tilde{O}(T'(n)) = o(\ell(n)^m)$. For typical linear codes, $T(n) = \ell(n)^{2+o(1)}$, because the generator and parity-check matrices have at most $\ell(n)^2$ entries, so that $T'(n) := 2\mu \cdot \ell(n)^{2\mu+1+o(1)}$; this gives what we need if $m > 2\mu + 2$ and $\ell(n) \geq \mu$.

We do not prove the theorem via a black box invocation of the interactive composition theorem because we would obtain sub-optimal parameters: the outer proof system is a robust local tester rather than a robust PCPP, so that we would obtain a 2-round IOP of proximity with one empty round. Thus, in the proof below, we perform interactive composition directly for [BS06, Vid15]'s robust local tester and [Mie09]'s PCPP, obtaining a 1-round IOP of proximity.

Below, we denote by $\mathrm{Rel}(\mathscr{C}, m)$ the relation of instance-witness pairs (n, w) such that $w \in C_n^{\otimes m}$. So far we have treated the positive integer m as a constant, so that choosing the trivial divisor $\mu = 1$ yields constant soundness error in the statement below. We can also think of m as non-constant, implicitly depending on n, in which case we can still obtain constant soundness by choosing a divisor $\mu = \Omega(m)$ while maintaining the condition that $\tilde{O}(T'(n)) = o(\ell(n)^m)$.

Theorem 5.3 (formal statement of Theorem 1.3)**.** *There exists $c > 0$ such that the following holds. Let $\mathscr{C} = \{C_n\}_{n \in \mathbb{N}}$ be a $T(\cdot)$-efficient linear code family with alphabet $\mathbb{F}(\cdot)$, block length $\ell(\cdot)$, and relative distance $\tau(\cdot)$; let m be a positive integer and μ a divisor of m. Suppose that $m \geq 3\mu$ and $T'(n) \cdot (\log T'(n))^c = o(\ell(n)^m)$ with $T'(n) := 2\mu \cdot \ell(n)^{2\mu-1} \cdot T(n)$. Define $\mathscr{R} := \mathrm{Rel}(\mathscr{C}, m)$ and note that $\delta_{\mathscr{R}} = \tau^m$. For every $\delta \in (0, \frac{1}{2}\delta_{\mathscr{R}})$ there exist $q_0 > 0$ and a public-coin IOPP system (P, V) that puts $\mathscr{R}$ in the complexity class*

$$\textbf{IOPP}\left[\begin{array}{llcl} \text{rounds} & \mathsf{k}(n) & = & 1 \\ \text{answer alphabet} & \mathsf{a}(n) & = & \mathbb{F}_2 \\ \text{proof length} & \mathsf{l}(n) & = & o(\ell(n)^m \cdot \log |\mathbb{F}|) \\ \text{randomness} & \mathsf{r}(n) & = & O(m \cdot \log \ell(n)) \\ \text{query complexity} & \mathsf{q}(n) & = & q_0 \\ \text{soundness error} & \varepsilon(n) & = & 1 - \frac{\tau^{2(m/\mu)}}{4(m/\mu)^8} \cdot \delta \\ \text{proximity parameter} & \delta(n) & = & \delta \\ \text{prover time} & \mathsf{tp}(n) & = & o(\ell(n)^m) \\ \text{verifier time} & \mathsf{tv}(n) & = & \mathrm{poly}(m + \log \ell(n)) \end{array}\right].$$

Proof. We construct the IOPP system (P, V), then analyze its completeness, its soundness, and efficiency parameters.

Construction. Construct the IOPP system (P, V) as follows. Let $(\mathbb{x}, \mathbb{w}) = (n, w)$ be an instance-witness pair in the relation $\mathrm{Rel}(\mathscr{C}, m)$; the prover P receives the instance and witness as input, while the verifier V receives the instance as input and the witness as oracle. In the first round, the verifier V samples a random 2μ-dimensional plane H in $D(n)^m$ and then sends (the description of) H to P; the prover P then computes a proximity proof π attesting to the statement "$w|_H \in C_n^{\otimes 2\mu}$", and sends the proof string π to V. The verifier V checks π and accepts if the check passes.

We are left to specify which PCPP system to use for this task: we rely on the PCPP system for nondeterministic languages of Theorem 2.9. We apply the theorem to the relation $\mathscr{R}_{\mathsf{in}}$ of instance-witness pairs (n, f) such that $f \in C_n^{\otimes 2\mu}$; the relation $\mathscr{R}_{\mathsf{in}}$ can be decided in time $T'(n)$ with $\tilde{O}(T'(n)) = o(\ell(n)^m)$. By Theorem 2.9, we obtain that for every $\delta_{\mathsf{in}} \in (0, \frac{1}{2}\delta_{\mathscr{R}_{\mathsf{in}}})$ and $\varepsilon_{\mathsf{in}} > 0$ there exist $q_{\mathsf{in}} > 0$ and a PCPP system $(P_{\mathsf{in}}, V_{\mathsf{in}})$ for $\mathscr{R}_{\mathsf{in}}$ with the parameters described in the theorem statement. In this proof, we choose $\delta_{\mathsf{in}} < \min\{\frac{1}{2}\delta_{\mathscr{R}_{\mathsf{in}}}, \rho - \varepsilon'\}$ and $\varepsilon_{\mathsf{in}} := \frac{1}{100}$, where the constants $\rho, \varepsilon' \in (0, 1)$ are chosen in the soundness analysis below.

Completeness. Consider any instance-witness pair $(\mathbb{x}, \mathbb{w}) = (n, w)$ in the relation $\mathrm{Rel}(\mathscr{C}, m)$. By Lemma 2.8, $w|_H \in C^{\otimes 2\mu}$ for every 2μ-dimensional plane $H \in D(n)^m$; hence, $V_{\mathsf{in}}^{w|_H, \pi}(n)$ always accepts. We conclude that $P(\mathbb{x}, \mathbb{w})$ makes $V^{\mathbb{w}}(\mathbb{x})$ accept with probability 1.

Soundness. Consider any instance-witness pair $(\mathbb{x}, \mathbb{w}) = (n, w)$ and unbounded malicious prover $\tilde{P}$. Suppose that w is $\delta(n)$-far from any codeword in $C_n^{\otimes m}$. By Lemma 2.8, the expected distance of $w|_H$ to $C_n^{\otimes 2\mu}$ is at least $\rho := \frac{\tau^{2(m/\mu)}}{(m/\mu)^8} \cdot \delta(n)$. By [BGH$^+$06, Proposition 2.10], for any $\varepsilon' \leq \rho$, the distance is at most $\rho - \varepsilon'$ with probability at most $1 - \varepsilon'$. Call H bad if its distance to $C_n^{\otimes 2\mu}$ is at most $\rho - \varepsilon'$; else call H good. For any bad H, we only know that $V_{\mathsf{in}}^{w|_H, \pi}(n)$ accepts with probability at most 1; for any good H, we know that $V_{\mathsf{in}}^{w|_H, \pi}(n)$ accepts with probability at most $\varepsilon_{\mathsf{in}}$ because $\delta_{\mathsf{in}} \leq \rho - \varepsilon'$. Overall, we deduce that V accepts with probability at most $(1 - \varepsilon') \cdot 1 + \varepsilon' \cdot \varepsilon_{\mathsf{in}} = 1 - \varepsilon' \cdot (1 - \varepsilon_{\mathsf{in}})$. Now we pick $\varepsilon' := \rho/2$, and we conclude that V accepts with probability at most $1 - \rho/2 \cdot (1 - \varepsilon_{\mathsf{in}}) = 1 - \frac{1}{2} \cdot \frac{\tau^{2(m/\mu)}}{(m/\mu)^8} \cdot \delta(n) \cdot (1 - \frac{1}{100})$, and the claimed soundness follows.

Efficiency parameters. The constructed IOPP system has $\mathsf{k}(n) := 1$ rounds. The proof length is $\mathsf{l}(n) := \tilde{O}(T'(n)) = o(\ell(n)^m)$ because the prover sends the proximity proof π output by P_{in}. The randomness complexity is $\mathsf{r}(n) := O(m \cdot \log \ell(n))$ because the verifier samples a random axis-parallel 2μ-dimensional plane in $D(n)^m$ and then runs V_{in}. The query complexity is $\mathsf{q}(n) := q_{\mathsf{in}}$ because the verifier runs V_{in}. The prover running time is $\mathsf{tp}(n) := \tilde{O}(T'(n)) = o(\ell^m)$ because the prover runs P_{in}. The verifier running time is $\mathsf{tv}(n) := \mathrm{poly}(m + \log \ell(n))$ because the verifier samples a random axis-parallel 2μ-dimensional plane in $D(n)^m$ and then runs V_{in}. Finally, the protocol is clearly public coins. □

6 From circuit satisfiability to sumcheck

We prove that, with 1 round of IOP interaction, we can reduce circuit satisfiability to proximity testing to a linear code and a sumcheck over any degree-3 closure of it; moreover, the IOP introduces only constant overheads. We use this reduction in Section 7, along with other ingredients, to construct 3-round IOPs for circuit satisfiability with linear proof length and constant query complexity. We begin by recalling the notion of boolean circuits and their satisfiability.

Definition 6.1. *A* boolean circuit *ϕ with n gates and s inputs is a directed acyclic graph with n vertices of which s are sources and 1 is a sink; vertices represent gates while directed edges represent wires among them. We define ϕ's size to be n, and label the gates as $g_1, \ldots, g_n$ so that $g_1, \ldots, g_s$ are the input gates and g_n is the output gate. We assume that all gates (except input gates) are two-input NAND gates. Denote by $\ell \colon [n] \to [n]$ and $r \colon [n] \to [n]$ the functions such that $g_{\ell(i)}$ and $g_{r(i)}$ are the gates whose outputs are the left and right inputs of g_i; for $i \in [s]$, the value of $\ell(i)$ and $r(i)$ is arbitrary, e.g., 1. We say that $w \in \{0,1\}^n$ is a* satisfying assignment *for ϕ if $w_n = 0$ and, for every $i = s+1, \ldots, n$, w_i is the output of g_i when the input to the circuit is $w_1 \cdots w_s$.*

The relation $\mathscr{R}_{\mathrm{CSAT}}$ comprises all instance-witness pairs (ϕ, w) such that w is a satisfying assignment to ϕ.

In the statement and proof below we use the notion of a systematic code family, described in Section 2.4, and the notion of an *evading set*, described in Section 2.5. Given an n-systematic code family $\mathscr{C} = \{C_n\}_{n \in \mathbb{N}}$, we denote by:

- $\mathrm{Rel}(\mathscr{C})$ the relation of instance-witness pairs (n, w) such that $n \in \mathbb{N}$ and $w \in C_n$; and
- $\mathrm{Rel}(\mathrm{SC}, \mathscr{C})$ the relation of instance-witness pairs (n, w) such that $n \in \mathbb{N}$, $w \in C_n$, and $\sum_{i \in [n]} w(i) = 0$.

Theorem 6.2 (From CSAT to Sumcheck — formal statement of Lemma 1.6)**.** *Suppose that*

- *$\mathscr{C} = \{C_n\}_{n \in \mathbb{N}}$ is a $T_{\mathscr{C}}(\cdot)$-efficient n-systematic code family with alphabet $\mathbb{F}(\cdot)$,*
- *$\mathscr{D} = \{D_n\}_{n \in \mathbb{N}}$ is a $T_{\mathscr{D}}(\cdot)$-efficient n-systematic code family with relative distance $\tau_{\mathscr{D}}(\cdot)$ and is a degree-3 closure of $\mathscr{C}$,*
- *$\mathscr{S} = \{S_n\}_{n \in \mathbb{N}}$ is a $T_{\mathscr{S}}(\cdot)$-efficient $\gamma(\cdot)$-evading set family for $\mathbb{F}(\cdot)$ with $\frac{1}{|\mathbb{F}(n)|} \leq \gamma(n)$.*

Suppose further that the following two conditions hold with $\delta_{\mathscr{C}}(n) < \tau_{\mathscr{D}}(n)/16$:

***(1)** there exists an IOPP system $(P_{\mathscr{C}}, V_{\mathscr{C}})$ that puts* $\mathrm{Rel}(\mathscr{C})$ *in the complexity class*

$$\textbf{IOPP}\left[\begin{array}{ll} \text{rounds} & \mathsf{k}_{\mathscr{C}} \\ \text{answer alphabet} & \mathsf{a} \\ \text{proof length} & \mathsf{l}_{\mathscr{C}} \\ \text{randomness} & \mathsf{r}_{\mathscr{C}} \\ \text{query complexity} & \mathsf{q}_{\mathscr{C}} \\ \text{soundness error} & \varepsilon_{\mathscr{C}} \\ \text{proximity parameter} & \delta_{\mathscr{C}} \\ \text{prover time} & \mathsf{tp}_{\mathscr{C}} \\ \text{verifier time} & \mathsf{tv}_{\mathscr{C}} \end{array}\right]$$

and

***(2)** there exists an IOPP system $(P_{\mathscr{D}}, V_{\mathscr{D}})$ that puts* $\mathrm{Rel}(\mathrm{SC}, \mathscr{D})$ *in the complexity class*

$$\textbf{IOPP}\left[\begin{array}{ll} \text{rounds} & \mathsf{k}_{\mathscr{D}} \\ \text{answer alphabet} & \mathsf{a} \\ \text{proof length} & \mathsf{l}_{\mathscr{D}} \\ \text{randomness} & \mathsf{r}_{\mathscr{D}} \\ \text{query complexity} & \mathsf{q}_{\mathscr{D}} \\ \text{soundness error} & \varepsilon_{\mathscr{D}} \\ \text{proximity parameter} & \delta_{\mathscr{D}} \\ \text{prover time} & \mathsf{tp}_{\mathscr{D}} \\ \text{verifier time} & \mathsf{tv}_{\mathscr{D}} \end{array}\right]$$

Then there exists an IOP system (P, V) that puts the relation $\mathscr{R}_{\mathrm{CSAT}}$ in the complexity class

$$\textbf{IOP}\left[\begin{array}{lll} \text{rounds} & \mathsf{k}(n) & = \quad 1 + \max\{\mathsf{k}_{\mathscr{C}}(n), \mathsf{k}_{\mathscr{D}}(n)\} \\ \text{answer alphabet} & \mathsf{a}(n) & \\ \text{proof length} & \mathsf{l}(n) & = \quad 3 \cdot \mathsf{l}_{\mathscr{C}}(n) + \mathsf{l}_{\mathscr{D}}(n) + 3 \cdot \ell_{\mathscr{C}}(n) \cdot \log|\mathbb{F}(n)| \\ \text{randomness} & \mathsf{r}(n) & = \quad 3 \cdot \mathsf{r}_{\mathscr{C}}(n) + \mathsf{r}_{\mathscr{D}}(n) + 4 \cdot \log|S_n| + \log|\mathbb{F}(n)| \\ \text{query complexity} & \mathsf{q}(n) & = \quad 3 \cdot \mathsf{q}_{\mathscr{C}}(n) + \mathsf{q}_{\mathscr{D}}(n) \\ \text{soundness error} & \varepsilon(n) & = \quad \max\{\varepsilon_{\mathscr{C}}(n), \varepsilon_{\mathscr{D}}(n) + \gamma(n)\} \\ \text{prover time} & \mathsf{tp}(n) & = \quad 3 \cdot \mathsf{tp}_{\mathscr{C}}(n) + \mathsf{tp}_{\mathscr{D}}(n) + 8 \cdot T_{\mathscr{C}}(n) + 4 \cdot T_{\mathscr{S}}(n) + T_{\mathscr{D}}(n) + O(n \cdot \log|\mathbb{F}(n)|) \\ \text{verifier time} & \mathsf{tv}(n) & = \quad 3 \cdot \mathsf{tv}_{\mathscr{C}}(n) + \mathsf{tv}_{\mathscr{D}}(n) + 5 \cdot T_{\mathscr{C}}(n) + 4 \cdot T_{\mathscr{S}}(n) \end{array}\right]$$

Moreover, if $V_{\mathscr{C}}$'s and $V_{\mathscr{D}}$'s queries are non-adaptive so are V's queries; also, if $V_{\mathscr{C}}$ and $V_{\mathscr{D}}$ are public coin so is V.

We first give a simple lemma that says that a circuit's satisfiability can be represented as a set of low-degree constraints on three codewords; these codewords represent encodings of all gates' outputs, left inputs, and right inputs. Similar statements appear in several prior works that encode computation via, e.g., low-degree polynomials.

Lemma 6.3. *Let ϕ be a boolean circuit with n gates and s inputs, C an n-systematic code with alphabet $\mathbb{F}$, and $P_{\text{NAND}}\colon \mathbb{F}^3 \to \mathbb{F}$ the polynomial of total degree 2 that describes a NAND gate (for every $x, y, z \in \{0,1\}$, $P_{\text{NAND}}(x,y,z) = 0$ if and only if the NAND of x and y equals z). There exists a satisfying assignment for ϕ if and only if there exist codewords $W, W_{\mathsf{L}}, W_{\mathsf{R}} \in C$ that satisfy the following:*

- (booleanity constraints) *for every $i \in [n]$, $W(i)^2 = W(i)$;*
- (wiring constraints) *for every $i \in [n]$, $W_{\mathsf{L}}(i) = W(\ell(i))$ and $W_{\mathsf{R}}(i) = W(r(i))$;*
- (gate constraints) *for every $i \in [n] \setminus [s]$, $P_{\text{NAND}}(W_{\mathsf{L}}(i), W_{\mathsf{R}}(i), W(i)) = 0$;*
- (satisfiability constraint) $W(n) = 0$.

Proof. For every $w \in \{0,1\}^n \subseteq \mathbb{F}^n$, w is a satisfying assignment for ϕ if and only if $w_n = 0$ and $P_{\text{NAND}}(w_{\ell(i)}, w_{r(i)}, w_i) = 0$ for every $i \in [n] \setminus [s]$. With this in mind, we can argue the two sides.

Let $w \in \{0,1\}^n$ be a satisfying assignment for ϕ. Let $W, W_{\mathsf{L}}, W_{\mathsf{R}}$ be the codewords in C such that, for every $i \in [n]$, $W(i) = w_i$, $W_{\mathsf{L}}(i) = w_{\ell(i)}$, and $W_{\mathsf{R}}(i) = w_{r(i)}$; such codewords exist because C is n-systematic. One can verify that this choice of codewords fulfills the constraints in the statement.

Conversely, let $W, W_{\mathsf{L}}, W_{\mathsf{R}}$ be codewords in C that satisfy the constraints in the statement. Let w be the assignment that equals the codeword W restricted to $[n]$. One can verify that this choice of assignment is satisfying for ϕ. □

We now return to the proof of the theorem.

Proof of Theorem 6.2. We construct the IOP system (P, V), then analyze its completeness, its soundness, and efficiency parameters.

Construction. Construct the IOP system (P, V) as follows. Let (ϕ, w) be an instance-witness pair in the circuit-satisfiability relation $\mathscr{R}_{\text{CSAT}}$; the prover P receives the instance and witness as input, while the verifier V receives the instance as input.

- In the first round, the verifier V sends an empty message to P; next, the prover P proceeds as follows: (i) compute $W, W_{\mathsf{L}}, W_{\mathsf{R}} \in C_n$ from the assignment $w \in \{0,1\}^n$ so that, for every $i \in [n]$, $W(i) = w_i$, $W_{\mathsf{L}}(i) = w_{\ell(i)}$, and $W_{\mathsf{R}}(i) = w_{r(i)}$ (as in the proof of Lemma 6.3); (ii) send the proof string $(W, W_{\mathsf{L}}, W_{\mathsf{R}})$ to the verifier V.

- In the second round, the verifier chooses uniformly (and independently) at random $r, r', r'', r''' \in S_n$ and $\alpha \in \mathbb{F}(n)$, and sends these to P; here S_n is the $\gamma(n)$-evading set for $\mathbb{F}(n)^n$ in the family $\mathscr{S}$. We use this randomness to define the codewords $R_{\mathsf{W}}, R_{\mathsf{L}}, R_{\mathsf{R}}, R_{\mathsf{N}}, R_{\mathsf{B}} \in C_n$ as specified below:

$$\begin{aligned}
\forall\, i \in [n-1]\,, \quad & R_{\mathsf{W}}(i) = -\left(\sum_{j \in \ell^{-1}(i)} r_j + \sum_{j \in r^{-1}(i)} r'_j\right) \quad \text{and} \quad R_{\mathsf{W}}(n) = \alpha \\
\forall\, i \in [n]\,, \quad & R_{\mathsf{L}}(i) = r_i \\
\forall\, i \in [n]\,, \quad & R_{\mathsf{R}}(i) = r'_i \\
\forall\, i \in [n] \setminus [s]\,, \quad & R_{\mathsf{N}}(i) = r''_i \quad \text{and} \quad \forall\, i \in [s] \quad R_{\mathsf{N}}(i) = 0 \\
\forall\, i \in [n]\,, \quad & R_{\mathsf{B}}(i) = r'''_i
\end{aligned}$$

  The prover P and verifier V may compute these codewords, which in turn induce the codeword H in D_n defined as

$$H := R_{\mathsf{W}} \cdot W + R_{\mathsf{L}} \cdot W_{\mathsf{L}} + R_{\mathsf{R}} \cdot W_{\mathsf{R}} + R_{\mathsf{N}} \cdot P_{\text{NAND}}(W_{\mathsf{L}}, W_{\mathsf{R}}, W) + R_{\mathsf{B}} \cdot (W^2 + W)$$

  Indeed, note that H equals $Q(R_{\mathsf{W}}, R_{\mathsf{L}}, R_{\mathsf{R}}, R_{\mathsf{N}}, R_{\mathsf{B}}, W, W_{\mathsf{L}}, W_{\mathsf{R}})$ for a polynomial Q of total degree 3.

  In parallel, the prover P and verifier V engage in several interactive oracle proofs:

  - an IOPP $(P_{\mathscr{C}}, V_{\mathscr{C}})$ to prove proximity of (n, W) to $\mathrm{Rel}(\mathscr{C})$;
  - an IOPP $(P_{\mathscr{C}}, V_{\mathscr{C}})$ to prove proximity of (n, W_{L}) to $\mathrm{Rel}(\mathscr{C})$;
  - an IOPP $(P_{\mathscr{C}}, V_{\mathscr{C}})$ to prove proximity of (n, W_{R}) to $\mathrm{Rel}(\mathscr{C})$;
  - an IOPP $(P_{\mathscr{D}}, V_{\mathscr{D}})$ to prove proximity of (n, H) to $\mathrm{Rel}(\mathrm{SC}, \mathscr{D})$.

Completeness. Completeness of (P, V) follows from that of $(P_{\mathscr{C}}, V_{\mathscr{C}})$ and $(P_{\mathscr{D}}, V_{\mathscr{D}})$, and the fact that the sum of H over $[n]$ equals 0. Namely, for every instance-witness pair (ϕ, w) in the relation $\mathscr{R}_{\mathrm{CSAT}}$ it holds that:

- In the first round, the prover P sends three codewords $W, W_{\mathsf{L}}, W_{\mathsf{R}}$ that are in the code C_n; by construction, and since w is a satisfying assignment for ϕ, we know that $W, W_{\mathsf{L}}, W_{\mathsf{R}}$ satisfy the conditions in Lemma 6.3.
- In the second round, for any choice of verifier randomness, the codewords $R_{\mathsf{W}}, R_{\mathsf{L}}, R_{\mathsf{R}}, R_{\mathsf{N}}, R_{\mathsf{B}}$ are in the code C_n.
- The codeword H is derived from the above codewords via a polynomial Q of total degree 3 so that H is in the code D_n, because D_n is a degree-3 closure of C_n. Moreover, recalling that $R_{\mathsf{W}}|_{[n]}$ is related to $R_{\mathsf{L}}|_{[n]}$ and $R_{\mathsf{R}}|_{[n]}$ by the wiring constraints, H sums to 0 on $[n]$ because:

$$\begin{aligned}\sum_{i\in[n]} H(i) &= \sum_{i\in[n]} R_{\mathsf{W}}(i)\cdot W(i) + R_{\mathsf{L}}(i)\cdot W_{\mathsf{L}}(i) + R_{\mathsf{R}}(i)\cdot W_{\mathsf{R}}(i) + R_{\mathsf{N}}(i)\cdot P_{\mathrm{NAND}}(W_{\mathsf{L}}(i), W_{\mathsf{R}}(i), W(i)) + R_{\mathsf{B}}(i)\cdot (W^2(i) + W(i)) \\ &= \alpha\cdot W(n) + \sum_{i\in[n]} r_i\cdot\Big(W_{\mathsf{L}}(i) - W(\ell(i))\Big) + \sum_{i\in[n]} r'_i\cdot\Big(W_{\mathsf{R}}(i) - W(r(i))\Big) \\ &\quad + \sum_{i\in[n]\setminus[s]} r''_i\cdot P_{\mathrm{NAND}}\Big(W_{\mathsf{L}}(i), W_{\mathsf{R}}(i), W(i)\Big) + \sum_{i\in[n]} r'''_i\cdot\Big(W(i)^2 - W(i)\Big) \\ &= \alpha\cdot 0 + \sum_{i\in[n]} r_i\cdot 0 + \sum_{i\in[n]} r'_i\cdot 0 + \sum_{i\in[n]\setminus[s]} r''_i\cdot 0 + \sum_{i\in[n]} r'''_i\cdot 0 = 0 \enspace .\end{aligned}$$

- Hence, also in the second round (and any later rounds): the use of the IOPP system $(P_{\mathscr{C}}, V_{\mathscr{C}})$ to separately prove proximity of $W, W_{\mathsf{L}}, W_{\mathsf{R}}$ to C_n and the use of the IOPP system $(P_{\mathscr{D}}, V_{\mathscr{D}})$ to prove proximity of H to the subcode of D_n of codewords that sum to 0 on $[n]$ results in the verifier V accepting with probability 1.

Soundness. Consider any unsatisfiable boolean circuit ϕ, and unbounded malicious prover $\tilde{P}$. We distinguish among the following cases:

- *Case 1: one of $W, W_{\mathsf{L}}, W_{\mathsf{R}}$ sent by $\tilde{P}$ in the first round is $\delta_{\mathscr{C}}(n)$-far from C_n.*

  In this case, the IOPP verifier $V_{\mathscr{C}}$ accepts with probability at most $\varepsilon_{\mathscr{C}}(n)$.

- *Case 2: the above case does not hold.*

  Observe that H is a random variable that depends on the verifier randomness $\chi := (r, r', r'', r''', \alpha)$. Let A be the set of χ for which H is $\delta_{\mathscr{D}}(n)$-far from D_n.

  For any $\chi \in A$, the IOPP verifier $V_{\mathscr{D}}$ accepts with probability at most $\varepsilon_{\mathscr{D}}(n)$.

  For any $\chi \notin A$, let $\hat{W}, \hat{W}_{\mathsf{L}}, \hat{W}_{\mathsf{R}}$ be the unique codewords in C_n that are closest to $W, W_{\mathsf{L}}, W_{\mathsf{R}}$ (respectively); also, let $\hat{H}$ be the unique codeword in D_n that is closest to H. (Recall that proximity parameters are less than the unique-decoding radius, so such codewords exist; see Section 2.2.) By hypothesis, $\delta_{\mathscr{C}}(n) < \tau(D_n)/16$; hence, by Claim 2.2 (invoked for C_n, D_n, and $m = 8$), we deduce that $\hat{H} = Q(R_{\mathsf{W}}, R_{\mathsf{L}}, R_{\mathsf{R}}, R_{\mathsf{N}}, R_{\mathsf{B}}, \hat{W}, \hat{W}_{\mathsf{L}}, \hat{W}_{\mathsf{R}})$. Denote by $\varepsilon_{\hat{H}}$ the probability over χ, conditioned on $\chi \notin A$, that $\hat{H}$ sums to 0 on $[n]$.

  If ε is the probability that the verifier accepts in this case, we can write

$$\begin{aligned}\varepsilon &= \Pr[\chi\in A]\cdot\varepsilon_{\mathscr{D}}(n) + \Pr[\chi\notin A]\cdot\big(\varepsilon_{\hat{H}}\cdot 1 + (1-\varepsilon_{\hat{H}})\cdot\varepsilon_{\mathscr{D}}(n)\big) \\ &\le \max\big\{\varepsilon_{\mathscr{D}}(n), \varepsilon_{\hat{H}}\cdot 1 + (1-\varepsilon_{\hat{H}})\cdot\varepsilon_{\mathscr{D}}(n)\big\} \\ &\le \max\big\{\varepsilon_{\mathscr{D}}(n), \varepsilon_{\mathscr{D}}(n) + \varepsilon_{\hat{H}}\big\} = \varepsilon_{\mathscr{D}}(n) + \varepsilon_{\hat{H}}\end{aligned}$$

  so we are left to upper bound $\varepsilon_{\hat{H}}$.

  Since ϕ is unsatisfiable, $\hat{W}, \hat{W}_{\mathsf{L}}, \hat{W}_{\mathsf{R}}$ do not satisfy the conditions in Lemma 6.3. Therefore

$$\begin{aligned}\sum_{i\in[n]} \hat{H}(i) = \alpha\cdot\hat{W}(n) &+ \sum_{i\in[n]} r_i\cdot\Big(\hat{W}_{\mathsf{L}}(i) - \hat{W}(\ell(i))\Big) + \sum_{i\in[n]} r'_i\cdot\Big(\hat{W}_{\mathsf{R}}(i) - \hat{W}(r(i))\Big) \\ &+ \sum_{i\in[n]\setminus[s]} r''_i\cdot P_{\mathrm{NAND}}\Big(\hat{W}_{\mathsf{L}}(i), \hat{W}_{\mathsf{R}}(i), \hat{W}(i)\Big) + \sum_{i\in[n]} r'''_i\cdot\Big(\hat{W}(i)^2 - \hat{W}(i)\Big)\end{aligned}$$

  is zero with probability at most $\gamma(n)$, as we now explain. Consider two sub-cases:

- *Case 2.a:* $\hat{W}(n) \neq 0$. In this case $\varepsilon_{\hat{H}} \leq \frac{1}{|\mathbb{F}(n)|}$ because α is uniformly random in $\mathbb{F}(n)$.
- *Case 2.b:* $\hat{W}(n) = 0$. One of the four sums is an inner product of a non-zero vector with a uniformly random element in the $\gamma(n)$-evading set S_n; hence, $\varepsilon_{\hat{H}} \leq \gamma(n)$.

Recalling that $\frac{1}{|\mathbb{F}(n)|} \leq \gamma(n)$ by hypothesis, we deduce that $\varepsilon_{\hat{H}} \leq \gamma(n)$.

We deduce that the verifier accepts with probability that is at most the maximum of the acceptance probability across the two cases, namely, $\max\{\varepsilon_{\mathscr{C}}(n), \varepsilon_{\mathscr{D}}(n) + \gamma(n)\}$.

Efficiency parameters. The constructed IOP system has $\mathsf{k}(n) := 1 + \max\{\mathsf{k}_{\mathscr{C}}(n), \mathsf{k}_{\mathscr{D}}(n)\}$ rounds. The proof length is $\mathsf{l}(n) := 3 \cdot \mathsf{l}_{\mathscr{C}}(n) + \mathsf{l}_{\mathscr{D}}(n) + 3 \cdot \ell_{\mathscr{C}}(n)$ because the prover sends the three codewords $W, W_{\mathsf{L}}, W_{\mathsf{R}}$ in C_n, and also runs three invocations of $P_{\mathscr{C}}$ and one invocation of $P_{\mathscr{D}}$. The randomness complexity is $\mathsf{r}(n) := 3 \cdot \mathsf{r}_{\mathscr{C}}(n) + \mathsf{r}_{\mathscr{D}}(n) + 4 \cdot \log|S_n| + \log|\mathbb{F}(n)|$ because the verifier runs three invocations of $V_{\mathscr{C}}$ and one invocation of $V_{\mathscr{D}}$, samples four elements from the evading set S_n, and also one element from $\mathbb{F}(n)$. The query complexity is $\mathsf{q}(n) := 3 \cdot \mathsf{q}_{\mathscr{C}}(n) + \mathsf{q}_{\mathscr{D}}(n)$ because the verifier runs three invocations of $V_{\mathscr{C}}$ and one invocation of $V_{\mathscr{D}}$, and makes no other queries otherwise. The prover running time is $\mathsf{tp}(n) := 3 \cdot \mathsf{tp}_{\mathscr{C}}(n) + \mathsf{tp}_{\mathscr{D}}(n) + 8 \cdot T_{\mathscr{C}}(n) + T_{\mathscr{D}}(n) + O(n \cdot \log|\mathbb{F}(n)|)$ because the prover encodes the circuit assignment to obtain the three codewords $W, W_{\mathsf{L}}, W_{\mathsf{R}}$ in C_n, computes the four vectors sampled by the verifier from the evading set, encodes these four vectors, along with a fifth derived from them, to obtain the five codewords $R_{\mathsf{W}}, R_{\mathsf{L}}, R_{\mathsf{R}}, R_{\mathsf{N}}, R_{\mathsf{B}}$ in C_n, coordinate-wise computes a message and encodes it to obtain the codeword H in D_n, and then runs three invocations of $P_{\mathscr{C}}$ and one invocation of $P_{\mathscr{D}}$. The verifier running time is $\mathsf{tv}(n) := 3 \cdot \mathsf{tv}_{\mathscr{C}}(n) + \mathsf{tv}_{\mathscr{D}}(n) + 5 \cdot T_{\mathscr{C}}(n) + 4 \cdot T_{\mathscr{S}}(n)$ because the verifier samples four vectors from the evading set, also computes the codewords $R_{\mathsf{W}}, R_{\mathsf{L}}, R_{\mathsf{R}}, R_{\mathsf{N}}, R_{\mathsf{B}}$, and then runs three invocations of $V_{\mathscr{C}}$ and one invocation of $V_{\mathscr{D}}$. □

7 IOP for circuit satisfiability

We show that for circuit satisfiability we can obtain IOPs with linear proof length and constant query complexity; moreover, 3 rounds of interaction and public coins suffice.

Theorem 7.1 (formal statement of Theorem 1.1)**.** *Let $\mathscr{R}_{\mathrm{CSAT}}$ be the relation consisting of instance-witness pairs (ϕ, w) where ϕ is a boolean circuit (of two-input NAND gates) and w is a binary input that satisfies ϕ; we use n to denote the number of gates in ϕ. For every $\eta \in (0,1)$ there exists a public-coin IOP system that puts $\mathscr{R}_{\mathrm{CSAT}}$ is in the complexity class*

$$\mathbf{IOP}\left[\begin{array}{lccl} \text{rounds} & \mathsf{k}(n) & = & 3 \\ \text{answer alphabet} & \mathsf{a}(n) & = & \mathbb{F}_2 \\ \text{proof length} & \mathsf{l}(n) & = & O(n) \\ \text{randomness} & \mathsf{r}(n) & = & \mathrm{polylog}(n) \\ \text{query complexity} & \mathsf{q}(n) & = & O(1) \\ \text{soundness error} & \varepsilon(n) & = & 1/2 \\ \text{prover time} & \mathsf{tp}(n) & = & O(n^{1+\eta}) \\ \text{verifier time} & \mathsf{tv}(n) & = & O(n^{1+\eta}) \end{array}\right].$$

We first show how to combine our Sublinear Sumcheck Theorem (Theorem 4.1) and our 2-round IOPPs for tensor product codes (Theorem 5.3) to obtain a 2-round IOPP for the sumcheck relation corresponding to the $O(1)$-wise tensor product of a 'good' systematic linear code.

Lemma 7.2. *Let $\mathscr{C} = \{C_k\}_{k \in \mathbb{N}}$ be a family of k-systematic linear codes with a constant-size alphabet, constant pseudorate $\hat{\rho}(\mathscr{C}) > 0$, constant relative distance $\tau(\mathscr{C}) > 0$, and efficiency $T_{\mathscr{C}}(k) = O(k^c)$ for constant c. Then, for any integer constant $c' > c$,*

$$\mathrm{Rel}(\mathrm{SC}, \mathscr{C}^{\otimes 2c'}) \in \mathbf{IOPP}\left[\begin{array}{lccl} \text{rounds} & \mathsf{k}(n) & = & 2 \\ \text{answer alphabet} & \mathsf{a}(n) & = & \mathbb{F}_2 \\ \text{proof length} & \mathsf{l}(n) & = & o(n) \\ \text{randomness} & \mathsf{r}(n) & = & \mathrm{polylog}(n) \\ \text{query complexity} & \mathsf{q}(n) & = & O(1) \\ \text{soundness error} & \varepsilon(n) & = & 1/2 \\ \text{proximity parameter} & \delta(n) & = & \delta \\ \text{prover time} & \mathsf{tp}(n) & = & o(n) \\ \text{verifier time} & \mathsf{tv}(n) & = & \mathrm{polylog}(n) \end{array}\right],$$

where $n := k^{2c'}$ denotes the systematicity of $\mathrm{Rel}(\mathrm{SC}, \mathscr{C}^{\otimes 2c'})$.

Proof. Define the linear code family $\mathscr{D} := \mathscr{C}^{\otimes c'}$; note that $\mathscr{D}$ is $k^{c'}$-systematic and $\mathscr{D}^{\otimes 2} = \mathscr{C}^{\otimes 2c'}$. Define the subset family $\mathscr{H} := \{H_k = [k^{c'}]\}_{k \in \mathbb{N}}$ where $H_k := [k^{c'}]$; note that H_k is a subset of D_k's domain. By properties of tensor codes (see Section 2.4), $\mathscr{D}$ is $T_{\mathscr{D}}(\cdot)$-efficient with $T_{\mathscr{D}}(k) = O(k^{c'+c}) = O(k^{2c'-1})$, so the relation $\mathrm{Rel}(\mathscr{D}, 1, \mathscr{H})$ can be decided in $\mathbf{NTIME}(k^{2c'-1})$.

We now construct proof systems for the two relations $\mathrm{Rel}(\mathscr{D}, 1, \mathscr{H})$ and $\mathrm{Rel}(\mathscr{C}^{\otimes 2c'})$:

- By Theorem 2.9 (PCPPs for nondeterministic languages), there exists a PCPP (which is a 1-round IOPP) for $\mathrm{Rel}(\mathscr{D}, 1, \mathscr{H})$ with proof length $\tilde{O}(k^{2c'-1}) = o(n)$ (since $n = k^{2c'}$) and constant query complexity.
- By Theorem 5.3 there exists a 1-round IOPP for $\mathrm{Rel}(\mathscr{C}^{\otimes 2c'})$ with proof length $o(n)$ and constant query complexity. Here we use the fact that the pseudorate $\hat{\rho}(\mathscr{C})$ is constant, so that the block length $\ell(C_k)$ of C_k is $\Theta(k)$, and thus the block length $\ell(C_k^{\otimes 2c'})$ of $C_k^{\otimes 2c'}$ is $\ell(C_k)^{2c'} = \Theta(k^{2c'}) = \Theta(n)$.

We use these two IOPPs to fulfill the conditions of Theorem 4.1 (sublinear sumcheck) when the 'sumcheck dimension' is $m = 2$, which implies that there exists a 2-round IOPP for $\mathrm{Rel}(\mathscr{D}, 2, \mathscr{H})$ with the above parameters. The theorem follows by observing that H_k^2 corresponds to the systematic part of the domain of $C_k^{\otimes 2c'}$, and so $\mathrm{Rel}(\mathscr{D}, 2, \mathscr{H})$ equals $\mathrm{Rel}(\mathrm{SC}, \mathscr{C}^{\otimes 2c'})$. □

We now prove the theorem by combining the above lemma with our reduction from circuit satisfiability to sumcheck (Theorem 6.2), instantiating both with an efficient construction of algebraic-geometry codes (see Theorem 2.3).

Proof of Theorem 7.1. Let $\mathscr{A}$, $\mathscr{B}$ be the code classes ('AG codes') from Theorem 2.3, with alphabet $\mathbb{F}_q$ for some $q \geq 5$, and $\mathscr{B}$ being the degree-3 closure of $\mathscr{A}$. Then:

- By Theorem 5.3 there exists a 1-round IOPP for $\mathrm{Rel}(\mathscr{A}^{\otimes 2c'})$ with proof length $o(n)$ and constant query complexity.
- Applying Lemma 7.2 to $\mathscr{B}$ with $c' := \lceil 3/\eta + 1 \rceil$ (in particular, $c' > 4$), we obtain an IOPP for $\mathrm{Rel}(\mathrm{SC}, \mathscr{B}^{\otimes 2c'})$.
- By Lemma 2.5, there exists an $\tilde{O}(n)$-efficient $(2/5)$-evading set family $\mathscr{S}$ for $\mathbb{F}_q$.

Combining these ingredients with Theorem 6.2 (CSAT to sumcheck reduction) yields the theorem statement.

We only highlight one technicality: Theorem 6.2 requires that $\mathscr{A}^{\otimes 2c'} = \{A_k^{\otimes 2c'}\}_{k \in \mathbb{N}}$ be k-systematic, but this code family is $k^{2c'}$-systematic. We therefore 'scale' the code family $\mathscr{A}^{\otimes 2c'}$ to a family $\mathscr{A}' = \{A'_n\}_{n \in \mathbb{N}}$ where $A'_n := A_k^{\otimes 2c'}$ for the smallest integer k such that $k^{2c'} \geq n$. We apply to same 'scaling' to $\mathscr{B}^{\otimes 2c'}$ to obtain the code family $\mathscr{B}'$. It is easily seen that both $\mathscr{A}'$ and $\mathscr{B}'$ are n-systematic, as required. □

Acknowledgments

This work has received funding from the Israel Science Foundation (grant 1501/14), the UC Berkeley Center for Long-Term Cybersecurity, and the United States – Israel Binational Science Foundation (grant 2021036).

References

[AGHP92] Noga Alon, Oded Goldreich, Johan Håstad, and René Peralta. Simple construction of almost k-wise independent random variables. *Random Structures and Algorithms*, 3(3):289–304, 1992.

[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.

[AS03] Sanjeev Arora and Madhu Sudan. Improved low-degree testing and its applications. *Combinatorica*, 23(3):365–426, 2003. Preliminary version appeared in STOC '97.

[Bab85] László Babai. Trading group theory for randomness. In *Proceedings of the 17th Annual ACM Symposium on Theory of Computing*, STOC '85, pages 421–429, 1985.

[BBC$^+$17] 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. In *Proceedings of the 36th Annual International Conference on Theory and Application of Cryptographic Techniques*, EUROCRYPT '17, pages 551–579, 2017.

[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, pages 585–594, 2013.

[BCGV16] Eli Ben-Sasson, Alessandro Chiesa, Ariel Gabizon, and Madars Virza. Quasilinear-size zero knowledge from linear-algebraic PCPs. In *Proceedings of the 13th Theory of Cryptography Conference*, TCC '16-A, pages 33–64, 2016.

[BCS16] Eli Ben-Sasson, Alessandro Chiesa, and Nicholas Spooner. Interactive oracle proofs. In *Proceedings of the 14th Theory of Cryptography Conference*, TCC '16-B, pages 31–60, 2016.

[BFL90] László Babai, Lance Fortnow, and Carsten Lund. Nondeterministic exponential time has two-prover interactive protocols. In *Proceedings of the 31st Annual Symposium on Foundations of Computer Science*, SFCS '90, 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$^+$05] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil Vadhan. Short PCPs verifiable in polylogarithmic time. In *Proceedings of the 20th Annual IEEE Conference on Computational Complexity*, CCC '05, pages 120–134, 2005.

[BGH$^+$06] Eli Ben-Sasson, Oded Goldreich, Prahladh Harsha, Madhu Sudan, and Salil P. Vadhan. Robust PCPs of proximity, shorter PCPs, and applications to coding. *SIAM Journal on Computing*, 36(4):889–974, 2006.

[BKK$^+$13] Eli Ben-Sasson, Yohay Kaplan, Swastik Kopparty, Or Meir, and Henning Stichtenoth. Constant rate PCPs for Circuit-SAT with sublinear query complexity. In *Proceedings of the 54th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '13, pages 320–329, 2013.

[BS06] Eli Ben-Sasson and Madhu Sudan. Robust locally testable codes and products of codes. *Random Structures and Algorithms*, 28(4):387–402, 2006.

[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.

[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.

[CC88] D. V. Chudnovsky and G. V. Chudnovsky. Algebraic complexities and algebraic curves over finite fields. *Journal of Complexity*, 4(4):285–316, 1988.

[CMS17] Alessandro Chiesa, Peter Manohar, and Igor Shinkar. On axis-parallel tests for tensor product codes. In *Proceedings of the 2017 Conference on Approximation, Randomization, and Combinatorial Optimization*, RANDOM '17, pages 39:1–39:22, 2017.

[DH13] Irit Dinur and Prahladh Harsha. Composition of low-error 2-query PCPs using decodable PCPs. *SIAM Journal on Computing*, 42(6):2452–2486, 2013. Preliminary version appeared in Property Testing '10.

[DHK15] Irit Dinur, Prahladh Harsha, and Guy Kindler. Polynomially low error PCPs with polyloglog n queries via modular composition. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, STOC '15, pages 267–276, 2015.

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

[DL78] Richard A. DeMillo and Richard J. Lipton. A probabilistic remark on algebraic program testing. *Information Processing Letters*, 7(4):193–195, 1978.

[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.

[DSW06] Irit Dinur, Madhu Sudan, and Avi Wigderson. Robust local testability of tensor products of LDPC codes. In *Proceedings of the 9th International Workshop on Approximation Algorithms for Combinatorial Optimization Problems, and of the 10th International Workshop on Randomization and Computation*, APPROX-RANDOM '06, pages 304–315, 2006.

[Efr12] Klim Efremenko. 3-query locally decodable codes of subexponential length. *SIAM Journal on Computing*, 41(6):1694–1703, 2012. Preliminary version appeared in STOC '09.

[FGL$^+$91] Uriel Feige, Shafi Goldwasser, László Lovász, Shmuel Safra, and Mario Szegedy. Approximating clique is almost NP-complete (preliminary version). In *Proceedings of the 32nd Annual Symposium on Foundations of Computer Science*, SFCS '91, pages 2–12, 1991.

[For65] David G. Forney. Concatenated codes. Technical report, MIT, Cambridge, MA, USA, 1965.

[FRS88] Lance Fortnow, John Rompel, and Michael Sipser. On the power of multi-prover interactive protocols. In *Theoretical Computer Science*, pages 156–161, 1988.

[FS86] Amos Fiat and Adi Shamir. How to prove yourself: practical solutions to identification and signature problems. In *Proceedings of the 6th Annual International Cryptology Conference*, CRYPTO '86, pages 186–194, 1986.

[GGR11] Parikshit Gopalan, Venkatesan Guruswami, and Prasad Raghavendra. List decoding tensor products and interleaved codes. *SIAM Journal on Computing*, 40(5):1432–1462, 2011. Preliminary version appeared in STOC '09.

[GI05] Venkatesan Guruswami and Piotr Indyk. Linear-time encodable/decodable codes with near-optimal rate. *IEEE Transactions on Information Theory*, 51(10):3393–3400, 2005. Preliminary version appeared in STOC '03.

[GIMS10] Vipul Goyal, Yuval Ishai, Mohammad Mahmoody, and Amit Sahai. Interactive locking, zero-knowledge PCPs, and unconditional cryptography. In *Proceedings of the 30th Annual Conference on Advances in Cryptology*, CRYPTO'10, pages 173–190, 2010.

[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.

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1):186–208, 1989. Preliminary version appeared in STOC '85.

[Gop81] Valery Denisovich Goppa. Codes on algebraic curves. *Soviet Mathematics — Doklady*, 1(24):170–172, 1981.

[GS96] Arnaldo Garcia and Henning Stichtenoth. On the asymptotic behaviour of some towers of function fields over finite fields. *Journal of Number Theory*, 61(2):248–273, 1996.

[GS06] Oded Goldreich and Madhu Sudan. Locally testable codes and PCPs of almost-linear length. *Journal of the ACM*, 53:558–655, July 2006. Preliminary version in STOC '02.

[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 '01.

[KMRS16] Swastik Kopparty, Or Meir, Noga Ron-Zewi, and Shubhangi Saraf. High-rate locally-correctable and locally-testable codes with sub-polynomial query complexity. In *Proceedings of the 48th ACM Symposium on the Theory of Computing*, STOC '16, pages 202–215, 2016.

[KR08] Yael Kalai and Ran Raz. Interactive PCP. In *Proceedings of the 35th International Colloquium on Automata, Languages and Programming*, ICALP '08, pages 536–547, 2008.

[KSY14] Swastik Kopparty, Shubhangi Saraf, and Sergey Yekhanin. High-rate codes with sublinear-time decoding. *Journal of the ACM*, 61(5):28:1–28:20, 2014. Preliminary version appeared in STOC '11.

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

[Mei12] Or Meir. Combinatorial PCPs with short proofs. In *Proceedings of the 26th Annual IEEE Conference on Computational Complexity*, CCC '12, 2012.

[Mei13] Or Meir. IP = PSPACE using error-correcting codes. *SIAM Journal on Computing*, 42(1):380–403, 2013.

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

[Mie09] Thilo Mie. Short PCPPs verifiable in polylogarithmic time with o(1) queries. *Annals of Mathematics and Artificial Intelligence*, 56:313–338, 2009.

[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.

[NN90] Joseph Naor and Moni Naor. Small-bias probability spaces: efficient constructions and applications. In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing*, STOC '90, pages 213–223, 1990.

[PF79] Nicholas Pippenger and Michael J. Fischer. Relations among complexity measures. *Journal of the ACM*, 26(2):361–381, 1979.

[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.

[PS96] David Pointcheval and Jacques Stern. Security proofs for signature schemes. In *Proceedings of the 14th Annual International Conference on Theory and Application of Cryptographic Techniques*, EUROCRYPT '96, pages 387–398, 1996.

[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, pages 49–62, 2016.

[RS97] Ran Raz and Shmuel Safra. A sub-constant error-probability low-degree test, and a sub-constant error-probability PCP characterization of NP. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing*, STOC '97, pages 475–484, 1997.

[RS06] Ron M. Roth and Vitaly Skachek. Improved nearly-MDS expander codes. *IEEE Transactions on Information Theory*, 52(8):3650–3661, 2006.

[SAK$^+$01] Kenneth W. Shum, Ilia Aleshnikov, P. Vijay Kumar, Henning Stichtenoth, and Vinay Deolalikar. A low-complexity algorithm for the construction of algebraic-geometric codes better than the Gilbert–Varshamov bound. *IEEE Transactions on Information Theory*, 47(6):2225–2241, 2001.

[Sch80] Jacob T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *Journal of the ACM*, 27(4):701–717, 1980.

[Sha92] Adi Shamir. IP = PSPACE. *Journal of the ACM*, 39(4):869–877, 1992.

[Spi96] Daniel A. Spielman. Linear-time encodable and decodable error-correcting codes. *IEEE Transactions on Information Theory*, 42(6):1723–1731, 1996. Preliminary version appeared in STOC '95.

[Sti08] Henning Stichtenoth. *Algebraic function fields and codes*. Springer Publishing Company, 2nd edition, 2008.

[Tan81] Robert Michael Tanner. A recursive approach to low complexity codes. *IEEE Transactions on Information Theory*, 27(5):533–547, 1981.

[Val08] Paul Valiant. Incrementally verifiable computation or proofs of knowledge imply time/space efficiency. In *Proceedings of the 5th Theory of Cryptography Conference*, TCC '08, pages 1–18, 2008.

[Vid10] Michael Viderman. A note on high-rate locally testable codes with sublinear query complexity, 2010. ECCC TR10-171.

[Vid15] Michael Viderman. A combination of testability and decodability by tensor products. *Random Structures and Algorithms*, 46(3):572–598, 2015. Preliminary version appeared in APPROX-RANDOM '12.

[WE63] Jack Keil Wolf and Bernard Elspas. Error-locating codes - a new concept in error control. *IEEE Transactions on Information Theory*, 9(2):113–117, 1963.

[Wol65] Jack Keil Wolf. On codes derivable from the tensor product of check matrices. *IEEE Transactions on Information Theory*, 11(2):281–284, 1965.

[Yek08] Sergey Yekhanin. Towards 3-query locally decodable codes of subexponential length. *Journal of the ACM*, 55(1), 2008. Preliminary version appeared in STOC '07.

[Zip79] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *Proceedings of the 1979 International Symposium on Symbolic and Algebraic Computation*, EUROSAM '79, pages 216–226, 1979.